DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

“The role of emotion identification in empathy: evidence from psychopharmacological, neuroimaging, and effective connectivity approaches”

verfasst von / submitted by

Yili Zhao

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doktorin der Naturwissenschaften (Dr.rer.nat.)

Wien, 2021/ Vienna 2021

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 796 605 298
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Psychologie
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Claus Lamm

Contents

Chapter 1 – General introduction

Overview

Human affective states are influenced by others' feelings. When seeing another person suffering either physically or mentally, we not only feel sympathetic towards that person but can also, to some extent, feel and share their affect as if it were our own. This ability is a key component of the multi-faceted experience of empathy (Decety & Hodges, 2006; Decety & Jackson, 2006). In our daily life, empathy is vital for interpersonal interaction. Maintaining the ability to empathize with others can promote prosocial behaviors and ultimately contributes to social coordination and possibly also harmony. Therefore, it is important from both scientific and applied perspectives to disentangle which factors can influence empathic responses so that people who show empathy deficits can be helped to overcome them. Emotion identification, the ability to identify another's emotion, has recently been proposed as a necessary step, in addition to affect sharing, for successfully empathizing with others (Coll et al., 2017). If a person shows little empathy for another's suffering, it could be due to a reduced ability or tendency to share another person's affect, or it could be because this person cannot appropriately identify that person's emotion. Approximately 50% of people who receive a diagnosis of autism spectrum disorder score rather high on measures of alexithymia (Geoffrey; Bird & Viding, 2014). For those individuals who show a reduction in their ability to share the affect of others, a big contributor is that they cannot appropriately "read" others' affective states due to alexithymia rather than autism itself (G. Bird & Cook, 2013). As a result, it is important to investigate the mechanisms of emotion identification and its relationship with empathy for both basic research and clinical practice. Yet, studies on this topic are rather scarce.

In this doctoral thesis, I focused on the following research questions: 1) whether and how the recognition of painful emotional expressions was modulated by the opioidergic system (Chapter 2), 2) what neural activation dissociated empathic responses to seeing others genuinely in pain from seeing others pretending to be in pain (Chapter 3), and 3) what modality-independent and modality-dependent neural signatures underly seeing others genuinely experiencing disgust vs. pretending to experience disgust as compared to the findings of pain (Chapter 4). Accordingly, three studies were performed. In Chapter 2, using psychopharmacological manipulation and functional magnetic resonance imaging (fMRI), we investigated how the opioid antagonist naltrexone influenced the discrimination of painful facial expressions from its morphed expressions with disgust. In Chapter 3, using fMRI and dynamical causal modeling (DCM) analysis, we studied the neural networks that modulated the affective responses to genuine pain experienced by others as compared to pretended pain. In Chapter 4, building upon and extending our findings from Chapter 3, we investigated what were the shared and distinct neural mechanisms between seeing others experiencing genuine disgust vs. pretended disgust and the results of seeing others experiencing genuine pain vs. pretended pain in Chapter 3. The theoretical and methodological bases of these studies will be elaborated on in the next sections of this chapter.

Empathy

The definition of empathy

Empathy, the capacity to understand and share the thoughts and feelings of others, plays a crucial role in social interactions and promotes prosocial behaviors (Decety & Jackson, 2006; Lamm et al., 2019). Crucially, empathy requires knowing that another person is the origin of our affective state (Decety & Jackson, 2004; de Vignemont & Singer, 2006). When we talk about empathy, it is inevitable to mention two related terms: sympathy and emotional contagion. The term "sympathy" refers to “feelings of concern about the welfare of others” (P. 204) but not necessarily to resonate with their affective states (Decety, 2010); whereas for emotional contagion, we match the affective state of another, but this affective state does not require to be “explicitly recognized as being experienced by the other” (P. 521) (Geoffrey; Bird & Viding, 2014). Empathy is composed of multiple perceptual, cognitive, and affective processes, including observation, memory, emotion, and reasoning (Ickes, 1997). It is regarded as a multidimensional construct with three main components: (1) identifying others’ emotions and taking the perspective of others, (2) sharing other peoples’ affective states, and (3) regulatory mechanisms for maintaining a clear distinction between the self and others (Decety & Jackson, 2006).

The neural network of empathy and shared representations

To study the neural underpinnings of empathy, many social neuroscience studies have used pain as a means to evoke participants’ empathic responses and implemented fMRI as a technical approach to record the corresponding brain activations. The main neuroimaging findings are that empathy for pain activates parts of the neural network that are also involved in the first-hand processing of pain, namely, the anterior insula cortex (aIns) and the anterior-mid cingulate cortex (aMCC) (Singer et al., 2004; Botvinick et al., 2005; Jackson et al., 2005; Lamm et al., 2011; Rütgen et al., 2015b; Jauniaux et al., 2019; Xiong et al., 2019; Fallon et al., 2020; Zhou et al., 2020). These two regions are typically considered to represent the affective-motivational components of empathy (Lamm et al., 2016). In a pioneering study, Singer et al. (2004) showed activation in a conjunctive network, including aIns and anterior cingulate cortex, when participants received painful stimulation themselves as well as when they saw their partner experiencing pain.

This neural overlap between the first-hand experience of pain and empathy for pain could be explained by the theory of shared representations. Initialized from the perception-action model proposed by Preston & De Waal (2002), this view proposes that the perception of another person's affective state automatically activates the observer's own representation of that state, followed by the associated autonomic and somatic responses. Therefore, an intriguing and hotly-debated question is to what extent shared brain activations indicate shared representations (Decety, 2010; Lamm & Majdandžić, 2015; Lamm et al., 2016). It is well known that in fMRI research, one and the same brain structure can be activated by a variety of tasks and functions, which means an overlap of neural activations does not necessarily imply that the same psychological representations are engaged. Areas such as aIns and aMCC, for example, are not only activated by pain, but also by numerous other processes, such as arousal/salience, attention, and cognitive control (Zaki et al., 2016). Similar results have been found using other approaches, such as electroencephalography (EEG) and Transcranial magnetic stimulation (TMS) (Lamm & Majdandžić, 2015).

Recent technological and methodological advances have resulted in considerable progress on this issue. Corradi-Dell'Acqua and colleagues (2011) applied multivariate pattern analysis (MVPA) showing overlapped activation patterns between the first-hand experience of pain and the pain perceived in others. Another series of studies (Rütgen et al., 2015a; Rütgen et al., 2015b; Rütgen et al., 2018; Rütgen et al., 2021), using placebo analgesia manipulation on the first-hand pain, found reduced subjective ratings of pain intensity and unpleasantness for both self-oriented and other-oriented pain. Furthermore, this approach line revealed decreased activation in aIns and aMCC led by the placebo analgesia during both self-directed and other-directed painful stimulation (Rütgen et al., 2015b). In the same study, the researchers also found the blockage of the placebo effect by the opioid antagonist naltrexone seemed to eliminate the placebo analgesia induced beforehand for both first-hand pain and empathy for pain (Rütgen et al., 2015b). This effect is supported by numerous studies on the essential engagement of the endogenous opioid system in the modulation of the placebo analgesic effect (Petrovic et al., 2002; Wager et al., 2004; Benedetti et al., 2005; Zubieta et al., 2005). These findings indicate that the effect applied to the first-hand experience of pain could be "transferred" to the vicarious pain, hinting that the placebo analgesic effect works on the shared representations. Therefore, empathy for pain might in fact be grounded in first-hand pain.

A new theoretical framework of empathy: the role of emotion identification

Recently, our lab has collaborated with colleagues from the UK to address the relationship between emotion identification and affective sharing during the experience of empathy (Coll et al., 2017). Specifically, researchers have introduced a new measurement framework of empathy, in which empathy is not just defined as affective sharing – "the degree to which identification of another's state causes a corresponding state in the self", but also emphasized the role of emotion identification – "the ability to identify another's emotional state" (P. 132). Therefore, when measuring an individual's degree of empathic response, one should consider both affective sharing and emotion identification. If the empathizers do not correctly identify someone else's emotion, they cannot show the same empathic response as another person who is able to identify the emotion correctly, even though they might both have the same underlying degree of affect sharing. In this context, people of the former group cannot be defined as "low-empathizers". In other words, only after we are certain that all empathizers have identified the emotion of others correctly, can it be assumed that the variation of their empathic responses is due to differences in affect sharing. On top of that, researchers re-analyzed the behavioral data of the previous placebo analgesia study from our lab (Rütgen et al., 2015b), and reported that the effect of placebo analgesia on empathic responses was almost fully mediated by the ratings of the intensity of pain attributed to the other person. This result seems to suggest that the effect of placebo analgesia on empathy might mostly act on the process of emotion identification rather than on affective sharing, and further stresses the importance of studying emotion identification during empathy.

However, two major limitations must be addressed before making these statements on the role of emotion identification in empathy more definite. Firstly, the original study of placebo analgesia was not designed to test the relationship between emotion identification and affective sharing, and as a result, the conclusions drawn from the re-analysis of the data seem quite indirect and more convincing evidence is needed. Secondly, the ratings of pain intensity and unpleasantness used in the previous study (Rütgen et al., 2015b), which were, respectively, regarded as measurements of emotion identification and affective sharing (Coll et al., 2017), were presented in a fixed order that the unpleasantness ratings always came after the pain intensity ratings, therefore, we cannot exclude the potential influence of order effects. Altogether, it is necessary to collect new evidence and to further investigate the role of emotion identification during empathy.

Before going into more details about the studies, I would like to further clarify and underline the research goals of this thesis. Considering the complexity of emotion identification and the difficulty of operating and measuring emotion identification directly, the current thesis did not aim to test the whole model of Coll et al. (Coll et al., 2017), namely how emotion identification and affect sharing independently contribute to the consequences of empathy. Rather, it mainly focused on those processes closely related to emotion identification but can be more easily and directly manipulated and measured, and how the manipulation of these processes influenced empathic responses to others. In the following section, I will elaborate on emotion identification and those relevant processes so that readers can have a better understanding of how the present studies subserve decoding the role of emotion identification in empathy.

Emotion identification and its relevant processes

Definitions and the underlying neural mechanisms

Emotion identification, the ability to be aware of and identify the emotional state of the target, is fundamental in social communication and interpersonal interactions (Schutte et al., 2001). Several different psychological processes – emotion perception, recognition, and categorization - are considered to be comprised of emotion identification (Coll et al., 2017). Schimer and Adolphs (2017) have defined these relevant but still distinct processes: emotion perception focuses on detecting and discriminating an emotion in an early stage, during which stage the emotion is internally conceptualized in preparation for the subsequent processes of emotion recognition and emotion categorization. Emotion recognition is related to infer others' internal states and retrieve the conceptual knowledge about an emotion; while for emotion categorization, it is more relevant to sorting an emotion into specific categories implicated in labeling that emotion. These processes often elicit one's own emotional responses. Therefore, neural activations underpinning functions from primary visual perception to higher-order cognitive manipulation and affective response are all potentially engaged in emotion identification (Gur et al., 2002; Srinivasan & Hanif, 2010). In experimental research, pictures or video clips of facial emotional expressions are the most frequently employed materials in the tasks related to emotion identification (Berthoz et al., 2002; Moriguchi et al., 2006; Clark et al., 2008; Bal et al., 2010). Additionally, other modalities, such as vocal and tactile stimulation can also be applied to study processes related to emotion identification (Schirmer & Adolphs, 2017). In terms of studies on the recognition of facial emotions, the main neuroimaging findings indicate that key brain areas involved in processes relevant to emotion identification include the fusiform face area (FFA), the superior temporal sulcus (STS), amygdala, the frontal lobes, and the parietal lobes (Calder, 1996; Adolphs, 2002; Gur et al., 2002; Grill-Spector et al., 2004; Heberlein et al., 2008). Particularly, among the two visual cortices - the FFA and STS, the former area is claimed to be more engaged in processing static facial expressions whereas the latter one is considered to be more engaged in processing dynamic facial expressions (Pitcher et al., 2011; De Winter et al., 2015).

Relationship with the opioid system

Activation of neurochemical systems regulates emotions. Among these neurochemical systems, the opioid system has been shown to be generally engaged in modulating different types of emotions. Recently, opioid agonists and antagonists have been reported to alter individuals' responses to emotional facial expressions of others (Ipser et al., 2013; Schmidt et al., 2014; Bershad et al., 2016; Meier et al., 2016; Wardle et al., 2016). For instance, Wardle et al. (2016) found 50 mg naltrexone (an opioid antagonist) generally increased attention to emotional faces (anger, fear, happiness, or sadness) and prolonged the recognition of sad and fearful facial expressions, whereas Meier et al. (2016), in an electromyography study, showed that the same dose of naltrexone resulted in increased corrugator and depressor activity to happiness expressions. Both muscles are associated with negatively valenced emotions. Furthermore, Ipser et al. (2013) observed that the administration of 0.2 mg of buprenorphine (an opioid agonist) reduced sensitivity to fear expressions in a recognition task. These mixed results suggest that drawing firm conclusions regarding the role of the opioid system in recognizing the different emotions of others is still not possible.

Studies on how the opioid system modulates the perception of others' pain are rather rare, though there is a broad consensus that the endogenous opioid system essentially modulates the direct experience of pain in terms of both sensory and affective components (Zubieta et al., 2001; Wager et al., 2004; Wager et al., 2007). Using positron emission tomography and fMRI, Karjalainen and colleagues (2017) revealed the involvement of μ-opioid receptors when watching others in pain and the association of μ-opioid receptor availability and brain activations during these painful scenes. The psychopharmacological and neuroimaging findings of our lab (Rütgen et al., 2015b; Rütgen et al., 2018) indicate a causal role of the opioid system in empathy for pain, that the relieved empathy for pain by the placebo analgesic effect seems to be "renormalized" after applying the opioid antagonist naltrexone. Recent mediation analyses (details can be found in a former section) suggest that this was possibly explained by effects on emotion identification (Coll et al., 2017). Using supervised machine-learning classification, Haaker et al. (2017) showed that the blockade of opioid receptors enhanced social threat learning from observing others in pain. Additionally, related research revealed that other neurochemical mechanisms, for instance, acetaminophen, might also modulate empathy for another's pain (Mischkowski et al., 2016). However, since none of these studies had a specific focus on the visual perception and the processing of painful facial expressions, it remains an open question on the specific relationship between the opioid system and the perception of other's pain expressions. Therefore, in Chapter 2, we aimed to investigate whether the endogenous opioid system was able to modulate the recognition of painful facial expressions in others and the corresponding neural mechanisms.

Experimental paradigms related to emotion identification

Several experimental paradigms have been used to measure participants' ability to discriminate or recognize emotional facial expressions. Ipser et al. (2013) performed an emotion recognition task in which participants watched short video clips (1-3 s) that started from a neutral face and dynamically morphed into one of the four facial expressions (anger, fear, happiness, and sadness) at different intensities (20% - 100%, in 10% step). Participants' task was to label which emotion this video belonged to among all four emotions and the error rate was respectively computed for each emotion. In another study, Brewer et al. (2015) used a paradigm in which participants watched several full-blown facial expressions (happiness, sadness, disgust, anger, surprise, fear, and pain) that were obscured by different levels of visual noise (10%, 30%, 50%, 70%, and 90%). Participants were required to judge whether the presented emotion belonged to a prompted emotion (e.g., "pain: yes or no") and the maximum level of noise tolerance (e.g., 60% maximum level of noise tolerance means participants could reliably identify an emotion when the level of noise is 60% at highest) was individually calculated. However, both paradigms are generally appropriate for studying the recognition of multiple emotions rather than focusing on one or two specific emotions. Cook and colleagues (2013) have used continually morphed pictures of two attributes – expression (e.g., disgust and anger) and identity (e.g., Harold and Felix) – to determine the relative contribution of autism and alexithymia on participant's ability to attribute facial identity and emotion. For instance, from a morphed continuum derived from Harold expressing anger to Felix expressing disgust, subjects need to attribute either its expression (e.g., "disgust or anger?") or its identity (e.g., "Harold or Felix?") following each image. This paradigm could allow for studying emotion discrimination between two emotions. To further investigate whether autism or alexithymia was actually correlated with the perceptual ability of emotion detection, in the same study Cook and colleagues performed a second experiment in which participants were asked to merely judge whether two stimuli were identical (expression and identity were morphed independently). However, the latter paradigm could not allow us to test high-order cognitive and affective processing related to emotion identification. Altogether, in Chapter 2 the former paradigm of Cook et al. (2013) was adapted to investigate emotion discrimination/recognition of painful facial expressions under psychopharmacological opioidergic manipulation. Considering the focus of my study is the processes relate to emotion identification, we only considered manipulating the attribution of expression, rather than identity, of the people in the morphed pictures. Here we used disgust as well as pain, and I justify these choices in the next section.

Pain and disgust

Pain and disgust are two aversive experiences that can signal potential threats to one's physical integrity and health. From an evolutionary perspective, they are mental representations that signal human physiological states (Damasio & Carvalho, 2013). Specifically, pain is a sensory and affective experience due to actual or potential somatic damage (Merskey & Bogduk, 1994), which is usually linked to tissue or organ damage or disease (Innes, 2005), whereas disgust is conceptualized as a "motivational system" to particularly detect signs of and avoid potential pathogens, contaminants, and toxins (Kavaliers et al., 2019).

Both cross-modal and modality-dependent functional mechanisms have been found for these two affective states. One study showed that preceding either high painful or high disgusted cues induced higher unpleasantness for the following thermal and olfactory stimulation, as opposed to low valenced cues. Furthermore, this effect was stronger when the modality of the cue and the stimulus was consistent compared to that of inconsistent (Sharvit et al., 2015). Another study demonstrated that priming thermal pain or olfactory disgust biased the classification of the subsequent hybrid expressions as pain or disgust, but this effect did not extend to other emotions like surprise or neutral emotion (Antico et al., 2019).

There have been consistent claims that the movements of our facial muscles are largely overlapping for facial expressions of pain and disgust (Ekman & Friesen, 1978; Kappesser & Williams, 2002; Simon et al., 2008). That said, studies suggest that the facial expressions of pain and disgust are still distinct enough that individuals can dissociate them from each other (Kunz et al., 2013; Kunz & Lautenbacher, 2015; Zhao et al., 2021). For instance, Kunz et al. (2013) showed that pain expressions are mostly encoded with muscles movements surrounding the eyes and/or together with the contraction of eyebrows, whereas disgust is expressed by both the contraction of the eyebrows and also the raising of the upper lip. What seems clear though is that both experiences engage a protective withdrawal response, and that is also expressed on as well as communicated by the face.

In terms of the neural substrates, similar to pain, aIns and aMCC, but especially aIns, are also claimed as core areas for the direct and vicarious experiences of disgust, and in particular its affective components (Wicker et al., 2003; Botvinick et al., 2005; Gottfried & Zald, 2005; Corradi-Dell'Acqua et al., 2016). On top of the overlaps, domain-specific activity is engaged for these two affective states. For the experience of pain, the primary and secondary somatosensory cortex are strongly engaged (Peyron et al., 2000), and seem to encode predominantly the sensory-discriminative component of pain. For the emotion of disgust, the primary olfactory cortex, amygdala, orbital frontal cortex, and the striatum are reliably activated, tracking predominantly sensory aspects as well (Phillips et al., 1997; Phillips et al., 1998; Gottfried & Zald, 2005).

Since the direct and vicarious experiences of pain and disgust are fundamentally implicated in brain activations in aIns and aMCC, one may wonder whether these overlapping activations represent cross-modal or modality-specific processes. Previous studies have shown evidence of both features. Corradi-Dell'Acqua et al. (2016) using multivoxel pattern analysis demonstrated that in the left aIns and aMCC, the shared encoding was detected between the direct and vicarious experiences of pain and disgust, regardless of the same or different modality. In contrast, in the right aIns, sensory-specific rather than modality-independent patterns were more plausible for processing the direct and vicarious pain and disgust. Furthermore, Sharvit and colleagues (2018) found that different subdivisions of aIns were likely to distinctly influence the sensory-specific expectancy of pain and disgust. Specifically, an intermediate section of aIns could mediate the effect of expectancy on the following stimuli which was consistent with the priming cue, while a more anterior portion of aIns suppressed this effect.

On the neurochemical level, different endogenous mechanisms could underlie the experiences of pain and disgust. A number of studies suggest that the opioid system modulates nociceptive processing, and a few studies imply its engagement in the vicarious experience of pain as well (see previous research in the above section). Whereas for disgust, there is little evidence showing a link with the opioid system, while other systems, e.g., the oxytocin system, are more likely to play a role (see Kavaliers et al., 2019).

To summarize, disgust is an emotion that shares a lot of similarities with pain, in terms of their evolutionary meaning, their corresponding facial expressions, and affect-related neural underpinnings. But they do differ in their sensory-specific activations and in their reliance on opioidergic mechanisms. Pain and disgust are expressed similarly in the faces of those experiencing these emotions (but they can still be dissociated from each other) and disgust is likely to be distinctly influenced by the opioid system as compared to pain. Altogether, this makes disgust an appropriate counterpart to pain for studying the opioidergic modulation of pain in an emotion discrimination/recognition task among healthy participants (Chapter 2). Furthermore, it is meaningful for us to investigate the neural underpinnings of how those processes implicated in emotion identification influenced empathic responses across different aversive experiences, namely, pain and disgust (Chapter 3 and Chapter 4).

Self-other distinction and empathy

Self-other distinction refers to the ability to distinguish the affect experienced by oneself from that of others (Lamm et al., 2016). According to the notion of shared representations, there is a shared network between the self- and other-related processes underlying the functional mechanism of empathy. To share the feelings and emotions of another person, empathy requires us to switch between and integrate the mental states of oneself and that person. During this process, it still holds a regulatory mechanism to distinguish between those emotions deriving from one's own to those from another person (Decety & Hodges, 2006; Lamm et al., 2016). Failure to dissociate the mental states of ourselves' from others' might impact regular social interactions (Decety & Sommerville, 2003;

Decety & Jackson, 2004), and lead to abnormal imitative behaviors (Steinbeis et al., 2015) and personal distress (Decety & Lamm, 2011).

The right inferior parietal lobule is considered to play a vital role in distinguishing oneself from others (Decety & Sommerville, 2003). Recent research has claimed a functional subdivision within this region. An anterior portion, namely, the right supramarginal gyrus (rSMG), is suggested to act as a major hub selectively engaged in self-other distinction regarding others' affective states (Decety & Sommerville, 2003; Silani et al., 2013; Steinbeis et al., 2015; Hoffmann et al., 2016; Bukowski et al., 2020), and a posterior subregion - the right temporoparietal junction (rTPJ), is more frequently engaged in cognitive self-other distinction, including inferring and reasoning about others' beliefs and thoughts (Steinbeis et al., 2015). Evidence on functional connectivity has exhibited distinguished connectivity patterns of these two regions, that the "cognitive area" rTPJ is more strongly connected to regions implicated in perspective-taking (e.g., posterior cingulate, temporal pole, and medial prefrontal cortex and posterior cingulate), while the "affective area" rSMG is more connected to the regions linked to the affective components of empathy (i.e., aIns and aMCC) (Mars et al., 2011).

Regarding how the self- and other-related processes are represented in these two regions, Cheng et al. (2010) employed a task in which participants were instructed either to take a stranger's or take their partner's perspective when observing pictures exhibiting painful scenes. They found increased activations in rTPJ from the stranger's perspective; moreover, the closer a relationship the participant showed with their partner, the greater deactivation was detected in rTPJ. Silani et al. (2013) performed an experiment in which participants were required to rate the affective states of themselves or another person when their affective experiences were congruent (i.e., both pleasant or both unpleasant) or incongruent (i.e., one pleasant and one unpleasant). Results showed increased activations in rSMG for overcoming emotional egocentricity bias, that is, the larger difference in the judgment between the other's affect state with that of one's own, the stronger activation in rSMG. Using transcranial magnetic stimulation (TMS), Bukowski et al. (2020) have further confirmed the findings of the former study.

Though previous studies have indicated increased functional connectivity between rSMG and areas associated with affect processing (Mars et al., 2011; Bukowski et al., 2020), the mechanism underlying how this affective self-other distinction participates in empathy still requires more nuanced insights. In Chapter 3, we investigated effective connectivity by means of dynamical causal modeling (DCM) to explore the "crosstalk" between areas related to affect processing and self-other distinction that was modulated by the recognition of others' pain in different contexts. In Chapter 4, similar research questions were investigated in the emotion of disgust.

Empathy, emotion identification, and contextual information

Contextual information influences how we perceive and understand the suffering of others, and thus interacts with the consequences of empathy. Studies have shown that when seeing others in a plausible painful situation, contextual reality (e.g., an injection on the cheek with a needle of a syringe or with a Q-tip, Gu & Han, 2007) and situational rationality (e.g., an injection on the cheek with a needle of a syringe or with a Q-tip, Han et al., 2009) significantly influence the consequences of empathy. Furthermore, being aware of whether the suffering of a person is real or just acted also influences the empathic response to that person. A study revealed that when participants watched video clips showing either fictional victims of violence or actual people being injured or killed, they demonstrated more empathic responses to the victims' suffering when they knew these videos were real violence as compared to fictional violence (Ramos et al., 2013). In Chapter 3 and Chapter 4, we designed a novel experimental paradigm in which participants were presented with video clips of two conditions, a genuine condition and a pretended condition. In the genuine condition, participants were informed that they were watching some people displaying painful (Chapter 3) or disgusted (Chapter 4) facial expressions when they were genuinely experiencing pain or disgust. Whereas, in the pretended condition, participants observed another group of people showing similar painful or disgusted expressions but they were informed that these people were merely acting. Particularly, we matched the perceptual salience between the genuine and pretended conditions, aiming to disentangle those behavioral responses and neural signatures related to perceptual saliency from actual affect sharing. With this design, we expected participants to show a higher recognized affect of others and higher empathy in the genuine condition as opposed to lower recognized affect and empathy in the pretended condition. Hence, we could investigate how processes closely related to emotion identification modulated by distinct contextual information resulted in different consequences of empathy.

References

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology, 12*(2), 169-177. doi: https://doi.org/10.1016/S0959-4388(02)00301-X

Antico, L., Cataldo, E., & Corradi-Dell'Acqua, C. (2019). Does my pain affect your disgust? Cross-modal influence of first-hand aversive experiences in the appraisal of others' facial expressions. *Eur J Pain, 23*(7), 1283-1296. doi: http://doi.org/10.1002/ejp.1390

Bal, E., Harden, E., Lamb, D., Van Hecke, A. V., Denver, J. W., & Porges, S. W. (2010). Emotion Recognition in Children with Autism Spectrum Disorders: Relations to Eye Gaze and Autonomic State. *Journal of Autism and Developmental Disorders, 40*(3), 358-370. doi: http://doi.org/10.1007/s10803-009-0884-3

Benedetti, F., Mayberg, H. S., Wager, T. D., Stohler, C. S., & Zubieta, J.-K. (2005). Neurobiological Mechanisms of the Placebo Effect. *The Journal of Neuroscience, 25*(45), 10390-10402. doi: http://doi.org/10.1523/jneurosci.3458-05.2005

Bershad, A. K., Seiden, J. A., & de Wit, H. (2016). Effects of buprenorphine on responses to social stimuli in healthy adults. *Psychoneuroendocrinology, 63*, 43-49. doi: https://doi.org/10.1016/j.psyneuen.2015.09.011

Berthoz, S., Artiges, E., Van De Moortele, P.-F., Poline, J.-B., Rouquette, S., Consoli, S. M., & Martinot, J.-L. (2002). Effect of impaired recognition and expression of emotions on frontocingulate cortices: an fMRI study of men with alexithymia. *The American journal of psychiatry, 159*(6), 961-967. doi: http://doi.org/10.1176/appi.ajp.159.6.961

Bird, G., & Cook, R. (2013). Mixed emotions: the contribution of alexithymia to the emotional symptoms of autism. *Transl Psychiatry, 3*(7), e285-e285. doi: 10.1038/tp.2013.61

Bird, G., & Viding, E. (2014). The self to other model of empathy: Providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience & Biobehavioral Reviews, 47*, 520-532. doi: https://doi.org/10.1016/j.neubiorev.2014.09.021

Botvinick, M., Jha, A. P., Bylsma, L. M., Fabian, S. A., Solomon, P. E., & Prkachin, K. M. (2005). Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *NeuroImage, 25*(1), 312-319. doi: https://doi.org/10.1016/j.neuroimage.2004.11.043

Brewer, R., Cook, R., Cardi, V., Treasure, J., & Bird, G. (2015). Emotion recognition deficits in eating disorders are explained by co-occurring alexithymia. *Royal Society Open Science, 2*(1), 140382. doi: http://doi.org/doi:10.1098/rsos.140382

Bukowski, H., Tik, M., Silani, G., Ruff, C. C., Windischberger, C., & Lamm, C. (2020). When differences matter: rTMS/fMRI reveals how differences in dispositional empathy translate to distinct neural underpinnings of self-other distinction in empathy. *Cortex, 128*, 143-161. doi: https://doi.org/10.1016/j.cortex.2020.03.009

Calder, A. J. (1996). Facial Emotion Recognition after Bilateral Amygdala Damage: Differentially Severe Impairment of Fear. *Cognitive Neuropsychology, 13*(5), 699-745. doi: http://doi.org/10.1080/026432996381890

Cheng, Y., Chen, C., Lin, C.-P., Chou, K.-H., & Decety, J. (2010). Love hurts: An fMRI study. *NeuroImage, 51*(2), 923-929. doi: https://doi.org/10.1016/j.neuroimage.2010.02.047

Clark, T. F., Winkielman, P., & McIntosh, D. (2008). Autism and the extraction of emotion from briefly presented facial expressions: stumbling at the first step of empathy. *Emotion, 8 6*, 803-809. doi: https://doi.org/10.1176/appi.ajp.159.6.961

Coll, M.-P., Viding, E., Rütgen, M., Silani, G., Lamm, C., Catmur, C., & Bird, G. (2017). Are we really measuring empathy? Proposal for a new measurement framework. *Neuroscience & Biobehavioral Reviews, 83*, 132-139. doi: https://doi.org/10.1016/j.neubiorev.2017.10.009

Cook, R., Brewer, R., Shah, P., & Bird, G. (2013). Alexithymia, Not Autism, Predicts Poor Recognition of Emotional Facial Expressions. *Psychological Science, 24*(5), 723-732. doi: https://doi.org/10.1177/0956797612463582

Corradi-Dell'Acqua, C., Hofstetter, C., & Vuilleumier, P. (2011). Felt and Seen Pain Evoke the Same Local Patterns of Cortical Activity in Insular and Cingulate Cortex. *The Journal of Neuroscience, 31*(49), 17996-18006. doi: http://doi.org/10.1523/jneurosci.2686-11.2011

Corradi-Dell’Acqua, C., Tusche, A., Vuilleumier, P., & Singer, T. (2016). Cross-modal representations of first-hand and vicarious pain, disgust and fairness in insular and cingulate cortex. *Nature Communications, 7*(1), 10904. doi: 10.1038/ncomms10904

Damasio, A., & Carvalho, G. B. (2013). The nature of feelings: evolutionary and neurobiological origins. *Nat Rev Neurosci, 14*(2), 143-152. doi: http://doi.org/10.1038/nrn3403

de Vignemont, F., & Singer, T. (2006). The empathic brain: how, when and why? *Trends in cognitive sciences, 10*(10), 435-441. doi: https://doi.org/10.1016/j.tics.2006.08.008

De Winter, F.-L., Zhu, Q., Van den Stock, J., Nelissen, K., Peeters, R., de Gelder, B., Vanduffel, W., & Vandenbulcke, M. (2015). Lateralization for dynamic facial expressions in human superior temporal sulcus. *NeuroImage, 106*, 340-352. doi: https://doi.org/10.1016/j.neuroimage.2014.11.020

Decety, J. (2010). To What Extent is the Experience of Empathy Mediated by Shared Neural Circuits? *Emotion Review, 2*(3), 204-207. doi: http://doi.org/10.1177/1754073910361981

Decety, J., & Hodges, S. D. (2006). The social neuroscience of empathy *Bridging social psychology* (pp. 121-128): Psychology Press.

Decety, J., & Jackson, P. L. (2004). The Functional Architecture of Human Empathy. *Behavioral and Cognitive Neuroscience Reviews, 3*(2), 71-100. doi: http://doi.org/10.1177/1534582304267187

Decety, J., & Jackson, P. L. (2006). A social-neuroscience perspective on empathy. *Current Directions in Psychological Science, 15*(2), 54-58. doi: https://doi.org/10.1111/j.0963-7214.2006.00406.x

Decety, J., & Lamm, C. (2011). Empathy versus Personal Distress: Recent Evidence from Social Neuroscience. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 199 - 213): MIT Press.

Decety, J., & Sommerville, J. A. (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends in cognitive sciences, 7*(12), 527-533. doi: https://doi.org/10.1016/j.tics.2003.10.004

Ekman, P., & Friesen, W. (1978). *Action coding system: a technique for the measurement of facial movement*. Palo Alto: Consulting Psychologists Press.

Fallon, N., Roberts, C., & Stancak, A. (2020). Shared and distinct functional networks for empathy and pain processing: A systematic review and meta-analysis of fMRI studies. *Social Cognitive and Affective Neuroscience*. doi: https://doi.org/10.1093/scan/nsaa090

Gottfried, J. A., & Zald, D. H. (2005). On the scent of human olfactory orbitofrontal cortex: Meta-analysis and comparison to non-human primates. *Brain Research Reviews, 50*(2), 287-304. doi: https://doi.org/10.1016/j.brainresrev.2005.08.004

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience, 7*(5), 555-562. doi: http://doi.org/10.1038/nn1224

Gu, X., & Han, S. (2007). Attention and reality constraints on the neural processes of empathy for pain. *NeuroImage, 36*(1), 256-267. doi: https://doi.org/10.1016/j.neuroimage.2007.02.025

Gur, R. C., Schroeder, L., Turner, T., McGrath, C., Chan, R. M., Turetsky, B. I., Alsop, D., Maldjian, J., & Gur, R. E. (2002). Brain Activation during Facial Emotion Processing. *NeuroImage, 16*(3, Part A), 651-662. doi: https://doi.org/10.1006/nimg.2002.1097

Haaker, J., Yi, J., Petrovic, P., & Olsson, A. (2017). Endogenous opioids regulate social threat learning in humans. *Nature Communications, 8*(1), 15495. doi: http://doi.org/10.1038/ncomms15495

Han, S., Fan, Y., Xu, X., Qin, J., Wu, B., Wang, X., Aglioti, S. M., & Mao, L. (2009). Empathic neural responses to others' pain are modulated by emotional contexts. *Human Brain Mapping, 30*(10), 3227-3237. doi: https://doi.org/10.1002/hbm.20742

Heberlein, A. S., Padon, A. A., Gillihan, S. J., Farah, M. J., & Fellows, L. K. (2008). Ventromedial frontal lobe plays a critical role in facial emotion recognition. *J Cogn Neurosci, 20*(4), 721-733. doi: http://doi.org/10.1162/jocn.2008.20049

Hoffmann, F., Koehne, S., Steinbeis, N., Dziobek, I., & Singer, T. (2016). Preserved Self-other Distinction During Empathy in Autism is Linked to Network Integrity of Right Supramarginal Gyrus. *Journal of Autism and Developmental Disorders, 46*(2), 637-648. doi: http://doi.org/10.1007/s10803-015-2609-0

Ickes, W. J. (1997). *Empathic accuracy*. New York: Guilford Press.

Innes, S. I. (2005). Psychosocial factors and their role in chronic pain: A brief review of development and current status. *Chiropractic & Osteopathy, 13*(1), 6. doi: http://doi.org/10.1186/1746-1340-13-6

Ipser, J. C., Terburg, D., Syal, S., Phillips, N., Solms, M., Panksepp, J., Malcolm-Smith, S., Thomas, K., Stein, D. J., & van Honk, J. (2013). Reduced fear-recognition sensitivity following acute buprenorphine administration in healthy volunteers. *Psychoneuroendocrinology, 38*(1), 166-170. doi: https://doi.org/10.1016/j.psyneuen.2012.05.002

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *NeuroImage, 24*(3), 771-779. doi: https://doi.org/10.1016/j.neuroimage.2004.09.006

Jauniaux, J., Khatibi, A., Rainville, P., & Jackson, P. L. (2019). A meta-analysis of neuroimaging studies on pain empathy: investigating the role of visual information and observers' perspective. *Social Cognitive and Affective Neuroscience, 14*(8), 789-813. doi: https://doi.org/10.1093/scan/nsz055

Kappesser, J., & Williams, A. C. d. C. (2002). Pain and negative emotions in the face: judgements by health care professionals. *Pain, 99*(1), 197-206. doi: https://doi.org/10.1016/S0304-3959(02)00101-X

Karjalainen, T., Karlsson, H. K., Lahnakoski, J. M., Glerean, E., Nuutila, P., Jääskeläinen, I. P., Hari, R., Sams, M., & Nummenmaa, L. (2017). Dissociable roles of cerebral μ-opioid and type 2 dopamine receptors in vicarious pain: a combined PET–fMRI study. *Cerebral Cortex, 27*(8), 4257-4266. doi: https://doi.org/10.1093/cercor/bhx129

Kavaliers, M., Ossenkopp, K.-P., & Choleris, E. (2019). Social neuroscience of disgust. *Genes, Brain and Behavior, 18*(1), e12508. doi: https://doi.org/10.1111/gbb.12508

Kunz, M., & Lautenbacher, S. (2015). Improving recognition of pain by calling attention to its various faces. *European Journal of Pain, 19*(9), 1350-1361. doi: https://doi.org/10.1002/ejp.666

Kunz, M., Peter, J., Huster, S., & Lautenbacher, S. (2013). Pain and Disgust: The Facial Signaling of Two Aversive Bodily Experiences. *PLoS One, 8*(12), e83277. doi: https://doi.org/10.1371/journal.pone.0083277

Lamm, C., Bukowski, H., & Silani, G. (2016). From shared to distinct self-other representations in empathy: evidence from neurotypical function and socio-cognitive disorders. *Philos Trans R Soc Lond B Biol Sci, 371*(1686), 20150083. doi: http://doi.org/10.1098/rstb.2015.0083

Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage, 54*(3), 2492-2502. doi: https://doi.org/10.1016/j.neuroimage.2010.10.014

Lamm, C., & Majdandžić, J. (2015). The role of shared neural activations, mirror neurons, and morality in empathy – A critical comment. *Neuroscience Research, 90*, 15-24. doi: https://doi.org/10.1016/j.neures.2014.10.008

Lamm, C., Rütgen, M., & Wagner, I. C. (2019). Imaging empathy and prosocial emotions. *Neuroscience Letters, 693*, 49-53. doi: https://doi.org/10.1016/j.neulet.2017.06.054

Mars, R. B., Sallet, J., Schüffelgen, U., Jbabdi, S., Toni, I., & Rushworth, M. F. S. (2011). Connectivity-Based Subdivisions of the Human Right "Temporoparietal Junction Area": Evidence for Different Areas Participating in Different Cortical Networks. *Cerebral Cortex, 22*(8), 1894-1903. doi: http://doi.org/10.1093/cercor/bhr268

Meier, I. M., Bos, P. A., Hamilton, K., Stein, D. J., van Honk, J., & Malcolm-Smith, S. (2016). Naltrexone increases negatively-valenced facial responses to happy faces in female participants. *Psychoneuroendocrinology, 74*, 65-68. doi: https://doi.org/10.1016/j.psyneuen.2016.08.022

Merskey, H., & Bogduk, N. (1994). *Classification of chronic pain: Descriptions of chronic pain syndromes and definitions of pain terms*. Seattle: IASP Press.

Mischkowski, D., Crocker, J., & Way, B. M. (2016). From painkiller to empathy killer: acetaminophen (paracetamol) reduces empathy for pain. *Social Cognitive and Affective Neuroscience, 11*(9), 1345-1353. doi: https://doi.org/10.1093/scan/nsw057

Moriguchi, Y., Ohnishi, T., Lane, R. D., Maeda, M., Mori, T., Nemoto, K., Matsuda, H., & Komaki, G. (2006). Impaired self-awareness and theory of mind: An fMRI study of mentalizing in alexithymia. *NeuroImage, 32*(3), 1472-1482. doi: https://doi.org/10.1016/j.neuroimage.2006.04.186

Petrovic, P., Kalso, E., Petersson, K. M., & Ingvar, M. (2002). Placebo and opioid analgesia--imaging a shared neuronal network. *Science, 295*(5560), 1737-1740. doi: http://doi.org/10.1126/science.1067176

Peyron, R., Laurent, B., & García-Larrea, L. (2000). Functional imaging of brain responses to pain. A review and meta-analysis (2000). *Neurophysiologie Clinique/Clinical Neurophysiology, 30*(5), 263-288. doi: https://doi.org/10.1016/S0987-7053(00)00227-6

Phillips, M. L., Young, A. W., Scott, S. K., Calder, A. J., Andrew, C., Giampietro, V., Williams, S. C. R., Bullmore, E. T., Brammer, M., & Gray, J. A. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 265*(1408), 1809-1817. doi: http://doi.org/doi:10.1098/rspb.1998.0506

Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., Bullmore, E. T., Perrett, D. I., Rowland, D., Williams, S. C. R., Gray, J. A., & David, A. S. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature, 389*(6650), 495-498. doi: http://doi.org/10.1038/39051

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

Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 25*(1), 1-20. doi: http://doi.org/10.1017/S0140525X02000018

Ramos, R. A., Ferguson, C. J., Frailing, K., & Romero-Ramirez, M. (2013). Comfortably numb or just yet another movie? Media violence exposure does not reduce viewer empathy for victims of real violence among primarily Hispanic viewers. *Psychology of Popular Media Culture, 2*(1), 2-10. doi: http://doi.org/10.1037/a0030119

Rütgen, M., Seidel, E.-M., Riečanský, I., & Lamm, C. (2015a). Reduction of Empathy for Pain by Placebo Analgesia Suggests Functional Equivalence of Empathy and First-Hand Emotion Experience. *The Journal of Neuroscience, 35*(23), 8938-8947. doi: http://doi.org/10.1523/jneurosci.3936-14.2015

Rütgen, M., Seidel, E. M., Pletti, C., Riecansky, I., Gartus, A., Eisenegger, C., & Lamm, C. (2018). Psychopharmacological modulation of event-related potentials suggests that first-hand pain and empathy for pain rely on similar opioidergic processes. *Neuropsychologia, 116*(Pt A), 5-14. doi: https://doi.org/10.1016/j.neuropsychologia.2017.04.023

Rütgen, M., Seidel, E. M., Silani, G., Riecansky, I., Hummer, A., Windischberger, C., Petrovic, P., & Lamm, C. (2015b). Placebo analgesia and its opioidergic regulation suggest that empathy for pain is grounded in self pain. *Proceedings of the National Academy of Sciences, 112*(41), E5638-E5646. doi: https://doi.org/10.1073/pnas.1511269112

Rütgen, M., Wirth, E.-M., Riečanský, I., Hummer, A., Windischberger, C., Petrovic, P., Silani, G., & Lamm, C. (2021). Beyond Sharing Unpleasant Affect—Evidence for Pain-Specific Opioidergic

Modulation of Empathy for Pain. *Cerebral Cortex, 31*(6), 2773-2786. doi: http://doi.org/10.1093/cercor/bhaa385

Schirmer, A., & Adolphs, R. (2017). Emotion Perception from Face, Voice, and Touch: Comparisons and Convergence. *Trends in cognitive sciences, 21*(3), 216-228. doi: http://doi.org/10.1016/j.tics.2017.01.001

Schmidt, A., Borgwardt, S., Gerber, H., Wiesbeck, G. A., Schmid, O., Riecher-Rössler, A., Smieskova, R., Lang, U. E., & Walter, M. (2014). Acute Effects of Heroin on Negative Emotional Processing: Relation of Amygdala Activity and Stress-Related Responses. *Biological Psychiatry, 76*(4), 289-296. doi: https://doi.org/10.1016/j.biopsych.2013.10.019

Schutte, N. S., Malouff, J. M., Bobik, C., Coston, T. D., Greeson, C., Jedlicka, C., Rhodes, E., & Wendorf, G. (2001). Emotional Intelligence and Interpersonal Relations. *The Journal of Social Psychology, 141*(4), 523-536. doi: http://doi.org/10.1080/00224540109600569

Sharvit, G., Corradi-Dell'Acqua, C., & Vuilleumier, P. (2018). Modality-specific effects of aversive expectancy in the anterior insula and medial prefrontal cortex. *Pain, 159*(8), 1529-1542. doi: http://doi.org/10.1097/j.pain.0000000000001237

Sharvit, G., Vuilleumier, P., Delplanque, S., & Corradi-Dell'Acqua, C. (2015). Cross-modal and modality-specific expectancy effects between pain and disgust. *Scientific Reports, 5*(1), 17487. doi: http://doi.org/10.1038/srep17487

Silani, G., Lamm, C., Ruff, C. C., & Singer, T. (2013). Right Supramarginal Gyrus Is Crucial to Overcome Emotional Egocentricity Bias in Social Judgments. *The Journal of Neuroscience, 33*(39), 15466-15476. doi: http://doi.org/10.1523/jneurosci.1488-13.2013

Simon, D., Craig, K. D., Gosselin, F., Belin, P., & Rainville, P. (2008). Recognition and discrimination of prototypical dynamic expressions of pain and emotions. *PAIN®, 135*(1), 55-64. doi: https://doi.org/10.1016/j.pain.2007.05.008

Singer, T., Seymour, B., O'doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303*(5661), 1157-1162. doi: https://doi.org/10.1126/science.1093535

Srinivasan, N., & Hanif, A. (2010). Global-happy and local-sad: Perceptual processing affects emotion identification. *Cognition and Emotion, 24*(6), 1062-1069. doi: http://doi.org/10.1080/02699930903101103

Steinbeis, N., Bernhardt, B. C., & Singer, T. (2015). Age-related differences in function and structure of rSMG and reduced functional connectivity with DLPFC explains heightened emotional egocentricity bias in childhood. *Social Cognitive and Affective Neuroscience, 10*(2), 302-310. doi: https://doi.org/10.1093/scan/nsu057

Wager, T. D., Rilling, J. K., Smith, E. E., Sokolik, A., Casey, K. L., Davidson, R. J., Kosslyn, S. M., Rose, R. M., & Cohen, J. D. (2004). Placebo-Induced Changes in fMRI in the Anticipation and Experience of Pain. *Science, 303*(5661), 1162-1167. doi: http://doi.org/10.1126/science.1093065

Wager, T. D., Scott, D. J., & Zubieta, J.-K. (2007). Placebo effects on human µ-opioid activity during pain. *Proceedings of the National Academy of Sciences, 104*(26), 11056-11061. doi: http://doi.org/10.1073/pnas.0702413104

Wardle, M. C., Bershad, A. K., & de Wit, H. (2016). Naltrexone alters the processing of social and emotional stimuli in healthy adults. *Social neuroscience, 11*(6), 579-591. doi: https://doi.org/10.1080/17470919.2015.1136355

Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V., & Rizzolatti, G. (2003). Both of Us Disgusted in My Insula: The Common Neural Basis of Seeing and Feeling Disgust. *Neuron, 40*(3), 655-664. doi: https://doi.org/10.1016/S0896-6273(03)00679-2

Xiong, R.-C., Fu, X., Wu, L.-Z., Zhang, C.-H., Wu, H.-X., Shi, Y., & Wu, W. (2019). Brain pathways of pain empathy activated by pained facial expressions: a meta-analysis of fMRI using the activation likelihood estimation method. *Neural regeneration research, 14*(1), 172-178. doi: http://doi.org/10.4103/1673-5374.243722

Zaki, J., Wager, T. D., Singer, T., Keysers, C., & Gazzola, V. (2016). The Anatomy of Suffering: Understanding the Relationship between Nociceptive and Empathic Pain. *Trends in cognitive sciences, 20*(4), 249-259. doi: https://doi.org/10.1016/j.tics.2016.02.003

Zhao, Y., Rütgen, M., Zhang, L., & Lamm, C. (2021). Pharmacological fMRI provides evidence for opioidergic modulation of discrimination of facial pain expressions. *Psychophysiology, 58*(2), e13717. doi: https://doi.org/10.1111/psyp.13717

Zhou, F., Li, J., Zhao, W., Xu, L., Zheng, X., Fu, M., Yao, S., Kendrick, K. M., Wager, T. D., & Becker, B. (2020). Empathic pain evoked by sensory and emotional-communicative cues share common and process-specific neural representations. *eLife, 9*, e56929. doi: http://doi.org/10.7554/eLife.56929

Zubieta, J.-K., Bueller, J. A., Jackson, L. R., Scott, D. J., Xu, Y., Koeppe, R. A., Nichols, T. E., & Stohler, C. S. (2005). Placebo Effects Mediated by Endogenous Opioid Activity on µ-Opioid Receptors. *The Journal of Neuroscience, 25*(34), 7754-7762. doi: http://doi.org/10.1523/jneurosci.0439-05.2005

Zubieta, J.-K., Smith, Y. R., Bueller, J. A., Xu, Y., Kilbourn, M. R., Jewett, D. M., Meyer, C. R., Koeppe, R. A., & Stohler, C. S. (2001). Regional mu opioid receptor regulation of sensory and affective dimensions of pain. *Science, 293*(5528), 311-315. doi: 10.1126/science.1060952

This manuscript is published in *Psychophysiology*

Chapter 2 - Pharmacological fMRI provides evidence for opioidergic modulation of discrimination of facial pain expressions

Yili Zhao[1], Markus Rütgen[1,2], Lei Zhang[1,3], Claus Lamm[1,2,3*]

[1] Social, Cognitive and Affective Neuroscience Unit, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria

[2] Vienna Cognitive Science Hub, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria

[3] Neuropsychopharmacology and Biopsychology Unit, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria

DOI: http://doi.org/10.1111/psyp.13717

Abstract

The endogenous opioid system is strongly involved in the modulation of pain. However, the potential role of this system in perceiving painful facial expressions from others has not been sufficiently explored as of yet. To elucidate the contribution of the opioid system to the perception of painful facial expressions, we conducted a double-blind, within-subjects pharmacological functional magnetic resonance imaging (fMRI) study, in which 42 participants engaged in an emotion discrimination task (pain vs. disgust expressions) in two experimental sessions, receiving either the opioid receptor antagonist naltrexone or an inert substance (placebo). On the behavioral level, participants less frequently judged an expression as pain under naltrexone as compared to placebo. On the neural level, parametric modulation of activation in the (putative) right fusiform face area (FFA), which was correlated with increased pain intensity, was higher under naltrexone than placebo. Regression analyses revealed that brain activity in the right FFA significantly predicted behavioral performance in disambiguating pain from disgust, both under naltrexone and placebo. These findings suggest that reducing opioid system activity decreased participants' sensitivity for facial expressions of pain, and that this was linked to possibly compensatory engagement of processes related to visual perception, rather than to higher-level affective processes, and pain regulation.

Keywords

Emotion discrimination; facial expression; pain; opioid system; right fusiform face area; naltrexone; fMRI; psychopharmacology

1 Introduction

The ability to perceive pain in others is fundamental in social interaction, as it strengthens social connections and promotes care for others' well-being. Studies have shown that when observing others in pain, brain regions involved in the processing of self-directed pain were also activated, but this mainly included parts of the affective-motivational component of pain processing, such as the anterior mid-cingulate cortex (aMCC) and the anterior insula (AI) (Singer et al., 2004; Botvinick et al., 2005; Lamm et al., 2011). Recent meta-analytic research has revealed additional activations shared by empathy for pain and self-direct pain in the inferior frontal gyrus and supramarginal gyri (Fallon et al., 2020). Besides, another meta-analysis research demonstrated that the core neural empathy network (aMCC and AI) was activated when observing others in painful states as well as in nonpain negative affective states (Timmers et al., 2018). Pain is commonly conveyed to others via facial expressions. In those studies where visual perception of other's facial expressions of pain was critical to recognize their emotional state, the visual cortex, especially the fusiform face area, has also been repeatedly found as activated (Botvinick et al., 2005; Simon et al., 2006; Lamm et al., 2007; Decety et al., 2013). Recent meta-analyses have reported involvement of the fusiform gyrus (along with aMCC and AI) in empathy paradigms employing facial expressions (Jauniaux et al., 2019; Xiong et al., 2019).

At the neurochemical level, the endogenous opioid system has been recognized to play an essential role in the experience of pain. Abundant studies have demonstrated that the opioid system is involved in the modulation of self-pain experience in both sensory and affective states (Singer et al., 2004; Botvinick et al., 2005; Lamm et al., 2011). Recently, research has reported that opioid antagonists (e.g., naltrexone) and agonists (e.g., buprenorphine) could alter individuals' sensitivity or responses to facial expressions of emotions such as anger, fear, happiness, or sadness, in others (Ipser et al., 2013; Meier et al., 2016; Wardle et al., 2016). However, the findings of these studies were rather inconsistent, suggesting that definite conclusions regarding the role of the opioid system in recognizing the emotions of others are still not possible.

Moreover, mechanistic evidence on whether and how the endogenous opioid system functions when perceiving others' pain is still lacking. Using positron emission tomography, Karjalainen et al. (2017) revealed the involvement of µ-opioid receptor in vicarious pain, and psychopharmacological and neuroimaging findings of our own lab (Rütgen et al., 2015a; Rütgen et al., 2015b; Rütgen et al., 2018) indicate a causal role of the opioid system in empathy for pain. Recent mediation analyses, though somewhat inconclusively, suggest that this may be explained by effects on emotion identification (Coll et al., 2017), while related research revealed that other neurochemical mechanisms might play a role in empathy for pain as well (Mischkowski et al., 2016). However, since none of these studies had a specific focus on the visual recognition and processing of others' facial pain expressions, it remains unclear whether there is a specific relationship between the opioid system and the perception of pain expressions.

To bridge this gap, we adopted an emotion discrimination paradigm (Young et al., 2002; Cook et al., 2013), to investigate whether the opioid system influences discrimination of morphed facial expressions between pain and another emotion (i.e., disgust in the present study). The reasons to morph disgust, instead of other emotions, with pain were twofold: 1) Pain and disgust facial expressions share some similarities but are still distinct enough to be distinguished from each other (Kunz et al., 2013; Sharvit et al., 2015); 2) According to a recent review (Nummenmaa & Tuominen, 2018), the endogenous opioid system is engaged in modulating a wide range of basic emotions (e.g., anger, fear, sadness, and pleasure), while there is only scarce evidence for the potential involvement of the opioid system in the modulation of disgust (which appears to be rather susceptible to other types of modulation, such as the oxytocin system; see Kavaliers et al., 2019). Disgust for these two main reasons thus appeared to be a reasonable choice, especially in comparison to other emotions. However, when designing the study and when discussing the results, we were aware that absence of evidence does not imply evidence of absence of effects on disgust.

Specifically, in a pharmaco-fMRI study, we applied an emotion discrimination task to examine whether administration of the opioid antagonist naltrexone influenced how painful facial expressions were discriminated from disgust expressions, and to which brain areas this was associated. In terms of our initial research interest, the focus was on the core empathy affective regions (i.e., aMCC and bilateral AI) and regions of interests (ROIs) were determined accordingly (based on the meta-analysis of Lamm et al., 2011; see also Rütgen et al., 2015). Besides, since results from the parametric modulation analysis showed significant activity in the FFA (especially in the right hemisphere), we identified this region as a fourth (post-hoc exploratory) ROI. Apart from FFA, the superior temporal sulcus (STS) is another region that was frequently reported in studies of facial expression processing (Narumoto et al., 2001; Engell & Haxby, 2007; Wegrzyn et al., 2015).However, we did not find significant parametric modulation of activation in the STS. Studies have shown that the STS is preferentially engaged during the processing of dynamic facial expressions; the FFA, on the other hand, is preferentially engaged during the processing of static facial expressions (Pitcher et al., 2011; De Winter et al., 2015). Given that only static facial expressions were used in the present study, this may be the reason why we only saw parametric modulation in FFA, which is why we focused on this area rather than STS. Based on the rather controversial previous findings, our hypothesis regarding behavioral discrimination was two-sided, while our analyses of the neural underpinnings focused on areas related to the discrimination of pain expressions, including the fusiform face area (FFA), the anterior mid-cingulate cortex (aMCC), and the anterior insular cortex (AI).

2 Method

2.1 Participants

Fifty-two participants (30 females; age: 24.06 ± 3.39 years) were recruited through online advertisements. Exclusion criteria were left-handedness and any history or presence of neurological and psychiatric disorders. In addition to the online MRI safety-check questionnaire, all participants were screened by a physician at the Faculty of Psychology, University of Vienna, including a medical history check, and basic physical examination. Nine participants who did not show up for the second session were excluded. One further exclusion was due to a consistent failure to discriminate pain and disgust expressions. The final dataset thus consisted of 42 participants (24 females; age: 24.12 ± 3.50 years). The study was approved by the ethics committee of the Medical University of Vienna and was conducted in line with the latest version of the Declaration of Helsinki (2013) of the World Medical Association (https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/). All participants provided written consent to participate.

2.2 Paradigm

While their brains were being scanned using fMRI, participants were engaged in a modified version of the emotion discrimination task (Cook et al., 2013), in which facial expressions were morphed along a continuum between pain and disgust from 20% to 80% in 10% steps. In the original version of Cook et al., a surprise-fear continuum was derived from the same person, comprising seven images morphing between surprise and fear while holding the face identity constant. Here, we adopted a pain-disgust continuum using pictures of pain and disgust expressions from the same person. In each trial, after 1500 ms fixation, a facial expression was presented on the screen, which consisted of a morph between two original images showing pain and disgust (e.g., 50% pain and 50% disgust) for 800 ms. The stimulus presentation time was chosen according to the one of the original task paradigm by Cook and colleagues (2013). Following each stimulus, a prompt asking “pain or disgust?” was presented on the screen, and participants were required to judge whether the previous expression was showing pain, or disgust (see Fig. 1). Studies have found that the gender of targets affected observer’s speed and accuracy as well as neural responses when detecting facial expressions of pain (Simon et al., 2006; Riva et al., 2011), which seemed to suggest a potential difference in the mechanism underlying processing pain expressions of male and female targets. Consequently, we decided to use faces from one gender (female here) to minimize this potential confound of the behavioral and neural responses.

The original facial expressions of pain and disgust had been extracted from the Montreal Pain and Affective Face Clips (Simon et al., 2008), using images extracted from two females' clips. Morphs were adopted and generated with FantaMorph 5 (Deluxe edition; http://www.fantamorph.com/), with each individual morph image using pain and disgust pictures from the same female.

2.3 Procedure

Participants were invited to visit the lab twice to take part in two functional magnetic resonance imaging (fMRI) sessions, separated by at least one week, to ensure complete drug washout (Bisaga et al., 2018). In each session, participants received either 50 mg naltrexone or an inert substance (i.e., placebo), in a double-blind fashion. Session order was counterbalanced, and pills were delivered with the cover story that they were "MRI signal enhancer pills", which was done in order to avoid that subjective beliefs about the study aims and opioid action would bias the results. However, participants were fully informed about the possible effects in the consent form, including possible side effects of naltrexone as part of the consent form. All participants signed the consent form as an agreement of participation after they completely understood and already evaluated any possible risk or adverse effects regarding the naltrexone administration.

Experimental sessions took place at the University of Vienna MR Centre. Before administering the pill, participants were screened for drug consumption again, using a urinary drug test. Then, participants were instructed to orally take a pill, which either contained naltrexone or the inert substance (placebo). They were further required to wait for 45 min for the drug to take effect (KatzenPerez et al., 2001; Price et al., 2016). After the waiting time, participants experienced a cold pressor test (CPT) in which they were asked to immerse one of their hands into cold water (1 ~ 5 °C) as long as they could. The CPT procedure could promote endogenous opiate activation (Jungkunz et al., 1983; Washington et al., 2000; Robertson et al., 2008). From an experimental design perspective, we feared that the effect of the opioid blockade would be negligible at a baseline level. By applying a cold pressor test (CPT), we sought to induce endorphin release (Casale et al., 1985; King et al., 2013) and, consequently, a greater difference between placebo and naltrexone sessions. Afterwards, participants were led into the scanner room and first underwent an empathy for pain task (Rütgen et al., 2015b), whose findings are outside the scope of the present paper, and then the emotion discrimination task. In the emotion discrimination task, participants were required to judge 140 morphed facial expressions presented in a pseudorandom order on whether it was pain or disgust, by pressing either the left or right button on the MRI-compatible button box. The left button represented the choice of pain expression, and the right button represented the choice of disgust expression. This button assignment was kept identical to the original version. Participants were instructed to respond as accurately and fast as possible. Though there was no time limit in the judgement phase, trials whose reaction time was longer than 4s were regarded as invalid, and the ratio of invalid trials was taken into account during data analysis. According to the histogram of reaction times (RT) we plotted, the bulk of RTs were below 4 s, and the RTs above 4 s were rather variable (RT range: 4.06 ~ 24.35 s). To reduce unsystematic RT variations, we excluded RTs above 4 s (3.5‰ of all data). No difference of trial numbers between sessions was found after excluding outliers, t_{41} = 1.07, p = .293. The jitter between trials varied at random between 3000 to 5000 ms (see Fig. 1). Following this run, an anatomical scan was performed. After scanning, participants reported on 51 potential side-effects of naltrexone in a binary fashion (yes/no).

Participants were scheduled for the second fMRI session at about the same time of day, but at minimum one week later. The procedure of the second session was the same as the first one, except that participants who received naltrexone in the first session were given the placebo in the second session and *vice versa*. At the end of the second session, participants were debriefed and received 90 EUR overall for their participation.

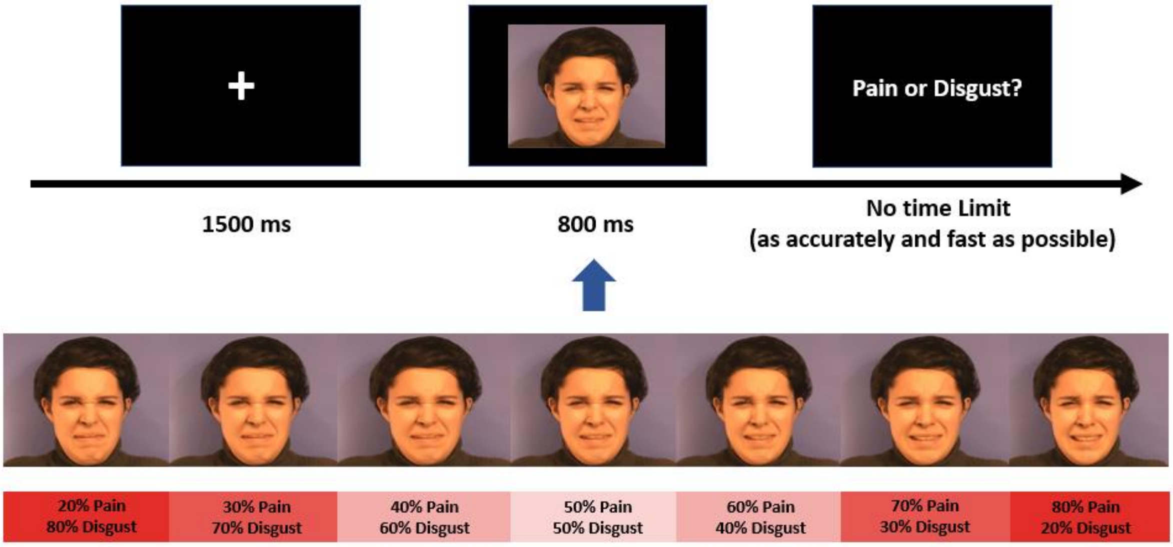

Fig. 1. **Experimental procedure.** Upper panel: Following a fixation cross (displayed for 1500 ms), participants were presented with a morphed facial expression image selected from one out of seven options differing in the composition of pain and disgust intensities (e.g., 50% pain and 50% disgust, as shown here) lasting for 800 ms. After each stimulus, participants needed to judge whether the presented expression was pain or disgust. Lower panel: The facial expressions continuum from one of the two female characters. The seven stimuli are shown here for illustration purposes, participants only saw the images in the upper row, at full-screen size.

2.4 Behavioral analysis

We examined whether naltrexone affected participants' discrimination performance of pain expressions by comparing the fitted trends of participants' pain choices at each pain intensity. According to previous studies, the relationship between categorical perception and the morphed stimuli could be fitted into a sigmoid function (McKone et al., 2001; McCullough & Emmorey, 2009; Granato et al., 2012). Following this procedure, we fitted a sigmoid model to each participant's pain choices (proportion of answering "pain") at seven intensities, and then extracted the fitted parameters: the slope of the sigmoid curve and the point of subjective equivalence (PSE) at which participants equally chose pain or disgust (Cook et al., 2013) using the Palamedes toolbox (http://www.palamedestoolbox.org/). Two-tailed paired *t*-tests were conducted to test drug effects in slope and PSE values.

We further tested whether, on average and on any pain intensity, there were drug effects on the proportion of pain choices across the seven intensities regardless of the slope. A linear mixed effect (LME) model (M1) with drug (Naltrexone vs. Placebo), pain intensity (20% to 80%), and their interaction as fixed factors and subject identity as the random intercept was created. Subject identity here refers to the identifiers (i.e., subject ID) that were used to encode and discriminate different subjects, and they were applied as random intercepts in LMEs. To set subject identity as random intercepts could effectively control the inter-subject variation merely related to sample selection itself instead of the experimental manipulation. This approach has more advantages than the traditional ANOVA method. As the interaction was not significant, it was removed from M1. To test whether there was an alternative model that better fitted the data, we estimated a second LME model (M2) with pain intensities as random slope and performed model comparison between M1 and M2. We did not include drug groups as another random slope because it is commonly not recommended to consider as random slopes when a factor only has two levels (Barr et al., 2013). In fact, this full model, pain choices ~ drug * pain intensity + (drug * pain intensity | subject identity) failed to converge. Thus, we compared M1 and M2, and results showed that M1 (AIC: 2641.7) better accounted for the data than M2 (AIC: 2650.4; $\chi^2 = 45.33$, $p = .015$). Therefore, the final LME model we chose included the two main effects of drug effect and pain intensity as fixed factors and subject identity as a random intercept, without any random slope.

Two additional linear mixed effect models were constructed to test whether brain activation in the visual region of interest (ROI, see below) could predict behavioral responses, and whether this differed between the naltrexone and the placebo session. For each session, a model with the proportion of pain choices at each pain intensity as the dependent variable, percent signal change (PSC) of the seven pain intensities in the visual ROI in the corresponding session as the fixed factor, and subject identity as a random intercept was set up. Fisher's *z* transformation was performed on the coefficient of determination (R^2) of the two sessions. Based on the transformed *z* scores, a one-sample *t*-test was conducted to assess drug effects. Statistical significance was calculated with Satterthwaite approximation for degrees of freedom and set as $p < .05$.

Lastly, we tested for differences in reported side-effects between sessions. A two-tailed paired *t*-test was performed using SPSS version 25 (IBM Corp, Armonk, NY, USA) on the sum score of all items. As nausea, one of the known potential side-effects of naltrexone, may likely interfere with the processing of disgust, we additionally conducted a two-tailed Fisher's exact test for this specific item.

2.5 MRI acquisition and data preprocessing

MRI data were acquired using a 3T Siemens Magnetom Skyra MRI scanner (Siemens, Erlangen, Germany) with a 32-channel head coil. Functional whole-brain scans were collected using a multiband accelerated T2*-weighted echoplanar imaging (EPI) sequence (32 slices, multiband acceleration factor = 4, TR = 704 ms, TE = 34 ms, flip angle = 50°, FOV = 210 × 210 mm, voxel size = 2.2 × 2.2 × 3.5 mm). Structural images were acquired with a magnetization-prepared rapid gradient-echo (MPRAGE) sequence (176 slices, TR = 2300ms, TE = 2.29 ms, flip angle = 8°, voxel size = 0.9 × 0.9 × 0.9 mm, FOV = 240 × 240 mm). Imaging data were preprocessed with Statistical Parametric Mapping (SPM12; Wellcome Trust Centre for Neuroimaging, London, UK, https://www.fil.ion.ucl.ac.uk/spm/software/spm12/).

Preprocessing included realignment to the first image of the first session for both sessions, co-registration to the T1 image, segmentation, normalization to MNI template space using Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL) toolbox (Ashburner, 2007), and smoothing with a 6 mm full width at half-maximum (FWHM) Gaussian kernel.

In order to improve data quality, functional scans were individually scrubbed with the frame-wise displacement (FD) over 0.5 mm (Power et al., 2012; Power et al., 2014). That is, we identified individual outlier scans and flagged the volume indices as nuisance regressors into the General Linear Model (GLM) of the first-level analysis.

Functional scans corresponding to trials whose reaction time was less than 100 ms or more than 4 s were identified as additional nuisance regressors (i.e., invalid trials; 7.1% of all trials, number of remaining trials (Mean ± *SD*) out of 140 trials: naltrexone session: 128.10 ± 15.97, placebo session: 132.05 ± 9.38, no significant difference between two sessions on average, $t_{41} = -1.693$, $p = .98$).

2.6 First-level analysis

Two design matrices were created. First, we aimed to ensure that our task widely activated the brain network underlying perception of facial expressions as a task manipulation check (GLM1). The following regressors were entered in the model for both sessions: picture onsets of valid trials, picture onsets of invalid trials (invalid as defined above, i.e., those trials whose reaction time < 100 ms or > 4 s; if any), judgment onsets of valid trials, and judgment onsets of invalid trials (if any). Six head motion parameters and the scrubbing regressors (FD > 0.5 mm; see above) were further entered as nuisance regressors. A contrast (pictures > baseline, across naltrexone and placebo sessions) was created out of our main interest; besides, a reversed contrast (baseline > pictures, across naltrexone and placebo sessions) was created as a comparison to the contrast of interest. Second, we sought to test wherein the brain showed parametric responses with pain intensity of and in the morphs (i.e., 20%, 30%, 40%, 50%, 60%, 70%, and 80%; GLM2). GLM2 included the following regressors: picture onsets of valid trials, pain intensity of valid trials (parametric modulator), picture onsets of invalid trials (if any), pain intensity of invalid trials (if any), judgment onsets of valid trials, and judgment onsets of invalid trials (if any). Six head motion parameters and the scrubbing regressors (FD > 0.5 mm; see above) were further entered as nuisance regressors. Note that in GLM2, we only focused on the picture onset regressor and the pain intensity parametric regressor; all the other regressors were included to account for variances of no interest. Although no jitter was implemented between imaging viewing and the judgement phase, no multicollinearity was observed between picture onsets and the parametric regressor of pain intensities (Spearman rank-order correlation: r = -.03, p = .77). Furthermore, the naltrexone and placebo sessions of each subject were entered separately into the same first-level GLMs. Contrasts (i.e., naltrexone: parametric modulator > baseline; placebo: parametric modulator > baseline) were generated to assess the effect of the increased pain intensities in each session against the implicit baseline.

2.7 Second-level analysis

On the group level, we implemented random-effects analyses across all subjects in SPM12. For GLM1, the threshold for the manipulation check was set at a whole-brain family-wise error (FWE) correction of $p < .05$, at the voxel level. For GLM2, we applied an initial threshold of $p < .001$ and FWE correction ($p < .05$) at the cluster level, which is a conventional threshold in general for fMRI analyses (Woo et al., 2014). The reason for applying an even stricter threshold for the former was that GLM1 was a manipulation check (i.e., pictures > baseline), in which very strong activation and thus higher effect sizes were expected. The extent threshold of GLM2 was determined by the SPM extension "cp_cluster_Pthresh.m" (https://goo.gl/kjVydz) with a cluster extent of $p < .05$ corrected for multiple comparisons across the whole brain. As a result, the cluster extent threshold for GLM2 was set at $k = 351$, with an initial selection threshold of $p < .001$.

2.8 Region of interest analysis

In addition to the whole-brain analyses, we further defined two types of Regions of Interest (ROIs) related to the discrimination of pain expressions, and then performed an ROI analysis. First, regions representing empathy for pain based on a meta-analysis (Lamm et al., 2011) were selected, including the anterior mid-cingulate cortex (aMCC, MNI peak: -2, 23, 40), the left anterior insula (lAI, MNI peak: -40, 22, 0), and the right anterior insula (rAI, MNI peak: 39, 23, -4). Spheres of 10 mm radius centered at each peak coordinate were created as ROI masks (Rütgen et al., 2015b). Second, regions from the significant activation of parametric modulators were identified: a significant cluster in the right visual association cortex (MNI peak: 18, -84, -12; Brodmann area 18, 19, and 37) was found to be positively correlated with pain intensity averaged across two sessions (i.e., naltrexone vs. placebo; contrast weight [0.5, 0.5]; see Fig. 5a and 5b below). This visual ROI was orthogonal to the analyses that were subsequently performed.

Drug effects (i.e., naltrexone > placebo) were checked in terms of the mean activation in those ROIs (aMCC, lAI, and rAI), regardless of pain intensity. We did not perform this analysis on the visual ROI as the definition of this ROI implicitly contained pain intensity information. The mean signal values of each ROI were extracted with the REX toolbox (Massachusetts Institute of Technology, Cambridge, MA, USA). Three two-tailed paired *t*-tests on drug effects were conducted. Additionally, we investigated whether any drug effects occurred in all four ROIs as the pain intensity increased (i.e., drug effects of the parametric regressors). We first tested if there was a significant ROI activation of the parametric regressors averaged across the two sessions (i.e., naltrexone vs. placebo: contrast weight [0.5, 0.5]) after small volume correction (SVC). To achieve an estimate of the brain activity regarding each specific experimental condition, percent signal change (PSC) values were estimated and applied in the following analyses (Gläscher, 2009). As only the visual ROI passed SVC, we respectively extracted PSC values for each individual in the visual ROI on all seven pain intensities for both sessions, using the rfxplot toolbox (Gläscher 2009; http://rfxplot.sourceforge.net/).

We then examined putative drug effects in the PSC values of each pain intensity with a linear mixed effect (LME) model. The model was performed using the lme4 package (v 1.1-21; https://cran.r-project.org/web/packages/lme4/index.html) in R. The full model included drug (Naltrexone and Placebo), pain intensity (20%, 30%, 40%, 50%, 60%, 70%, and 80%), and their interaction as fixed factors and subject identity as a random intercept. As the interaction term was not significant, we removed it from the model and report results from the model which only included the main effects.

Statistical significance was calculated with Satterthwaite approximation for degrees of freedom and set as $p < .05$.

3 Results

3.1 Behavioral Results

3.1.1 Slope and point of subjective equivalence (PSE) of Sigmoid function

In general, sigmoid functions showed good individual fit for both the naltrexone and the placebo sessions (see Fig. 2 for an example subject's fitted curve). Paired *t*-tests showed that there were no significant drug effects, neither in slope (t_{41} = .46, *p* = .65) nor in terms of the PSE values (t_{41} = 1.26, *p* = .22) of the fitted Sigmoid functions.

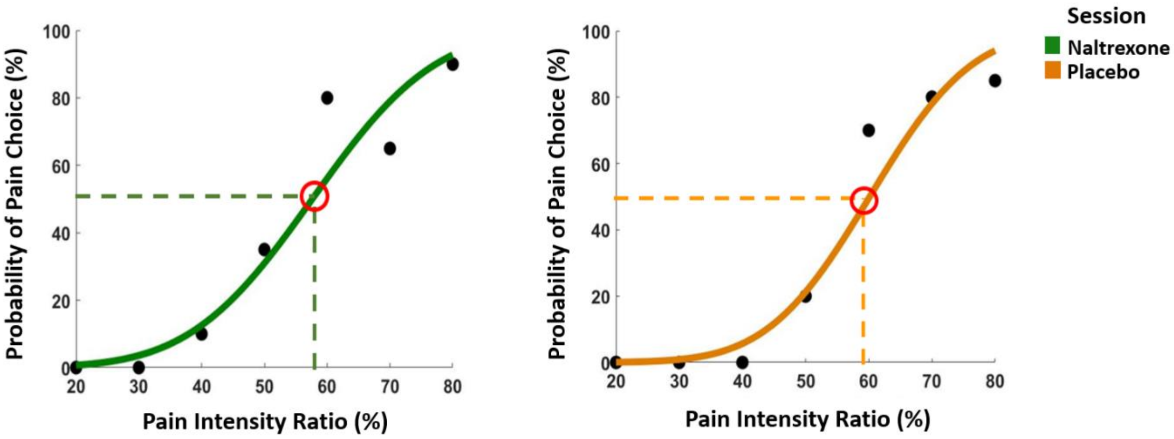

Fig. 2. **Example subject's Sigmoid function showing the relationship between pain intensity and pain response.** For the shown subject's data, there is no difference in either the point of subjective equivalence (PSE), or the slopes in the naltrexone session (left panel) and the placebo session (right panel). PSE is indicated by the red circle in each plot, and the average slope of the curve generally represents how accurate when participants judged an expression as pain or disgust. The steeper the slope, the better the performance.

3.1.2 Linear mixed effect (LME) model of pain choices

Results from the linear mixed effect model (main effects only model; see Methods) for drug and intensity showed that the main effect of drug ($F_{1,488}$ = 3.92, β = .47, *p* = .048) and intensity *(*$F_{6,489}$ = 497.41, the smallest β = .80*, p* < .0001) were both significant. The post-hoc Tukey test between different levels of pain intensities (across naltrexone and placebo sessions) showed that the majority of comparisons (except 30% vs. 20% pain, 70% vs. 60% pain, and 80% vs. 70% pain) were significant

(Supplementary Table 1). On average, participants made less frequent pain choices in the naltrexone session than in the placebo session, and generally (no interaction effect with drug) when pain intensity increased, participants showed an increasing probability of making a pain judgment (Fig. 3).

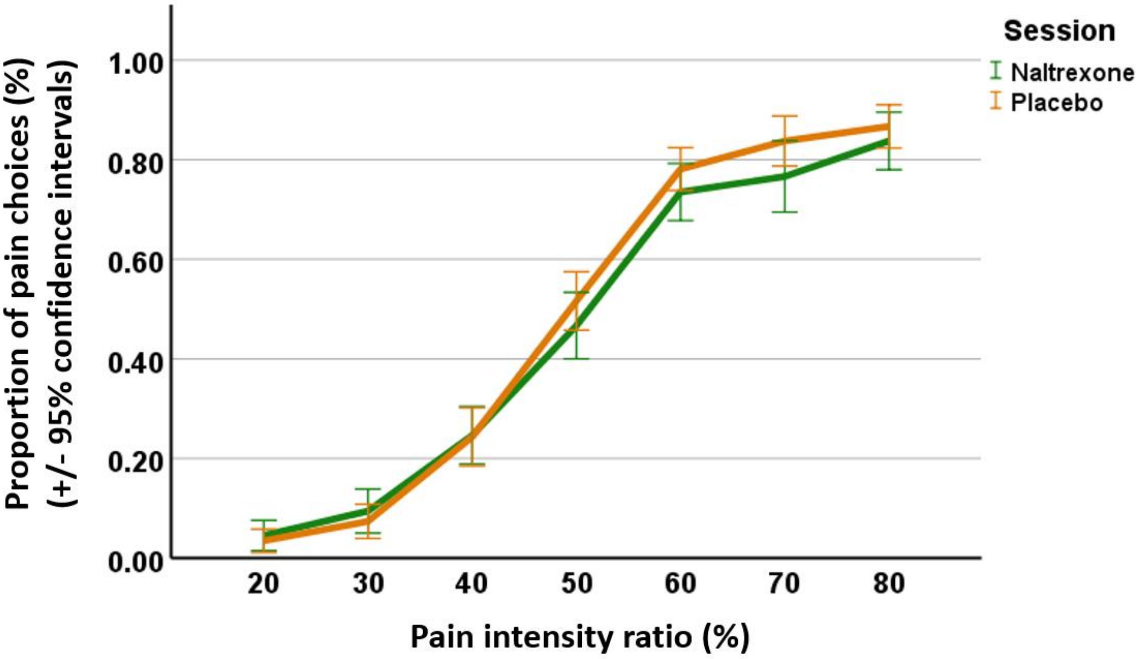

***Fig. 3*. Effects of drug and pain intensity on responses of pain expressions.** The proportion of judging an expression as pain at each pain intensity was illustrated for the naltrexone session (green) and the placebo session (orange). For the effect of drug, LME analysis indicated that the administration of naltrexone on average induced fewer pain choices compared with the placebo session. For the effect of pain intensity, in general, more pain choices were made as pain intensity increased. Error bars represent the 95% confidence interval.

3.1.3 Comparison of side effects

The two-tailed paired *t*-test on the sum score of potential side effects of naltrexone between sessions remained nonsignificant (t_{41} = 1.59, *p* = .12). The two-tailed Fisher’s exact test on differences in reported nausea between sessions was not significant either (*p* = 1.00). Specifically, only two participants reported nausea in the naltrexone session and one reported nausea in the placebo session.

3.2 Imaging results

3.2.1 Task manipulation check and parametric modulation

We performed two contrasts for the manipulation check: pictures > baseline, and the reverse contrast baseline > pictures. As expected, the brain network underlying perception of facial expressions was widely activated, including regions that we were interested in, namely, the fusiform face area, anterior insula, and anterior mid-cingulate cortex. See Fig. 4 for a graphical display of the two contrasts, and Table 1 for a summary of all findings.

Within the parametric modulation model, the contrast on naltrexone vs. placebo ([0.5, 0.5]) revealed a cluster in the right visual association cortex (MNI peak: 18, -84, -12), including what has been labeled in previous research as the fusiform face area (e.g., MNI local peak: 26, -54, -14); see Fig. 5a and 5b. In other words, this area showed a parametric increase of activation with increasing pain intensity, on average across both sessions.

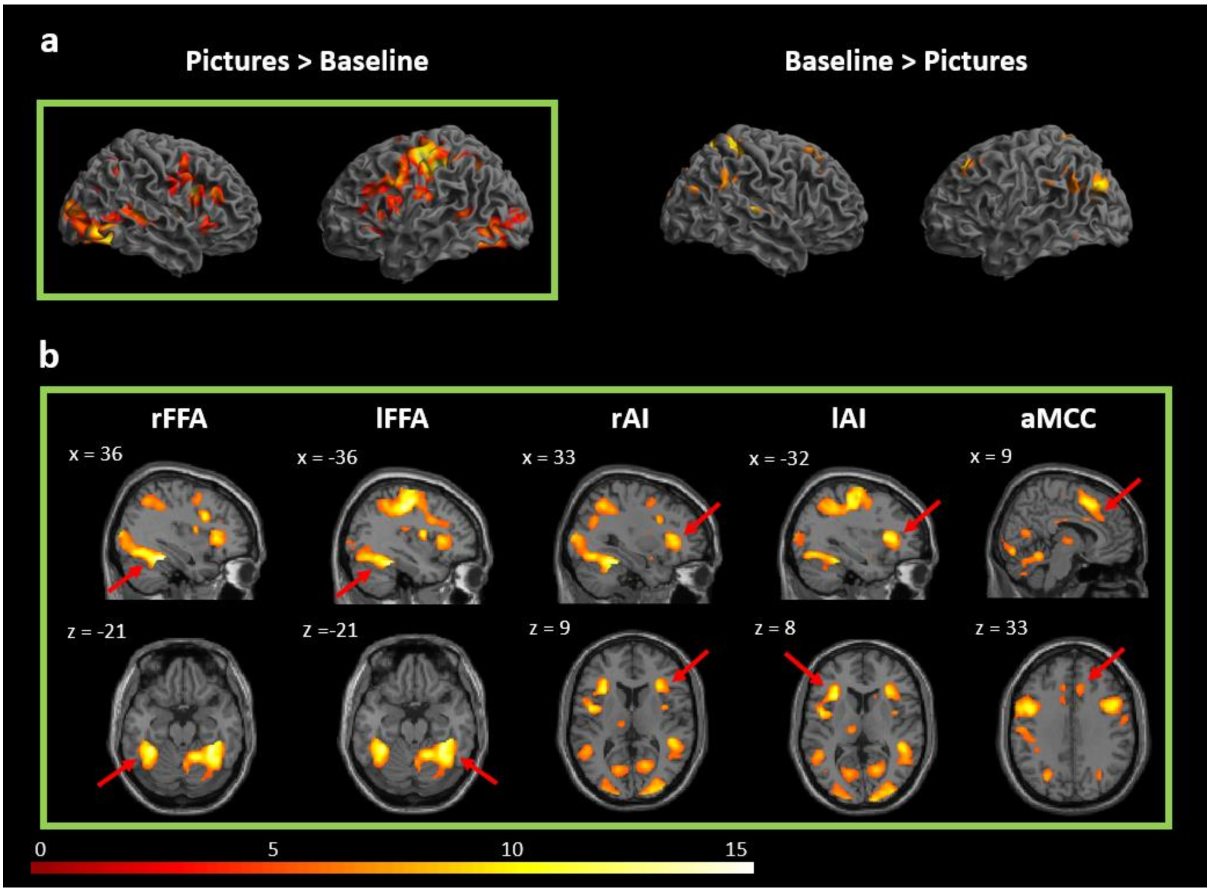

Fig. 4. **Neural correlates of the contrasts of pictures > baseline and baseline > pictures**. (a) In the upper panel, pictures > baseline mainly revealed activation in left postcentral gyrus, right fusiform gyrus, left inferior frontal gyrus, left supplementary motor area, right angular gyrus, and right calcarine sulcus (k > 1000); baseline > pictures, mainly activated right precuneus, left middle occipital cortex, left middle frontal cortex (k > 1000). More extensive and stronger activity was generally detected when comparing pictures vs. baseline than the reverse contrast. (b) In the lower panel, pictures > baseline showed that the paradigm led to significant activation in

regions that we were interested in. rFFA = right fusiform face area, lFFA = left fusiform face area, rAI = right anterior insula, lAI = left anterior insula, and aMCC = anterior mid-cingulate cortex. Thresholded at voxel-level FWE corrected *p* <.05.

Table 1. Results of the manipulation check for contrasts **pictures > baseline** and **baseline > pictures** (*p* < .05 voxel-level FWE-corrected), in *MNI* space. Comparing pictures vs. baseline revealed significantly stronger activation in right visual cortices. Besides, more brain areas and more voxels, in general, were activated with the contrast of pictures vs. baseline compared to baseline vs. pictures. These findings suggest that the task manipulation was successful. Region names were labeled with the AAL atlas. BA = Brodmann area, L = left hemisphere, R = right hemisphere.

Region label	**BA**	**Cluster size**	***x***	***y***	***z***	***t*-value**
Pictures > baseline						
Postcentral_L	1	18565	-38	-25	45	15.01
Fusiform_R	37	19239	36	-43	-21	14.08
Frontal_Inf_Oper_R	44	7770	44	11	24	13.38
Supp_Motor_Area_L	6	5379	-4	8	54	13.03
Thalamus_L	50	903	-12	-21	2	9.79
Angular_R	39	2238	33	-58	46	9.29
Calcarine_R	17	1051	16	-66	9	8.79
Insula_R	13	189	38	-1	12	8.39
Cingulum_Mid_R	24	123	6	6	30	8.32
Temporal_Mid_L	21	526	-48	-48	8	7.82
Calcarine_L	17	660	-14	-70	9	7.81
Lingual_R	30	204	16	-36	0	6.97
Amygdala_R	34	20	33	2	-18	6.77
Thalamus_R	50	202	8	-13	2	6.73
Cerebellum_L	18	15	-8	-75	-44	6.28
Paracentral_L	1	80	-4	-30	60	6.09
Cuneus_R	19	14	15	-66	34	5.36
Temporal_Sup_R	22	2	52	-9	-9	5.16
Baseline > pictures						
ParaHippocampal_L	36	830	-34	-39	-10	10.61
Precuneus_R	7	6835	4	-49	58	9.99
Temporal_Sup_R	41	423	59	-27	9	8.47
Occipital_Mid_L	39	1539	-40	-76	30	8.17
Cuneus_L	18	603	-2	-96	18	8.15

Frontal_Mid_L	8	1031	-22	24	46	7.99
Precuneus_L	23	822	-8	-54	9	7.9
Insula_R	13	362	42	-9	-4	7.23
Angular_R	39	589	57	-51	30	6.86
Frontal_Sup_R	8	712	26	30	50	6.79
Occipital_Mid_R	19	121	42	-76	26	6.63
Hippocampus_R	54	44	39	-24	-12	6.41
Frontal_Mid_L	6	78	-40	15	54	6.39
Frontal_Sup_Medial_L	10	378	-10	51	20	6.32
Frontal_Mid_R	9	25	29	51	39	6.19
Temporal_Inf_L	37	50	-58	-54	-14	6.13
Frontal_Sup_L	10	37	-24	63	26	6.12
Cerebelum_L	19	19	-30	-78	-39	5.66
Frontal_Sup_R	9	2	22	-10	69	5.36
SupraMarginal_R	40	9	48	-27	26	5.29
Occipital_Sup_R	19	18	24	-81	39	5.27

3.2.2 ROI results

We performed follow-up ROI analyses on the defined ROIs. First, three paired *t*-tests were performed on ROIs aMCC, lAI, and rAI to test whether there were significant drug effects on the means of seven pain intensity. None of the three regions showed significant difference between the naltrexone session and the placebo session (t_{41} = 0.10, - 0.003, and -0.20, respectively, all *p* values > .86). Next, SVC was performed for all four ROIs to investigate whether there was significant parametric activation across the two sessions. Results showed that only the visual ROI passed the SVC (*p* < .0001, cluster-level FWE-corrected) with the initial threshold of *p* < .001, uncorrected. Therefore, the following LME analyses were only performed with the visual ROI.

3.2.3 Linear mixed effect (LME) model of brain activation

Results showed that significant main effects of drug ($F_{1,484}$ = 4.57, β = -.02*,* *p* = .03) and intensity ($F_{6,486}$ = 3.48, the smallest β = .013*,* *p* = .002), see also Fig. 5c. As for drug, activation (percent signal change) was on average higher in the naltrexone than in the placebo session. As for intensity, a Tukey post-hoc showed that, when comparing higher pain intensity with lower pain intensity conditions, all *t* values were positive, and among these comparisons, significant higher PSC values were achieved in the comparisons of 70% vs. 20% pain and 70% vs. 40% pain (t_{485} = 3.89, *p* = .002; t_{488} = 3.14, *p* = .03), and trends in the same direction were observed in the comparisons of 70% vs. 30% pain and 80% vs.

20% pain (t_{487} = 2.85, p = .07; t_{485} = 2.80, p = .08). Overall, the results thus indicate that on average there was a higher activation in the visual ROI in the naltrexone session compared with the placebo session, and that the higher pain intensities in general, and irrespective of drug, showed stronger activation than the lower pain intensities.

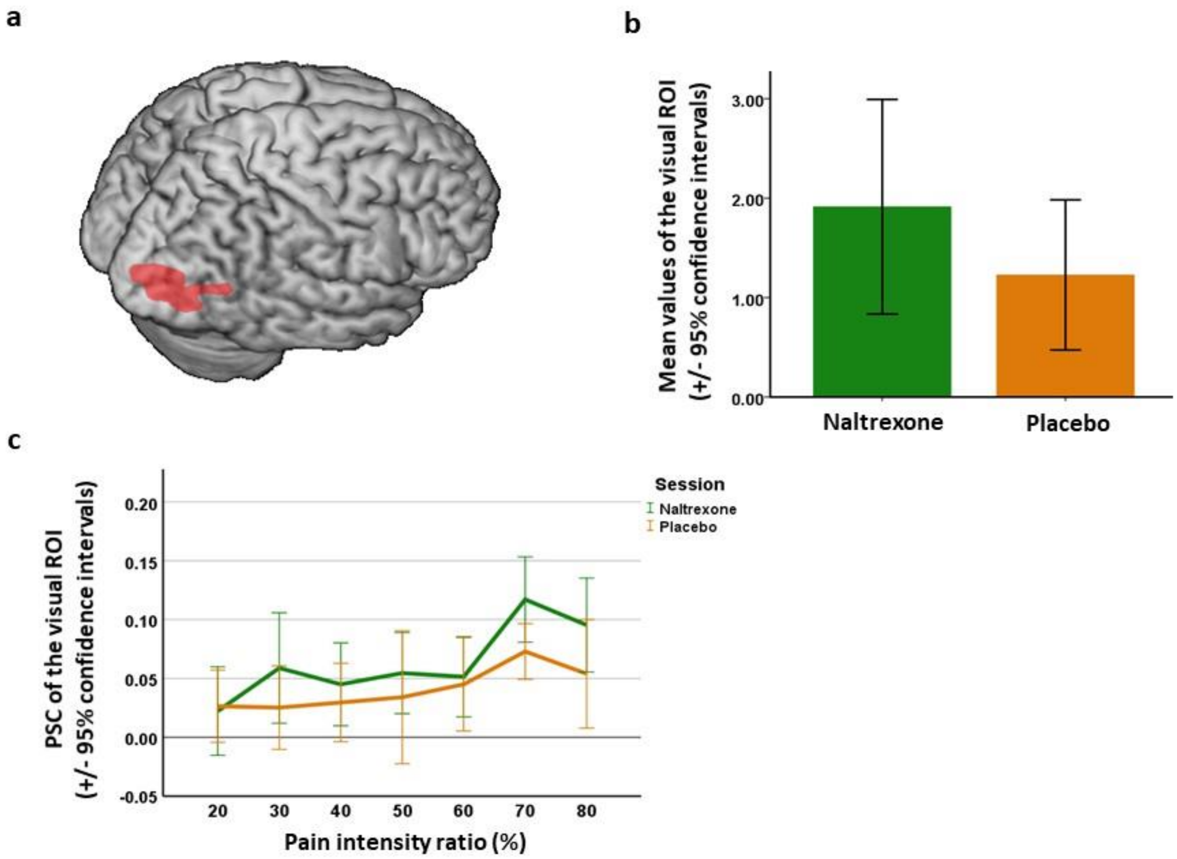

Fig. 5. **Neuroimaging findings of parametric activation in right visual cortex** (**a**) **Neural correlates of parametric effects of pain intensity averaged across sessions.** Activation was mainly localized to the right higher-order visual cortex (MNI peak: 18, -84, -12), and encompassed what has been referred to as the fusiform face area (FFA). This activity was observed under the threshold of p < .05 cluster-level FWE correction. **(b) Mean activations of the parametric modulators in the visual ROI in the naltrexone session and the placebo session**. Both sessions showed positive mean values of the parametric modulators in the visual ROI. This result indicates that as the intensity of expressed pain increased, activity in the ROI increases as well, in both sessions. **(c) Effects of drug and pain intensity on PSC in the visual ROI.** For the effect of drug, significant higher parametric modulation of the ROI activation on average was observed in the naltrexone session compared to the placebo session; for the effect of pain intensity, in general, high pain intensities (e.g., 60%, 70%, and 80%) showed increased activation in the visual ROI compared with low pain intensities (e.g., 20%, 30%, and 40%). Error bars show the 95% confidence interval.

3.3 Association between visual neural activity and pain choices

Two LME regression analyses were conducted to explore whether the neural activation in the visual ROI could predict behavioral responses in the naltrexone session and the placebo session, respectively. Results showed that in the naltrexone session, PSC in the visual ROI explained 4% of the variation in the proportion of pain choices, R^2 = .04, p = .001; and in the placebo session, PSC in the visual ROI explained 2% of the variation in the proportion of pain choices, R^2 = .02, p = .047. (Fig. 6). The one-sample *t*-test of Fisher's *z* scores showed there was no statistical difference in R^2 between the two sessions (p = .36). This implies that the visual ROI on average explained 3% of the variance of the behavioral choice, and that the naltrexone and placebo sessions did not differ in their brain-behavior predictions.

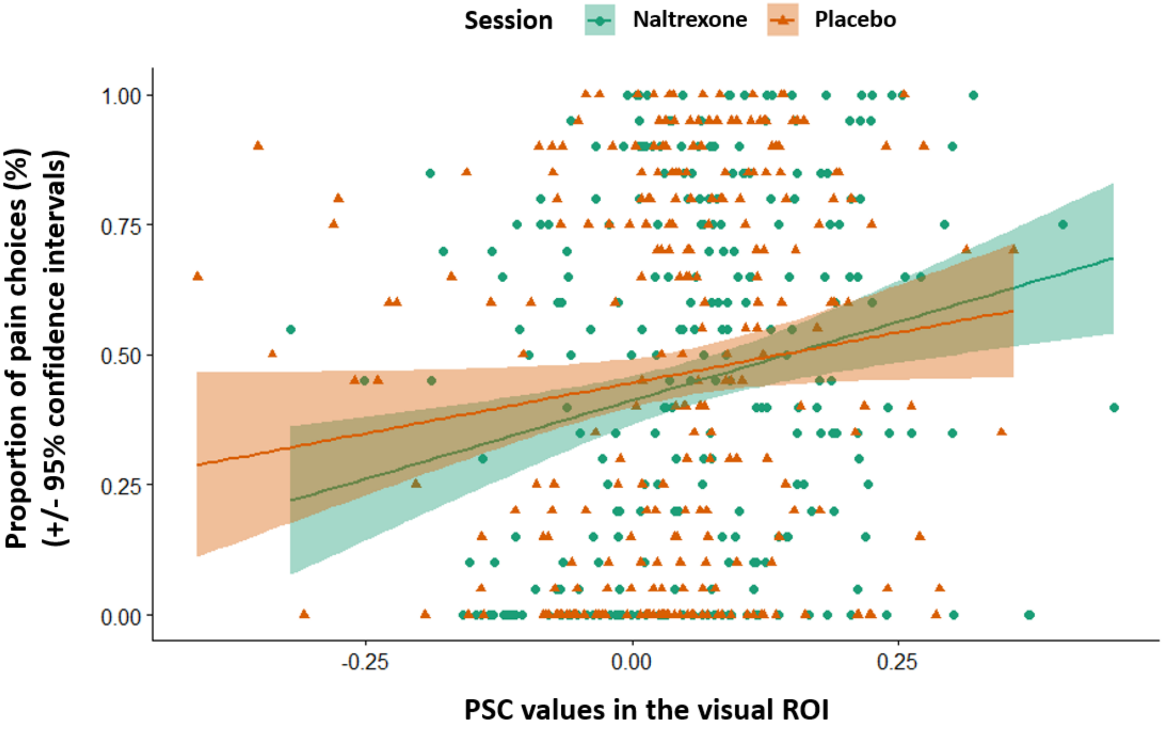

Fig. 6. **Association between brain and behavior.** LME regressions of PSC values in the visual ROI predicting the proportion of pain choices. Each data point represents an individual's proportion of pain choices and its corresponding PSC activation at a certain pain intensity in the naltrexone session (green) and the placebo session (orange). Shaded regions indicate the 95% confidence interval.

4 Discussion

The present study aimed to examine whether the opioid system influences the perception of pain expressions, using an emotion discrimination task. A double-blind cross-over within-subject design was applied, to investigate whether administration of the opioid antagonist naltrexone affected participants' judgments on whether a facial expression showed pain, and with what kind of neural activation this was associated. In brief, the results indicate that naltrexone decreased participants' discrimination on pain expressions, and more pain choices were generally made for higher pain intensities. The disturbance of the endogenous opioid system by naltrexone induced more activation in the right visual association cortex including FFA, and higher visual activity was detected with the high pain intensities (i.e., 70% and 80%) compared with the low pain intensities (i.e., 20%, 30%, and 40%).

We find that naltrexone decreased the probability for participants to judge an expression as pain. This finding provides evidence of the effects of the opioid antagonist on the discrimination of facial pain expression, suggesting a lower sensitivity to painful facial expressions resulting from the decrease in opioid system activity. Studies have demonstrated a reduction in seeking certain social cues under antagonism of the opioid system by naltrexone (Chelnokova et al., 2016; Wardle et al., 2016). The evolutionary meaning of pain is believed to work as a cue to aversive stimuli that constitute a potential threat to the individual (Kavaliers, 1988; Broom, 2001). In this respect, pain and in particular expressions that can be perceived by others also has an important social and communicative function, allowing the person in pain to signal that they are in need of help, apart from signaling potential dangers and threat to others. Thus, seeing others in pain is likely to induce empathy and concern, and to increase the intention of showing prosocial behaviors, including the provision of psychological comfort and concrete helping behaviors (Goubert et al., 2005; Hein et al., 2011; Masten et al., 2011; van der Meulen et al., 2016). In this respect, it is important to note that the opioid system has been linked to different facets of prosociality. For instance, release of endogenous opioids has been associated with social bonding, attachment, and empathy (Machin & Dunbar, 2011; Rütgen et al., 2015b; Nummenmaa & Karjalainen, 2018; Nummenmaa & Tuominen, 2018; Rütgen et al., 2018). Therefore, the decreases in discriminating others' facial expressions as showing pain under opioidergic blockade might suggest an attenuation in the sensitivity of pain expression perception, which ultimately may result in a decrease in the social-affective link between persons, and a corresponding reduction in prosociality. This needs to remain a speculation, and the current data are not conclusive in this respect, as activation in the insular and cingulate cortex, i.e., areas previously linked to empathy and prosocial concern, were not affected by the opioid antagonist.

The observation that higher-order visual cortex was the only area that showed effects of naltrexone thus poses the question what kind of processes were affected by the opioid system in the current study. The LME results show that stronger blood-oxygen-level-dependent (BOLD) signals of the parameter pain intensity in the visual ROI were detected in the naltrexone session compared with the placebo session, indicating a significant distinction in the visual activity specifically related to discriminating pain expressions. One possibility is that the increased visual activation under naltrexone might reflect a compensatory effect in coping with the reduced visual sensitivity to pain expressions. Previous studies have stated that the opioid system was engaged in maintaining visual perceptual processing, for example, visual attention (Dalley et al., 2005; Chelnokova et al., 2016). In this study, the normal visual perceptual processing was disturbed by naltrexone. As a consequence, the visual sensitivity to the facial expressions of pain was affected. In order to recover or compensate for the blunting of visual sensitivity, there may have been an increase in activity in the visual cortex (possibly of neurons that do not use opioidergic neurotransmission) in an attempt to overcome the reduced visual sensitivity. This idea of the visual compensatory effect is supported by research on old adults (Riis et al., 2008) and patients with Alzheimer's disease (Bokde et al., 2009). However, it should be noted that this compensatory effect is merely a fine adjustment and cannot reverse the whole activation pattern in the visual cortex.

The results of how the visual ROI activation predicted behavioral responses demonstrate that the visual activation detected in this study is necessarily associated with discriminating pain expression of varied intensities. It showed that irrespective of the pharmacological manipulation, visual activation in that area was always positively correlated with and predicted, though with a rather low effect size, the increased intensities of pain expression. The location of the visual ROI is centered in the extrastriate cortex (i.e., V2, V3, and V4), and extends to the middle fusiform gyrus. This subregion of the fusiform gyrus has been repeatedly referred to as the "fusiform face area" (Kanwisher et al., 1997; McCarthy et al., 1997; Dubois et al., 1999; Halgren et al., 1999; Haxby et al., 1999; Tong & Nakayama, 1999). Studies have robustly revealed that the FFA exhibits a stronger activation to faces rather than nonface stimuli (Kanwisher et al., 1997; McCarthy et al., 1997; Halgren et al., 1999; Haxby et al., 1999). Even though it has been suggested that FFA is an area linked to perceptual identification of the face, some studies have also indicated its involvement in processing facial expressions (Ganel et al., 2005; Fox et al., 2009). On top of that, the function of processing facial information in FFA has been found to exhibit a right hemisphere dominance (Haxby et al., 1999; Barton et al., 2002; Rossion et al., 2003a; Rossion et al., 2003b; Ganel et al., 2005; Fox et al., 2009). This is confirmed in our study for that only activation in the right FFA instead of the left was observed when facial expressions were perceived with increased pain intensity. In accordance with previous studies, our findings verify that the right

FFA not only gets engaged in general facial recognition but also modulates perception on facial expression features such as the intensity of pain expressions (Calder & Young, 2005; Eichmann et al., 2008; Loughead et al., 2008). Furthermore, this activity found in the right FFA was modulated by the opioid system, suggesting an underlying opioidergic mechanism engaged in the facial processing of emotional expressions via the ventral stream.

As pain intensities increased, we did not find significant parametric increases in activity of the anterior insula and the anterior mid-cingulate cortex. This could be related to the aversive nature of both pain and disgust. However, it is important to note that our manipulation check analyses showed significant activity in the bilateral anterior insular cortex and the anterior mid-cingulate cortex when comparing the task with baseline, as expected. Therefore, the absence of parametric modulation in these regions might imply naltrexone had no unique modulatory effect on the affective components of perceiving pain expressions.

We would also like to address the potential clinical implications of our work, which may have significant meaning with respect to the diagnosis and therapeutic interventions of pain. First, patients with chronic pain, especially those who were high in fear of (re)injury, showed strong attention bias towards painful facial expressions (Khatibi et al., 2009). Furthermore, participants with high catastrophizing personality were found to exhibit longer gaze duration for both pain and neutral expressions (Vervoort et al., 2013). Observation of others' pain expressions, however, has been detected to increase pain perception in self (Mailhot et al., 2012; Vachon-Presseau et al., 2012; Reicherts et al., 2013; Khatibi et al., 2014; Khatibi et al., 2015). It has thus been argued that the facilitation of pain perception induced by vicarious pain might be modulated by top-down attentional processes (Khatibi et al., 2014). In terms of our findings, it indicates that when administrating opioidergic analgesics (e.g., painkillers) to diminish patients' pain, the analgesic effect might be counteracted by the enhanced attention towards others' pain expressions. Second, our findings also raise the issue that applying opioidergic analgesics might affect how accurately and efficiently we detect pain in others. More specifically, the reduction in our capacity or efficiency to accurately identify pain-related expressions may affect the diagnosis and treatment of patients' pain by medical personnel under the influence of (opioid) painkillers. This is especially important considering the current "opioid endemic", with about 11.5 million people in the USA alone showing a misuse of prescription opioids in 2016 (National Center for Health Statistics, 2017; Substance Abuse and Mental Health Services Administration, 2017).

The endogenous opioid system is engaged in modulating many affective and cognitive functions, such as visual perception, emotion, and reward (Liberzon et al., 2002; Dalley et al., 2005; Koepp et al., 2009;

Colasanti et al., 2012; Rütgen et al., 2015b; Chelnokova et al., 2016). Studying the neurochemical underpinnings of emotion perception may also have important implications in understanding the experience and processes of empathy. A recent framework of empathy considers that emotion identification and affect sharing are two separate processes that independently contribute to empathic responses (Coll et al., 2017). While emotion identification in this framework is closely related to but not synonymous with emotion discrimination, failure to clarify the possibly distinct effects of emotion discrimination and affect sharing might lead to inaccurate characterization of the experience of empathy. The findings from the present study are thus particularly relevant for the further development and validation of this framework. They suggest that the opioid system not only plays a role in higher-order affective processes underpinning empathy, and in particular affect sharing, but that causally manipulating opioidergic activity also modulates the perception and judgment of pain in others. This, we hope, will inspire further exploration of the relationship between the opioid system and socio-perceptual processes, and how they inform higher-level processes and aspects of the multi-faceted experience of empathy. However, in making these connections, it needs to be considered that the present study used an opioid antagonist to investigate the role of the opioidergic activity on pain discrimination. Interestingly, our findings seem at odds with the predictions and analyses reported in Coll et al., which would seem to suggest an increase rather than a decrease in pain discrimination with reduced opioid activity. Future studies with opioid agonists (Chelnokova et al., 2016). or other types of painkillers (Mischkowski et al., 2016) are thus needed to verify the possible links between analgesics and pain perception, in both self and others.

Finally, some limitations of this study deserve discussion. First, the different degrees of pain/disgust in the stimuli were determined by the morphing software, and may thus not accurately reflect the subjective experience of these degrees by study participants. Nevertheless, this method is frequently used in studies on emotion identification (Young et al., 1997; Averbeck et al., 2011; Wells et al., 2016), and our behavioral results suggest that the morphed degrees used in our study corresponded approximately with subjective experience. Second, variable inter-trial intervals could be implemented between viewing pictures and the judgement phase in future studies, though this had little influence on our current findings due to parametric analysis approach and the documented quasi-absence of multicollinearity (r = -.03). Third, naltrexone was suggested to be associated with increased attention to the negative valence of stimuli in general (Murray et al., 2014; Meier et al., 2016). Despite of the main focus of the current study being pain, further experiments are thus required to carefully test whether naltrexone similarly induces increased attention to both pain and disgust expressions and whether this general effect affects our current results. However, given that we detected differential effects for pain and disgust, we can rather safely say that in those aspects the drug had selective effects; this is however not to say that in other domains, additional effects could have been detected but were occluded due to a lack of selectivity associated with opioid blockade.

In conclusion, the behavioral and neural findings of this psychopharmacological fMRI study shed light on a causal role of the opioid system in the discrimination of painful facial expressions, paving the way for further exploration of clinical implications in the domains of pain diagnosis and treatment on the one hand, and future research on the relationship between basic socio-perceptual processing and empathy on the other.

5 Author Notes

This work was supported by Chinese Scholarship Council (CSC) Grant #201604910515 to Y.Z.; the Vienna Science and Technology Fund (WWTF VRG13-007) and the Austrian Science Fund (FWF P 32686) to C.L.. We thank Elena Navratil, Kathrin Steinmüller and Pablo Reinhardt for help with data acquisition. The authors declare no competing financial interests.

Reprints: Claus Lamm, claus.lamm@univie.ac.at

6 References

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage, 38*(1), 95-113. doi: https://doi.org/10.1016/j.neuroimage.2007.07.007

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. doi: https://doi.org/10.1016/j.jml.2012.11.001

Barton, J. J., Press, D. Z., Keenan, J. P., & O'Connor, M. (2002). Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology, 58*(1), 71-78. doi: https://doi.org/10.1212/wnl.58.1.71

Bisaga, A., Mannelli, P., Sullivan, M. A., Vosburg, S. K., Compton, P., Woody, G. E., & Kosten, T. R. (2018). Antagonists in the medical management of opioid use disorders: Historical and existing treatment strategies. *The American journal on addictions, 27*(3), 177-187. doi: https://doi.org/10.1111/ajad.12711

Bokde, A. L. W., Lopez-Bayo, P., Born, C., Ewers, M., Meindl, T., Teipel, S. J., . . . Hampel, H. (2009). Alzheimer Disease: Functional Abnormalities in the Dorsal Visual Pathway. *Radiology, 254*(1), 219-226. doi: https://doi.org/10.1148/radiol.2541090558

Botvinick, M., Jha, A. P., Bylsma, L. M., Fabian, S. A., Solomon, P. E., & Prkachin, K. M. (2005). Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *Neuroimage, 25*(1), 312-319. doi: https://doi.org/10.1016/j.neuroimage.2004.11.043

Broom, D. M. (2001). Evolution of pain. *Vlaams Diergeneeskundig Tijdschrift, 70*(1), 17-21.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience, 6*(8), 641-651. doi: https://doi.org/10.1038/nrn1724

Casale, G., Pecorini, M., Cuzzoni, G., & de Nicola, P. (1985). Beta-Endorphin and Cold Pressor Test in the Aged. *Gerontology, 31*(2), 101-105. doi: http://doi.org/10.1159/000212687

Chelnokova, O., Laeng, B., Løseth, G., Eikemo, M., Willoch, F., & Leknes, S. (2016). The µ-opioid system promotes visual attention to faces and eyes. *Social Cognitive and Affective Neuroscience, 11*(12), 1902-1909. doi: https://doi.org/10.1093/scan/nsw116

Colasanti, A., Searle, G. E., Long, C. J., Hill, S. P., Reiley, R. R., Quelch, D., . . . Lingford-Hughes, A. R. (2012). Endogenous opioid release in the human brain reward system induced by acute amphetamine administration. *Biological Psychiatry, 72*(5), 371-377. doi: https://doi.org/10.1016/j.biopsych.2012.01.027

Coll, M.-P., Viding, E., Rütgen, M., Silani, G., Lamm, C., Catmur, C., & Bird, G. (2017). Are we really measuring empathy? Proposal for a new measurement framework. *Neuroscience and Biobehavioral Reviews, 83*, 132-139. doi: https://doi.org/10.1016/j.neubiorev.2017.10.009

Cook, R., Brewer, R., Shah, P., & Bird, G. (2013). Alexithymia, Not Autism, Predicts Poor Recognition of Emotional Facial Expressions. *Psychological Science, 24*(5), 723-732. doi: https://doi.org/10.1177/0956797612463582

Dalley, J. W., Lääne, K., Pena, Y., Theobald, D. E., Everitt, B. J., & Robbins, T. W. (2005). Attentional and motivational deficits in rats withdrawn from intravenous self-administration of cocaine or heroin. *Psychopharmacology, 182*(4), 579-587. doi: https://doi.org/10.1007/s00213-005-0107-3

De Winter, F.-L., Zhu, Q., Van den Stock, J., Nelissen, K., Peeters, R., de Gelder, B., . . . Vandenbulcke, M. (2015). Lateralization for dynamic facial expressions in human superior temporal sulcus. *Neuroimage, 106*, 340-352. doi: https://doi.org/10.1016/j.neuroimage.2014.11.020

Decety, J., Skelly, L. R., & Kiehl, K. A. (2013). Brain Response to Empathy-Eliciting Scenarios Involving Pain in Incarcerated Individuals With Psychopathy. *JAMA Psychiatry, 70*(6), 638-645. doi: https://doi.org/10.1001/jamapsychiatry.2013.27

Dubois, S., Rossion, B., Schiltz, C., Bodart, J.-M., Michel, C., Bruyer, R., & Crommelinck, M. (1999). Effect of familiarity on the processing of human faces. *Neuroimage, 9*(3), 278-289. doi: https://doi.org/10.1006/nimg.1998.0409

Eichmann, M., Kugel, H., & Suslow, T. (2008). Difficulty Identifying Feelings and Automatic Activation in the Fusiform Gyrus in Response to Facial Emotion. *Perceptual and Motor Skills, 107*(3), 915-922. doi: https://doi.org/10.2466/pms.107.3.915-922

Engell, A. D., & Haxby, J. V. (2007). Facial expression and gaze-direction in human superior temporal sulcus. *Neuropsychologia, 45*(14), 3234-3241. doi: https://doi.org/10.1016/j.neuropsychologia.2007.06.022

Fallon, N., Roberts, C., & Stancak, A. (2020). Shared and distinct functional networks for empathy and pain processing: A systematic review and meta-analysis of fMRI studies. *Social Cognitive and Affective Neuroscience*. doi: https://doi.org/10.1093/scan/nsaa090

Fox, C. J., Moon, S. Y., Iaria, G., & Barton, J. J. S. (2009). The correlates of subjective perception of identity and expression in the face network: an fMRI adaptation study. *Neuroimage, 44*(2), 569-580. doi: https://doi.org/10.1016/j.neuroimage.2008.09.011

Ganel, T., Valyear, K. F., Goshen-Gottstein, Y., & Goodale, M. A. (2005). The involvement of the “fusiform face area” in processing facial expression. *Neuropsychologia, 43*(11), 1645-1654. doi: https://doi.org/10.1016/j.neuropsychologia.2005.01.012

Gläscher, J. (2009). Visualization of group inference data in functional neuroimaging. *Neuroinformatics, 7*(1), 73-82. doi: https://doi.org/10.1007/s12021-008-9042-x

Goubert, L., Craig, K. D., Vervoort, T., Morley, S., Sullivan, M. J. L., Williams, d. C. A. C., . . . Crombez, G. (2005). Facing others in pain: the effects of empathy. *Pain, 118*(3), 285-288. doi: https://doi.org/10.1016/j.pain.2005.10.025

Granato, P., Vinekar, S., Gansberghe, J.-P. V., & Bruyer, R. (2012). Evidence of Impaired Facial Emotion Recognition in Mild Alzheimer's Disease: A Mathematical Approach and Application. *Open Journal of Psychiatry, 2*(3), 171-186. doi: https://doi.org/10.4236/ojpsych.2012.23023

Halgren, E., Dale, A. M., Sereno, M. I., Tootell, R. B., Marinkovic, K., & Rosen, B. R. (1999). Location of human face-selective cortex with respect to retinotopic areas. *Human Brain Mapping, 7*(1), 29-37. doi: https://doi.org/10.1002/(SICI)1097-0193(1999)7:1<29::AID-HBM3>3.0.CO;2-R

Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., & Martin, A. (1999). The Effect of Face Inversion on Activity in Human Neural Systems for Face and Object Perception. *Neuron, 22*(1), 189-199. doi: https://doi.org/10.1016/S0896-6273(00)80690-X

Hein, G., Lamm, C., Brodbeck, C., & Singer, T. (2011). Skin Conductance Response to the Pain of Others Predicts Later Costly Helping. *PloS One, 6*(8), e22759. doi: https://doi.org/10.1371/journal.pone.0022759

Ipser, J. C., Terburg, D., Syal, S., Phillips, N., Solms, M., Panksepp, J., . . . van Honk, J. (2013). Reduced fear-recognition sensitivity following acute buprenorphine administration in healthy volunteers. *Psychoneuroendocrinology, 38*(1), 166-170. doi: https://doi.org/10.1016/j.psyneuen.2012.05.002

Jauniaux, J., Khatibi, A., Rainville, P., & Jackson, P. L. (2019). A meta-analysis of neuroimaging studies on pain empathy: investigating the role of visual information and observers' perspective. *Social Cognitive and Affective Neuroscience, 14*(8), 789-813. doi: https://doi.org/10.1093/scan/nsz055

Jungkunz, G., Engel, R. R., King, U. G., & Kuss, H. J. (1983). Endogenous opiates increase pain tolerance after stress in humans. *Psychiatry Research, 8*(1), 13-18. doi: https://doi.org/10.1016/0165-1781(83)90133-6

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*(11), 4302-4311. doi: https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Karjalainen, T., Karlsson, H. K., Lahnakoski, J. M., Glerean, E., Nuutila, P., Jääskeläinen, I. P., . . . Nummenmaa, L. (2017). Dissociable roles of cerebral µ-opioid and type 2 dopamine receptors in vicarious pain: a combined PET–fMRI study. *Cerebral Cortex, 27*(8), 4257-4266. doi: https://doi.org/10.1093/cercor/bhx129

KatzenPerez, K. R., Jacobs, D. W., Lincoln, A., & Ellis, R. J. (2001). Opioid blockade improves human recognition memory following physiological arousal. *Pharmacology Biochemistry and Behavior, 70*(1), 77-84. doi: https://doi.org/10.1016/s0091-3057(01)00589-5

Kavaliers, M. (1988). Evolutionary and comparative aspects of nociception. *Brain Research Bulletin, 21*(6), 923-931. doi: https://doi.org/10.1016/0361-9230(88)90030-5

Kavaliers, M., Ossenkopp, K.-P., & Choleris, E. (2019). Social neuroscience of disgust. *Genes, Brain and Behavior, 18*(1), e12508. doi: https://doi.org/10.1111/gbb.12508

Khatibi, A., Dehghani, M., Sharpe, L., Asmundson, G. J., & Pouretemad, H. (2009). Selective attention towards painful faces among chronic pain patients: evidence from a modified version of the dot-probe. *Pain, 142*(1-2), 42-47. doi: https://doi.org/10.1016/j.pain.2008.11.020

Khatibi, A., Schrooten, M., Bosmans, K., Volders, S., Vlaeyen, J. W. S., & Van den Bussche, E. (2015). Sub-optimal presentation of painful facial expressions enhances readiness for action and pain perception following electrocutaneous stimulation. *Frontiers in Psychology, 6*(913), 1-9. doi: https://doi.org/10.3389/fpsyg.2015.00913

Khatibi, A., Vachon-Presseau, E., Schrooten, M., Vlaeyen, J., & Rainville, P. (2014). Attention effects on vicarious modulation of nociception and pain. *PAIN®, 155*(10), 2033-2039. doi: https://doi.org/10.1016/j.pain.2014.07.005

King, C. D., Goodin, B., Kindler, L. L., Caudle, R. M., Edwards, R. R., Gravenstein, N., . . . Fillingim, R. B. (2013). Reduction of conditioned pain modulation in humans by naltrexone: an exploratory study of the effects of pain catastrophizing. *Journal of Behavioral Medicine, 36*(3), 315-327. doi: https://doi.org/10.1007/s10865-012-9424-2

Koepp, M. J., Hammers, A., Lawrence, A. D., Asselin, M.-C., Grasby, P. M., & Bench, C. (2009). Evidence for endogenous opioid release in the amygdala during positive emotion. *Neuroimage, 44*(1), 252-256. doi: https://doi.org/10.1016/j.neuroimage.2008.08.032

Kunz, M., Peter, J., Huster, S., & Lautenbacher, S. (2013). Pain and Disgust: The Facial Signaling of Two Aversive Bodily Experiences. *PloS One, 8*(12), e83277. doi: https://doi.org/10.1371/journal.pone.0083277

Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience, 19*(1), 42-58. doi: https://doi.org/10.1162/jocn.2007.19.1.42

Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage, 54*(3), 2492-2502. doi: https://doi.org/10.1016/j.neuroimage.2010.10.014

Liberzon, I., Zubieta, J. K., Fig, L. M., Phan, K. L., Koeppe, R. A., & Taylor, S. F. (2002). µ-Opioid receptors and limbic responses to aversive emotional stimuli. *Proceedings of the National Academy of Sciences, 99*(10), 7084-7089. doi: https://doi.org/10.1073/pnas.102174799

Loughead, J., Gur, R. C., Elliott, M., & Gur, R. E. (2008). Neural circuitry for accurate identification of facial emotions. *Brain Research, 1194*, 37-44. doi: https://doi.org/10.1016/j.brainres.2007.10.105

Machin, A. J., & Dunbar, R. I. (2011). The brain opioid theory of social attachment: a review of the evidence. *Behaviour, 148*(9-10), 985-1025. doi: https://doi.org/10.1163/000579511X596624

Mailhot, J. P., Vachon-Presseau, E., Jackson, P. L., & Rainville, P. (2012). Dispositional empathy modulates vicarious effects of dynamic pain expressions on spinal nociception, facial responses and acute pain. *European Journal of Neuroscience, 35*(2), 271-278. doi: https://doi.org/10.1111/j.1460-9568.2011.07953.x

Masten, C. L., Morelli, S. A., & Eisenberger, N. I. (2011). An fMRI investigation of empathy for 'social pain' and subsequent prosocial behavior. *Neuroimage, 55*(1), 381-388. doi: https://doi.org/10.1016/j.neuroimage.2010.11.060

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience, 9*(5), 605-610. doi: https://doi.org/10.1162/jocn.1997.9.5.605

McCullough, S., & Emmorey, K. (2009). Categorical perception of affective and linguistic facial expressions. *Cognition, 110*(2), 208-221. doi: https://doi.org/10.1016/j.cognition.2008.11.007

McKone, E., Martini, P., & Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *Journal of Experimental Psychology: Human Perception and Performance, 27*(3), 573-599. doi: https://doi.org/10.1037//0096-1523.27.3.573

Meier, I. M., Bos, P. A., Hamilton, K., Stein, D. J., van Honk, J., & Malcolm-Smith, S. (2016). Naltrexone increases negatively-valenced facial responses to happy faces in female participants. *Psychoneuroendocrinology, 74*, 65-68. doi: https://doi.org/10.1016/j.psyneuen.2016.08.022

Mischkowski, D., Crocker, J., & Way, B. M. (2016). From painkiller to empathy killer: acetaminophen (paracetamol) reduces empathy for pain. *Social Cognitive and Affective Neuroscience, 11*(9), 1345-1353. doi: https://doi.org/10.1093/scan/nsw057

Narumoto, J., Okada, T., Sadato, N., Fukui, K., & Yonekura, Y. (2001). Attention to emotion modulates fMRI activity in human right superior temporal sulcus. *Cognitive Brain Research, 12*(2), 225-231. doi: https://doi.org/10.1016/S0926-6410(01)00053-2

National Center for Health Statistics. (2017). *Wide-ranging online data for epidemiologic research (WONDER)*. Atlanta, GA: CDC.National Center for Health Statistics.

Nummenmaa, L., & Karjalainen, T. (2018). Opioidergic regulation of pain and pleasure in human social relationships. *Neuropsychopharmacology, 43*(1), 217-218. doi: https://doi.org/10.1038/npp.2017.200

Nummenmaa, L., & Tuominen, L. (2018). Opioid system and human emotions. *British Journal of Pharmacology, 175*(14), 2737-2749. doi: https://doi.org/10.1111/bph.13812

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

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage, 59*(3), 2142-2154. doi: https://doi.org/10.1016/j.neuroimage.2011.10.018

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. doi: https://doi.org/10.1016/j.neuroimage.2013.08.048

Price, R. C., Christou, N. V., Backman, S. B., Stone, L., & Schweinhardt, P. (2016). Opioid-receptor antagonism increases pain and decreases pleasure in obese and non-obese individuals. *Psychopharmacology, 233*(23), 3869-3879. doi: https://doi.org/10.1007/s00213-016-4417-4

Reicherts, P., Gerdes, A. B., Pauli, P., & Wieser, M. J. (2013). On the mutual effects of pain and emotion: facial pain expressions enhance pain perception and vice versa are perceived as more arousing when feeling pain. *PAIN®, 154*(6), 793-800. doi: https://doi.org/10.1016/j.pain.2013.02.012

Riis, J. L., Chong, H., Ryan, K. K., Wolk, D. A., Rentz, D. M., Holcomb, P. J., & Daffner, K. R. (2008). Compensatory neural activity distinguishes different patterns of normal cognitive aging. *Neuroimage, 39*(1), 441-454. doi: https://doi.org/10.1016/j.neuroimage.2007.08.034

Robertson, L. J., Hammond, G. R., & Drummond, P. D. (2008). The Effect of Subcutaneous Naloxone on Experimentally Induced Pain. *The Journal of Pain, 9*(1), 79-87. doi: https://doi.org/10.1016/j.jpain.2007.08.008

Rossion, B., Caldara, R., Seghier, M., Schuller, A. M., Lazeyras, F., & Mayer, E. (2003a). A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain, 126*(11), 2381-2395. doi: https://doi.org/10.1093/brain/awg241

Rossion, B., Schiltz, C., & Crommelinck, M. (2003b). The functionally defined right occipital and fusiform “face areas” discriminate novel from visually familiar faces. *Neuroimage, 19*(3), 877-883. doi: https://doi.org/10.1016/S1053-8119(03)00105-8

Rütgen, M., Seidel, E. M., Pletti, C., Riecansky, I., Gartus, A., Eisenegger, C., & Lamm, C. (2018). Psychopharmacological modulation of event-related potentials suggests that first-hand pain and empathy for pain rely on similar opioidergic processes. *Neuropsychologia, 116*(Pt A), 5-14. doi: https://doi.org/10.1016/j.neuropsychologia.2017.04.023

Rütgen, M., Seidel, E. M., Riecansky, I., & Lamm, C. (2015a). Reduction of Empathy for Pain by Placebo Analgesia Suggests Functional Equivalence of Empathy and First-Hand Emotion Experience. *The Journal of Neuroscience, 35*(23), 8938-8947. doi: https://doi.org/10.1523/jneurosci.3936-14.2015

Rütgen, M., Seidel, E. M., Silani, G., Riecansky, I., Hummer, A., Windischberger, C., . . . Lamm, C. (2015b). Placebo analgesia and its opioidergic regulation suggest that empathy for pain is grounded in self pain. *Proceedings of the National Academy of Sciences, 112*(41), E5638-E5646. doi: https://doi.org/10.1073/pnas.1511269112

Sharvit, G., Vuilleumier, P., Delplanque, S., & Corradi-Dell'Acqua, C. (2015). Cross-modal and modality-specific expectancy effects between pain and disgust. *Scientific Reports, 5*, 17487. doi: https://doi.org/10.1038/srep17487

Simon, D., Craig, K. D., Gosselin, F., Belin, P., & Rainville, P. (2008). Recognition and discrimination of prototypical dynamic expressions of pain and emotions. *PAIN®, 135*(1), 55-64. doi: https://doi.org/10.1016/j.pain.2007.05.008

Simon, D., Craig, K. D., Miltner, W. H. R., & Rainville, P. (2006). Brain responses to dynamic facial expressions of pain. *Pain, 126*(1), 309-318. doi: https://doi.org/10.1016/j.pain.2006.08.033

Singer, T., Seymour, B., O'doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303*(5661), 1157-1162. doi: https://doi.org/10.1126/science.1093535

Substance Abuse and Mental Health Services Administration. (2017). *Key substance use and mental health indicators in the United States: Results from the 2016 National Survey on Drug Use and Health (HHS Publication No. SMA 17-5044, NSDUH Series H-52)*. Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration.

Timmers, I., Park, A. L., Fischer, M. D., Kronman, C. A., Heathcote, L. C., Hernandez, J. M., & Simons, L. E. (2018). Is Empathy for Pain Unique in Its Neural Correlates? A Meta-Analysis of Neuroimaging Studies of Empathy. *Frontiers in Behavioral Neuroscience, 12*(289). doi: http://doi.org/10.3389/fnbeh.2018.00289

Tong, F., & Nakayama, K. (1999). Robust representations for faces: evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 25*(4), 1016. doi: https://doi.org/10.1037//0096-1523.25.4.1016

Vachon-Presseau, E., Roy, M., Martel, M.-O., Albouy, G., Chen, J., Budell, L., . . . Rainville, P. (2012). Neural processing of sensory and emotional-communicative information associated with the perception of vicarious pain. *Neuroimage, 63*(1), 54-62. doi: https://doi.org/10.1016/j.neuroimage.2012.06.030

van der Meulen, M., van Ijzendoorn, M. H., & Crone, E. A. (2016). Neural Correlates of Prosocial Behavior: Compensating Social Exclusion in a Four-Player Cyberball Game. *PloS One, 11*(7), e0159045. doi: https://doi.org/10.1371/journal.pone.0159045

Vervoort, T., Trost, Z., Prkachin, K. M., & Mueller, S. C. (2013). Attentional processing of other's facial display of pain: An eye tracking study. *PAIN®, 154*(6), 836-844. doi: https://doi.org/10.1016/j.pain.2013.02.017

Wardle, M. C., Bershad, A. K., & de Wit, H. (2016). Naltrexone alters the processing of social and emotional stimuli in healthy adults. *Social Neuroscience, 11*(6), 579-591. doi: https://doi.org/10.1080/17470919.2015.1136355

Washington, L. L., Gibson, S. J., & Helme, R. D. (2000). Age-related differences in the endogenous analgesic response to repeated cold water immersion in human volunteers. *Pain, 89*(1), 89-96. doi: https://doi.org/10.1016/S0304-3959(00)00352-3

Wegrzyn, M., Riehle, M., Labudda, K., Woermann, F., Baumgartner, F., Pollmann, S., . . . Kissler, J. (2015). Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. *Cortex, 69*, 131-140. doi: https://doi.org/10.1016/j.cortex.2015.05.003

Xiong, R.-C., Fu, X., Wu, L.-Z., Zhang, C.-H., Wu, H.-X., Shi, Y., & Wu, W. (2019). Brain pathways of pain empathy activated by pained facial expressions: a meta-analysis of fMRI using the activation likelihood estimation method. *Neural regeneration research, 14*(1), 172-178. doi: http://doi.org/10.4103/1673-5374.243722

Young, A. W., Perrett, D., Calder, A., Sprengelmeyer, R., & Ekman, P. (2002). *Facial expressions of emotion: Stimuli and tests (FEEST)*. Bury St. Edmunds, Suffolk, England: Thames Valley Test Company.

7 Supplemental information

Supplementary Table 1. Post-hoc Tukey's test: *t* values of the pairwise comparisons between pain intensities across naltrexone and placebo sessions

Pain intensity	***20%***	***30%***	***40%***	***50%***	***60%***	***70%***
20%						
30%	1.85					
40%	9.59***	7.46***				
50%	20.37***	18.04***	10.89***			
60%	33.07***	30.30***	23.09***	11.59***		
70%	34.68***	31.91***	24.81***	13.44***	2.08	
80%	36.69***	33.86***	26.86***	15.41***	4.17**	2.06

p < .05*, *p* < .001**, *p* < .0001**

The manuscript is published in *eLife*

Chapter 3 - Neural dynamics between anterior insular cortex and right supramarginal gyrus dissociate genuine affect sharing from perceptual saliency of pretended pain

Yili Zhao[1], Lei Zhang[1], Markus Rütgen[1,2], Ronald Sladky[1], Claus Lamm[1,2*]

[1] Social, Cognitive and Affective Neuroscience Unit, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria

[2] Vienna Cognitive Science Hub, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria

[3] Neuropsychopharmacology and Biopsychology Unit, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria

DOI: http://doi.org/10.7554/eLife.69994

Abstract

Empathy for pain engages both shared affective responses and self-other distinction. In this study, we addressed the highly debated question of whether neural responses previously linked to affect sharing could result from the perception of salient affective displays. Moreover, we investigated how the brain network involved in affect sharing and self-other distinction underpinned our response to a pain that is either perceived as genuine or pretended (while in fact both were acted for reasons of experimental control). We found stronger activations in regions associated with affect sharing (anterior insula, aIns, and anterior mid-cingulate cortex, aMCC) as well as with affective self-other distinction (right supramarginal gyrus, rSMG), in participants watching video clips of genuine vs. pretended facial expressions of pain. Using dynamic causal modeling (DCM), we then assessed the neural dynamics between the right aIns and rSMG in these two conditions. This revealed a reduced inhibitory effect on the aIns to rSMG connection for genuine compared to pretended pain. For genuine pain only, brain-to-behavior regression analyses highlighted a linkage between this inhibitory effect on the one hand, and pain ratings as well as empathic traits on the other. These findings imply that if the pain of others is genuine and thus calls for an appropriate empathic response, neural responses in the aIns indeed seem related to affect sharing and self-other distinction is engaged to avoid empathic over-arousal. In contrast, if others merely pretend to be in pain, the perceptual salience of their painful expression results in neural responses that are down-regulated to avoid inappropriate affect sharing and social support.

Introduction

As social beings, our own affective states are influenced by other people's feelings and affective states. The facial expression of pain by others acts as a distinctive cue to signal their pain to others, and thus results in sizeable affective responses in the observer. Certifying such responses as evidence for empathy, however, requires successful self-other distinction, the ability to distinguish the affective response experienced by ourselves from the affect experienced by the other person.

Studies using a wide variety of methods convergently have shown that observing others in pain engages neural responses aligning with those coding for the affective component of self-experienced pain, with the anterior insula (aIns) and the anterior mid-cingulate cortex (aMCC) being two key areas in which such an alignment has been detected (Lamm et al., 2011; Rütgen et al., 2015; Jauniaux et al., 2019; Xiong et al., 2019; Zhou et al., 2020; Fallon et al., 2020, for meta-analyses). However, there is consistent debate on whether activity observed in these areas should indeed be related to the sharing of pain affect, or whether it may not rather result from automatic responses to salient perceptual cues - with pain vividly expressed on the face being one particularly prominent example (Zaki et al., 2016, for review). It was thus one major aim of our study to address this question. In this respect, contextual factors, individuals' appraisals, and attentional processes would all impact their exact response to the affective states of others (Gu & Han, 2007; Hein & Singer, 2008, for review; Lamm et al., 2010; Forbes & Hamilton, 2020; Zhao et al., 2021). Recently, Coll et al. (2017) have thus proposed a framework that attempts to capture these influences on affect sharing and empathic responses. This model posits that individuals who see identical negative facial expressions of others may have different empathic responses due to distinct contextual information, and that this may depend on identification of the underlying affective state displayed by the other. In the current functional magnetic resonance imaging (fMRI) study, we therefore created a situation where we varied the genuineness of the pain affect felt by participants while keeping the perceptual saliency (i.e., the quality and strength of pain expressions) identical. To this end, participants were shown video clips of other persons who supposedly displayed genuine pain on their face vs. merely pretended to be in pain. Note that for reasons of experimental control, all painful expressions on the videos had been acted. This enabled us to interpret possible differences between conditions to the observers' appraisal of the situation rather than to putative visual and expressive differences. This way, we sought to identify the extent to which responses in affective nodes (such as the aIns and the aMCC) genuinely track the pain of others, rather than resulting predominantly from the salient facial expressions associated with the pain.

Another major aim of our study was to assess how self-other distinction allowed individuals to distinguish between the sharing of actual pain vs. regulating an inappropriate and potentially misleading “sharing” of what in reality is only a pretended affective state. We focused on the right supramarginal gyrus (rSMG), which has been suggested to act as a major hub selectively engaged in affective self-other distinction (Silani et al., 2013; Steinbeis et al., 2015; Hoffmann et al., 2016; Bukowski et al., 2020). Though previous studies have indicated that rSMG is functionally connected with areas associated with affect processing (Mars et al., 2011; Bukowski et al., 2020), we lack more nuanced insights into how exactly rSMG interacts with these areas, and thus how it supports accurate empathic responses. Hence, we used dynamic causal modeling (DCM) to investigate the hypothesized brain patterns of affective responses and self-other distinction for the genuine and pretended pain situations, focusing on the aIns, aMCC, and their interaction with rSMG. Furthermore, we investigated the relationship between neural activity and behavioral responses as well as empathic traits. In line with the literature reviewed above, we expected that, on the behavioral level, genuine pain would result in – alongside the obvious other-oriented higher pain ratings – higher self-oriented unpleasantness ratings. On the neural level, we predicted aIns and aMCC to show a stronger response to the genuine expressions of pain, but that these areas would also respond to the pretended pain, but to a lower extent. Differences in rSMG engagement and distinct patterns of this area’s effective connectivity with aIns and aMCC were expected to relate to self-other distinction, and thus to explain the different empathic responses to genuine vs. pretended pain.

Results

Behavioral results

Three repeated-measures ANOVAs were performed with the factors *genuineness* (genuine vs. pretended and *pain* (pain vs. no pain), for each of the three behavioral ratings. For ratings of painful *expressions* in others (Figure 1C, left), there was a main effect of the factor genuineness: participants showed higher ratings for the genuine vs. pretended conditions, $F_{\text{genuineness}}$ (1, 42) = 8.816, p = 0.005, η^2 = 0.173. There was also a main effect of pain: participants showed higher ratings for the pain vs. no pain conditions, F_{pain} (1,42) = 1718.645, p < 0.001, η^2 = 0.976. The interaction term was significant as well, $F_{\text{interaction}}$ (1, 42) = 7.443, p = 0.009, η^2 = 0.151, and this was related to higher ratings of painful expressions in others for the genuine pain compared to the pretended pain condition. For ratings of painful feelings in others (Figure 1C, middle), there was a main effect of genuineness: participants showed higher ratings for the genuine vs. pretended conditions, $F_{\text{genuineness}}$ (1, 42) = 770.140, p < 0.001, η^2 = 0.948. There was also a main effect of pain, as participants showed higher ratings for the pain vs. no pain conditions, F_{pain} (1,42) = 1544.762, p < 0.001, η^2 = 0.974. The interaction for painful feelings ratings was significant as well, $F_{\text{interaction}}$ (1, 42) = 752.618, p < 0.001, η^2 = 0.947, and this was related to higher ratings of painful feelings in others for the genuine pain compared to the pretended pain condition. For ratings of unpleasantness in self (Figure 1C, right), there was a main effect of genuineness: participants showed higher ratings for the genuine vs. pretended conditions, $F_{\text{genuineness}}$ (1, 42) = 74.989, $p < 0.001$, $\eta^2 = 0.641$. There was also a main effect of pain: participants showed higher ratings for the pain vs. no pain conditions, F_{pain} (1,42) = 254.709, $p < 0.001$, $\eta^2 = 0.858$. The interaction for unpleasantness ratings was significant as well, $F_{\text{interaction}}$ (1, 42) = 73.620, $p < 0.001$, $\eta^2 = 0.637$, and this was related to higher ratings of unpleasantness in self for the genuine pain compared to the pretended pain condition. In sum, the behavioral data indicated higher ratings and large effect sizes of painful feelings in others and unpleasantness in self for the genuine compared to the pretended pain condition. Ratings of pain expressions also differed in terms of genuineness, at comparably low effect size, though they were expected to not show a difference by way of our experimental design and the pilot study.

We also found a significant correlation between behavioral ratings of painful feelings in others and unpleasantness in self in the genuine pain condition, $r = 0.691$, $p < 0.001$; while in the pretended pain condition, the correlation was not significant, $r = 0.249$, $p = 0.107$ (Figure 1D). A bootstrapping comparison showed a significant difference between the two correlation coefficients, $p = 0.002$, 95% Confidence Interval (CI) = [0.230, 1.060].

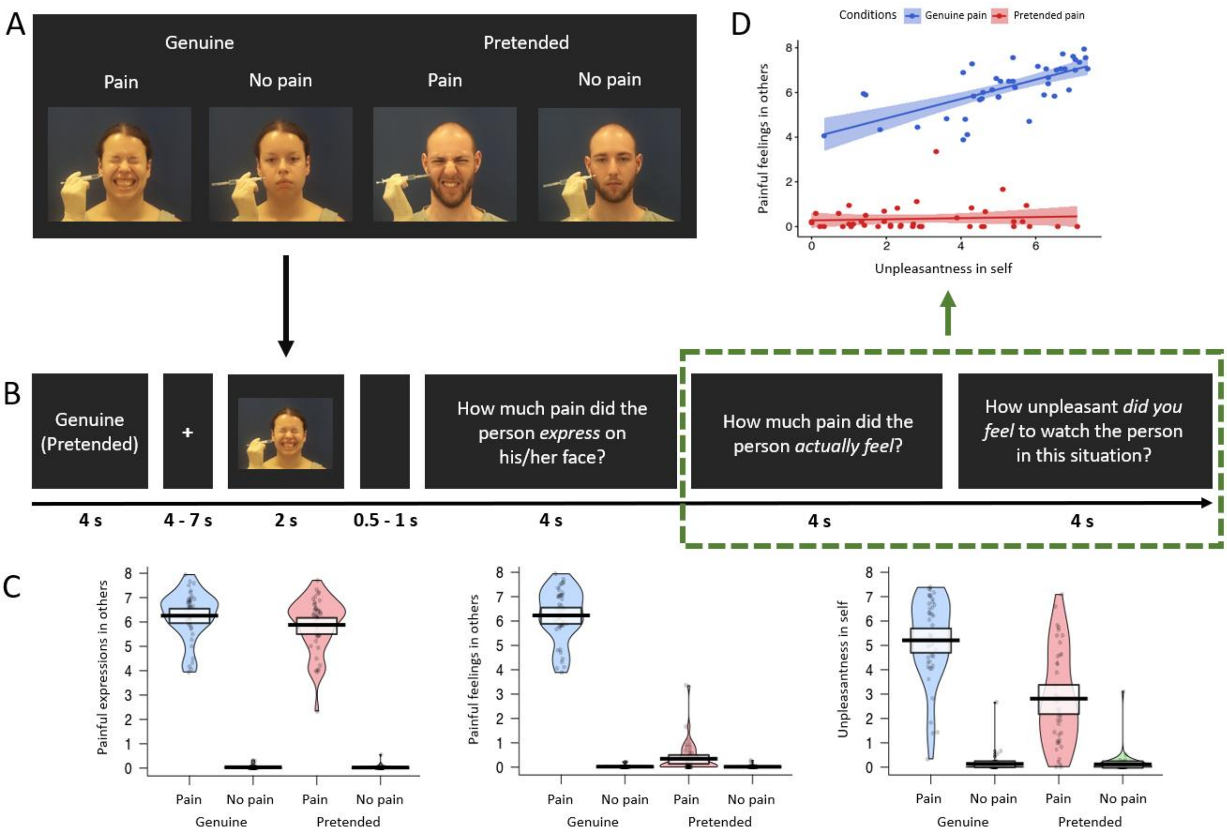

Figure 1. fMRI experimental design and behavioral results. (A) Overview of the experimental design with the four conditions genuine vs. pretended, pain vs. no pain. Examples show static images, while in the experiment participants were shown video clips. (B) Overview of experimental timeline. At the outset of each block, a reminder of “genuine” or “pretended” was shown (both terms are shown here for illustrative purposes, in the experiment either genuine or pretended was displayed). After a fixation cross, a video in the corresponding condition appeared on the screen. Followed by a short jitter, three questions about the video were separately presented and had to be rated on a visual analogue scale. These would then be followed by the next video clip and questions (not shown). (C) Violin plots of the three types of ratings for all conditions. Participants generally demonstrated higher ratings for painful expressions in others, painful feelings in others, and unpleasantness in self in the genuine pain condition than in the pretended pain condition. Ratings of all three questions were higher in the painful situation than in the neutral situation, regardless of whether in the genuine or pretended condition. The thick black lines illustrate mean values, and the white boxes indicate a 95% CI. The dots are individual data, and the “violin” outlines illustrate their estimated density at different points of the scale. (D) Correlations of painful feelings in others and unpleasantness in self for the genuine pain and the pretended pain (the relevant questions were highlighted with a green rectangular). Results revealed a significant Pearson correlation between the two questions in the genuine pain condition, but no correlation in the pretended pain condition. The lines represent the fitted regression lines, bands indicate a 95% CI.

fMRI results: mass-univariate analyses

Three contrasts were computed: 1) genuine: pain – no pain, 2) pretended: pain – no pain, and 3) genuine (pain – no pain) – pretended (pain – no pain). Across all three contrasts, we found activations as hypothesized in bilateral aIns, aMCC, and rSMG (Figure 2A and Table 1).

To identify whether or which brain activity was selectively related to the behavioral ratings described above, we performed a multiple regression analysis where we explored the relationship of activation in the contrast genuine pain – pretended pain with the three behavioral ratings. We found significant clusters in bilateral aIns, visual cortex, and cerebellum that could be selectively explained by the ratings of self-unpleasantness rather than ratings of painful expressions in others or painful feelings in others (Figure 2B).

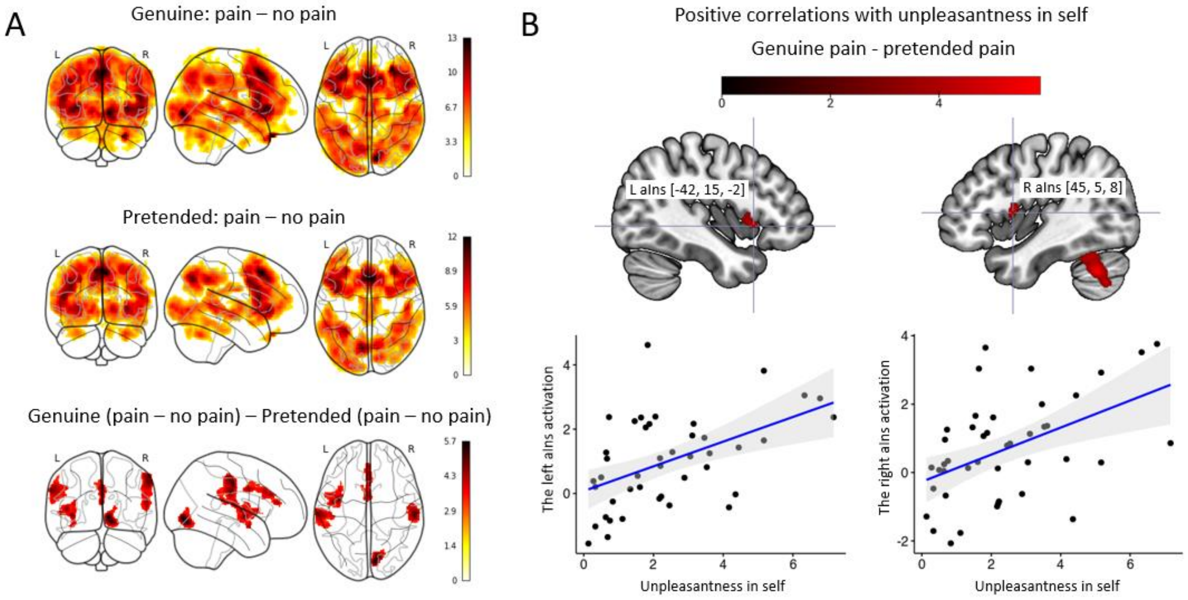
A
Genuine: pain – no pain
L
R
13
10
6.7
3.3
0
Pretended: pain – no pain
12
8.9
5.9
3
0
Genuine (pain – no pain) – Pretended (pain – no pain)
5.7
4.3
2.9
1.4
0
B
Positive correlations with unpleasantness in self
Genuine pain - pretended pain
0
2
4
L aIns [-42, 15, -2]
R aIns [45, 5, 8]
The left aIns activation
The right aIns activation
Unpleasantness in self
4
2
0
-2
6

Figure 2. Neuroimaging results: Mass-univariate analyses. (A) Activation maps of genuine: pain – no pain (top), pretended: pain - no pain (middle), and genuine (pain – no pain) – pretended (pain – no pain) (bottom). As expected, we found brain activations in the bilateral aIns, aMCC, and rSMG in all three contrasts (except for the bottom contrast, where the right aIns is only close to the significance threshold). (B) The multiple regression analysis demonstrated significant clusters in the left (peak: [-42, 15, -2]) and right anterior insular cortex (peak: [45, 5, 8]) that were positively correlated with the ratings of unpleasantness in self comparing genuine pain vs. pretended pain. All activations are thresholded with cluster-level FWE correction, $p < 0.05$ ($p < 0.001$ uncorrected initial selection threshold). The lines of the scatterplots represent the fitted regression lines, bands indicate a 95% CI.

Table 1. Results of mass-univariate functional segregation analyses in the MNI space. Region names were labeled with the AAL atlas, threshold $p < 0.05$ cluster-wise FWE correction (initial selection threshold $p < 0.001$, uncorrected). BA = Brodmann area, L = left hemisphere, R = right hemisphere.

Region label	**BA**	**Cluster size**	***x***	***y***	***z***	***t*-value**
Genuine: pain - no pain						
Lingual_R	18	183732	11	-84	-3	13.38
Temporal_Pole_Sup_R	38		30	33	-33	13.31
Supp_Motor_Area_R	8		5	15	51	12.96
Supp_Motor_Area_R	8		3	17	50	12.92
Supp_Motor_Area_L	8		-5	17	48	12.56
Insula_L	45		-32	26	6	12.32
Insula_R	45		33	29	3	12.09
Frontal_Inf_Oper_R	44		51	14	15	12.01
Frontal_Inf_Oper_R	44		50	12	18	11.79
Precentral_L	6		-42	3	39	11.72
Fusiform_R	20	463	36	-5	-41	5.58
Pretended: pain - no pain						
Supp_Motor_Area_R	8	59665	5	20	48	11.80
Supp_Motor_Area_L	8		-6	18	50	11.14
Frontal_Inf_Oper_L	44		-50	15	15	10.39
Insula_R	45		33	29	0	9.81
Insula_L	45		-29	30	0	9.60
Frontal_Inf_Tri_R	44		47	15	26	9.21
Precuneus_L	7	35136	-9	-71	41	10.27
Parietal_Inf_L	39		-32	-51	41	9.39
Precuneus_R	7		9	-69	38	8.44
Temporal_Mid_L	21		-53	-47	5	7.67
Occipital_Mid_L	19		-44	-78	2	7.47
Parietal_Inf_R	39		39	-50	41	7.25
Temporal_Mid_R	22	12970	51	-20	-6	7.70
Lingual_R	17		12	-86	-2	7.40
Fusiform_R	37		47	-33	-27	5.32
Occipital_Mid_R	18		33	-86	3	5.23
Cingulum_Mid_R	23	1666	-3	-14	27	6.35
Cingulum_Mid_L	23		-3	-24	32	5.57
Temporal_Pole_Sup_R	47	589	32	35	-33	7.18
Frontal_Sup_Orb_R	11		17	41	-24	3.36
Genuine (pain – no pain) – pretended (pain – no pain)						

SupraMarginal_L	40	1877	-66	-21	32	4.94
Postcentral_L	1		-50	-21	26	3.75
SupraMarginal_R	40	1833	63	-20	42	5.09
Rolandic_Oper_R	40		59	-15	14	4.47
Insula_L	13	1299	-38	-3	-2	5.01
Rolandic_Oper_L	4		-45	-6	8	4.8
Cingulum_Ant_L	32	1138	0	41	17	4.54
Cingulum_Mid_R	32		2	24	32	4.45
Cingulum_Mid_L	24		0	2	35	4.43
Cingulum_Ant_R	8		2	32	27	4.42
Lingual_R	18	1003	9	-84	-3	5.72
Calcarine_R	17		18	-78	8	3.61
Insula_R	13	225	39	8	-3	3.91
Rolandic_Oper_R	13		41	0	11	3.77

DCM results

We performed DCM analysis to specifically examine the modulatory effect of genuineness on the effective connectivity between the right aIns and rSMG. More specifically, we sought to assess whether the experimental manipulation of genuine pain vs. pretended pain tuned the bidirectional neural dynamics from aIns to rSMG and *vice versa*, in terms of both directionality (sign of the DCM parameter) and intensity (magnitude of the DCM parameter). If the experimental manipulation modulated the effective connectivity, we would observe a strong posterior probability (*pp* > 0.95) of the modulatory effect. Our original analysis plan was to include aMCC in the DCM analyses, but based on the fact that aMCC did not show as strong evidence (in terms of the multiple regression analysis) as the aIns of being involved in our task, we decided to use a more parsimonious DCM model without the aMCC.

We found strong evidence of inhibitory effects on the aIns to rSMG connection both in the genuine pain condition and in the pretended pain condition (Figure 3A, 3B and 3C). Comparing the strength of these modulatory effects on the aIns to rSMG connection revealed a reduced inhibitory effect for genuine pain as opposed to pretended pain, t_{41} = 2.671, p = 0.011 (Mean $_{\text{genuine pain}}$ = -0.821, 95% CI = [-0.878, -0.712]; Mean $_{\text{pretended pain}}$ = -0.934, 95% CI = [-1.076, -0.822]; Figure 3C). There was no evidence of a modulatory effect on the rSMG to aIns connection.

Individual associations between modulatory effects, behavioral ratings and questionnaires

To examine how the modulatory effects from the DCM were related to the behavioral ratings, we computed two multiple linear regression models for each condition. For the genuine pain condition, we find that the modulatory effect was significantly related to the rating of painful feelings in others ($t = 2.317$, $p = 0.026$) but not related to the rating of either painful expressions in others ($t = -1.492$, $p = 0.144$) or unpleasantness in self ($t = 0.058$, $p = 0.954$). For the pretended pain condition, none of the ratings was significantly related to the modulatory effect (Figure 3D). The variance inflation factors (*VIF*s) for three ratings in both models were calculated to diagnose collinearity, showing no severe collinearity problem (all *VIFs* < 5; the smallest *VIF* =1.132 and the largest *VIF* = 4.387).

In addition, we tested two multiple stepwise linear regression models to investigate whether subscales of all three questionnaires could explain modulatory effects for genuine pain and pretended pain. In the genuine pain condition, we found that the modulatory effect was significantly explained by scores of two subscales, i.e., affective ability and affective reactivity of the ECQ: $F_{\text{model}}\ (1, 39) = 6.829$, $p = 0.003$, $R^2 = 0.270$; $B_{\textit{affective ability}} = 0.052$, *beta* = 0.497, $p = 0.002$; $B_{\textit{affective reactivity}} = -0.040$, *beta* = -0.421, $p = 0.008$. No significant predictor was found with the other questionnaires (i.e., IRI and TAS). In the pretended pain condition, none of the three questionnaires significantly predicted variations of the modulatory effect. No severe collinearity problem was detected for either regression model (all *VIFs* < 2; the smallest *VIF* =1.011 and the largest *VIF* = 1.600).

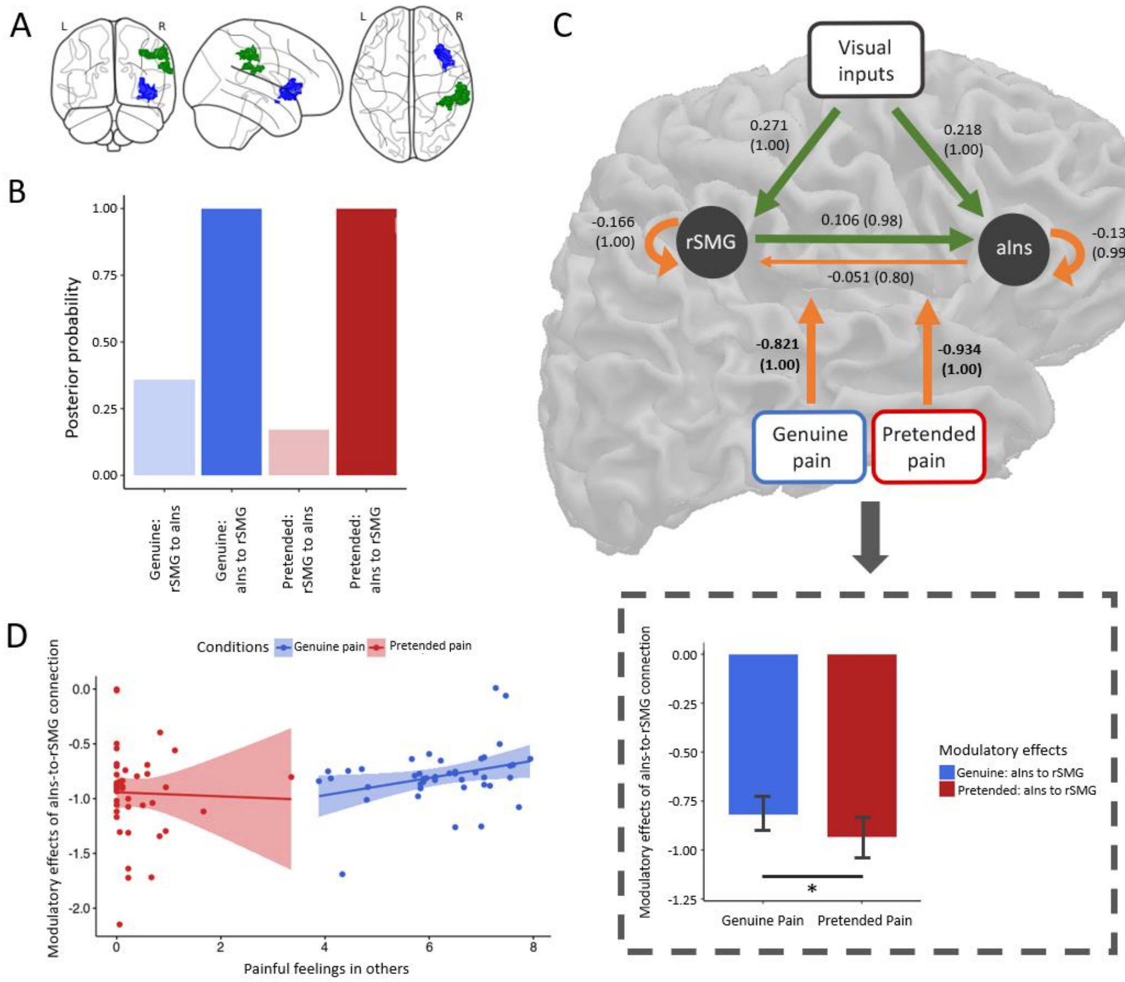

Figure 3. DCM results and brain-behavior analyses. (A) ROIs included in the DCM: aIns (blue; peak: [33, 29, 2]) and rSMG (green; peak: [41, -39, 42]). (B) Posterior probability of modulatory effects for the genuine pain and the pretended pain. (C) The group-average DCM model. Green arrows indicate neural excitation, and orange arrows indicate neural inhibition. Importantly, we found strong evidence of inhibitory effects on the connection of aIns to rSMG for both the genuine pain condition and the pretended pain condition. Values without the bracket quantify the strength of connections and values in the bracket indicate the posterior probability of connections. All DCM parameters of the optimal model showed greater than a 95% posterior probability (i.e., strong evidence) except for the intrinsic connection of aIns to rSMG (*pp* = 0.80). Paired sample t-test showed less inhibitory effects of the aIns-to-rSMG connection for the genuine pain than the pretended pain. This result is highlighted with a grey rectangular. Data are mean ± 95% CI. (D) The multiple linear regression model revealed a positive correlation between the inhibitory effect and painful feelings in others and not with the other two ratings for genuine pain but no correlation for pretended pain.

Discussion

In this study, we developed and used a novel experimental paradigm in which participants watched video clips of persons who supposedly either genuinely experienced or merely pretended to be in strong pain. Combining mass-univariate analysis with effective connectivity (DCM) analyses, our study provides evidence on the distinct neural dynamics between regions suggestive of affect processing (i.e., aIns and aMCC) and self-other distinction (i.e., rSMG) for genuinely sharing vs. responding to pretended, non-genuine pain. With this, we aimed to clarify two main questions: First, whether neural responses in areas such as the aIns and aMCC to the pain of others are indeed related to a veridical sharing of affect, as opposed to simply tracking automatic responses to salient affective displays. And second, how processes related to self-other distinction, implemented in the rSMG, enable appropriate empathic responses to genuine vs. merely pretended affective states.

The mass-univariate analyses suggest that the increased activity in aIns for genuine pain as opposed to pretended pain properly reflects affect sharing. As aforementioned, the network of affective sharing and certain domain-general processes (e.g., salience detection and automatic emotion processing) overlap in aIns and aMCC (Zaki et al., 2016, for review). This indicates that indeed, part of the activation in these areas could be related to perceptual salience, which is why it has been widely debated as a potential confound of empathy and affect sharing models (Zaki et al., 2016, for review; Lamm et al., 2019, for review). However, when comparing genuine pain versus pretended pain, activity in these areas was not only found to be stronger in response to genuine pain, but the increased activation in aIns was also selectively correlated with ratings of self-oriented unpleasantness and was not correlated with either other-related painful expressions or painful feelings in terms of the regression analysis. That only aIns and not also aMCC shows such correlation may be explained by previous studies, according to which aIns is more specifically associated with affective representations, while the role of aMCC rather seems to evaluate and regulate emotions that arise due to empathy (Fan et al., 2011; Lamm et al., 2011; Jauniaux et al., 2019). Taken together, the activation and brain-behavior findings provide evidence that responses in aIns (and to a lesser extent also the aMCC) are not simply automatic responses triggered by perceptually salient events (otherwise the increased aIns activation should also be explained by other behavioral ratings in the sense of shared influence by domain-general effects). Rather, they seem to track the actual affective states of the other person, and thus the shared neural representation of that response (see Zhou et al., 2020, for similar recent conclusions based on multi-voxel pattern analyses). Our findings are also in line with the proposed model of Coll et al. (2017), which suggests that affect sharing is the consequence of emotion identification. More specifically, while part of the activation in the aIns and aMCC is indeed related to an (presumably earlier) automatic response, the added engagement of these areas once they have identified the pain as genuine shows that only in this condition, they then also engage in proper affect sharing. Ideally, one should be able to discern these processes in time, but neither the temporal resolution of our fMRI measurements nor the paradigm in which we always announced the conditions beforehand would have been sensitive enough to do so. Thus, future studies including complementary methods such as EEG and MEG, and tailored experimental designs are needed to pinpoint the exact sequence of processes engaged in automatic affective responses vs. proper affect sharing.

Beyond higher activation in affective nodes supporting (pain) empathy, increased activation was also found in rSMG. The inferior parietal lobule was shown to be generally engaged in selective attention, action observation and imitating emotions (Bach et al., 2010; Pokorny et al., 2015; Gola et al., 2017; Hawco et al., 2017). Importantly, a specific role in affective rather than cognitive self-other distinction has been identified for rSMG (Silani et al., 2013; Steinbeis et al., 2015; Bukowski et al., 2020). Based on such findings, it has been proposed that the rSMG allows for a rapid switching between or the integration of self- and other-related representations, as two processes that may underpin the functional basis of successful self-other distinction (Lamm et al., 2016, for review). Theoretical models of empathy and related socio-affective responses suggest that such regulation is especially important to avoid so-called empathic over-arousal, which would shift the focus away from empathy and the other’s needs, towards taking care of one’s own personal distress (Batson et al., 1987, for review; Decety & Lamm, 2011, for review). Concerning the current findings, we thus propose that the higher rSMG engagement in the genuine pain condition reflects an increasing demand for self-other distinction imposed by the stronger shared negative affect experienced in this condition.

Beyond these differences in the magnitude of rSMG activation, the DCM analysis demonstrated less inhibition on the aIns-to-rSMG connection for genuine pain compared to pretended pain. Note that our focus on the right aIns rather than bilateral aIns was because it is located in the same hemisphere as the right SMG. Various theoretical accounts suggest that areas such as the aIns and rSMG may play a key role in comparing self-related information with the sensory evidence (Decety & Lamm, 2007, for review; Seth, 2013, for review). According to recent theories on predictive processing (Clark, 2013, for review) and active inference (Friston, 2010, for review), the brain can be regarded as a “prediction machine”, in which the top-down signals pass over predictions and the bottom-up signals convey prediction errors across different levels of cortical hierarchies (Chen et al., 2009; Friston, 2010, for review; Bastos et al., 2015). It is suggested that these top-down predictions are mediated by inhibitory neural connections (Zhang et al., 2008; Bastos et al., 2015; Miska et al., 2018). Our findings align with such views, by suggesting that the inhibitory connection from aIns to rSMG can be explained as the predictive mismatch between the top-down predictions of self-related information (e.g., personal affect) and sensory inputs (e.g., pain facial expressions). This suppression of neural activity leads to an explaining away of incoming bottom-up prediction error. This is reflected by the absence of any condition-dependent modulatory effects on the rSMG to aIns connection, suggesting that the influence of the task conditions is sufficiently modeled by the predictions from aIns to rSMG. Therefore, the stronger inhibition for pretended pain, compared to genuine pain, could indicate a higher demand to overcome the mismatch between the visual inputs and the agent's prior beliefs and contextual information about the situation (i.e., "this person looks like in pain, but I know he/she does not actually feel it"). We speculate that a dynamic interaction between sensory-driven and control processes is underlying the modulatory effect: when individuals realized after an initial sensory-driven response to the facial expression that it was not genuinely expressing pain, control and appraisal processes led to a reappraisal of the triggered emotional response, and thus a dampening of the unpleasantness. The reduced inhibition in the genuine pain condition could moreover be a mechanism that explains the higher rSMG activation in this condition.

Model comparison showed that the best model to explain the inhibitory effect with the behavioral ratings for both the genuine and pretended pain is the model without interactions between ratings. That is, if any behavioral rating contributed to the modulation of aIns to rSMG, the effect would be more likely coming from single ratings rather than their interactions. Specifically, we found the strength of the inhibitory effect in the genuine pain condition to correlate with ratings of painful feelings in others, but not with the ratings of pain expression in others or unpleasantness in self. For the pretended pain condition none of the ratings showed a correlation. The latter could in principle be due to a lack of variation in the ratings (which by way of the design were mostly close to zero or one). We deem it more plausible, though, that the correlation findings provide further evidence that the modulation of aIns to rSMG is implicated in encoding others' emotional states, which serves as a functional foundation for self-other processing when participants engaged in genuine affect sharing. This regulation cannot be totally attributed to domain-general processes, otherwise other ratings should have also explained this variation. It is also interesting to note that the found correlation relates to cognitive evaluations of the other's pain rather than to own affect, as tracked by the unpleasantness in self-ratings. This would to some extent be in line with DCM findings by Kanske et al. (2016). These authors found that the inhibition of the temporoparietal junction (TPJ) by the aIns was linked to interactions between Theory of Mind (ToM) and empathic distress, i.e., the interaction of "cognitive" vs. "affective" processes engaged in understanding others' cognitive and affective states. Note that the right TPJ is an overarching area involved in self-other distinction of which rSMG is considered a part or at least closely connected to (Decety & Lamm, 2007, for review).

The correlations between the DCM inhibitory effect and empathic traits assessed via questionnaires provide further refinements for the relevance of rSMG in implementing self-other distinction to allow for an appropriate empathic response. When participants shared genuine affect, the inhibitory effect on the aIns to rSMG connection was positively correlated with affective ability and negatively correlated with affective reactivity. Affective ability reflects the capacity to subjectively share emotions with others, while affective reactivity plays a role in the susceptibility to vicarious distress and thus to more automatic responses to another's emotion (Batchelder et al., 2017). Again, as for the correlations with the three rating scales, we did not find correlations of empathic traits for the pretended pain condition. Taken together, the DCM results and their qualification by the correlation findings suggest that in the genuine pain condition, which requires an accurate sharing of pain, rSMG interacts with aIns to achieve "affective-to-affective" self-other distinction – i.e., disambiguating affective signals originating in the self from those attributable to the other person. The aIns to rSMG connection in the pretended pain condition may reflect a related, yet slightly distinct mechanism. Here, it seems that "cognitive-to-affective" self-other distinction is at play, which helps resolve conflicting information between the top-down contextual information (i.e., that the demonstrator is not actually in pain) from what seems an unavoidable affective response to the highly salient perceptual cue of the facial expression of pain. Given our behavioral and trait data did not allow us to distinguish more precisely between these different types of self-other distinction, this however remains an interpretation and a hypothesis that will require further investigation. This inhibitory effect might be related to socioemotional disturbances of individuals with autism spectrum disorders (ASD), who show impairments in social cognition, including self- and other-related processing (Hoffmann et al., 2016; Lamm et al., 2016, for review). It is thus likely that ASD individuals exhibit distinct inhibition of the aIns-to-rSMG connectivity pattern compared to healthy controls. Further research with ASD individuals is required.

One potential limitation of the study could be the slightly higher ratings of other-oriented pain expressions for genuine pain, which were hypothesized to have no difference, as compared to pretended pain. As we found the enhanced aIns activation in the genuine pain condition mainly tracked personal unpleasantness rather than perceptually domain-general processes, and because the effect size of the pain expression difference was much smaller than for the affect ratings, we consider this difference did not fundamentally influence the interpretation of our findings. An additional limitation was that our study design did not aim to explicitly quantify self-other distinction. Rather, in line with previous research and based on our theoretical framework and rationale, we inferred the engagement of this process from the experimental conditions and the associated behavioral and neural responses. We expect our findings to prompt and inform future research designed to quantify and experimentally disentangle self- and other-related processes more explicitly. Note though that our participants and the targets shown in the videos were balanced with respect to their sex/gender. Yet sex/gender effects (for a review, though, see Christov-Moore et al., 2014) were outside the scope of the current study, and we thus did not perform any sex/gender-related analyses.

In conclusion, the current study advances our understanding of two main aspects of empathy. First, we provide evidence that empathy-related responses in the aIns can indeed be linked to affective sharing, rather than attributing them to responses triggered only by perceptual saliency. Second, we show how aIns and rSMG are orchestrated to track what another person really feels, thus enabling us to appropriately respond to their actual needs. Beyond these basic research insights, our study provides novel avenues for clinical application, and the investigation of contextual and interpersonal factors in the accurate diagnosis of pain and its expression.

Materials and Methods

Participants

Forty-eight participants took part in the study. Five of them were excluded because of excessive head motion (> 15% scans with the frame-wise displacement over 0.5 mm in one session). Data of the remaining 43 participants (21 females; age: Mean = 26.72 years, S.D. = 4.47) were entered into analyses. This sample size was determined on a priori power analysis in Gpower 3.1 (Faul et al., 2007). We assumed a medium effect size of Cohen's d = 0.5. After calculation, the minimum sample size statistically required for this study was 34 ($\alpha = 0.05$, two-tailed, $1-\beta = 0.80$). Participants were pre-screened by an MRI safety-check questionnaire, assuring normal or corrected to normal vision and no presence or history of neurologic, psychiatric, or major medical disorders. All participants were being right-handed (self-reported) and provided written consent including post-disclosure of any potential deception. The study was approved by the ethics committee of the Medical University of Vienna and was conducted in line with the latest version of the Declaration of Helsinki (2013).

Manipulation of facial expressions

As part of our study we developed a novel experimental design and corresponding stimuli, which consisted of video clips showing different demonstrators ostensibly in four different situations: 1) Genuine pain: the demonstrator's right cheek was penetrated by a hypodermic needle attached to a syringe, and the demonstrator's facial expression changed from neutral to a strongly painful facial expression. 2) Genuine no pain: the demonstrator maintained a neutral facial expression when a Q-tip fixed on the backend of the same syringe touched their right cheek. 3) Pretended pain: the demonstrator’s right cheek was approached by the same syringe and the hypodermic needle, with the latter covered by a protective cap; upon touch by the cap, the demonstrator’s facial expression changed from neutral to a strongly painful facial expression. 4) Pretended no pain: the demonstrator maintained a neutral facial expression when a Q-tip fixed on the backend of the same syringe touched their right cheek.

To create these stimuli, we recruited 20 demonstrators (10 females), with experience in acting, and filmed them in front of a dark blue background. An experimenter who stood on the right side of the demonstrators, but of whom only the right hand holding the syringe could be seen, administered the injections and touches. Unbeknownst to the participants, all painful expressions were acted, as the needle was a telescopic needle (i.e., a needle that seemed to enter the cheek upon contact, but in reality, was invisibly retracting into the syringe). The reason for using a protective cap in the pretended pain condition was to match the perceptual situation that an aversive object was approaching a body part in both pain conditions. In all situations, the demonstrator was instructed to look naturally towards the camera 1.5 m in front of them. As soon as the needle or the cap touched the demonstrator’s cheek, the demonstrator made a painful facial expression, as naturally and vividly as possible. In the neutral control conditions, demonstrators maintained a neutral facial expression when a Q-tip fixed at the backend of the syringe touched their cheek. Again, a syringe with a needle attached to the other end was used to perceptually control for the presence of an aversive object in all four conditions. Note that in another set of conditions, demonstrators showed disgusted or neutral expressions. Data from these conditions will be reported elsewhere. All demonstrators signed an agreement that their video clips and static images could be used for scientific purposes.

Stimulus validation and pilot study

To validate the stimuli, we performed an online validation study with N = 110 participants, who were asked to rate a total of 120 video clips of 2 s duration of the two conditions (60 of each condition) showing painful expressions (i.e., the genuine and the pretended pain conditions). The main aim of the validation study was to identify a set of demonstrators that expressed pain with comparable intensity and quality, and whose pain expressions in the genuine and pretended conditions were comparable. After each video clip, participants rated three questions on a visual analog scale with 9 tick-marks and the two end-points marked as “almost not at all” to “unbearable”: 1) How much pain did the person express on his/her face? 2) How much pain did the person actually feel? 3) How unpleasant did you feel to watch the person in this situation? The order of these three questions was pseudo-randomized. Moreover, eight catch trials randomly interspersed across the validation study to test whether participants maintained attention to the stimuli. Here, participants were asked to correctly select the demonstrator they had seen in the last video, between two static images of the correct and a distractor demonstrator displayed side by side, both showing neutral facial expressions.

The validation study was implemented within the online survey platform SoSci Survey (https://www.soscisurvey.de), with a study participation invite published on Amazon Mechanical Turk (https://www.mturk.com/), a globally commercial platform allowing for online testing. Survey data of 62 out of 110 participants (34 females; age: Mean = 28.71 years, S.D. =10.11) were entered into analysis (inclusion criteria: false rate for the test questions < 2/8, survey duration > 20 min and < 150 min, and the maximum number of continuous identical ratings < 5). Based on this validation step, we had to exclude videos of 6 demonstrators (3 females) for which participants showed a significant difference in painful expressions in others between the genuine pain and the pretended pain conditions. As a result of this validation, videos of 14 demonstrators (7 females), which showed no difference in the pain expression rating between genuine and pretended conditions, and which overall showed comparable mean ratings in all three ratings, were selected for the subsequent pilot study.

In the pilot study, 47 participants (24 females; age: Mean = 26.28 years, S.D. = 8.80) were recruited for a behavioral experiment in the behavioral laboratory. The aim was to verify the experimental effects and the feasibility of the experimental procedures that we intended to use in the main fMRI experiment, as well as to identify video stimuli that may not yield the predicted responses. Thus, all four conditions described above were presented to the participants. Participants were explicitly instructed that they would watch other persons' genuine painful expressions in some blocks, while in other blocks, they would see other persons acting out painful expressions (recall that in reality, all demonstrators had been actors, and the information about this type of necessary deception was conveyed to participants at the debriefing stage). They would see all demonstrators' neutral expressions as well. Participants were instructed to rate the three questions mentioned above. Upon screening for video clips that showed aberrant responses, we excluded videos of two demonstrators (1 female), for whom the pain *expression* rating difference between the pretended vs. genuine expressions was large. 48 videos of 12 demonstrators entered the following analyses. Three separate repeated-measures ANOVAs were respectively performed for the three rating questions. For the main effect of *genuineness* (genuine vs. pretended), it was not significant and low in effect size for painful expressions in others ($F_{\text{genuineness}}$ (1, 46) = 2.939, p = 0.093, η^2 = 0.060), but was significant with high effect size for the painful feelings in others ($F_{\text{genuineness}}$ (1, 46) = 280.112, p < 0.001, η^2 = 0.859) as well as the unpleasantness in self ($F_{\text{genuineness}}$ (1, 46) =43.143, p < 0.001, η^2 = 0.484). The main effects of *pain* (pain vs. no pain) for all three questions were found significant with high effect size (the smallest effect size was for the rating of unpleasantness in self, F_{pain} (1, 46) = 82.199, p < 0.001, η^2 = 0.641). Our pilot study thus a) provided assuring evidence that the novel experimental paradigm worked as expected, and b) made it possible to select video clips that we could match for the two conditions (i.e., genuine pain and pretended pain). More specifically, as expected and required for the main study, participants rated the painfulness of the demonstrators to be substantially higher when it was genuine as compared to those that were pretended, and this also resulted in much higher unpleasantness experienced in the self. It is worth noting that, the two conditions did not differ with respect to the ratings of the painful facial expressions, implying that putative differences in ratings as well as the subsequent brain imaging data could only be attributed to the contextual appraisal of the demonstrators' actual painful states, rather than the differences in facial pain perception. Based on this pilot study, we thus decided on video clips of 12 demonstrators (6 females) in the main fMRI experiment.

Experimental design and procedure of the fMRI study

The experiment was implemented using Cogent 2000 (version 1.33; http://www.vislab.ucl.ac.uk/cogent_2000.php). MRI scanning took place at the University of Vienna MRI Center. Once participants arrived at the scanner site, an experimenter instructed them that they would watch videos from the four conditions outlined above. Participants were explicitly instructed to recreate the feelings of the demonstrators shown in the videos as vividly and intensely as possible. Based on the validation and pilot study, the painful expressions for the genuine and pretended conditions were matched. We also counterbalanced the demonstrators appearing in the genuine and pretended conditions across participants, thus controlling for differences in behavioral and brain response that could be explained by differences between the stimulus sets. Note that, all video clips were validated and piloted multiple times to ensure the experimental effect (details can be found in the section above).

The participant performed the fMRI experiment in two runs (Figure 1A and 1B). Each run was composed of two blocks showing genuine pain and two blocks showing pretended pain. In each block, the participant watched nine video clips containing both painful and neutral videos. To remind participants' the condition of the upcoming block, a label of 4 s duration appeared at the beginning of each block, showing either "genuine" or "pretended" (in German). Each trial started with a fixation cross (+) presented for 4 – 7 s (in steps of 1.5 s, Mean = 5.5 s). After that, the video (duration = 2 s) was played. A short jitter was inserted after the video for 0.5 – 1.0 s (in steps of 0.05 s, Mean = 0.75 s). After the jitter, the following three questions were displayed (in German) one after the other in a pseudo-randomized order: 1) How much pain did the person express on his/her face? 2) How much pain did the person actually feel? 3) How unpleasant did you feel to watch the person in this situation? Beneath each question, a visual analog scale ranging from 0 (not at all) to 8 (unbearable) with 9 tick-marks was positioned. The participant moved the marker along the scale by pressing the left or right keys on the button box, and they pressed the middle key to confirm their answer. The marker initially was always located at the midpoint ("4") of the scale. When the confirmed key was pressed, the marker turned from black to red. All ratings lasted for 4 s even when the participant pressed the confirmed key before the end of this period. Between the two runs, the participant had a short break (1-2 min).

Before entering the scanner, participants conducted practice trials on the computer to get familiarized with the button box and the experimental interface. After that, participants were moved into the scanner and performed the task. Following the functional imaging runs, a 6.5 min structural scanning was employed. When participants finished the scanning session, they were scheduled for a date to complete three questionnaires in the lab: the Empathy Components Questionnaire (ECQ) (Batchelder, 2015; Batchelder et al., 2017), the Interpersonal Reactivity Index (IRI) (Davis, 1980), and the Toronto Alexithymia Scale (TAS) (Bagby et al., 1994). For the ECQ, there are 27 items in total to be categorized into five subscales: cognitive ability, cognitive drive, affective ability, affective drive, and affective reactivity, using a 4-point Likert scale ranging from 1 ("strongly disagree") to 4 ("strongly agree") (Batchelder, 2015; Batchelder et al., 2017). For the IRI, there are 28 items divided into four subscales: perspective taking, fantasy, empathic concern, and personal distress, using a 5-point Likert scale ranging from 0 ("does not describe me well") to 4 ("describes me very well") (Davis, 1980). For the TAS, there are 20 items and three subscales - difficulty describing feelings, difficulty identifying feelings, and externally oriented thinking, using a 5-point Likert scale ranging from 1 ("strongly disagree") to 5 ("strongly agree") (Bagby et al., 1994). The average interval between the scanning session and the lab survey was one week. The participant was debriefed after completing the whole study.

Behavioral data analysis

We applied repeated-measures ANOVAs to investigate the main effects and the interaction of the two factors genuine vs. pretended and pain vs. no pain, using SPSS (version 26.0; IBM). Furthermore, we conducted Pearson correlations to examine whether ratings of painful feelings in others were correlated with unpleasantness in self for the genuine pain and the pretended pain. The correlation coefficients were further compared using a bootstrap approach with the R package bootcorci (https://github.com/GRousselet/bootcorci).

fMRI data acquisition

fMRI data were collected using a Siemens Magnetom Skyra MRI scanner (Siemens, Erlangen, Germany) with a 32-channel head coil. Functional whole-brain scans were collected using a multiband-accelerated T2*-weighted echoplanar imaging (EPI) sequence (multiband acceleration factor = 4, interleaved ascending acquisition in multi-slice mode, 52 slices co-planar to the connecting line between anterior and posterior commissure, TR = 1200 ms, TE = 34 ms, acquisition matrix = 96 × 96 voxels, FOV = 210 × 210 mm2, flip angle = 66°, inter-slice gap = 0.4 mm, voxel size = 2.2 × 2.2 × 2 mm3). Two functional imaging runs, each lasting around 16 min (~800 images per run), were performed. Structural images were acquired with a magnetization-prepared rapid gradient-echo (MPRAGE) sequence (TE/TR = 2.43/2300 ms, flip angle = 8°, ascending acquisition, single-shot multi-slice mode, FOV= 240 × 240 mm2, voxel size = 0.8×0.8×0.8 mm3, 208 sagittal slices, slice thickness = 0.8 mm).

fMRI data processing and mass-univariate functional segregation analyses

Imaging data were preprocessed with a combination of Nipype (Gorgolewski et al., 2011) and MATLAB (version R2018b 9.5.0; MathWorks) with Statistical Parametric Mapping (SPM12; https://www.fil.ion.ucl.ac.uk/spm/software/spm12/). Raw data were imported into BIDS format (http://bids.neuroimaging.io/). Functional data were subsequently preprocessed using slice timing correction to the middle slice (Sladky et al., 2011), realignment to the first image of each session, co-registration to the T1 image, segmentation between grey matter, white matter and cerebrospinal fluid (CSF), normalization to MNI template space using Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL) toolbox (Ashburner, 2007), and smoothing with a 6 mm full width at half-maximum (FWHM) three-dimensional Gaussian kernel.

To improve data quality, we performed data scrubbing of the functional scans for those whose frame-wise displacements (FD) were over 0.5 mm (Power et al., 2012; Power et al., 2014). In other words, we identified individual outlier scans and flagged the volume indices as nuisance regressors in the general linear model (GLM) for the first-level analysis.

In order to perform mass-univariate functional segregation analyses, a first-level GLM design matrix was created and composed of two identically modeled runs for each participant. Seven regressors of interest were entered in each model: stimulation phase of the four conditions (i.e., genuine pain, genuine no pain, pretended pain, pretended no pain; 2000 ms), rating phase of the three questions (i.e., painful expressions in others, painful feelings in others, and unpleasantness in self; 12000 ms). Six head motion parameters and the scrubbing regressors (FD > 0.5 mm; if applicable) were additionally entered as nuisance regressors. Individual contrasts of the four conditions and the three ratings (all across the two runs) against implicit baseline were respectively created.

On the second level, a flexible factorial design was employed to perform the group-level analysis. The design included three factors: a between-subject factor (i.e., subject) that was specified independent and with equal variance, a within-subject factor (i.e., genuine or pretended) that was specified dependent and with equal variance, and a second within-subject factor (i.e., pain or no pain) that was specified dependent and with equal variance (Gläscher & Gitelman, 2008). Three contrasts were computed: (1) main effect of genuine: pain – no pain, (2) main effect of pretended: pain – no pain, and (3) interaction: genuine (pain – no pain) – pretended (pain – no pain). We applied an initial threshold of $p < 0.001$ (uncorrected) at the voxel level and a family-wise error (FWE) correction ($p < 0.05$) at the cluster level. The cluster extent threshold was determined by the SPM extension "cp_cluster_Pthresh.m" (https://goo.gl/kjVydz).

Brain-behavior relationships

A multiple regression model was built on the group level to investigate the relationship between specific brain activations and behavioral ratings. In this model, the contrast genuine pain – pretended pain was set as the dependent variable, and differences between conditions for three behavioral ratings were specified as independent variables. The reason that we used the comparison between conditions for both brain signals and behavioral ratings was to control for potential effects of perceptual salience. All covariates were mean-centered. An intercept was added in the model. To test whether the order of entering ratings into the regression model influence the results, we performed five additional regression analyses with all possible orders of three ratings. The results were consistent across all six regression models, and we only showed the result for one regression (i.e., expression + feeling + unpleasantness) in the Results section. Note that, we performed the regression model with the contrast genuine pain – pretended pain instead of the more exhaustive contrast genuine (pain - no pain) - pretended (pain – no pain), and this was because the genuine and the pretended pain conditions were the main focus of our work. Moreover, the pain contrast showed more robust (in terms of statistical effect size) and widespread activations across the brain, making it more likely to pick up possible brain-behavior relationships. The same threshold as above was applied in this analysis.

We aimed to assess these brain-behavior relationships for the following regions of interest (ROI): 1) aIns and aMCC, i.e., two regions associated with affective processes and specifically with empathy for pain, 2) rSMG, an area implicated in affective self-other distinction. The ROI masks were defined as the conjunction of the averaging contrast between genuine and pretended: pain – no pain (threshold:

voxel-wise FWE correction, $p < 0.05$) and the anatomical masks created by the Wake Forest University (WFU) Pick Atlas SPM toolbox (http://fmri.wfubmc.edu) with the automated anatomical atlas (AAL). The ROI masks were created with Marsbar ROI Toolbox implemented in SPM12 (Brett et al., 2002). Note that we specifically selected the ROIs this way, such that they were orthogonal (i.e., independent) to the subsequent analyses of interest. As exploratory analyses found significant correlations mainly in aIns, rather than in aMCC, we will focus in the results section on two ROIs: the right aIns and the rSMG. Focusing on the right aIns instead of the left one was because the right aIns is on the ipsilateral hemisphere as rSMG.

Analyses using dynamic causal modeling (DCM)

To investigate the functional network involved in affective processes and self-other distinction and how it was modulated by our experimental manipulations (i.e., genuine pain and pretended pain), we used DCM to estimate the effective connectivity between the ROIs based on the tasked-related brain responses (Stephan & Friston, 2010, for review). The DCM analyses were conducted with DCM12.5 implemented in SPM12 (v. 7771). Firstly, we extracted individual time series separately for each ROI. To ensure the selected voxels engaged in a task-relevant activity but not random signal fluctuations, we determined the voxels both on a group-level threshold and an individual-level threshold (Holmes et al., 2020). An initial threshold was set as $p < 0.05$, uncorrected. The significant voxels in the main effect of genuine pain and pretended pain were further selected by an individual threshold. For each participant, an individual peak coordinate within the ROI mask was searched and an individual mask was consequently defined using a sphere of the 6 mm radius around the peak. As a result, the individual time series for each ROI was extracted from the significant voxels of the individual mask and summarized by the first eigenvariate. One participant was excluded as no voxels survived significance testing. Secondly, we specified three regressors of interest: genuine pain, pretended pain, and the video input condition (the combination of genuine pain and pretended pain). That we did not specify no-pain conditions was because 1) the pain conditions were our main focus, and 2) adding no-interest conditions would inevitably increase the model complexity. Then, a fully connected DCM model for each participant was created. Three parameters were specified: 1) bidirectional connections between regions and self-connections (matrix A), 2) modulatory effects (i.e., genuine pain and pretended pain) on the between-region connections (matrix B), and 3) driving inputs (i.e., the video input condition) into the model on both regions (matrix C) (Zeidman et al., 2019a). To remain parsimonious, we did not set modulatory effects on the self-connections in matrix A. Then the full DCM model was individually estimated. Finally, group-level DCM inference was performed using parametric empirical Bayes (Zeidman et al., 2019b). We conducted an automatic search over the entire model space (max. n =256) using Bayesian model reduction (BMR) and random-effects Bayesian model averaging (BMA), resulting in a final group model that takes accuracy, complexity, and uncertainty into account (Zeidman et al., 2019b). The threshold of the Bayesian posterior probability was set to *pp* > 0.95 (i.e., strong evidence) but we reported all parameters above *pp* > 0.75 (i.e., positive evidence) for full transparency of the DCM results. Finally, a paired sample t-test was performed to compare modulatory effects between the genuine pain and the pretended pain conditions.

To probe whether task-related modulatory effects were associated with behavioral measurements, we performed multiple linear regression analyses of modulatory parameters with, 1) the three behavioral ratings, and 2) the empathy-related questionnaires (i.e., IRI, ECQ, and TAS). We set up regression models for the genuine pain condition and the pretended pain condition, respectively, in which the DCM parameters of modulatory effects were determined as dependent variables and the ratings of painful expressions in others, painful feelings in others, and unpleasantness in self as independent variables. Considering that interactions between behavioral ratings might contribute to the regression model, we tested five regression models (with and without interaction; See Supplementary Table 1) for both genuine pain and pretended pain. Results showed that for both genuine pain and pretend pain, the model without any interaction outperformed other models. The results of the winning multiple regression model are reported in the Results section. We performed additional two regression models for both conditions in which DCM modulatory effects were set as dependent variables and scores of each subscale of all questionnaires were set as independent variables, respectively. Considering the number of independent variables was relatively large (>10), we performed the analyses for questionnaires using a stepwise regression approach. As two participants did not complete all three questionnaires, we excluded their data from the regression analyses. The statistical significance of the regression analysis was set to *p* < 0.05. The multicollinearity for independent variables was diagnosed using the variance inflation factor (VIF) that measures the correlation among independent variables, in the R package car (https://cran.r-project.org/web/packages/car/index.html). Here we used a rather conservative threshold of VIF < 5 as a sign of no severe multicollinearity (Menard, 2002; James et al., 2013).

Acknowledgements

This work was supported by Chinese Scholarship Council (CSC) Grant (201604910515) and Vienna Doctoral School in Cognition, Behavior and Neuroscience (VDS CoBeNe) completion grant fellowship to Y.Z.; the Vienna Science and Technology Fund (WWTF VRG13-007) to C.L., and the Austrian Science Fund (FWF P 32686) to C.L. and M.R.. We thank Michael Schnödt, Lukas Repnik, Elisa Warmuth, Betty

Geidel, Phan Ri, Sven Sander, Gvantsa Gogisvanidze, Robert Meyka, Anja Tritt, Tim Reinboth, Boryana Todorova, and Sophia Shea for help with data acquisition.

Conflicts of interest

The authors declare no competing financial interests.

References

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage, 38*(1), 95-113. doi:https://doi.org/10.1016/j.neuroimage.2007.07.007

Bach, P., Peelen, M. V., & Tipper, S. P. (2010). On the Role of Object Information in Action Observation: An fMRI Study. *Cerebral Cortex, 20*(12), 2798-2809. doi:http://doi.org/10.1093/cercor/bhq026

Bagby, R. M., Taylor, G. J., & Parker, J. D. A. (1994). The twenty-item Toronto Alexithymia Scale: II. Convergent, discriminant, and concurrent validity. *Journal of Psychosomatic Research, 38*(1), 33-40. doi:http://doi.org/10.1016/0022-3999(94)90006-X

Bastos, A. M., et al. (2015). A DCM study of spectral asymmetries in feedforward and feedback connections between visual areas V1 and V4 in the monkey. *NeuroImage, 108*, 460-475. doi:https://doi.org/10.1016/j.neuroimage.2014.12.081

Batchelder, L. (2015). *Characterising the components of empathy: implications for models of autism.* University of Bath.

Batchelder, L., Brosnan, M., & Ashwin, C. (2017). The Development and Validation of the Empathy Components Questionnaire (ECQ). *PLOS ONE, 12*(1), e0169185. doi:http://doi.org/10.1371/journal.pone.0169185

Batson, C. D., Fultz, J., & Schoenrade, P. A. (1987). Distress and Empathy: Two Qualitatively Distinct Vicarious Emotions with Different Motivational Consequences. *Journal of Personality, 55*(1), 19-39. doi:https://doi.org/10.1111/j.1467-6494.1987.tb00426.x

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). *Region of interest analysis using an SPM toolbox.* Paper presented at the 8th international conference on functional mapping of the human brain.

Bukowski, H., et al. (2020). When differences matter: rTMS/fMRI reveals how differences in dispositional empathy translate to distinct neural underpinnings of self-other distinction in empathy. *Cortex, 128*, 143-161. doi:https://doi.org/10.1016/j.cortex.2020.03.009

Chen, C. C., Henson, R. N., Stephan, K. E., Kilner, J. M., & Friston, K. J. (2009). Forward and backward connections in the brain: A DCM study of functional asymmetries. *NeuroImage, 45*(2), 453-462. doi:https://doi.org/10.1016/j.neuroimage.2008.12.041

Christov-Moore, L., et al. (2014). Empathy: Gender effects in brain and behavior. *Neuroscience & Biobehavioral Reviews, 46*, 604-627. doi:https://doi.org/10.1016/j.neubiorev.2014.09.001

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181-204. doi:http://doi.org/10.1017/S0140525X12000477

Coll, M.-P., et al. (2017). Are we really measuring empathy? Proposal for a new measurement framework. *Neuroscience & Biobehavioral Reviews, 83*, 132-139. doi:https://doi.org/10.1016/j.neubiorev.2017.10.009

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

Decety, J., & Lamm, C. (2007). The Role of the Right Temporoparietal Junction in Social Interaction: How Low-Level Computational Processes Contribute to Meta-Cognition. *The Neuroscientist, 13*(6), 580-593. doi:http://doi.org/10.1177/1073858407304654

Decety, J., & Lamm, C. (2011). Empathy versus Personal Distress: Recent Evidence from Social Neuroscience. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 199 - 213): MIT Press.

Fallon, N., Roberts, C., & Stancak, A. (2020). Shared and distinct functional networks for empathy and pain processing: A systematic review and meta-analysis of fMRI studies. *Social cognitive and affective neuroscience*. doi:https://doi.org/10.1093/scan/nsaa090

Fan, Y., Duncan, N. W., de Greck, M., & Northoff, G. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neuroscience & Biobehavioral Reviews, 35*(3), 903-911. doi:https://doi.org/10.1016/j.neubiorev.2010.10.009

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. doi:http://doi.org/10.3758/BF03193146

Forbes, P. A. G., & Hamilton, A. F. d. C. (2020). Brief Report: Autistic Adults Assign Less Weight to Affective Cues When Judging Others' Ambiguous Emotional States. *Journal of Autism and Developmental Disorders, 50*(8), 3066-3070. doi:http://doi.org/10.1007/s10803-020-04410-w

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127-138. doi:http://doi.org/10.1038/nrn2787

Gläscher, J., & Gitelman, D. (2008). Contrast weights in flexible factorial design with multiple groups of subjects. *SPM@ JISCMAIL. AC. UK) Sml, editor*, 1-12.

Gola, K. A., et al. (2017). A neural network underlying intentional emotional facial expression in neurodegenerative disease. *NeuroImage: Clinical, 14*, 672-678. doi:https://doi.org/10.1016/j.nicl.2017.01.016

Gorgolewski, K., et al. (2011). Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Frontiers in Neuroinformatics, 5*(13). doi:http://doi.org/10.3389/fninf.2011.00013

Gu, X., & Han, S. (2007). Attention and reality constraints on the neural processes of empathy for pain. *NeuroImage, 36*(1), 256-267. doi:https://doi.org/10.1016/j.neuroimage.2007.02.025

Hawco, C., et al. (2017). Neural Activity while Imitating Emotional Faces is Related to Both Lower and Higher-Level Social Cognitive Performance. *Scientific Reports, 7*(1), 1244. doi:http://doi.org/10.1038/s41598-017-01316-z

Hein, G., & Singer, T. (2008). I feel how you feel but not always: the empathic brain and its modulation. *Current Opinion in Neurobiology, 18*(2), 153-158. doi:https://doi.org/10.1016/j.conb.2008.07.012

Hoffmann, F., Koehne, S., Steinbeis, N., Dziobek, I., & Singer, T. (2016). Preserved Self-other Distinction During Empathy in Autism is Linked to Network Integrity of Right Supramarginal Gyrus. *Journal of Autism and Developmental Disorders, 46*(2), 637-648. doi:http://doi.org/10.1007/s10803-015-2609-0

Holmes, E., Zeidman, P., Friston, K. J., & Griffiths, T. D. (2020). Difficulties with Speech-in-Noise Perception Related to Fundamental Grouping Processes in Auditory Cortex. *Cerebral Cortex, 31*(3), 1582-1596. doi:http://doi.org/10.1093/cercor/bhaa311

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112): Springer.

Jauniaux, J., Khatibi, A., Rainville, P., & Jackson, P. L. (2019). A meta-analysis of neuroimaging studies on pain empathy: investigating the role of visual information and observers' perspective. *Social cognitive and affective neuroscience, 14*(8), 789-813. doi:https://doi.org/10.1093/scan/nsz055

Kanske, P., Böckler, A., Trautwein, F.-M., Parianen Lesemann, F. H., & Singer, T. (2016). Are strong empathizers better mentalizers? Evidence for independence and interaction between the routes of social cognition. *Social cognitive and affective neuroscience, 11*(9), 1383-1392. doi:http://doi.org/10.1093/scan/nsw052

Lamm, C., Bukowski, H., & Silani, G. (2016). From shared to distinct self-other representations in empathy: evidence from neurotypical function and socio-cognitive disorders. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 371*(1686), 20150083. doi:http://doi.org/10.1098/rstb.2015.0083

Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage, 54*(3), 2492-2502. doi:https://doi.org/10.1016/j.neuroimage.2010.10.014

Lamm, C., Meltzoff, A. N., & Decety, J. (2010). How do we empathize with someone who is not like us? A functional magnetic resonance imaging study. *J. Cognitive Neuroscience, 22*(2), 362–376. doi:http://doi.org/10.1162/jocn.2009.21186

Lamm, C., Rütgen, M., & Wagner, I. C. (2019). Imaging empathy and prosocial emotions. *Neuroscience Letters, 693*, 49-53. doi:https://doi.org/10.1016/j.neulet.2017.06.054

Mars, R. B., et al. (2011). Connectivity-Based Subdivisions of the Human Right "Temporoparietal Junction Area": Evidence for Different Areas Participating in Different Cortical Networks. *Cerebral Cortex, 22*(8), 1894-1903. doi:http://doi.org/10.1093/cercor/bhr268

Menard, S. (2002). *Applied logistic regression analysis* (Vol. 106): Sage.

Miska, N. J., Richter, L. M. A., Cary, B. A., Gjorgjieva, J., & Turrigiano, G. G. (2018). Sensory experience inversely regulates feedforward and feedback excitation-inhibition ratio in rodent visual cortex. *eLife, 7*, e38846. doi:http://doi.org/10.7554/eLife.38846

Pokorny, J. J., et al. (2015). The Action Observation System when Observing Hand Actions in Autism and Typical Development. *Autism Research, 8*(3), 284-296. doi:https://doi.org/10.1002/aur.1445

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage, 59*(3), 2142-2154. doi:https://doi.org/10.1016/j.neuroimage.2011.10.018

Power, J. D., et al. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage, 84*, 320-341. doi:https://doi.org/10.1016/j.neuroimage.2013.08.048

Rütgen, M., et al. (2015). Placebo analgesia and its opioidergic regulation suggest that empathy for pain is grounded in self pain. *Proceedings of the National Academy of Sciences, 112*(41), E5638-E5646. doi:https://doi.org/10.1073/pnas.1511269112

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences, 17*(11), 565-573. doi:https://doi.org/10.1016/j.tics.2013.09.007

Silani, G., Lamm, C., Ruff, C. C., & Singer, T. (2013). Right Supramarginal Gyrus Is Crucial to Overcome Emotional Egocentricity Bias in Social Judgments. *The Journal of Neuroscience, 33*(39), 15466-15476. doi:http://doi.org/10.1523/jneurosci.1488-13.2013

Sladky, R., et al. (2011). Slice-timing effects and their correction in functional MRI. *NeuroImage, 58*(2), 588-594. doi:https://doi.org/10.1016/j.neuroimage.2011.06.078

Steinbeis, N., Bernhardt, B. C., & Singer, T. (2015). Age-related differences in function and structure of rSMG and reduced functional connectivity with DLPFC explains heightened emotional egocentricity bias in childhood. *Social cognitive and affective neuroscience, 10*(2), 302-310. doi:https://doi.org/10.1093/scan/nsu057

Stephan, K. E., & Friston, K. J. (2010). Analyzing effective connectivity with functional magnetic resonance imaging. *WIREs Cognitive Science, 1*(3), 446-459. doi:https://doi.org/10.1002/wcs.58

Xiong, R.-C., et al. (2019). Brain pathways of pain empathy activated by pained facial expressions: a meta-analysis of fMRI using the activation likelihood estimation method. *Neural regeneration research, 14*(1), 172-178. doi:http://doi.org/10.4103/1673-5374.243722

Zaki, J., Wager, T. D., Singer, T., Keysers, C., & Gazzola, V. (2016). The Anatomy of Suffering: Understanding the Relationship between Nociceptive and Empathic Pain. *Trends in Cognitive Sciences, 20*(4), 249-259. doi:https://doi.org/10.1016/j.tics.2016.02.003

Zeidman, P., et al. (2019a). A guide to group effective connectivity analysis, part 1: First level analysis with DCM for fMRI. *NeuroImage, 200*, 174-190. doi:https://doi.org/10.1016/j.neuroimage.2019.06.031

Zeidman, P., et al. (2019b). A guide to group effective connectivity analysis, part 2: Second level analysis with PEB. *NeuroImage, 200*, 12-25. doi:https://doi.org/10.1016/j.neuroimage.2019.06.032

Zhang, M., et al. (2008). A Self-Regulating Feed-Forward Circuit Controlling C. elegans Egg-Laying Behavior. *Current Biology, 18*(19), 1445-1455. doi:https://doi.org/10.1016/j.cub.2008.08.047

Zhao, Y., Rütgen, M., Zhang, L., & Lamm, C. (2021). Pharmacological fMRI provides evidence for opioidergic modulation of discrimination of facial pain expressions. *Psychophysiology, 58*(2), e13717. doi:https://doi.org/10.1111/psyp.13717

Zhou, F., et al. (2020). Empathic pain evoked by sensory and emotional-communicative cues share common and process-specific neural representations. *eLife, 9*, e56929. doi:http://doi.org/10.7554/eLife.56929

Supplementary information

Supplementary Table 1. Model comparison of linear regression models with three behavioral ratings (independent variables) and the inhibitory effect (dependent variable) for genuine pain and pretended pain. Smaller AIC/BIC indicates better model fit. Results showed that M1 (without interaction; highlighted with underlining) was the best fitting model for both genuine pain and pretended pain.

Regression Model	**AIC**	**BIC**
Genuine pain		
M1 (expression + feeling + unpleasantness)	10.069	18.757
M2 (expression * feeling + unpleasantness)	11.342	21.768
M3 (expression + feeling * unpleasantness)	11.195	21.621
M4 (expression * unpleasantness + feeling)	11.126	21.552
M5 (expression * feeling * unpleasantness)	16.153	31.792
Pretended pain		
M1 (expression + feeling + unpleasantness)	51.041	59.919
M2 (expression * feeling + unpleasantness)	51.230	61.467
M3 (expression + feeling * unpleasantness)	52.643	63.069
M4 (expression * unpleasantness + feeling)	52.584	63.010
M5 (expression * feeling * unpleasantness)	55.200	70.839

The manuscript is submitted to *the Journal of Neuroscience*

Chapter 4 - Effective connectivity reveals distinctive patterns in response to others' genuine affective experience of disgust as compared to pain

Yili Zhao[1], Lei Zhang[1], Markus Rütgen[1,2], Ronald Sladky[1], Claus Lamm[1,2*]

[1] Social, Cognitive and Affective Neuroscience Unit, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria

[2] Vienna Cognitive Science Hub, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria

[3] Neuropsychopharmacology and Biopsychology Unit, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria

Abstract

Empathy is significantly influenced by the identification of others' emotions. In a recent study, we have found increased activation in the anterior insular cortex (aIns) that could be attributed to affect sharing rather than perceptual saliency, when seeing another person genuinely experiencing pain as opposed to merely acting to be in pain. This study further revealed effective connectivity between aIns and the right supramarginal gyrus (rSMG) to track what another person really feels. In the present study, we used a similar paradigm to investigate the corresponding neural signatures in the domain of empathy for disgust - with participants seeing others genuinely sniffing unpleasant odors as compared to pretending to smell something disgusting. Consistent with the previous findings on pain, we found stronger activations in aIns associated with affect sharing for genuine disgust compared with pretended disgust. However, instead of rSMG we found engagement of the olfactory cortex. Using dynamic causal modeling (DCM), we estimated the neural dynamics of aIns and the olfactory cortex between the genuine and pretended conditions. This revealed an increased excitatory modulatory effect for genuine disgust compared to pretended disgust. For genuine disgust only, brain-to-behavior regression analyses highlighted a link between the observed modulatory effect and the perspective-taking empathic trait. Altogether, the current findings complement and expand our previous work, by showing that perceptual saliency alone does not explain responses in the insular cortex. Moreover, it reveals that different brain networks are implicated in a modality-specific way when sharing the affective experiences associated with pain vs. disgust.

Significant statement

Others' feelings influence our own feelings, no matter whether these feelings are genuine or merely pretended. Prior research provided a window into the brain networks allowing us to empathize with and share the genuine pain of others. Here, we tested whether similar or distinct mechanisms are at play in the domain of disgust. Similar to pain, sharing genuine disgust specifically recruited the anterior insular cortex, confirming the specific role of this area as a key node for affect sharing. However, dissociating genuine from pretended disgust was linked to interactions between insular and the olfactory cortex, rather than the right supramarginal gyrus found for pain. These findings thus suggest both similar and distinct mechanisms in empathy for others' affective experiences.

Introduction

Our affective states are remarkably affected by the perceived feelings of others. A theoretical framework of empathy proposed by Coll et al. (2017) states that identification of another's emotion crucially contributes to the consequential sharing of feelings with that person. A recent study by our group has revealed that when individuals witness another person genuinely experiencing pain as compared to merely acting to be in pain, they attribute more painful feelings to that person and report experiencing stronger self-unpleasantness in response to the other's genuine pain (Zhao et al., 2021b).

On the neural level, that study found increased brain activations for the genuine compared to the pretended pain in the anterior insular cortex (aIns) and the anterior mid-cingulate cortex (aMCC), i.e., a network that has been consistently associated with affective responding in studies on self-experienced pain as well as empathy for pain (Lamm et al., 2011; Rütgen et al., 2015; Jauniaux et al., 2019; Xiong et al., 2019; F. Zhou et al., 2020; Fallon et al., 2020, for meta-analyses). One major contribution of our previous study is that we have shown aIns, a key node of this neural network, is indeed associated with affect sharing, rather than being driven by the perceptual saliency of the facial expressions of pain. Moreover, by means of dynamic causal modeling (DCM) analyses, distinctive effective connectivity of genuine pain vs. pretended pain has been found on the connection between aIns and the right supramarginal gyrus (rSMG), a region selectively related to affective self-other distinction (Silani et al., 2013; Steinbeis et al., 2015; Hoffmann et al., 2016; Bukowski et al., 2020). This suggests that the interaction of aIns and rSMG tracks how we identify and share the actual feelings of another person, allowing an observer to engage in appropriate affect sharing rather than simply responding to salient, yet possibly non-genuine displays of pain.

What remains an open question is whether these findings are specific to pain or could be extended to other aversive experiences. Among the array of aversive experiences, the emotion of disgust partially overlaps with pain regarding its neural mechanisms (Corradi-Dell'Acqua et al., 2016). Also, disgust and pain share similarities with respect to their facial expression (Zhao et al., 2021a) and are similarly important for survival and somatic protection (Sharvit et al., 2015; Sharvit et al., 2020). Particularly, research using multi-voxel pattern analysis (MVPA) shows overlapping brain maps in aIns and aMCC not only for self-experienced but also vicarious experiences of pain and disgust, suggesting a modality-independent representation of the unpleasantness shared by self-experienced aversive affect and empathy for such affect (Corradi-Dell'Acqua et al., 2016).

The aim of the present study was, thus, to replicate and expand the findings of our previous study on pain (Zhao et al., 2021b), but targeting the emotion of disgust. Specifically, participants watched video clips either presenting a person showing a disgust expression when sniffing something unpleasant, or merely displaying a disgust expression without genuinely smelling any unpleasant odor. We expected to find that 1) on the behavioral level, genuine disgust would result in higher other-oriented disgust ratings and self-oriented unpleasantness ratings; 2) on the neural level, aIns, aMCC, and rSMG would show stronger responses to the genuine disgust, as compared to pretended disgust; and 3) distinct patterns of aIns' effective connectivity with rSMG would be found, and explain the different empathic responses to genuine vs. pretended disgust in a similar way as for pain.

Materials and Methods

To maximize comparability, data collection for the current study had been planned and performed together with the study focusing on pain (Zhao et al., 2021b). Thus, all procedures of both studies (i.e., creation and validation of stimuli, the pilot study, and the main fMRI experiment) were exactly conducted in the same sessions and with the identical participant sample. We decided to analyze and report them separately for reasons of reporting complexity and as the two reports have a different focus. While the details about all procedures are fully documented in (Zhao et al., 2021b), for ease of access, we also summarize the main points relevant to the current study herein.

Participants

Forty-eight participants participated in this study. This sample size was estimated a priori using Gpower 3.1 (Faul et al., 2007), for which a minimum sample size statistically required for this study was 34 with a medium effect size of Cohen's $d = 0.5$ ($\alpha = 0.05$, two-tailed, $1-\beta = 0.80$). Three participants (only for the current study) were excluded because of excessive head motion (> 15% scans with the frame-wise displacement over 0.5 mm in one session; same criteria as the pain study). Data of the remaining 45 participants (21 females; age: Mean = 26.76 years, S.D. = 4.58) were entered into analyses. Participants had normal or corrected to normal vision and were pre-screened by an MRI safety-check questionnaire to assure no presence or history of neurologic, psychiatric, or major medical disorders. All participants reported being right-handed and signed the informed consent. The study was approved by the ethics committee of the Medical University of Vienna and was conducted in accordance with the latest revision of the Declaration of Helsinki (2013).

Manipulation of facial expressions

In strict analogy to the stimuli we created for pain, the stimuli we created for this study consisted of video clips showing different demonstrators ostensibly in four different situations: 1) Genuine disgust: the demonstrator sniffed dog feces in an opened bottle with a picture depicting dog feces on it; the demonstrator's facial expression changed from neutral to strongly disgusted. 2) Genuine no disgust:

the demonstrator sniffed cotton balls in an opened bottle with a picture depicting cotton balls on it; the demonstrator's facial expression maintained neutral. 3) Pretended disgust: the demonstrator sniffed dog feces in a closed bottle (covered by a cap) with a picture depicting dog feces on it; the demonstrator's facial expression changed from neutral to strongly disgusted. 4) Pretended no disgust: the demonstrator sniffed cotton balls in an opened bottle with a picture depicting cotton balls on it; the demonstrator's facial expression maintained neutral.

Twenty demonstrators (10 females), with experience in acting, were recruited for creating the stimuli of the current study. Each demonstrator signed the agreement of using their videos clips and static images for scientific purposes. An experimenter who stood on the right side of the demonstrators, of whom only the right hand holding the bottle could be seen, moved the bottle from the demonstrator's right side and stopped it just below the demonstrator's nose. Unbeknownst to the participants, all disgusted expressions were acted and the so-called "dog feces" were actually an odor-neutral object that resembled dog feces. As soon as the bottle was close enough to the demonstrator's nose (just below the right nostril), the demonstrator started to make a disgusted facial expression along with a slightly avoidant movement of their head, as naturally and vividly as possible. In the neutral control conditions, demonstrators maintained a neutral facial expression during the whole process of the bottle movement. Note that, the reason for presenting the pictures and supposed content of dog feces in both disgust conditions was because we deemed it essential to match the conditions in terms of the presence and visibility of an aversive disgusting object approaching the other person's face. Otherwise, any difference between conditions could be confounded by responses of participants to the presence vs. absence of a disgusting object and its explicit photographic display. Note that the pain condition of our previous work also followed this logic, with a needle covered by a plastic cap approaching the cheek.

Stimulus validation and pilot study

To validate the stimuli, 110 participants (59 females; age: Mean = 29.32 years, S.D. =10.17) were recruited and asked to rate a total of 120 video clips of 2 s duration of the two conditions (60 of each condition) showing disgusted facial expressions (i.e., the genuine and pretended disgust conditions). The main aim of the validation study was to identify a set of demonstrators that expressed disgust with comparable intensity and quality, and whose expressions of disgust in the genuine and pretended conditions were comparable. After each video clip, participants rated three questions on a visual analog scale with 9 tick-marks and the two end-points marked as "almost not at all" to "unbearable": 1) How much disgust did the person express on his/her face? 2) How much disgust did the person actually feel? 3) How unpleasant did you feel to watch the person in this situation? These questions were presented in a pseudo-randomized order. Moreover, we set eight catch trials to test whether participants maintained attention to the stimuli, in which participants were required to correctly choose the demonstrator they had seen in the last video, from two static images showing either the correct demonstrator's or a distractor's neutral facial expression side by side.

Data collection was performed with the online survey platform SoSci Survey (https://www.soscisurvey.de), and participants got access to the survey through a participation invite published on Amazon Mechanical Turk (https://www.mturk.com/). Survey data of 62 out of 110 participants (34 females; age: Mean = 28.71 years, S.D. =10.11) were entered into the analysis (inclusion criteria: false rate for the test questions < 2/8, survey duration > 20 min and < 150 min, and the maximum number of continuous identical ratings < 5). According to the validation analysis, videos of 6 demonstrators (3 females) were excluded for which participants showed a significant difference in perceived disgust expressions in others between genuine disgust and pretended disgust. As a result of this validation, videos of 14 demonstrators (7 females) were selected for the subsequent pilot study.

In the pilot study, a separate group (N =47, 24 females; age: Mean = 26.28 years, S.D. = 8.80) were recruited for a behavioral experiment in the behavioral laboratory. All conditions including the neutral conditions described above were presented to the participants to test the feasibility of the procedures that we intended to use in the following fMRI experiment. Participants were explicitly instructed that they would watch other persons' genuine expressions of disgust in some blocks, while in other blocks, they would see other persons acting out disgust expressions (recall that in reality, all demonstrators had been actors). All demonstrators showed neutral expressions as well. The three questions mentioned above were required to be rated. According to the video screening, we excluded videos of two demonstrators (1 female) for whom participants showed a large difference in ratings of *expression* of disgust between pretended vs. genuine conditions. Three separate repeated-measures ANOVAs were respectively performed for the three rating questions regarding the remaining videos. For the disgusted expressions in others, the main effect of genuineness (genuine vs. pretended) was not significant and was low in effect size ($F_{\text{genuineness}}$ (1, 46) = 0.867, p = 0.357, η^2 = 0.018), but it was significant and showed high effect size for the disgusted feelings in others ($F_{\text{genuineness}}$ (1, 46) = 207.225, p < 0.001, η^2 = 0.818) as well as for the unpleasantness in self ($F_{\text{genuineness}}$ (1, 46) =21.360, p < 0.001, η^2 = 0.317). The main effects of disgust (disgust vs. no disgust) for all three ratings were significant with high effect size (the smallest effect size was for the rating of unpleasantness in self, F_{disgust} (1, 46) = 44.489, p < 0.001, η^2 = 0.492). The findings of our pilot study for the domain of disgust were thus very much in line with the findings of the same pilot study for the domain of pain (see Zhao et al., 2021). Finally, video clips of 12 demonstrators (6 females) were determined for the main fMRI experiment.

Experimental design and procedures of the fMRI study

The experimental design and procedures are sketched in Figure 1A and 1B. The fMRI experiment was performed in two runs, and each run consisted of two blocks showing genuine disgust and two blocks showing pretended disgust. In each block, participants watched nine video clips containing both disgusted and neutral videos.

After the scanner session, participants came on another day to complete three questionnaires in the lab: the Empathy Components Questionnaire (ECQ), the Interpersonal Reactivity Index (IRI), and the Toronto Alexithymia Scale (TAS). The ECQ is categorized into Five subscales with 27 items (i.e., cognitive ability, cognitive drive, affective ability, affective drive, and affective reactivity), using a 4-point Likert scale ranging from 1 ("strongly disagree") to 4 ("strongly agree") (Batchelder, 2015; Batchelder et al., 2017). The IRI is divided into four subscales with 28 items (i.e., perspective taking, fantasy, empathic concern, and personal distress), using a 5-point Likert scale ranging from 0 ("does not describe me well") to 4 ("describes me very well") (Davis, 1980). The TAS is composed of three subscales with 20 items (i.e., difficulty describing feelings, difficulty identifying feelings, and externally oriented thinking), using a 5-point Likert scale ranging from 1 ("strongly disagree") to 5 ("strongly agree") (Bagby et al., 1994). Participants were debriefed at the end of the whole study.

Behavioral data analysis

Repeated-measures ANOVAs were run in SPSS (version 26.0; IBM) to investigate the main effects and the interaction of the two factors genuine vs. pretended and disgust vs. no disgust. Furthermore, we conducted Pearson correlations to examine whether ratings of disgust feelings in others were correlated with unpleasantness in self for the genuine disgust and the pretended disgust. The comparison of the correlation coefficients was performed using a bootstrap approach with the R package bootcorci (https://github.com/GRousselet/bootcorci).

fMRI data acquisition

We used a Siemens Magnetom Skyra MRI scanner (Siemens, Erlangen, Germany) with a 32-channel head coil to collect fMRI data. A multiband-accelerated T2*-weighted echoplanar imaging (EPI) sequence was applied to collect functional whole-brain scans (TR = 1200 ms, TE = 34 ms, acquisition matrix = 96 × 96 voxels, FOV = 210 × 210 mm2, flip angle = 66°, inter-slice gap = 0.4 mm, voxel size = 2.2 × 2.2 × 2 mm3, multiband acceleration factor = 4, interleaved ascending acquisition in multi-slice mode, 52 slices co-planar to the connecting line between anterior and posterior commissure). Each of the two functional imaging runs lasted around 16 min (~800 images per run). A magnetization-prepared rapid gradient-echo (MPRAGE) sequence was implemented to acquire structural images (TE/TR = 2.43/2300 ms, FOV= 240 × 240 mm2, flip angle = 8°, voxel size = 0.8×0.8×0.8 mm3, slice thickness = 0.8 mm, ascending acquisition, 208 sagittal slices, single-shot multi-slice mode).

fMRI data processing and mass-univariate functional segregation analyses

Imaging data preprocessing was performed with a combination of Nipype (Gorgolewski et al., 2011) and MATLAB (version R2018b 9.5.0; MathWorks) with Statistical Parametric Mapping (SPM12; https://www.fil.ion.ucl.ac.uk/spm/software/spm12/). Raw data were arranged into BIDS format (http://bids.neuroimaging.io/; Gorgolewski et al., 2016). Functional data were 1) slice time corrected to the middle slice (Sladky et al., 2011), realigned to the first image of each session, 3) co-registered to the T1 image, 4) segmented between grey matter, white matter, and cerebrospinal fluid (CSF), 5) normalized to MNI template space using Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL) toolbox (Ashburner, 2007), and 6) smoothed using a 6 mm full width at half-maximum (FWHM) of the Gaussian kernel. To improve data quality, scrubbing was performed when the frame-wise displacement (FD) of a scan was larger than 0.5 mm (Power et al., 2012; Power et al., 2014).

In order to perform mass-univariate functional segregation analyses, we created a first-level GLM design matrix composed of two identically modeled runs for each participant. Seven regressors of interest were entered in each model: stimulation phase of the four conditions (i.e., genuine disgust, genuine no disgust, pretended disgust, pretended no disgust; 2000 ms), rating phase of the three questions (i.e., disgusted expressions in others, disgusted feelings in others, and unpleasantness in self; 12000 ms). Six head motion parameters and the scrubbing regressors (FD > 0.5 mm; if applicable) were additionally entered as nuisance regressors.

On the second level, we used a flexible factorial design for the group-level analysis. Three factors were included: a between-subject factor (i.e., subject) that was specified independent and with equal variance, a within-subject factor (i.e., genuine or pretended) that was specified dependent and with equal variance, and a second within-subject factor (i.e., disgust or no disgust) that was specified dependent and with equal variance were included in the design (Gläscher & Gitelman, 2008). Four contrasts were computed: 1) genuine: disgust – no disgust, 2) pretended: disgust – no disgust, 3) genuine disgust – pretended disgust, and 4) genuine (disgust – no disgust) – pretended (disgust – no disgust). An initial threshold of $p < 0.001$ (uncorrected) at the voxel level and a family-wise error (FWE) correction ($p < 0.05$) at the cluster level were applied. The cluster extent threshold was determined by the SPM extension “cp_cluster_Pthresh.m” (https://goo.gl/kjVydz).

Brain-behavior relationships

We built a multiple regression model on the group level to investigate the relationship between specific brain activations and behavioral ratings. In this model, the contrast genuine disgust – pretended disgust was set as the dependent variable, and differences between conditions for three behavioral ratings were specified as independent variables. The reason that we entered the condition differences for both brain signals and behavioral ratings into the analyses was to control for the potential effects of perceptual salience. Moreover, we used the contrast genuine disgust – pretended disgust instead of the more exhaustive contrast genuine (disgust – no disgust) – pretended (disgust – no disgust), because our aim was to focus on the genuine and pretended disgust conditions rather than the neutral conditions. In addition, the disgust contrast showed more robust (in terms of statistical effect size) and widespread activations across the brain, making it more likely to pick up possible brain-behavior relationships. The same threshold as above (i.e., cluster-wise FWE correction, $p < 0.05$) was applied in this analysis. All covariates were mean-centered. An intercept was added to the model.

Analyses using dynamic causal modeling (DCM)

We considered the following regions of interest (ROI) for the DCM model space: the right aIns and rSMG according to the previous study of pain (Zhao et al., 2021) and the left (primary) olfactory cortex according to the exploratory analyses. As for the latter, the results showed no evidence that effective connectivity between aIns and rSMG for genuine disgust vs. pretended disgust was distinct. Therefore, we extended the analysis to the primary olfactory cortex, which was not hypothesized when planning this study but highly plausible given the employed task and the specific link between olfaction and disgust. In fact, previous studies indeed demonstrated the olfactory cortex was not only engaged in perceptual processes (e.g., odor perception and recognition), but also in affective processing of disgust-related experiences (Gottfried et al., 2002; Zelano et al., 2011; Alessandrini et al., 2016; Schulze et al., 2017; Schienle et al., 2020). We additionally defined an ROI of the right olfactory cortex as a comparison to the left olfactory cortex. The ROI masks were defined as the anatomical masks created by the Wake Forest University (WFU) Pick Atlas SPM toolbox (http://fmri.wfubmc.edu) with the automated anatomical atlas (AAL). Note that the olfactory cortex mask defined in AAL largely overlaps with the primary olfactory cortex that we were interested in (Desikan et al., 2006).

Three DCM analyses were performed based on different considerations. Firstly, to investigate if the distinct effective connectivity between aIns and rSMG we have found between genuine pain and pretend pain could be observed in disgust as well, we performed a DCM analysis between the right aIns and rSMG under the manipulation of genuine disgust and pretended disgust. Secondly, to explore if other brain patterns could dissociate genuine disgust and pretended disgust, we performed a second DCM analysis between the right aIns and the left olfactory cortex. Finally, to validate the second DCM model, we performed a third DCM analysis between the right aIns and the right olfactory cortex.

All DCM analyses were performed with DCM12.5 implemented in SPM12 (v. 7771). As a first step, individual time series were extracted separately for each ROI. The voxels were determined both on a group-level and an individual-level threshold to ensure the selected voxel were indeed engaged in a task-relevant activity instead of random signal fluctuations (Holmes et al., 2020). The initial threshold was set as p < 0.05, uncorrected. The significant voxels in the main effect of genuine disgust and pretended disgust were further selected by an individual threshold. An individual peak coordinate within the ROI mask was searched for each participant and an individual mask was consequently defined using a sphere of the 6 mm radius around the peak. The individual time series for each ROI was subsequently extracted from the significant voxels of the individual mask and summarized by the first eigenvariate. For the second and third DCM analyses, seven participants were excluded as no voxels survived significance testing in either the left or right olfactory cortex. In the next step, three regressors of interest were specified: genuine disgust, pretended disgust, and the video input condition (the combination of genuine disgust and pretended disgust). The reasons for not specifying the no-disgust conditions were that 1) disgust conditions were our main focus, and 2) adding effects of non-interest would inevitably increase the model complexity. In DCM, three sets of parameters were estimated: bidirectional connections between the regions and their self-connections (matrix A), modulatory effects (i.e., genuine disgust and pretended disgust) on the between-region connections (matrix B), and driving inputs (i.e., the video input condition) on both regions (matrix C) (Zeidman et al., 2019a). Finally, we performed group-level DCM inference using parametric empirical Bayes (Zeidman et al., 2019b). An automatic search was conducted over the entire model space (max. n =256) using Bayesian model reduction (BMR) and random-effects Bayesian model averaging (BMA), resulting in a final group model that takes accuracy, complexity, and uncertainty into account (Zeidman et al., 2019b). This procedure was similarly performed for all three DCM analyses. We reported all parameters with positive evidence on the posterior probability (pp > 0.75). Finally, modulatory effects of the genuine and pretended disgust conditions were compared using a paired sample t-test for each group-averaged model.

To probe whether task-related modulatory effects were associated with behavioral measurements, we performed multiple linear regression analyses of modulatory parameters with, 1) the three behavioral ratings, and 2) the empathy-related questionnaires (i.e., IRI, ECQ, and TAS). We set up two regression models for the genuine and pretended disgust conditions, respectively, in which the DCM parameters of modulatory effects were determined as dependent variables and the three ratings as independent variables. Considering that interactions between behavioral ratings might contribute to the regression model, five regression models (with and without interaction) were tested for both conditions. Results showed the model without any interaction outperformed other models for both genuine disgust (AIC = -27.422, BIC = -19.234) and pretended disgust (AIC= -10.697, BIC= -2.509). Smaller AIC/BIC indicates better model fit. The model with an interaction of disgusted expressions and disgusted feelings in others showed the smallest AIC and BIC among all models with interactions for both conditions: for genuine disgust, AIC = -25.547, BIC = -15.722; for pretended disgust, AIC = -10.650, BIC = -0.824. We will thus report the results of the winning multiple regression model in the results section. We performed two additional regression models for both conditions in which DCM modulatory effects were set as dependent variables and scores of each questionnaire subscale were set as independent variables, respectively. Given the number of independent variables was considerable (>10), we used a stepwise regression approach to perform the analyses for questionnaires. As two participants did not complete all three questionnaires, we excluded their data from the regression analyses. The statistical significance of the regression analysis was set to $p < 0.05$. The multicollinearity for independent variables was diagnosed using the variance inflation factor (VIF) that measures the correlation among independent variables, in the R package car (https://cran.r-project.org/web/packages/car/index.html). Here a rather conservative threshold of $VIF < 5$ was adapted as an indication of no severe multicollinearity (Menard, 2002; James et al., 2013).

Results

Behavioral results

We performed three repeated-measures ANOVAs with the factors *genuineness* (genuine vs. pretended) and *disgust* (disgust vs. no disgust), for each of the three behavioral ratings. For ratings of disgusted *expressions* in others (Figure 1C, left), the main effect of the factor genuineness was not significant: $F_{\text{genuineness}}$ (1, 44) = 1.861, p = 0.179, η^2 = 0.041. There was a main effect of disgust: participants showed higher ratings for the disgust vs. no disgust conditions, F_{disgust} (1, 44) = 1769.396, p < 0.001, η^2 = 0.976. The interaction term was not significant, $F_{\text{interaction}}$ (1, 44) = 2.270, p = 0.139, η^2 = 0.049. For ratings of disgusted feelings in others (Figure 1C, middle), there was a main effect of genuineness: participants showed higher ratings for the genuine vs. pretended conditions, $F_{\text{genuineness}}$ (1, 44) = 510.686, p < 0.001, η^2 = 0.921. There was also a main effect of disgust, as participants showed higher ratings for the disgust vs. no disgust conditions, F_{disgust} (1, 44) = 854.136, p < 0.001, η^2 = 0.951. The interaction was significant as well, $F_{\text{interaction}}$ (1, 44) = 360.516, p < 0.001, η^2 = 0.891, and this was related to higher ratings of disgusted feelings in others for the genuine disgust compared to the pretended disgust condition. For ratings of unpleasantness in self (Figure 1C, right), there was a main effect of genuineness: participants showed higher ratings for the genuine vs. pretended conditions, $F_{\text{genuineness}}$ (1, 44) = 37.694, $p < 0.001$, $\eta^2 = 0.461$. There was also a main effect of disgust: participants showed higher ratings for the disgust vs. no disgust conditions, F_{disgust} (1, 44) = 141.277, $p < 0.001$, $\eta^2 = 0.763$. The interaction was significant as well, $F_{\text{interaction}}$ (1, 44) = 32.341, $p < 0.001$, $\eta^2 = 0.424$, and this was related to higher ratings of unpleasantness in self for the genuine disgust compared to the pretended disgust condition. In sum, the behavioral data indicated that there was no difference in ratings of disgusted expression in others between the genuine and pretended disgust conditions, while higher ratings and large effect sizes of disgusted feelings in others and unpleasantness in self for the genuine disgust condition as compared to the pretended disgust condition. These results were perfectly in line with our hypotheses and what we found in the pilot study.

We also found significant correlations between behavioral ratings of disgusted feelings in others and unpleasantness in self for the genuine disgust condition, $r = 0.548$, $p < 0.001$; while for the pretended disgust condition, the correlation was not significant, $r = 0.051$, $p = 0.740$ (Figure 1D). A bootstrapping comparison showed a significant difference between the two correlation coefficients, $p = 0.025$, 95% Confidence Interval (CI) = [0.073, 0.860].

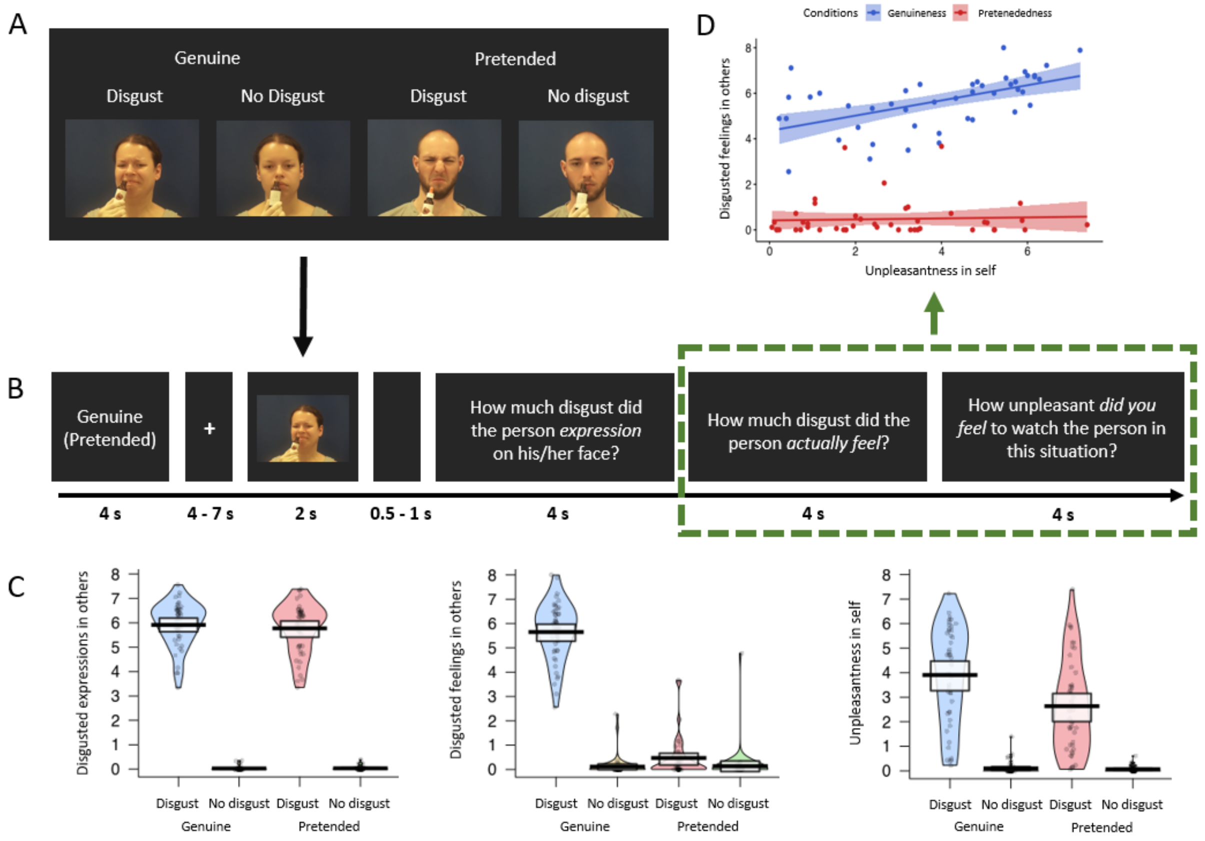

Figure 1. fMRI experimental design and behavioral results. (A) Overview of the experimental design with the four conditions genuine vs. pretended, disgust vs. no disgust. Examples show static images, while in the

experiment, participants were shown video clips. (B) Overview of experimental timeline. At the outset of each block, a reminder of “genuine” or “pretended” was shown (both terms are shown here for illustrative purposes, in the experiment either genuine or pretended was displayed). After a fixation cross, a video in the corresponding condition appeared on the screen. Followed by a short jitter, three questions about the video were separately presented and had to be rated on a visual analog scale. These would then be followed by the next video clip and questions (not shown). (C) Violin plots of the three types of ratings for all conditions. No difference was found for the rating of disgusted expressions in others between the genuine disgust condition and the pretended disgust condition. For the ratings of disgusted feelings in others and unpleasantness in self, participants demonstrated higher ratings for genuine disgust than pretended disgust. Ratings of all three questions were higher in the disgusted situation than in the neutral situation, regardless of whether in the genuine or pretended condition. The thick black lines illustrate mean values, and the white boxes indicate a 95% CI. The dots are individual data, and the “violin” outlines illustrate their estimated density at different points of the scale. (D) Correlations of disgusted feelings in others and unpleasantness in self for the genuine disgust and the pretended disgust (the relevant questions were highlighted with a green rectangular). Results revealed a significant Pearson correlation between the two questions for the genuine disgust condition, but no correlation in the pretended disgust condition. The lines represent the fitted regression lines, bands indicate a 95% CI.

fMRI results: mass-univariate analysis

We computed four contrasts: 1) genuine: disgust – no disgust, 2) pretended: disgust – no disgust, 3) genuine disgust – pretended disgust, and 4) genuine (disgust – no disgust) – pretended (disgust – no disgust). In the first two contrasts, we found the predicted activations in bilateral aIns, aMCC, and rSMG, as well as significant (not originally predicted) activation in the olfactory cortex; in the third contrast, we found significant activation in the right aIns, as well as strong activation (k = 255) in the left olfactory cortex; in the last contrast, the only significant activation was found in the right cerebellum (Figure 2A and Table 1).

To identify whether or which brain activity was selectively related to the behavioral ratings described above, we performed a multiple regression analysis where we explored the relationship of activation in the contrast genuine disgust – pretended disgust with the three behavioral ratings. The only significant cluster we found encompassed the right aIns, extending into the right inferior frontal gyrus, and this was selectively related to ratings of self-unpleasantness (Figure 2B) rather than the ratings of disgusted expressions in others or the disgusted feelings in others.

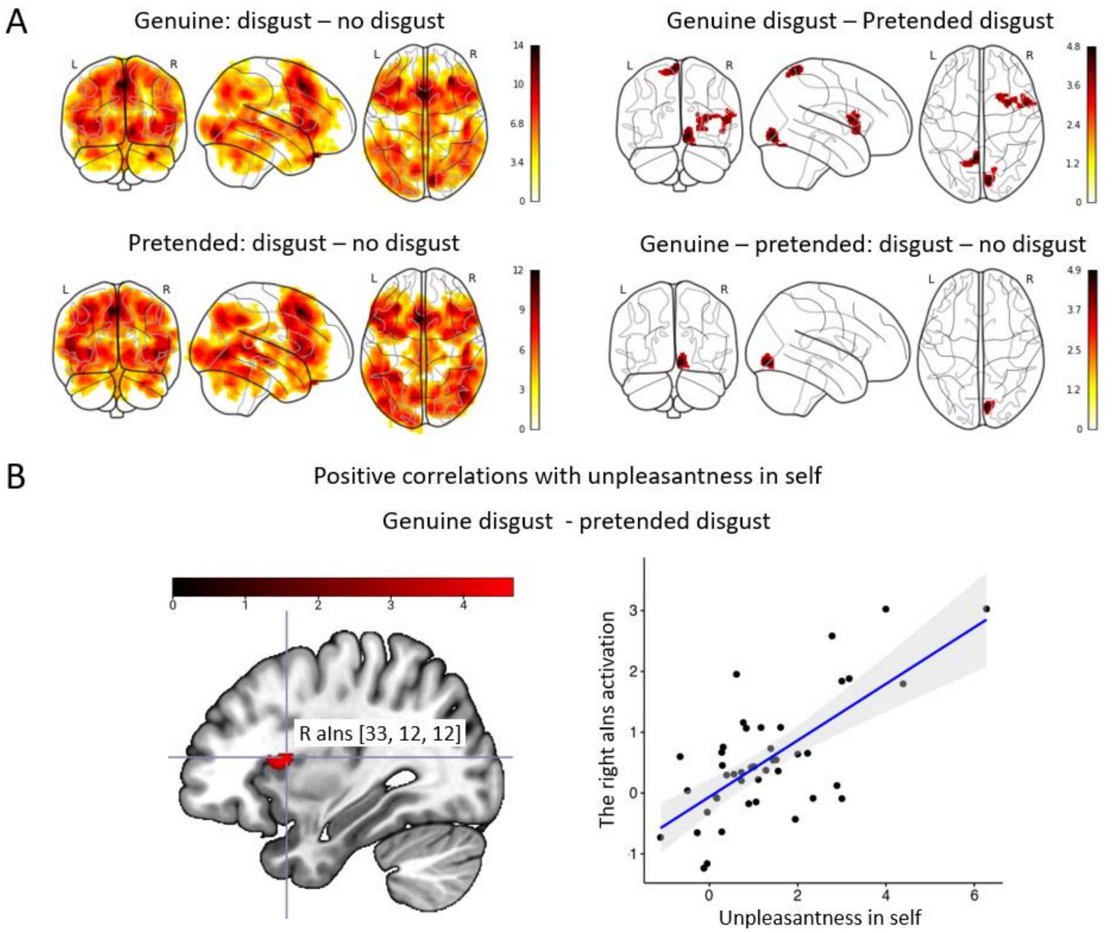
A
Genuine: disgust – no disgust
Genuine disgust – Pretended disgust
Pretended: disgust – no disgust
Genuine – pretended: disgust – no disgust
L
R
14
10
6.8
3.4
0
4.8
3.6
2.4
1.2
0
12
9
6
3
0
4.9
3.7
2.5
1.2
0
B
Positive correlations with unpleasantness in self
Genuine disgust - pretended disgust
0
1
2
3
4
R aIns [33, 12, 12]
The right aIns activation
-1
0
1
2
3
0
2
4
6
Unpleasantness in self

Figure 2. Neuroimaging results: Mass-univariate analyses. (A) Activation maps of genuine: disgust – no disgust (top left), pretended: disgust - no disgust (bottom left), genuine disgust – pretended disgust (top right), and genuine (disgust – no disgust) – pretended (disgust – no disgust) (bottom right). For contrasts of disgust –no disgust in both genuine and pretended conditions, we found expected brain activations in bilateral aIns, aMCC, and rSMG, and significant activation in the olfactory cortex; for the contrast of genuine disgust vs. pretended disgust, we found significant activation in the right aIns and strong activation in the left olfactory cortex (a cluster of k=255, though not pass the threshold); for the contrast of genuine (disgust – no disgust) vs. pretended (disgust – no disgust), the only significant activation was in the right cerebellum. (B) The multiple regression analysis demonstrated a significant cluster in the right aIns (peak: [33, 12, 12]) that was positively associated with the ratings of unpleasantness in self but not associated with the ratings of either disgusted expressions in others or disgusted feelings in others when comparing genuine disgust vs. pretended disgust. All activations are thresholded with cluster-level FWE correction, $p < 0.05$ ($p < 0.001$ uncorrected initial selection threshold). The lines of the scatterplots represent the fitted regression lines, bands indicate a 95% CI.

Table 1. Results of mass-univariate functional segregation analyses in the MNI space. Region names were labeled with the AAL atlas and thresholded with cluster-wise FWE correction, $p < 0.05$ (initial selection threshold $p < 0.001$, uncorrected). BA = Brodmann area, L = left hemisphere, R = right hemisphere.

Region label	BA	Cluster size	*x*	*y*	*z*	*t*-value
1) Genuine: disgust - no disgust						
Temporal_Pole_Sup_R	38	164988	32	33	-33	13.57
Supp_Motor_Area_L	8		-3	16	50	13.31
Lingual_R	18		9	-84	-6	11.89
Frontal_Sup_Medial_R	8		6	21	45	10.95
Insula_L	45		-32	27	4	10.81
Frontal_Inf_Oper_L	44		-50	15	6	10.80
Insula_R	13		33	27	4	10.17
Frontal_Inf_Tri_L	45		-30	32	0	10.12
Frontal_Inf_Oper_R	44		52	15	15	9.80
Frontal_Inf_Orb_R	47		32	32	-3	9.69
Lingual_L	17	422	-21	-66	4	5.10
2) Pretended: disgust - no disgust						
Supp_Motor_Area_L	8	137060	-4	16	50	11.95
Frontal_Sup_Medial_L	8		-8	26	44	10.69
Temporal_Mid_R	19		44	-68	2	10.35
Frontal_Inf_Oper_L	44		-51	16	6	10.01
Insula_L	45		-30	30	2	9.77
Frontal_Inf_Tri_L	45		-56	20	12	9.68
Cingulum_Mid_R	8		8	20	45	9.47
Parietal_Inf_L	39		-32	-51	40	9.37
Temporal_Pole_Sup_R	38		32	34	-33	9.35
Frontal_Mid_L	6		-27	2	54	9.26
Cingulate_Post_L	23	1821	-4	-42	22	5.71
Cingulate_Mid_R	23		-3	-26	27	5.58
Cingulate_Post_R	23		8	-39	22	5.33
Cingulate_Mid_L	24		-3	-12	30	5.23
Vermis_9	37	522	2	-57	-39	5.27
Cerebelum_9_L	18		-2	-60	-46	4.55
Cerebelum_9_R	37		10	-57	-51	3.96
Temporal_Inf_L	20	517	-40	-9	-38	4.80
Temporal_Pole_Mid_L	38		-28	8	-39	3.62
Cerebelum_Crus2_L	18	487	-8	-76	-34	5.15
Cerebelum_Crus1_L	18		-18	-80	-27	3.21

3) Genuine disgust – pretended disgust						
Insula_R	44	976	32	8	12	4.56
Rolandic_Oper_R	44		54	8	10	4.38
Caudate_R	48		21	14	15	4.05
Putamen_R	49		30	16	-2	3.89
Frontal_Inf_Oper_R	44		40	12	14	3.85
Lingual_R	18	550	10	-81	-9	4.72
Cerebelum_6_R	18		15	-70	-18	3.35
Precuneus_L	7	474	-6	-56	69	4.85
Parietal_Sup_L	7		-16	-63	64	3.95
4) Genuine (disgust – no disgust) – pretended (disgust – no disgust)						
Lingual_R	18	431	8	-81	-10	4.90

DCM results

We first performed a DCM analysis of the effective connectivity between the right aIns and rSMG to examine if the group-averaged model replicated what we found in our previous study on pain (Zhao et al., 2021). Specifically, we focused on the modulatory effect of genuineness, namely, whether the experimental manipulation of genuine disgust vs. pretended disgust tuned the bidirectional neural dynamics from aIns to rSMG and *vice versa*, in terms of both directionality (sign of the DCM posterior parameter) and intensity (magnitude of the DCM posterior parameter). If the experimental manipulation modulated the effective connectivity, we would observe a positive posterior probability (*pp* > 0.75) of the modulatory effect. The reasons that we did not include aMCC in this analysis were that 1) unlike aIns, in aMCC we did not find strong evidence for the task involvement (in the univariate and multiple regression analyses), 2) for comparability of the present model with the previous model on pain, where aMCC had not been included either.

Similar to what we found in the pain study, strong evidence (*pp* > 0.95; *pp* = 1.00) of inhibitory modulatory effects on the aIns-to-rSMG connection was shown for both the genuine disgust condition and the pretended disgust condition (see Figure 3A). However, we did not find a significant difference when comparing the strength of these two modulatory effects, t_{44} = -1.045, *p* = 0.302 (Mean $_{\text{genuine disgust}}$ = -1.214, 95% CI = [-1.462, -0.927]; Mean $_{\text{pretended disgust}}$ = -1.095, 95% CI = [-1.361, -0.799]. Note that the mean values we exhibit in the test could slightly differ from those shown in the DCM models of Figure 3, since we used frequentist statistics for comparison analysis rather than the Bayesian approach that was implemented to compute the parameters for the DCM model. We did not find robust evidence on the intrinsic connectivity either from aIns to rSMG or *vice versa*. Moreover, there was no evidence of a modulatory effect on the rSMG to aIns connection, which was in line with what we had found for pain. Taken together, the DCM analysis between aIns and rSMG partially replicated the results of the pain study, namely the inhibitory modulatory effect from aIns to rSMG for both genuine and pretended conditions; however, this inhibitory modulatory effect failed to dissociate the experimental manipulation of genuine disgust and pretended disgust, suggesting that a distinctive pattern or set of brain regions underpins how the genuineness of disgust is processed by our brains.

We therefore performed an exploratory DCM analysis to test whether distinct modulatory effects could be found for genuine and pretend disgust on the connection between the right aIns and the left olfactory cortex (see Figure 3B). As mentioned in the methods sections, while the involvement of the left olfactory cortex was mainly exploratory and data-driven, it was also plausible on theoretical grounds. Results showed a significant excitatory effect for both genuine disgust (strong evidence, *pp* = 1.00) and pretended disgust (positive evidence, *pp* = 0.93) on the connection of the left olfactory cortex to the right aIns. A further comparison analysis on the modulatory effect between conditions revealed a stronger excitatory modulatory effect for genuine disgust as opposed to pretended disgust, $t_{37} = 4.450$, $p < 0.001$ (Mean $_{\text{genuine disgust}}$ = 0.805, 95% CI = [0.755, 0.851]; Mean $_{\text{pretended disgust}}$ = 0.573, 95% CI = [0.517, 0.628]. We did not find any modulatory effect on the reverse connection of the right aIns to the left olfactory cortex.

Finally, to validate the modulatory effect we found in the DCM model above, we performed an additional DCM analysis on the connection between the right aIns and the right olfactory cortex (see Figure 3C). Results showed a similar group-average model to that of the right aIns and the left olfactory cortex. Importantly, we replicated the excitatory modulatory effect on the connection of the olfactory cortex to aIns in the sense of significant evidence for both conditions (genuine disgust: positive evidence, *pp* = 0.91; pretended disgust: positive evidence, *pp* = 0.88). No evidence of the modulatory effect on the connection of aIns to the right olfactory cortex was found, which was consistent with the group-average model with the left olfactory cortex. A further comparison analysis did not find significant difference on the strength of the two modulatory effects, $t_{37} = 0.595$, $p = 0.556$ (Mean $_{\text{genuine disgust}}$ = 0.624, 95% CI = [0.503, 0.738]; Mean $_{\text{pretended disgust}}$ = 0.577, 95% CI = [0.488, 0.678]).

Individual associations between modulatory effects, behavioral ratings, and questionnaires

Two linear regression models were computed to examine how the excitatory modulatory effect on the connection of the (left) olfactory cortex to aIns was related to behavioral ratings respectively for genuine disgust and pretended disgust. Results showed that none of the ratings was significant for either the genuine disgust model or the pretended disgust model. No severe collinearity problem was detected for either regression model (all *VIFs* < 4.500; the smallest *VIF* =1.229 and the largest *VIF* = 4.410).

Another two linear regression models were tested to investigate whether subscales of all three questionnaires could explain the excitatory modulatory effect for genuine disgust and pretended disgust. For the genuine disgust condition, we found that the modulatory effect was significantly explained by scores of the perspective-taking subscale of the IRI: F_{model} (1, 35) = 4.177, p = 0.049, R^2 = 0.109; B = 0.011, *beta* = 0.331, p = 0.049. No significant predictor was found with any subscale in the other two questionnaires (i.e., ECQ and TAS). None of the three questionnaires significantly explained variations of the modulatory effect for the pretended disgust condition. No severe collinearity problem was detected for either regression model (all *VIFs* < 2.500; the smallest *VIF* =1.000 and the largest *VIF* = 2.198).

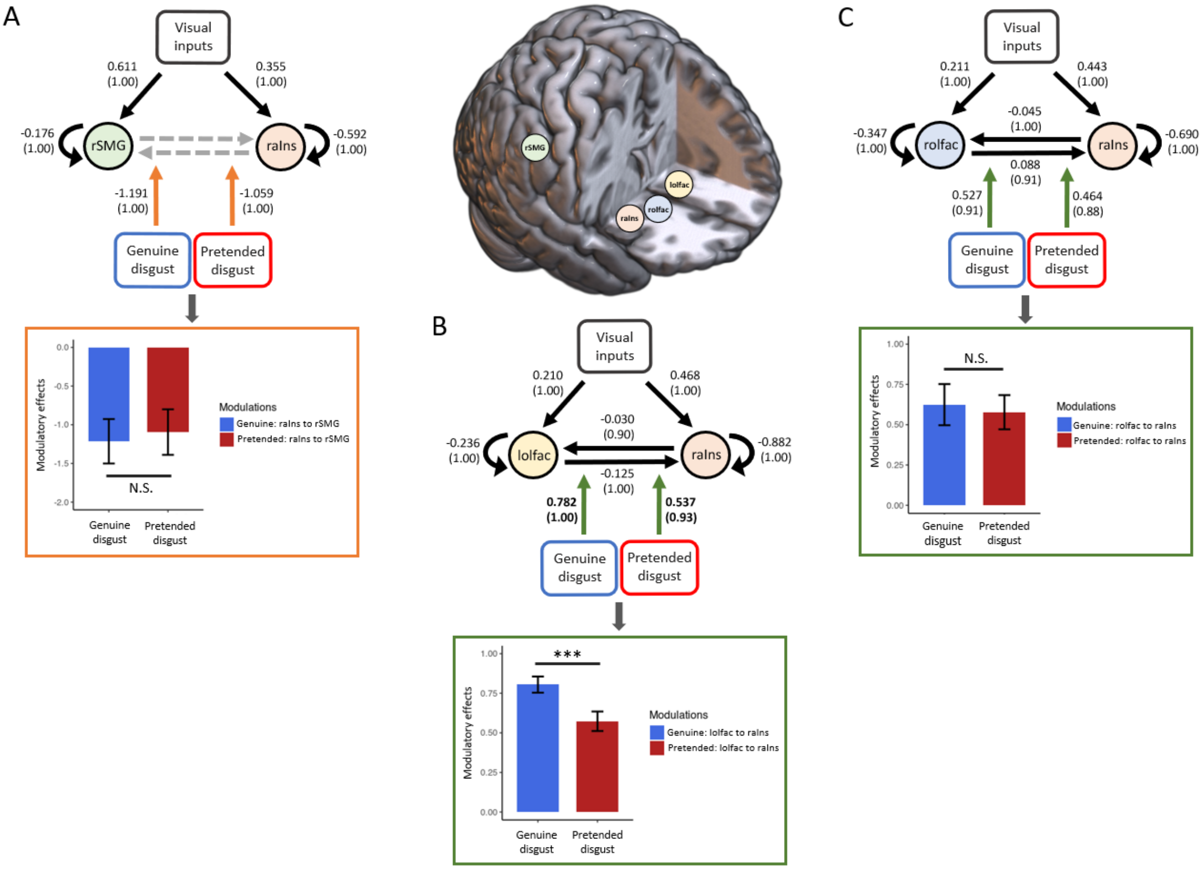

Figure 3. DCM results and brain-behavior analyses. Three-dimensional visualization of ROIs involved in the three DCM analyses is shown in the upper middle. (A) The group-average DCM model of the right anterior insula and the right supramarginal gyrus (rSMG) for genuine disgust and pretended disgust. We found an inhibitory modulatory effect (orange arrows) for both conditions. All DCM parameters of the optimal model showed greater than a 99% posterior probability (very strong evidence) except the bi-directionally intrinsic connectivity between the right aIns (raIns) and rSMG (grey dashed arrow; no evidence of existence, *pp* < 0.50). A paired

sample *t*-test showed no difference in the inhibitory modulatory effects on the raIns-to-rSMG connection between genuine disgust and pretended disgust. This result is highlighted with an orange rectangular. Data are mean ± 95% CI. (B) The group-average DCM model of the raIns and the left olfactory cortex (lolfac) for genuine disgust and pretended disgust. We found an excitatory modulatory effect (green arrows) for both conditions. All DCM parameters of the optimal model showed greater than or equal to a 90% posterior probability (*pp* > 0.75, positive evidence). A paired sample *t*-test showed a stronger excitatory modulatory effect of the lolfac-to-raIns connection for genuine disgust as compared to pretended disgust (*** *p* < 0.001). This result is highlighted with a green rectangular. Data are mean ± 95% CI. (C) The group-average DCM model of the raIns and the right olfactory cortex (rolfac) for genuine disgust and pretended disgust. We found an excitatory modulatory effect (green arrows) for both conditions. All DCM parameters of the optimal model showed greater than a 75% posterior probability (positive evidence). A paired sample *t*-test showed no difference in the inhibitory modulatory effect on the rolfac-to-raIns connection between genuine disgust and pretended disgust. This result is highlighted with a green rectangular. Data are mean ± 95% CI. For all DCM models, values without the bracket quantify the strength of connections; positive values indicate neural excitation and negative values indicate neural excitation. Values in the bracket indicate the posterior probability of connections.

Discussion

Using a paradigm matched to our previously published study on pain (Zhao et al., 2021), and in the same sample and experimental session, we here report how participants responded to video clips presenting people who supposedly either genuinely experienced disgust or merely pretended to feel disgusted. Combining mass-univariate analysis with effective connectivity (DCM) analyses, we aimed to clarify two main questions: 1) whether neural responses in areas such as aIns and aMCC to the disgust of others were indeed related to a veridical sharing of affect, as opposed to simply tracking sensory-driven responses to salient affective displays, and 2) whether the effective connectivity between aIns and rSMG that we previously found to disentangle genuine pain from pretended pain also enabled the dissociation of genuine disgust vs. pretended disgust.

We found increased activations in the right aIns for genuine disgust as compared to pretended disgust that was selectively associated with the unpleasantness in self. These findings are in line with what we have found in pain (Zhao et al., 2021), implying an essential role of aIns in the processing of shared feelings with others for both pain and disgust. However, an intriguing question is if the aIns activation we observed in these two aversive states reflects a form of cross-modal affective processing, or rather modality-dependent affective experiences? A study using multi-voxel pattern analysis (MVPA) has shown both cross-modal and modality-specific evidence in terms of the subfields of aIns for pain and disgust: in the left aIns (and aMCC), the shared encoding was detected for first-hand and vicarious pain and disgust, regardless of the same or different modality; while in the right aIns, sensory-specific rather than modality-independent patterns were more plausible for processing first-hand and vicarious pain and disgust (Corradi-Dell'Acqua et al., 2016). Taken together, the aIns activation we found suggests the engagement of affective processing that was related to others' pain and disgust, while future research that explicitly matches pain and disgust salience is required to further investigate whether this activation indicates cross-modal or modality-dependent affective experiences.

We found significant inhibitory modulatory effects on the connection of aIns to rSMG for both genuine and pretended disgust, but these effects did not differ significantly. This implies that we only partially replicate the findings of the pain study: while we reproduce a role of the aIns and rSMG connectivity, their crosstalk does not explain the distinction between genuine and pretended expressions of disgust, as is the case for pain (Zhao et al., 2021). We speculate that the absence of differences between two conditions in rSMG activation as well as the inhibitory modulatory effect could be related to generally lower salience of aversive experiences in the disgust task compared to that of pain. Further investigation is required to test this assumption.

Contrary to our expectations, we did not find any significant activation in rSMG for genuine disgust as compared to pretended disgust; instead, we showed a relatively stronger engagement of the primary olfactory cortex between conditions. As we mentioned beforehand, we did not target the latter area when planning the study but later included it inspired by the exploratory analysis. The (primary) olfactory cortex has been considered to mainly comprise the anterior olfactory nucleus, the olfactory tubercle, piriform cortices, and subregions of amygdala and entorhinal cortex (Savic et al., 2000; Tzourio-Mazoyer et al., 2002; Zhou et al., 2019). Studies have found that this region is recruited not only for direct olfactory sensations but also for the indirect experience of olfactory processing, such as odor imagery (Djordjevic et al., 2005; Bensafi et al., 2007) and odor prediction (Zelano et al., 2011). Olfactory priming could facilitate the identification of the emotion of disgust (Seubert et al., 2010a; Seubert et al., 2010b); in turn, priming with a disgusted face compared to a happy face enhanced activation in the olfactory cortex when processing pleasant odors (Schulze et al., 2017). These findings imply the engagement of the olfactory cortex in integrating olfactory processes with visually conveyed affective information. Furthermore, the primary olfactory cortex has been suggested to participate in processing the emotion of disgust, without necessarily experiencing sensory-related disgust. Compared with healthy controls, patients with reduced olfactory function (e.g., anosmia and hyposmia) have been found to identify less disgust for facial expressions of disgust and show greater activations in the primary olfactory cortex, suggestive of a compensatory effect, for disgusting scenes (Schienle et al., 2020). Altogether, the stronger engagement of the olfactory cortex might thus be related to the higher level of disgust identified in others who are genuinely rather than merely pretending to be disgusted.

The exploratory DCM analysis of the right aIns and the left olfactory cortex demonstrated a stronger excitatory modulatory effect on the olfactory cortex to aIns connection for genuine disgust as opposed to pretended disgust. Studies from nonhuman primates and humans using tractography have shown structural connections (for human: functional connectivity as well, see Deen et al., 2010) between the olfactory cortex and a partial region of aIns (Mufson & Mesulam, 1982; Carmichael et al., 1994; Ghaziri et al., 2015; Ghaziri et al., 2018). Specifically, as an important part of the secondary olfactory cortex, aIns is considered to engage in receiving and integrating the primary olfactory-affective information conveyed by the primary olfactory cortex. Moreover, the external sensory and affective messages seem to be already preprocessed in the primary olfactory cortex before they are conveyed to the secondary cortices (Soudry et al., 2011, for review; Seubert et al., 2013, for meta-analyses). Together with the evidence of higher brain activation in the right aIns and left olfactory cortex for genuine disgust compared to pretended disgust, we speculate that the increased excitatory modulatory effect on the olfactory-to-aIns connection may be related to the processing of passing messages of higher disgust emotion identified in others to activations related to affect processing, which may constitute the neural underpinning of the increased shared unpleasantness with others. This idea would also be in line with the theoretical framework proposed by Coll et al. (2017), that higher identified emotion in others contributes to stronger shared affect, and that the fully-fledged empathic response may be an integrated consequence of (at least) these two processes. We performed another DCM analysis between the right aIns and the right olfactory cortex to validate the modulatory effect we detected in the DCM model with the left olfactory cortex. Results showed a very similar pattern to the DCM model with the left olfactory cortex, in the sense of replicating the excitatory modulatory effect on the connection of the olfactory cortex to aIns for both conditions and the absence of any significant condition-dependent modulatory effect on the connection in the opposite direction. Even though for this model we did not find a significant difference in the modulatory effects between genuine disgust and pretended disgust, these results at least attest to the robustness of the modulatory effect from the olfactory cortex, regardless of the left or right hemisphere, to the right aIns.

We found the excitatory modulatory effect for genuine disgust was positively related to individual perspective-taking scores. This finding demonstrated that the connection of the olfactory cortex to aIns for genuine disgust was related to the tendency of adopting the psychological point of view of others, which would finally contribute to the level of one’s own affective responses to the emotion felt by another person. This view is supported by the evidence that aIns, especially the right aIns, is implicated in distinct neural patterns of representing self- and other-related aversive states for disgust as well as pain (Corradi-Dell'Acqua et al., 2016). No association with any questionnaire was found for pretended disgust. Taking all these results together, we would speculate that for the genuine disgust condition, the olfactory cortex interacts with aIns to achieve genuine identification of disgust in others. This would call for a higher demand to take the other's perspective, and in this way may contribute to the higher shared affect. For the pretended pain condition, sensory-driven ("automatic") emotion processing induced by the saliency of disgust expression interacts with the cognitive processes (i.e., knowing this person was merely acting out and did not feel any disgust at all), resulting in both a low level of identified disgust and shared affect. In this case, it may be less important to recruit the function of perspective-taking to share the emotions of others. However, further investigation is required to test these interpretations.

In conclusion, the current study largely replicates, as well as expands our previously reported findings on pain. Firstly, and similar to what we have shown for empathy for pain using the same experimental approach and within the same study, we provide evidence that responses related to empathy for disgust in aIns can indeed be linked to the affective sharing rather than merely perceptual saliency. Secondly, we show how aIns and the olfactory cortex, instead of aIns and rSMG that we previously found in pain, orchestrate the tracking of disgust felt by another person. Taken together, these findings indicate that similar as well as distinct brain networks are engaged in processing different affective experiences, in this case pain and disgust, experienced by others. This refines and expands our understanding of the neural bases of empathy, from a dynamic and multi-modal perspective.

Acknowledgements

This work was supported by Chinese Scholarship Council (CSC) Grant (201604910515) and Vienna Doctoral School in Cognition, Behavior and Neuroscience (VDS CoBeNe) completion grant fellowship to Y.Z.; the Vienna Science and Technology Fund (WWTF VRG13-007) to C.L., and the Austrian Science Fund (FWF P 32686) to C.L. and M.R..

Conflicts of interest

The authors declare no competing financial interests.

References

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage, 38*(1), 95-113. doi: https://doi.org/10.1016/j.neuroimage.2007.07.007

Bagby, R. M., Taylor, G. J., & Parker, J. D. A. (1994). The twenty-item Toronto Alexithymia Scale: II. Convergent, discriminant, and concurrent validity. *Journal of Psychosomatic Research, 38*(1), 33-40. doi: http://doi.org/10.1016/0022-3999(94)90006-X

Batchelder, L. (2015). *Characterising the components of empathy: implications for models of autism.* University of Bath.

Batchelder, L., Brosnan, M., & Ashwin, C. (2017). The Development and Validation of the Empathy Components Questionnaire (ECQ). *PLOS ONE, 12*(1), e0169185. doi: http://doi.org/10.1371/journal.pone.0169185

Bensafi, M., Sobel, N., & Khan, R. M. (2007). Hedonic-Specific Activity in Piriform Cortex During Odor Imagery Mimics That During Odor Perception. *Journal of Neurophysiology, 98*(6), 3254-3262. doi: http://doi.org/10.1152/jn.00349.2007

Bukowski, H., Tik, M., Silani, G., Ruff, C. C., Windischberger, C., & Lamm, C. (2020). When differences matter: rTMS/fMRI reveals how differences in dispositional empathy translate to distinct neural underpinnings of self-other distinction in empathy. *Cortex, 128*, 143-161. doi: https://doi.org/10.1016/j.cortex.2020.03.009

Carmichael, S. T., Clugnet, M.-C., & Price, J. L. (1994). Central olfactory connections in the macaque monkey. *Journal of Comparative Neurology, 346*(3), 403-434. doi: https://doi.org/10.1002/cne.903460306

Coll, M.-P., Viding, E., Rütgen, M., Silani, G., Lamm, C., Catmur, C., & Bird, G. (2017). Are we really measuring empathy? Proposal for a new measurement framework. *Neuroscience & Biobehavioral Reviews, 83*, 132-139. doi: https://doi.org/10.1016/j.neubiorev.2017.10.009

Corradi-Dell'Acqua, C., Tusche, A., Vuilleumier, P., & Singer, T. (2016). Cross-modal representations of first-hand and vicarious pain, disgust and fairness in insular and cingulate cortex. *Nature Communications, 7*(1), 10904. doi: 10.1038/ncomms10904

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

Deen, B., Pitskel, N. B., & Pelphrey, K. A. (2010). Three Systems of Insular Functional Connectivity Identified with Cluster Analysis. *Cerebral Cortex, 21*(7), 1498-1506. doi: http://dor.org/10.1093/cercor/bhq186

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage, 31*(3), 968-980. doi: https://doi.org/10.1016/j.neuroimage.2006.01.021

Djordjevic, J., Zatorre, R. J., Petrides, M., Boyle, J. A., & Jones-Gotman, M. (2005). Functional neuroimaging of odor imagery. *NeuroImage, 24*(3), 791-801. doi: https://doi.org/10.1016/j.neuroimage.2004.09.035

Fallon, N., Roberts, C., & Stancak, A. (2020). Shared and distinct functional networks for empathy and pain processing: A systematic review and meta-analysis of fMRI studies. *Social Cognitive and Affective Neuroscience*. doi: https://doi.org/10.1093/scan/nsaa090

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. doi: 10.3758/BF03193146

Ghaziri, J., Tucholka, A., Girard, G., Boucher, O., Houde, J.-C., Descoteaux, M., Obaid, S., Gilbert, G., Rouleau, I., & Nguyen, D. K. (2018). Subcortical structural connectivity of insular subregions. *Scientific Reports, 8*(1), 8596. doi: http://doi.org/10.1038/s41598-018-26995-0

Ghaziri, J., Tucholka, A., Girard, G., Houde, J.-C., Boucher, O., Gilbert, G., Descoteaux, M., Lippé, S., Rainville, P., & Nguyen, D. K. (2015). The Corticocortical Structural Connectivity of the Human Insula. *Cerebral Cortex, 27*(2), 1216-1228. doi: http://doi.org/10.1093/cercor/bhv308

Gläscher, J., & Gitelman, D. (2008). Contrast weights in flexible factorial design with multiple groups of subjects. *SPM@ JISCMAIL. AC. UK) Sml, editor*, 1-12.

Gorgolewski, K., Auer, T., Calhoun, V., Craddock, R. C., Das, S., Duff, E., Flandin, G., Ghosh, S., Glatard, T., Halchenko, Y., Handwerker, D., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T., Pellman, J., Poline, J.-B., Rokem, A., Schaefer, G., Sochat, V., Triplett, W., Turner, J., Varoquaux, G., & Poldrack, R. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data, 3*(1), 160044. doi: http://doi.org/10.1038/sdata.2016.44

Gorgolewski, K., Burns, C., Madison, C., Clark, D., Halchenko, Y., Waskom, M., & Ghosh, S. (2011). Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Frontiers in Neuroinformatics, 5*(13). doi: http://doi.org/10.3389/fninf.2011.00013

Hoffmann, F., Koehne, S., Steinbeis, N., Dziobek, I., & Singer, T. (2016). Preserved Self-other Distinction During Empathy in Autism is Linked to Network Integrity of Right Supramarginal Gyrus. *Journal of Autism and Developmental Disorders, 46*(2), 637-648. doi: http://doi.org/10.1007/s10803-015-2609-0

Holmes, E., Zeidman, P., Friston, K. J., & Griffiths, T. D. (2020). Difficulties with Speech-in-Noise Perception Related to Fundamental Grouping Processes in Auditory Cortex. *Cerebral Cortex, 31*(3), 1582-1596. doi: http://doi.org/10.1093/cercor/bhaa311

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112): Springer.

Jauniaux, J., Khatibi, A., Rainville, P., & Jackson, P. L. (2019). A meta-analysis of neuroimaging studies on pain empathy: investigating the role of visual information and observers' perspective. *Social Cognitive and Affective Neuroscience, 14*(8), 789-813. doi: https://doi.org/10.1093/scan/nsz055

Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage, 54*(3), 2492-2502. doi: https://doi.org/10.1016/j.neuroimage.2010.10.014

Menard, S. (2002). *Applied logistic regression analysis* (Vol. 106): Sage.

Mufson, E. J., & Mesulam, M.-M. (1982). Insula of the old world monkey. II: Afferent cortical input and comments on the claustrum. *Journal of Comparative Neurology, 212*(1), 23-37. doi: https://doi.org/10.1002/cne.902120103

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage, 59*(3), 2142-2154. doi: https://doi.org/10.1016/j.neuroimage.2011.10.018

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. doi: https://doi.org/10.1016/j.neuroimage.2013.08.048

Rütgen, M., Seidel, E. M., Silani, G., Riecansky, I., Hummer, A., Windischberger, C., Petrovic, P., & Lamm, C. (2015). Placebo analgesia and its opioidergic regulation suggest that empathy for pain is grounded in self pain. *Proceedings of the National Academy of Sciences, 112*(41), E5638-E5646. doi: https://doi.org/10.1073/pnas.1511269112

Savic, I., Gulyas, B., Larsson, M., & Roland, P. (2000). Olfactory Functions Are Mediated by Parallel and Hierarchical Processing. *Neuron, 26*(3), 735-745. doi: https://doi.org/10.1016/S0896-6273(00)81209-X

Schienle, A., Höfler, C., Keck, T., & Wabnegger, A. (2020). Neural underpinnings of perception and experience of disgust in individuals with a reduced sense of smell: An fMRI study. *Neuropsychologia, 141*, 107411. doi: https://doi.org/10.1016/j.neuropsychologia.2020.107411

Schulze, P., Bestgen, A.-K., Lech, R. K., Kuchinke, L., & Suchan, B. (2017). Preprocessing of emotional visual information in the human piriform cortex. *Scientific Reports, 7*(1), 9191. doi: http://doi.org/10.1038/s41598-017-09295-x

Seubert, J., Freiherr, J., Djordjevic, J., & Lundström, J. N. (2013). Statistical localization of human olfactory cortex. *NeuroImage, 66*, 333-342. doi: https://doi.org/10.1016/j.neuroimage.2012.10.030

Seubert, J., Kellermann, T., Loughead, J., Boers, F., Brensinger, C., Schneider, F., & Habel, U. (2010a). Processing of disgusted faces is facilitated by odor primes: A functional MRI study. *NeuroImage, 53*(2), 746-756. doi: https://doi.org/10.1016/j.neuroimage.2010.07.012

Seubert, J., Loughead, J., Kellermann, T., Boers, F., Brensinger, C. M., & Habel, U. (2010b). Multisensory integration of emotionally valenced olfactory-visual information in patients with schizophrenia and healthy controls. *Journal of psychiatry & neuroscience : JPN, 35*(3), 185-194. doi: 10.1503/jpn.090094

Sharvit, G., Lin, E., Vuilleumier, P., & Corradi-Dell'Acqua, C. (2020). Does inappropriate behavior hurt or stink? The interplay between neural representations of somatic experiences and moral decisions. *Science Advances, 6*(42), eaat4390. doi: http://doi.org/10.1126/sciadv.aat4390

Sharvit, G., Vuilleumier, P., Delplanque, S., & Corradi-Dell'Acqua, C. (2015). Cross-modal and modality-specific expectancy effects between pain and disgust. *Sci Rep, 5*, 17487. doi: https://doi.org/10.1038/srep17487

Silani, G., Lamm, C., Ruff, C. C., & Singer, T. (2013). Right Supramarginal Gyrus Is Crucial to Overcome Emotional Egocentricity Bias in Social Judgments. *The Journal of Neuroscience, 33*(39), 15466-15476. doi: http://doi.org/10.1523/jneurosci.1488-13.2013

Sladky, R., Friston, K. J., Tröstl, J., Cunnington, R., Moser, E., & Windischberger, C. (2011). Slice-timing effects and their correction in functional MRI. *NeuroImage, 58*(2), 588-594. doi: https://doi.org/10.1016/j.neuroimage.2011.06.078

Soudry, Y., Lemogne, C., Malinvaud, D., Consoli, S. M., & Bonfils, P. (2011). Olfactory system and emotion: Common substrates. *European Annals of Otorhinolaryngology, Head and Neck Diseases, 128*(1), 18-23. doi: https://doi.org/10.1016/j.anorl.2010.09.007

Steinbeis, N., Bernhardt, B. C., & Singer, T. (2015). Age-related differences in function and structure of rSMG and reduced functional connectivity with DLPFC explains heightened emotional egocentricity bias in childhood. *Social cognitive and affective neuroscience, 10*(2), 302-310. doi: https://doi.org/10.1093/scan/nsu057

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated Anatomical Labeling of Activations in SPM Using a

Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage, 15*(1), 273-289. doi: https://doi.org/10.1006/nimg.2001.0978

Xiong, R.-C., Fu, X., Wu, L.-Z., Zhang, C.-H., Wu, H.-X., Shi, Y., & Wu, W. (2019). Brain pathways of pain empathy activated by pained facial expressions: a meta-analysis of fMRI using the activation likelihood estimation method. *Neural regeneration research, 14*(1), 172-178. doi: http://doi.org/10.4103/1673-5374.243722

Zeidman, P., Jafarian, A., Corbin, N., Seghier, M. L., Razi, A., Price, C. J., & Friston, K. J. (2019a). A guide to group effective connectivity analysis, part 1: First level analysis with DCM for fMRI. *NeuroImage, 200*, 174-190. doi: https://doi.org/10.1016/j.neuroimage.2019.06.031

Zeidman, P., Jafarian, A., Seghier, M. L., Litvak, V., Cagnan, H., Price, C. J., & Friston, K. J. (2019b). A guide to group effective connectivity analysis, part 2: Second level analysis with PEB. *NeuroImage, 200*, 12-25. doi: https://doi.org/10.1016/j.neuroimage.2019.06.032

Zelano, C., Mohanty, A., & Gottfried, Jay A. (2011). Olfactory Predictive Codes and Stimulus Templates in Piriform Cortex. *Neuron, 72*(1), 178-187. doi: https://doi.org/10.1016/j.neuron.2011.08.010

Zhao, Y., Rütgen, M., Zhang, L., & Lamm, C. (2021a). Pharmacological fMRI provides evidence for opioidergic modulation of discrimination of facial pain expressions. *Psychophysiology, 58*(2), e13717. doi: https://doi.org/10.1111/psyp.13717

Zhao, Y., Zhang, L., Rütgen, M., Sladky, R., & Lamm, C. (2021b). Neural dynamics between anterior insular cortex and right supramarginal gyrus dissociate genuine affect sharing from perceptual saliency of pretended pain. *eLife, 10*, e69994. doi: http://doi.org/10.7554/eLife.69994

Zhou, F., Li, J., Zhao, W., Xu, L., Zheng, X., Fu, M., Yao, S., Kendrick, K. M., Wager, T. D., & Becker, B. (2020). Empathic pain evoked by sensory and emotional-communicative cues share common and process-specific neural representations. *eLife, 9*, e56929. doi: http://doi.org/10.7554/eLife.56929

Zhou, G., Lane, G., Cooper, S., Kahnt, T., & Zelano, C. (2019). Characterizing functional pathways of the human olfactory system. *eLife, 8*, e47177. doi: http://doi.org/10.7554/eLife.47177

Chapter 5 – General discussion

Summary of research

In this thesis, inspired by the theoretical model proposed by Coll et al. (2017), I employed behavioral, psychopharmacological, and functional brain imaging perspectives to investigate the potential role of emotion identification in empathy. Given the complexity of emotion identification and it is difficult to be directly operated and measured, here we focus on the processes essentially related to emotion identification (e.g., emotion recognition) and investigated the underlying opioidergic mechanism and how different levels of the recognized affect in others influenced empathic responses. In Chapter 2, we studied whether and how the opioid system influenced participants' judgments of painful facial expressions in an emotion discrimination/recognition task, with morphed facial expressions of pain and disgust at different intensities (Chapter 2). Specifically, participants, in a double-blind design, were allocated to two separate sessions, for which they were administered with either a naltrexone pill or a placebo pill in one session and the other pill in another session. Participants then performed an emotion discrimination/recognition task in the fMRI scanner, in which they were required to choose whether a mixed facial expression was showing either pain or disgust. On the behavioral level, we found participants less frequently to choose a morphed expression as pain during the naltrexone session as compared to the placebo session. On the neural level, participants showed parametrically enhanced activations in cortical visual association areas in the right hemisphere, including the fusiform face area, with increased pain intensity for both sessions; that is, the higher the intensity of pain was in a mixed expression, the stronger visual activation was in this region. Importantly, this parametric activation was generally stronger in the naltrexone session than in the placebo session. We considered this increased visual activation in the naltrexone session might reflect a compensatory effect in coping with the decreased sensitivity for pain expressions as compared to the placebo session. Further regression analysis showed that the parametric activation was individually associated with pain choices in both sessions. This study provided evidence 1) for the engagement of the opioid system in the discrimination/recognition of painful facial expressions, and 2) that during this process, the opioid system seemed to have more influence on visual-perceptive processing rather than affect processing, as we would have predicted based on previous research on empathy for pain (Rütgen et al., 2015). However, this finding does not necessarily suggest the opioidergic modulation merely had an effect on the visual-perceptive components and had no effect on the affect processing. I would argue that the lack of affective engagement due to the morphed expressions as well as the rather cognitive focus of the task on emotion recognition, may, at least to some extent, account for this.

In Chapter 3, we designed and used a novel paradigm to investigate the neural signatures underlying empathic responses induced by seeing others genuinely experiencing pain vs. the sensory-driven responses triggered by others' painful facial expressions though knowing the pain was merely acted. In the experiment, participants were instructed to observe video clips presenting a demonstrator's painful expressions when this person was receiving an injection on their cheek. Additionally, participants viewed videos of the same injection but the needle was covered by a protective cap (all conditions were thoroughly explained and clearly visible to participants). The demonstrator's facial expression, the demonstrator's feeling, and the participant's unpleasantness were rated following each video. Results showed higher ratings for all three aspects, indicating that participants identified higher pain in others and felt more unpleasantness for the genuine pain condition than the pretended pain condition. Moreover, stronger activations were found in the bilateral aIns, aMCC, and rSMG for genuine pain as compared to pretended pain, and the increased activity in aIns was unveiled to be selectively associated with the ratings of unpleasantness rather than ratings of pain expressions in others or painful feelings in others. A DCM analysis was implemented on the right aIns and rSMG, demonstrating reduced inhibitory modulatory effects on the aIns-to-rSMG connection for genuine pain as compared to pretended pain. In particular, the inhibitory modulatory effect detected for genuine pain was individually related to the ratings of painful feelings in others and empathic traits. No association between the modulatory effect and behavioral measurements was found for pretended pain. This study indicated that 1) the aIns activation induced by seeing others genuinely experiencing pain was indeed related to affect processing rather than merely perpetual saliency, and 2) the orchestration of aIns and rSMG tracked how other people really felt in painful situations.

Chapter 4 extended and built upon the study in Chapter 3. In this study, we focused on similar research questions but investigated it with the emotion of disgust; that is, which neural mechanisms engaged when seeing others genuinely experiencing disgust vs. pretending to be in disgust. Specifically, participants watched videos clips either showing someone who was genuinely sniffing the dog's faces, or merely pretending to smell something disgusting. We used the same ratings as Chapter 3 but were interested in disgust. Results showed increased activation in the right aIns and the left olfactory cortex for genuine disgust as opposed to pretended disgust. Similar to the pain findings, the increased aIns activation was particularly related to one's own unpleasantness rather than disgusted expressions or disgusted feelings in others. A DCM analysis was performed between the right aIns and rSMG, similar to what we did in the pain study, again showing inhibitory modulatory effects on the aIns-to-rSMG connection for both conditions; however, no difference was found between the genuine and pretended conditions. An exploratory DCM analysis was additionally performed between the right aIns and the left olfactory cortex. Results showed stronger excitatory modulatory effects on the olfactory-to-aIns connection for genuine disgust than pretended disgust, and this excitatory modulatory effect for genuine disgust was associated with an empathy-related trait, perspective-taking. No association with behavioral measurements was found for pretended disgust. Additionally, a third DCM analysis was performed between the right aIns and the right olfactory cortex, again we found excitatory modulatory effects on the connection from the olfactory cortex to aIns, though there was no difference between conditions. Combined with the findings in Chapter 3, this study thus suggests that: 1) the aIns is indeed an important region that can be specifically associated with affect sharing, of both pain and disgust, and 2) that different brain networks are engaged in responding to genuinely painful vs. disgusted experiences of others.

In the following sections, I will discuss in detail the scientific contributions, the limitations, and the suggested directions for further research based on the reported findings.

Implications and scientific contributions

In summary, my research has contributed to our understandings of how processes essentially related to emotion identification are engaged in and influence empathic responses in terms of 1) the theoretical implications, 2) the neural dynamics underlying affective and cognitive processes, and 3) cross-modal and modality-dependent evidence of different affective states. I will discuss these scientific contributions point-by-point as follows.

Firstly, all three studies have shown evidence to suggest that painful emotion recognition, an important process implicated in emotion identification, does play a role in empathy. According to the findings of Chapter 2, recognition of painful facial expressions can be modulated by the opioid system, for which the same system also shows the influence on empathy for pain (Rütgen et al., 2015; Karjalainen et al., 2017; Rütgen et al., 2018). The results on the neural mechanisms seem to be somehow against our expectation, that the opioidergic modulation is more likely to occur in the processing of visual perception rather than affect processing previously found by our lab in empathy for pain (Rütgen et al., 2015). I argue this “mismatch” may be attributed to different contexts (i.e., contents and instructions) of experimental paradigms implemented in the two studies. Using meta-analysis, Lamm et al. (2011) showed that empathy for pain induced by different paradigms engaged distinct neural networks that were related to the specific contexts: Observing pictures depicting another person’s limb in the painful situations strongly recruited activations in regions implicated in action understanding, whereas empathy elicited by visual cues (symbols) that indicated whether the target person was painful or not engaged more activations associated with representing self- and other-related processing. In addition, only the former paradigm induced activations in somatosensory areas. Moreover, many studies indicated that different instructions regarding the painful situations of others could significantly influence the consequence of empathic responses as well as the corresponding neural underpinnings (Gu & Han, 2007; Lamm et al., 2007a; Lamm et al., 2007b). Considering the distinct contexts of the two experimental paradigms, namely, observing others' morphed expressions of pain and disgust with the instruction to discriminate/recognize pain expression vs. watching symbolic cues representing others' states with the instruction to share others' painful affect, I would consider the "mismatched" results are not so unexpected/implausible as the two paradigms measure different subcomponents. Furthermore, absence of evidence is not evidence for absence; thus, a new experiment that better matches the contexts of the processes related to emotion identification (e.g., emotion recognition) and empathy is required to test this speculation and interpretation. Additionally, in Chapter 3 we detected a link between recognized pain in others and the modulatory effect from aIns to rSMG, a region that is particularly implicated in self- and other-related processing in the emotional domain, suggesting that the potential interaction of the recognition of others' pain and self-other processing might contribute to the empathic responses. In Chapter 4, we found the interplay between regions implicated in affective sharing and direct/indirect experience of disgust could dissociate genuine disgust vs. pretended disgust of others, implying the engagement of recognizing others' disgust in the modulation of empathic responses. All in all, these studies suggest that recognition of others' emotions, which is associated with different perceptual, cognitive, and affective processes related to emotion identification, indeed contributes to the consequences of empathizing with others.

Furthermore, our studies have demonstrated how affective and cognitive processes possibly interplay to contribute to the empathic response. In Chapter 3, we found modulatory effects (i.e., inhibition) between regions implicated in affect processing and affective self-other distinction when seeing others genuinely experiencing pain as well as acting to be in pain; specifically, the inhibitory modulatory effect from aIns to rSMG was higher for the pretended pain as compared to the genuine pain. We speculate two distinct functional mechanisms underlie these two modulations: for genuine pain, a modulation of "affective-to-affective" self-other distinction is achieved to disentangle the affective states originating from self from those attributed to others; for pretended pain, a "cognitive-to-affective" self-other distinction is required to resolve the conflicting information between the sensory-driven affective responses (i.e., "this person looks in pain") and the prior knowledge (i.e., "I know this person should not feel any pain at all"). In Chapter 4, we found stronger excitatory modulatory effects from regions implicated in the direct/indirect experiences of disgust (i.e., the primary olfactory cortex) to regions suggestive of affect processing (i.e., aIns), in response to seeing others genuinely experiencing disgust as opposed to acting disgusted. We argue this increased excitatory modulatory effect for genuine disgust may be related to the processing of conveying messages of more identified disgust in others to areas involved in affect processing, which may underpin the increased shared unpleasantness. This idea fits with the theoretical framework proposed by Coll et al. (2017), that the identified emotion in others contributes to shared affect, and the complete consequence of empathic responses is integrated by (at least) these two processes. Altogether, our studies shed light on the interactive activations underlying the cognitive and affective processes engaged in empathy, such as emotion recognition/identification, affect sharing, and self-other distinction. This thesis paves the way for further research to continue this work.

Last but not least, we have shown both cross-modal and modality-dependent neural signatures in terms of seeing others' pain (Chapter 3) and seeing others' disgust (Chapter 4). We found consistent evidence across these two studies that the increased aIns activation was always selectively associated with increased self-unpleasantness between the genuine and pretended conditions rather than perceived expressions or feelings in others. These findings have strongly supported the statement that aIns (perhaps also extended to aMCC) is essentially engaged in the affect processing of empathy rather than merely domain-general processes (e.g., automatic emotional processing and perceptual saliency). If one agrees with this interpretation, it triggers a follow-up question: whether this unpleasantness for genuine pain and disgust is generally induced by aversive experiences or specifically related to different emotions and their affective components. Albeit the current design does not allow us to answer this question directly, we did find some evidence from previous research and the current findings. Using multivariate pattern analysis (MVPA), Corradi-Dell'Acqua et al. (2016) found both cross-modal and modality-dependent brain patterns of the aIns activity for pain and disgust: in the left aIns (and aMCC), the shared coding has been suggested, regardless of the modality; whereas in the right aIns, a sensory-specific pattern has been more plausibly detected. In my studies, the aIns in the right hemisphere has been shown to be related to self-unpleasantness for both pain and disgust, and I speculate this activation to represent affect processing that is modality specific. Particularly, the DCM analyses we performed for pain and disgust demonstrated different connectivity: though inhibitory modulatory effects on the aIns-to-rSMG connection were found in the genuine and pretended conditions for both pain and disgust, the distinct modulations between conditions were only unveiled in pain but not in disgust. However, given the perceptual and affective salience of the pain and disgust stimuli was not precisely matched, we refrained from performing within-subject analyses of pain and disgust. It is thus hard to say whether the failure to completely replicate the findings across different affective experiences reflects modality-specific or domain-general distinct neural mechanisms, or not. Besides, we found the connectivity including the area specifically related to the sensory and affective components of disgust (i.e., the olfactory cortex) was able to be distinctly modulated by genuine disgust and pretended disgust. I would argue this modulatory effect might be particularly engaged in disgust emotion identification and may not occur for the identification of pain. All in all, these results indicate both cross-modal and modality-specific processes are implicated in processing empathy for different aversive experiences (i.e., pain and disgust), and future research is needed to further decode these overlapping and distinct neural underpinnings in empathy in terms of different affective experiences.

Limitations and future research

This thesis has contributed to our scientific understanding of how processes fundamentally related to emotion identification participated in and modulate empathy, yet several limitations remain for future research.

A first limitation is related to Chapter 2, that is, it is necessary to further clarify whether the effect of naltrexone manipulation on painful emotion recognition is mainly pain-specific or driven by features that only differ between pain and disgust but are not applied to other emotions. Since pain and disgust are closely related in many aspects, I would expect the results might be different when discriminating pain from another emotion, such as fear, with a rather similar paradigm. Therefore, it requires further experiments to investigate the discrimination/recognition of painful expressions with other negative emotions. Only in this way could we say whether the findings in Chapter 2 indicate the core opioidergic mechanisms underlying recognizing/identifying pain expressions of others, or not.

A second limitation regards Chapter 3, where we did not explicitly quantify the extent of self-other distinction. Albeit previous studies have shown converging evidence that self-other distinction is specifically related to the rSMG activation (Decety & Sommerville, 2003; Silani et al., 2013; Steinbeis et al., 2015; Hoffmann et al., 2016; Bukowski et al., 2020), this region is also engaged in some other processes, such as selective attention, action observation, and emotion imitation (e.g., Bach et al., 2010; Pokorny et al., 2015; Gola et al., 2017; Hawco et al., 2017). Therefore, I expect future research to explicitly quantify self- and other-related processes in an empathy-related task. For instance, one possible idea could be to test how people react to painful facial expressions (the intensity is matched beforehand) presented by their own and someone else in the painful situations (e.g., electrical stimulation), as well as by another person who also displays painful expressions but merely pretends to be in pain. By this design, one could directly quantify how participants reacted to self-directed genuine pain, other-directed genuine pain, and other-directed pretended pain.

A third limitation is about Chapters 3 and 4, where we failed to directly compare the responses of seeing others in pain and disgust together. The current discussion concerning the similarities and differences of seeing others in pain and disgust was not based on the direct findings of comparing these two affects. Instead, it was mostly inferred from the independent analyses of the two studies. As mentioned in the previous section, since the perceptual saliency of the vicarious experiences of pain and disgust in these two studies was not equal (i.e., higher behavioral ratings and stronger brain activations for pain in general as compared to disgust, according to preliminary analyses; the comparative results are not exhibited in this thesis), any variation detected in the comparison of behavioral measurements or brain activations could merely have been induced by the difference in stimuli salience or the responses to that salience. On top of that, another limitation is that the paradigms we used in Chapter 3 and 4 merely controlled for perceptual salience but not for affective salience, namely, empathic/vicarious responses to the perceptual salience. Therefore, though stronger activations are found in aIns and aMCC for genuine pain as compared to genuine disgust, we could not say pain induces more affect processing than disgust, as the perceptual salience of painful stimuli or the empathic responses induced by the perceptual salience may have been just generally higher than that of disgust. Further research is required to better match the saliency (both perceptual and affective) between pain and disgust so that researchers can perform a meaningfully conjunctive analysis to directly compare these two affects.

The last limitation concerns our research topic and research goals, that is, how much our studies could say about the role of emotion identification in empathy, and to what extent these findings test the theoretical model of Coll et al. (2017). As I have stated in the general introduction, it is hard to directly study emotion identification since the process of emotion identification is complex and it is difficult to be manipulated and measured. Furthermore, we did not aim to test the whole model, namely how emotion identification and affect sharing independently contribute to the consequences of empathy. Instead, we focused on the opioid mechanisms underlying recognition of painful expressions, since previous studies evidenced the link between the opioid system and empathy for pain, and how different levels of recognized affect in others (i.e., genuine pain/disgust vs. pretended pain/disgust) influenced the consequences of empathic responses. In a way, the current studies could not directly answer how emotion identification itself influences empathic responses, however, compared with previous research these findings have already provided more explicit and quantified evidence implying the engagement of emotion identification in empathy. To investigate the model more specifically, future research that more precisely quantifies the process of emotion identification is required.

References

Bach, P., Peelen, M. V., & Tipper, S. P. (2010). On the Role of Object Information in Action Observation: An fMRI Study. *Cerebral Cortex, 20*(12), 2798-2809. doi: http://doi.org/10.1093/cercor/bhq026

Bukowski, H., Tik, M., Silani, G., Ruff, C. C., Windischberger, C., & Lamm, C. (2020). When differences matter: rTMS/fMRI reveals how differences in dispositional empathy translate to distinct neural underpinnings of self-other distinction in empathy. *Cortex, 128*, 143-161. doi: https://doi.org/10.1016/j.cortex.2020.03.009

Coll, M.-P., Viding, E., Rütgen, M., Silani, G., Lamm, C., Catmur, C., & Bird, G. (2017). Are we really measuring empathy? Proposal for a new measurement framework. *Neuroscience & Biobehavioral Reviews, 83*, 132-139. doi: https://doi.org/10.1016/j.neubiorev.2017.10.009

Corradi-Dell'Acqua, C., Tusche, A., Vuilleumier, P., & Singer, T. (2016). Cross-modal representations of first-hand and vicarious pain, disgust and fairness in insular and cingulate cortex. *Nature Communications, 7*(1), 10904. doi: 10.1038/ncomms10904

Decety, J., & Sommerville, J. A. (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends in cognitive sciences, 7*(12), 527-533. doi: https://doi.org/10.1016/j.tics.2003.10.004

Gola, K. A., Shany-Ur, T., Pressman, P., Sulman, I., Galeana, E., Paulsen, H., Nguyen, L., Wu, T., Adhimoolam, B., Poorzand, P., Miller, B. L., & Rankin, K. P. (2017). A neural network underlying intentional emotional facial expression in neurodegenerative disease. *NeuroImage: Clinical, 14*, 672-678. doi: https://doi.org/10.1016/j.nicl.2017.01.016

Gu, X., & Han, S. (2007). Attention and reality constraints on the neural processes of empathy for pain. *NeuroImage, 36*(1), 256-267. doi: https://doi.org/10.1016/j.neuroimage.2007.02.025

Hawco, C., Kovacevic, N., Malhotra, A. K., Buchanan, R. W., Viviano, J. D., Iacoboni, M., McIntosh, A. R., & Voineskos, A. N. (2017). Neural Activity while Imitating Emotional Faces is Related to Both Lower and Higher-Level Social Cognitive Performance. *Scientific Reports, 7*(1), 1244. doi: http://doi.org/10.1038/s41598-017-01316-z

Hoffmann, F., Koehne, S., Steinbeis, N., Dziobek, I., & Singer, T. (2016). Preserved Self-other Distinction During Empathy in Autism is Linked to Network Integrity of Right Supramarginal Gyrus. *Journal of Autism and Developmental Disorders, 46*(2), 637-648. doi: http://doi.org/10.1007/s10803-015-2609-0

Karjalainen, T., Karlsson, H. K., Lahnakoski, J. M., Glerean, E., Nuutila, P., Jääskeläinen, I. P., Hari, R., Sams, M., & Nummenmaa, L. (2017). Dissociable roles of cerebral µ-opioid and type 2 dopamine receptors in vicarious pain: a combined PET–fMRI study. *Cerebral Cortex, 27*(8), 4257-4266. doi: https://doi.org/10.1093/cercor/bhx129

Lamm, C., Batson, C. D., & Decety, J. (2007a). The Neural Substrate of Human Empathy: Effects of Perspective-taking and Cognitive Appraisal. *Journal of Cognitive Neuroscience, 19*(1), 42-58. doi: 10.1162/jocn.2007.19.1.42

Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage, 54*(3), 2492-2502. doi: https://doi.org/10.1016/j.neuroimage.2010.10.014

Lamm, C., Nusbaum, H. C., Meltzoff, A. N., & Decety, J. (2007b). What are you feeling? Using functional magnetic resonance imaging to assess the modulation of sensory and affective responses during empathy for pain. *PLoS One, 2*(12). doi: http://doi.org/10.1371/journal.pone.0001292

Pokorny, J. J., Hatt, N. V., Colombi, C., Vivanti, G., Rogers, S. J., & Rivera, S. M. (2015). The Action Observation System when Observing Hand Actions in Autism and Typical Development. *Autism Research, 8*(3), 284-296. doi: https://doi.org/10.1002/aur.1445

Rütgen, M., Seidel, E. M., Pletti, C., Riecansky, I., Gartus, A., Eisenegger, C., & Lamm, C. (2018). Psychopharmacological modulation of event-related potentials suggests that first-hand pain and empathy for pain rely on similar opioidergic processes. *Neuropsychologia, 116*(Pt A), 5-14. doi: https://doi.org/10.1016/j.neuropsychologia.2017.04.023

Rütgen, M., Seidel, E. M., Silani, G., Riecansky, I., Hummer, A., Windischberger, C., Petrovic, P., & Lamm, C. (2015). Placebo analgesia and its opioidergic regulation suggest that empathy for pain is grounded in self pain. *Proceedings of the National Academy of Sciences, 112*(41), E5638-E5646. doi: https://doi.org/10.1073/pnas.1511269112

Silani, G., Lamm, C., Ruff, C. C., & Singer, T. (2013). Right Supramarginal Gyrus Is Crucial to Overcome Emotional Egocentricity Bias in Social Judgments. *The Journal of Neuroscience, 33*(39), 15466-15476. doi: http://doi.org/10.1523/jneurosci.1488-13.2013

Steinbeis, N., Bernhardt, B. C., & Singer, T. (2015). Age-related differences in function and structure of rSMG and reduced functional connectivity with DLPFC explains heightened emotional egocentricity bias in childhood. *Social Cognitive and Affective Neuroscience, 10*(2), 302-310. doi: https://doi.org/10.1093/scan/nsu057

Chapter 6 – Abstract

Our affective states are influenced by how we perceive others' feelings. Successful identification of others' emotions is vital for personal interaction, social cooperation, and harmony. With this doctoral thesis, I aimed to investigate the opioidergic mechanism underlying some processes essentially related to emotion identification and how these processes regulated the consequences of empathic responses. Three studies are reported, which involved behavioral measurements, neuroimaging, and psychopharmacological manipulation. In the first study, I investigated whether and how the opioid system influenced the discrimination of painful from disgusted facial expressions. Using the opioid antagonist naltrexone, I found that opioidergic modulation had a negative effect on the recognition of pain, and that this was linked to brain areas associated with visual-perceptive rather than affective neural processing. In the second study, I examined which neural networks dissociated the empathic responses when observing others genuinely experiencing pain vs. acting out pain. Using dynamic causal modeling (DCM), I found distinct effective connectivity between brain regions underlying affect processing (i.e., anterior insula, aIns) and self-other distinction (i.e., the right supramarginal gyrus, rSMG) for genuine vs. pretended pain, which seems to track how another person really feels. The third study aimed to complement and extend the second study on pain. Using a similar paradigm but with disgust (genuine disgust vs. pretended disgust), I replicated that activation in aIns was related to affect sharing rather than merely perceptual saliency. Connectivity between aIns and the olfactory cortex, rather than between aIns and rSMG, was found to track distinctive responses for genuine compared to pretended disgust. Taken together, this thesis has advanced our understanding of the neural networks involved in empathy and affect sharing, and in the potential role of emotion identification for these processes. It also highlights how empathy can be studied as an interactive and dynamic process comprised of different perceptive, cognitive, and affective processes.

Zusammenfassung

Wie wir die Gefühle Anderer wahrnehmen, beeinflusst unseren affektiven Zustand. Eine erfolgreiche Identifikation der Gefühle Anderer ist wichtig für persönliche Interaktionen, soziale Kooperation und Harmonie. Mein Ziel in dieser Doktorarbeit war es, sowohl opioiderge Mechanismen zu untersuchen, die denjenigen Prozessen zugrunde liegen, welche mit der Identifikation von Emotion in Verbindung stehen, als auch wie diese Prozesse die Konsequenzen empathischer Reaktionen regulierten. Diese Dissertation umfasst drei Studien, bei welchen behaviorale Maße, neuronale Bildgebung und psychopharmakologische Manipulation zum Einsatz kamen. In meiner ersten Studie untersuchte ich, inwiefern das Opioidsystem die Unterscheidung der Gesichtsausdrücke von Schmerz und Ekel beeinflusste. Durch die Verwendung des Opioidantagonisten Naltrexon konnte ich feststellen, dass opioiderge Modulation einen negativen Effekt auf die Erkennung von Schmerz hatte, welcher auf visuell-perzeptive Gehirnregionen zurückzuführen war, jedoch nicht auf affektive neuronale Prozesse. In meiner zweiten Studie untersuchte ich, welche neuronalen Netzwerke empathische Reaktionen bei der Beobachtung Anderer beeinflussten, wenn diese authentischen Schmerz empfinden oder Schmerz vortäuschen. Durch die Verwendung von „Dynamic Causal Modeling“ (DCM) konnte ich eine veränderte Konnektivität zwischen Gehirnregionen feststellen, welche affektiven Prozessen (anteriore Insula, aIns) und der Unterscheidung des Selbst von Anderen (rechter supramarginaler Gyrus, rSMG) unterliegen, je nachdem ob authentischer oder vorgetäuschter Schmerz beobachtet wurde. Änderungen in dieser Konnektivität tragen also augenscheinlich dazu bei zu erkennen, wie eine andere Person tatsächlich fühlt. Das Ziel der dritten Studie war es, die zweite Studie zu Schmerz zu replizieren und auszuweiten. Durch die Verwendung eines ähnlichen Paradigmas, diesmal bezogen auf Ekel (authentischer Ekel vs. vorgetäuschter Ekel), konnte ich replizieren, dass Aktivierung in der aIns eher mit geteiltem Affekt als reiner wahrgenommener Salienz in Zusammenhang stand. Um unterschiedliche Reaktionen für authentischen Ekel im Vergleich zu vorgetäuschtem Ekel zu erkennen, wurde die Konnektivität zwischen aIns und dem olfaktorischen Kortex, statt zwischen aIns und rSMG untersucht. Insgesamt konnte diese Arbeit unser Verständnis der neuronalen Netzwerke, welche in Empathie und geteilten Affekt involviert sind, verbessern, sowie die potenzielle Rolle der Emotionserkennung für diese Prozesse aufzeigen. Außerdem zeigt sie, wie Empathie als interaktiver und dynamischer Prozess untersucht werden kann, der aus unterschiedlichen perzeptiven, kognitiven und affektiven Prozessen besteht.

Chapter 7 – Curriculum Vitae

YILI ZHAO

Liebiggase 5, 1010 Vienna, Austria

(+43) 67761779997 ◇ yili.zhao@univie.ac.at

EDUCATION

10/2016 - present **PhD candidate, Psychology**
Social, Cognitive, and Affective Neuroscience Unit
Department of Cognition, Emotion, and Methods in Psychology
University of Vienna, Vienna, Austria
Supervisor: Prof. Claus Lamm

09/2013 - 06/2016 **MSc, Applied Psychology**
Institute of Psychology, Chinese Academy of Sciences, Beijing, China
Supervisor: Dr. Wencai Zhang

09/2009 - 06/2013 **BSc, Applied Psychology**
Department of Philosophy, Anhui University, Hefei, China
GPA: 3.75 /4.00 (ranked 1st in the class)

FELLOWSHIPS, HONORS & AWARDS

06/2021 **Best Poster Award** (only one winner)
*The European Society for Cognitive and Affective Neurosc*ience

01/2021 - 06/2021 **VDS CoBeNe Completion Grant**
Vienna Doctoral School in Cognition, Behavior and Neuroscience,
University of Vienna

10/2016 - 10/2020 **CSC Graduate Scholarship**
Chinese Scholarship Council

10/2015 **National Scholarship for Graduate Students** (2% of all students)
Chinese Ministry of Education

05/2015	**Best Debater Award of Life Sciences** *Beijing Institutes of Life Science, Chinese Academy of Sciences*
04/2015	**First-Class Researching Scholarship** (5% of all students) *Institute of psychology, Chinese Academy of Sciences*
04/2015	**Excellent Student Representative Award** (2% of all students) *Institute of psychology, Chinese Academy of Sciences*
11/2014	**Excellent Student Award** (15% of all students) *University of Chinese Academy of Sciences*
11/2012	**First-Class Learning Scholarship** (5% of all students) *Anhui University*
11/2012	**Excellent Student Award** (8% of all students) *Anhui University*
11/2011	**First-Class Learning Scholarship** (5% of all students) *Anhui University*
11/2011	**Excellent Student Award** (8% of all students) *Anhui University*
11/2010	**Tianda Scholarship** (1 out of 114) *Anhui University*
10/2010	**Third Prize of "FLTRP Cup" English Public Speaking Contest** *Anhui University*

PUBLICATIONS

Zhao, Y., Zhang, L., Rütgen, M., Sladky. R., & Lamm, C. (2021). Effective connectivity reveals distinctive patterns in response to others' genuine affective experience of disgust as compared to pain. *bioRxiv*. 2021.2009.2003.458875. https://doi.org/10.1101/2021.09.03.458875

Zhao, Y., Zhang, L., Rütgen, M., Sladky. R., & Lamm, C. (2021). Neural dynamics between anterior insular cortex and right supramarginal gyrus dissociate genuine affect sharing from perceptual saliency of pretended pain. *eLife, 10*, e69994. http://doi.org/10.7554/eLife.69994

Zhao, Y., Rütgen, M., Zhang, L., & Lamm, C. (2021). Pharmacological fMRI provides evidence for opioidergic modulation of discrimination of facial pain expressions.*Psychophysiology, 58* (2), e13717. http://doi.org/10.1111/psyp.13717

Zhao, Y., Liu, R., Zhang, J., Luo, J., & Zhang, W. (2020). Placebo Effect on Modulating Empathic Pain: Reduced Activation in Posterior Insula. *Frontiers in behavioral neuroscience, 14*, 8-8. doi: http://doi.org/10.3389/fnbeh.2020.00008

Zhao, Y., Zhang, J., Yuan, L., Luo, J., Guo, J., & Zhang, W. (2015). A transferable anxiolytic placebo effect from noise to negative effect. *Journal of Mental Health, 24*(4), 230-235. doi: http://doi.org/10.3109/09638237.2015.1021900

RESEARCH EXPERIENCES

Oral Presentations

10/2014 *The 17th National Academic Congress of Psychology, Beijing, China.* **Zhao, Y.,** Zhang, W. A transferable anxiolytic placebo effect from noise to negative emotions.

Poster Presentations

06/2021 *The European Society for Cognitive and Affective Neuroscience, Budapest, Hungary.* **Zhao, Y.**, Zhang, L., Rütgen, M., Sladky. R., & Lamm, C. (2021). Neural dynamics between anterior insular cortex and right supramarginal gyrus dissociate genuine affect sharing from automatic responses to pretended pain

05/2019 *The Organization for Human Brain Mapping (OHBM), Rome, Italy.* **Zhao, Y.**, Rütgen, M., Lamm, C. The role of the endogenous opioid system in emotion identification.

07/2018 *The European Society for Cognitive and Affective Neuroscience (ESCAN), Leiden, Netherland.* **Zhao, Y.**, Rütgen, M., Lamm, C. The role of the endogenous opioid system in emotion identification of painful expressions.

06/2015 *The 5th Mental Health Conference, Haikou, China.* **Zhao, Y.**, Yu, F., Zhang., W. An experimental study on the transferable placebo effect from pain to empathy for pain.

Workshops & Training Courses

11/2018	*Workshop "From Self-knowledge to Knowing Others: Insights from psychological and neuroscientific tools", Brussels, Belgium*
09/2017	*Python course, Vienna, Austria*
04/2017	*SPM course, Edingburgh, UK*

RESEARCH SKILLS

Research Methods

Behavioral measure, fMRI, computational modeling, and psychopharmacological administration

Programming

MATLAB, R, Python, and Shell scripting

Stimulus Presentation

Cogent Toolbox, Psychtoolbox, and E-Prime

Data Analysis

SPSS, SPM, R, and Python

INTERN SUPERVISIONS

2019	**Michael Schnödt, Elisa Warmuth** University of Vienna
	Lukas Repnik Sigmund Freud Private University
2018 - 2019	**Nhu Nguyen Phan, Sven Sander, Betty Geidel** University of Vienna
2018	**Luise Huybrechts, Robert Meyka, Gvantsa Gogisvanidze, Anja Tritt** University of Vienna

LANGUAGES

Mandarin (Native), English (Fluent), German (Basic)

Acknowledgements

It is hard to imagine I have been in Vienna for almost five years, and finally, the time to say goodbye comes. During this period of my doctoral research, I have suffered a lot but gained more. In particular, there are several important people I would like to express my great gratitude to. Without their support and help, I cannot get the current achievements and succeed to the end of this long journey.

Firstly, I would express my deepest thanks to my supervisor Prof. Claus Lamm, who is the main reason that makes me here in Austria for pursuing a doctorate degree and also gives me all-around support during the whole period. As the first doctoral student in the lab who comes from Asia, I did come across many obstacles, such as culture, language, and academia at the very beginning. I am so grateful that I got the full-heart support and help from Claus especially during my toughest time. Since he is such an outstanding researcher, I have learned so much from Claus concerning how to do research and the scientific way of thinking. Now it is almost till the end of the study, I could already say that it is one of the best decisions of my life to study with Claus for my doctoral research.

Furthermore, I would also thank my two (pre) co-supervisors, Dr. Lei Zhang and Dr. Markus Rütgen. Markus is the one who helped me to gradually get familiar with everything in the lab at the beginning, and who practically supervised me to start my first study here. Lei is also a fancy supervisor, who could always give me very practical suggestions whenever I ask for his opinion. From Lei, I do not only learn how to properly look for the solutions to specific scientific problems, but also gain a lot from how to build up my academic career and become a professional researcher. Here I would greatly thank these two guys, I could not have achieved so much without their help and support.

Moreover, I want to thank all the colleagues at the SCAN Unit, Ronny, Paul, Katja, Jonas, Sigrid … (here I only list some of these lovely people given there are so many names). It is a pleasure to work with you, and I'm grateful for all the kindness and help that you have given to me.

Particularly, I would be greatly thankful to my parents and my boyfriend, Peng Gao. Without their support, I can hardly imagine I could manage through my doctoral life. Even though we are so far away in distance (three different continents), their support is always the biggest strength and comfort for me. It's my parents and my boyfriend who have always accompanied me to get through those difficult times and make me a better person.

Last but not least, I want to thank the Chinese Scholarship Council (CSC) and the Vienna Doctoral School in Cognition, Behavior and Neuroscience (VDS CoBeNe), with whose financial support that I can finally make this thesis.

My journey in academia continues, and I would never forget these precious memories that I have gained in Vienna, with you.