**Review article**

**Social brain, social dysfunction and social withdrawal**

Stefano Porcelli, Nic Van Der Wee, Steven van der Werff, Moj Aghajani, Jeffrey C. Glennon, Sabrina van Heukelum, Floriana Mogavero, Antonio Lobo, Francisco Javier Olivera, Elena Lobo, Mar Posadas, Juergen Dukart, Rouba Kozak, Estibaliz Arce, Arfan Ikram, Jacob Vorstman, Amy Bilderbeck, Ilja Saris, Martien J. Kas, Alessandro Serretti

---

**ARTICLE INFO**

**Keywords:**
Social withdrawal
Neurobiology
Social brain
Social dysfunction
Social impairments
Social cognition
Social functioning
Schizophrenia
Alzheimer's disease
Major depression disorder

**ABSTRACT**

The human social brain is complex. Current knowledge fails to define the neurobiological processes underlying social behaviour involving the (patho-) physiological mechanisms that link system-level phenomena to the multiple hierarchies of brain function. Unfortunately, such a high complexity may also be associated with a high susceptibility to several pathogenic interventions. Consistently, social deficits sometimes represent the first signs of a number of neuropsychiatric disorders including schizophrenia (SCZ), Alzheimer’s disease (AD) and major depressive disorder (MDD) which leads to a progressive social dysfunction. In the present review we summarize present knowledge linking neurobiological substrates sustaining social functioning, social dysfunction and social withdrawal in major psychiatric disorders. Interestingly, AD, SCZ, and MDD affect the social brain in similar ways. Thus, social dysfunction and its most evident clinical expression (i.e., social withdrawal) may represent an innovative transdiagnostic domain, with the potential of being an independent entity in terms of biological roots, with the perspective of targeted interventions.

1. Background

The complexity of the processes that underlie social living is enormous, including processes such as the detection and processing of social stimuli, mentalizing activity, bond/relationships formation, social learning and so on (for detail see Cacioppo et al., 2014; Dunbar and Shultz, 2007; Dunbar, 2009). These processes are highly relevant in social species such as homo sapiens, to the point that some have suggested that complex social environments were the primary selective pressure for the human brain, being mediated by all the aspects of social problem solving (Dunbar and Shultz, 2007; Semendeferi et al., 2001, 2002). As a consequence of this 'social' evolutionary pressure, human brain shows a high degree of specialization for social stimuli processing, encompassing regulation from the neurotransmitter to the neural network level resulting in a 'social brain' (Dunbar, 2009). Economic processes underlie evolution with adaptation of the structures and neurotransmitters involved from their original general functions to the processing of social stimuli. Some structures (e.g., the Bed Nucleus of Stria Terminalis - BNST) and neurotransmitters (e.g., oxytocin - OXT), show a high degree of specialization for the processing of social stimuli. Unfortunately, such a high complexity may also be associated with a high susceptibility to several pathogenic interventions. Deficits in these processes may result in personal difficulties and interpersonal problems. The high vulnerability of the social brain is confirmed by the
clinical observation that social deficits can sometimes represent the first
signs of a number of neuropsychiatric disorders, manifesting far before
the full onset of the other symptoms (NICE, 2014). Social deficits could
be broadly defined as impairments in the subject’s capacity to integrate
behavioural, cognitive, and affective skills to flexibly adapt to diverse
social contexts and demands (Bierman and Welsh, 2000), resulting in
behavioural outcomes which are judged as negative according to the
standards of the specific social context (i.e. as impairments of the social
competence) (Dirks et al., 2007). Despite the fact that a large amount of
data about social dysfunction comes from studies on schizophrenia
(SCZ), where several deficits in social processes have been identified
(Addington and Addington, 2008; Fett et al., 2011; Green et al., 2015),
in recent years similar deficits have been described and recognized
more and more also in other neuropsychiatric disorders including social-
communication deficits in Autism spectrum disorders (ASD), empathy
dysregulation in Psychopathy, docility and visual agnosia in Kluver-Bucy Syndrome and social reclusion in Hikikomori Syndrome, all of which alter social functioning (Barak and Feng, 2016; Li and
Wong, 2015). However, deficits in social functioning have been increas-
ingly recognized in other neuropsychiatric disorders, such as Alzheimer’s Disease (AD) and other dementias (Dickson, 2015; Havins
et al., 2012), Major Depressive Disorder (MDD) (Bora and Berk, 2016; Kupferberg et al., 2016a), anxiety disorders (Plana et al., 2014), and
borderline and antisocial personality disorders (Beene et al., 2015;
Jeung and Herpertz, 2014; Patin and Hurlemann, 2015; Cotter et al.,
2018). Intriguingly, William’s Syndrome, is characterized by contrast-
ing patterns of deficits in social domains, resulting in hypersocia-
bility represented by an unusually cheerful demeanor and ease with
strangers (Barak and Feng, 2016). This provides a clear example of how
social dysfunction can result in different behavioural outcomes, ranging
from social avoidance to inappropriate friendly behaviours with strangers. However, all these behaviours may result in unsuccessful
social interactions. By causing repetitive inappropriate social beha-
viours, social dysfunction often results in a progressive withdrawal
from relationships and social living in general, which in turn contribute
to further worsening any psychiatric symptoms already present. The
deficits in social cognition (i.e. the ensemble of mental operations that
underlie social interactions, including perceiving, interpreting, and
generating responses to the intentions, dispositions, and behaviours of
others (Adolphs, 1999; Green et al., 2008; Kunda, 1999)) are reinforced
by social deprivation (Cacioppo and Hawkley, 2009; Cornwell and
Waite, 2009; El Haj et al., 2016; Kennedy and Adolphs, 2012; Tremeau
et al., 2016; Zhong et al., 2017; Hoffman, 2007). Clearly, social dys-
function as a whole is a complex phenotype, which is influenced by a
variety of socio-demographic features, as well as by basic domain def-
cits, in attention, working memory, and sensory processing. Alter-
natively, different neuropsychiatric disorders may share these impair-
ments (at least partially), which in turn may determine social dysfunction. Nonetheless, a growing amount of evidence suggests that
social dysfunction is partially independent from other symptoms/defi-
cits, as well as from cognitive and even from social cognitive impair-
ments. Therefore, the observed social dysfunction likely reflects (at
least partially) alterations in the social brain itself, which may be in-
dependent from other domains.

In the present review we will discuss how three different, frequent,
and highly impacting neuropsychiatric disorders (WHO, 2008; Wittchen
et al., 2011) (namely Alzheimer’s disease - AD, Schizophrenia - SCZ, and
Major Depressive Disorder - MDD) share a final common pathway that
affects the social brain, characterized by a similar social dysfunction
(although with different degrees of impairment), which often causes impairment in the ability to form/maintain social relationships
and networks, resulting in the final, deleterious, outcome of social
withdrawal (i.e., a disengagement from social activities that lead to
impoverished interpersonal relationships). These three neu-
ropsychiatric disorders were selected among the several ones char-
acterized by social dysfunction (Cotter et al., 2018) because of their
frequencies and heavy burden in Western countries (WHO, 2008;
Wittchen et al., 2011) (globally, they account for 31.5% of disability-
adjusted life years - DALYs - associated with neuropsychiatric disorders
and substance use disorders (Whiteford et al., 2015)) and because social
withdrawal often represents one of their first clinical features. For these
same reasons, these disorders will be investigated in the context of a
European founded project which aims to provide quantitative biological
measures for social and cognitive deficits, the PRISM project described
in this issue (E Kas et al., 2017) and Bilderbeck et al. in this issue.
However, they represent only three examples to show how different
psychopathological mechanisms could similarly affect social brain, re-
sulting in social dysfunction and eventually in social withdrawal.
Among the several behavioural outcomes associated with social dys-
function (e.g., socially disinhibition, inappropriate behaviour, etc.), we
will focus mainly on social withdrawal because it is an important source
of indirect costs and it has been identified as one of the main reasons
for mental health related disability benefit claims (UK Department for
Work and Pensions, 2013). Furthermore, it can be observed and mea-
sured in an objective way and it represents a real-world indicator of
social dysfunction (see Van der Wee et al. in this issue). This is not the
case for instance for social cognition impairments where a difference
between experimental performances and real-world functioning has
been repeatedly demonstrated (e.g., (Torralva et al., 2013), although
certainly some degree of correlation exist between them (Bierman and
Welsh, 2000; Cotter et al., 2018; Couture et al., 2011; Fett et al., 2011;
McKibbin et al., 2004)). It is beyond the aim of the present paper to
provide a comprehensive review of literature data on social func-
tioning, because of the enormous amount of data on this issue and the
several excellent reviews on single facets of this topic published so far
e.g. (Kennedy and Adolphs, 2012; Kupferberg et al., 2016a; Lewandowski et al., 2016; Macdonald and Leary, 2005; Mar, 2011;
Mercedes Perez-Rodriguez et al., 2015; Patin and Hurlemann, 2015;
Billing et al., 2008; Rocca et al., 2016; Shinagawa et al., 2015; Van
Overwalle, 2009)). Instead, we aim to provide a global view of the
neurobiological substrates of social functioning and their relationships
with basic cognitive domains, and to underline how those may be
aberrant in three among the most frequent and deleterious neu-
ropsychiatric disorders (i.e., AD, SCZ, and MDD) in a similar way
driving to social withdrawal. In doing so, we aim to suggest how social
dysfunction, and specifically social withdrawal, may represent an in-
novative transdiagnostic domain, with the potential of being an in-
dependent entity in terms of biological roots, with the perspective of
targeted interventions.

2. The social brain: neuroanatomical substrates

In the early nineties, the basic components of the “social brain” were
identified in the orbitofrontal cortex (OFC), amygdala, and temporal
cortex (mainly the superior temporal sulcus -STS) (Brothers, 1990). In
the later decade, other regions, such as the medial prefrontal cortex
(mPFC) and the anterior cingulate cortex (ACC), have been identified
as relevant for social functioning and were added to this original core
(Bickart et al., 2014b; Frith and Frith, 2006). Recent conceptualizations
of the social brain typically describe it as encompassing a dynamic and
hierarchical system of circuitry involved in simpler forms of more au-
tomated processing, like the detection of socially relevant stimuli, and
partially overlapping circuitry involved in higher order processes, like
reflecting on one’s own or others’ mental states.

In an influential recent review, Bickart et al. (2014b) reviewed and
summarized the large body of available functional, anatomical, and
neuropsychological data from rodents and primates on key regions, and
more importantly, on circuitry involved in the social brain. Based on
this extensive review, the authors delineated five large-scale brain
networks: three partially distinct brain networks anchored in the
amygdala (the so-called social perception network, social affiliation
network and social aversion network) (Bickart et al., 2014a,b), and two
other large-networks assemblies already extensively described, i.e. the mirror network (Rizzolatti and Craighero, 2004) and the mentalizing network (Frith and Frith, 2006) (Fig. 1). These networks overlap with the eight canonical brain networks (Yeo et al., 2011) (e.g., social perception and affiliation networks with the default mode network (Bickart et al., 2014b)), suggesting, as can be expected, that many of the structures involved play a number of roles also in other mental processes. Thus, mutual, many still to be elucidated, inter-dependencies likely exist across the different neural networks. In the present paper, we decided to focus on the networks recently identified by Bickart et al. (Fig. 1) because of the converging evidences supporting the presence of these networks also in humans (Kerestes et al., 2017) and their role in social functioning (Bickart et al., 2014a; Deuse et al., 2016; Hampton et al., 2016). We are aware that this may be an oversimplification of the reality, which is extremely complex as suggested by animal studies (see for example (Bergan, 2015; Newman, 1999)), but we believe that the networks identified by Bickart et al. may be a useful framework to discuss the other findings about the neurobiological basis of social dysfunction and to provide future research hypotheses. In the following paragraph we will briefly summarize the networks described in Bickart et al. (Bickart et al., 2014b), underlining the aspects of social functioning and social cognition which have been already associated with them. Since social cognition could be easily assessed in experimental conditions compared to other aspects of social functioning, the majority of the discussion will focus on this data. When possible, specific links to social withdrawal will be provided.

2.1. Social perception: detection and processing of social stimuli

Detection of social stimuli is pivotal in order to successfully engage in social interaction. This process is vastly integrated in memory systems to rapidly classify stimuli as salient based on previous experiences (Adolphs, 1999, 2009; Bickart et al., 2014b). Bickart et al. proposed that the amygdala acts as a central hub for supporting social perception, by orchestrating the perception network (Fig. 1) (Bickart et al., 2014b).

Information processing through this network occurs rapidly and automatically, and it is involved in vigilance for potentially salient stimuli (Herry et al., 2007; Whalen, 2007). In doing so, it is likely to interact with the salience network (Seeley et al., 2007), as suggested by the partial overlap between these networks. The salience network detects the valence of internally and externally relevant events, with subsequent activation of other neuro-circuitry and higher-order cognitive controls (Menon, 2011).

With respect to social perception, most relevant visual information can be derived from expressive aspects of the face and body of others (e.g., eye gaze). Amygdala seems to mediate the activity of the face perception network, which composes of brain areas that show preferential activity to faces compared to other stimuli, such as the fusiform face area (FFA), the posterior STS (pSTS), and the occipital face area (OFA) (Gobbini and Haxby, 2006; Haxby et al., 2000; Ho et al., 2011) and lesion studies show an impaired ability of facial emotional recognition as a result of amygdala damage (Adolphs et al., 1994; Vuilleumier et al., 2004). Finally, correlations between the strength of functional connectivity within the constituents of the face perception network (amygdala included) and face emotional process have been repeatedly found (Cohen Kadosh et al., 2011; Marsh, 2016; O’Neil et al., 2014; Wang et al., 2016; Zhu et al., 2011). Intuitively, impairments in emotion recognition could determine misinterpretation of social signals during interpersonal interactions (e.g. (Domes et al., 2009)), with deleterious consequences on social relationships, as we will discuss using as examples AD, SCZ, and MDD. However, individuals with damages within the amygdala show also an impaired ability to guide their visual attention to the region of the eyes (i.e. one of the most expressive part of the face), suggesting that the impairment in facial processing observed in these patients may be due to an inability to direct attentional resources to relevant social information rather than to a direct damage of the face perception network itself (Adolphs et al., 2014a).

Fig. 1. The five large-scale brain networks sustain processes important for social behavior. Figure adapted from (Bickart et al., 2014a).

*Reproduced from the “Atrophy in distinct corticolimbic networks in frontotemporal dementia relates to social impairments measured using the Social Impairment Rating Scale”, Vol 85, pages 438–448, 2014, with permission from BMJ Publishing Group Ltd.

Perception network: IOFC = lateral orbito frontal cortex; vTP = ventro lateral temporal pole; FG = fusiform gyrus; STS = superior temporal sulcus. Aversion network: dTP = dorso medial temporal pole; rACC = rostral anterior cingulate cortex; sgACC = subgenual anterior cingulate cortex; vmPFC = ventromedial prefrontal cortex; Ent = entorhinal cortex; PHip = para hippocampal cortex; vmSt = ventro medial striatum. Mentalizing network: dmPFC = dorso medial prefrontal cortex; PCC = posterior cingulate cortex; Ins = insula; SII = somatosensory operculum; vlSt = ventro lateral striatum. Mirror network: pSTS = posterior superior temporal sulcus; IPS = inferior parietal sulcus; AngG = angular gyrus (temporoparietal junction). Aversion network and mentalizing network: Amy = amygdala; MTL = medial temporal lobe; vm = ventromedial; dm = dorso medial; PHp = para hippocampus; ventro = ventral; vmSt = ventro medial striatum; dmPFC = dorso medial prefrontal cortex; PCC = posterior cingulate cortex; Ins = insula; SII = somatosensory operculum; vlSt = ventro lateral striatum. Mentalizing network: dmPFC = dorso medial prefrontal cortex; PCC = posterior cingulate cortex; Precun = precuneus; AngG = angular gyrus (temporoparietal junction). Mirror network: pSTS = posterior superior temporal sulcus; IPS = inferior parietal sulcus; PreMC = premotor cortex.

Amygdala networks
- Perception network
- Affiliation network
- Aversion network

Non-amygdala networks
- Mentalizing network
- Mirror network

8. Porcelli et al.

Neuroscience and Biobehavioral Reviews xxx (xxxx) xxx–xxx

rating within the constituents of the face perception network (amygdala included) and face emotional process have been repeatedly found (Cohen Kadosh et al., 2011; Marsh, 2016; O’Neil et al., 2014; Wang et al., 2016; Zhu et al., 2011). Intuitively, impairments in emotion recognition could determine misinterpretation of social signals during interpersonal interactions (e.g. (Domes et al., 2009)), with deleterious consequences on social relationships, as we will discuss using as examples AD, SCZ, and MDD. However, individuals with damages within the amygdala show also an impaired ability to guide their visual attention to the region of the eyes (i.e. one of the most expressive part of the face), suggesting that the impairment in facial processing observed in these patients may be due to an inability to direct attentional resources to relevant social information rather than to a direct damage of the face perception network itself (Adolphs et al., 2014a).
2.2. Social affiliation and social aversion networks

After and in reaction to detecting social stimuli, individuals may act either in pro-social or aversive ways. Pro-social behaviour refers to processes that are initiated as a result of compassion or empathy (Lieberman, 2007), whereas social aversive behaviours refer to processes that are a result of disgust or avoiding untrustworthy strangers (Bickart et al., 2014b; Cosmides and Tooby, 1992). Bickart et al. proposed the amygdala as an integrating hub also for the networks involved in both processes (Fig. 1).

The role of the social affiliation network is to form and maintain social bonds. It comprises, amongst others, the ventromedial PFC (vmPFC), ACC, and medial temporal cortices (Aron et al., 2005; Bickart et al., 2012; Moll et al., 2006). The amygdala is reciprocally connected to the vmPFC. Increased functional coupling between the amygdala and the vmPFC is related to an increased ability of emotion regulation, suggesting a regulatory effect of the vmPFC over the amygdala (Hariri et al., 2000; Morawetz et al., 2016; Ochsner et al., 2002; Phelps et al., 2004). Emotion regulation is fundamental in order to begin and maintain successful social interactions, as demonstrated by studies on individuals with lesions within vmPFC, who report a wide variety of impairments in social behaviours including violent outbursts, lack of empathy, lack of guilt, lack of remorse, apathy, indifference, disadvantageous (social) decision making (Anderson et al., 2005; Barrash et al., 2000; Damasio, 1996; Krajbich et al., 2009; Shamay-Tsoory et al., 2003).

When making pro-social decisions (e.g., when deciding to donate money) the ventral tegmental area (VTA) and striatal areas are activated (Inagaki et al., 2016; Moll et al., 2006). Also, pictures of loved ones elicit activation in the VTA and the caudate nucleus, areas overlapping with the mesolimbic reward circuitry, as confirmed by animal studies (O’Connell and Hofmann, 2011). It has been suggested that the reward network is activated to focus on a specific individual, such as when developing a romantic relationship (Aron et al., 2005). As a matter of fact, the modulation of the reward system by previous social experiences (e.g., copulation or co-habitation) seems to be deeply involved in social attachment process (for detail see (Coria-Avila et al., 2014)). Of note, these areas are key areas of the mesolimbic reward-related circuitry and overlap largely with the default mode network, supporting that many of these structures play a number of roles also in other mental processes (Andrews-Hanna et al., 2010; Yeo et al., 2011). Nonetheless, perturbations in the social affiliation network (Fig. 1) lead to emotional detachment, diminished responsiveness to feelings and warmth, which in turn often result in progressive social withdrawal, as can be seen in frontotemporal dementia patients (Solmberger et al., 2009). Thus, this network seems to play a relevant role in maintaining social interactions. In the following section we will describe the neurotransmitters involved in these processes.

The role of the aversion network (Fig. 1) on the other hand is to protect the subject from potentially harmful (social) interactions. The social aversion network, with the amygdala as a central hub, comprises the caudal ACC and the insula, as well as their connectional targets in the ventrolateral striatum, hypothalamus and brainstem. Feelings of social aversion such as disgust or anger activate areas in this network (Buckholtz et al., 2008; Moll et al., 2005). Levels of these feelings are variable, depending, as we will see in the following section, from many factors. As expected, also this network shares overlapping areas with the salience network (Seeley et al., 2007). Consistently, regions of the salience network are also involved in Pavlovian habit learning and evoke avoidance behaviour in response to somatosensory and social pain (Akitsuki and Decety, 2009; Balleine and O’Doherty, 2010). Lesions in areas of the social aversive circuitry of the amygdala, lead to impaired judgment of strangers (Kosik and Tranel, 2011), or to flirtatiousness and inappropriate familiarity (Adolphs et al., 1995). On the contrary, hyperactivity of the amygdala has been associated with increased social avoidance (Kaldeuwaij et al., 2017; Mikics et al., 2008). Intuitively, all these behaviours likely results in interpersonal difficulties and unsuccessful social interactions, which may progressive lead individuals to social withdrawal. Studies on animal models suggested that the Hypothalamic-Pituitary-Adrenal (HPA) axis acts as a final executor of the social aversion network, since manipulation of the stress response results in social avoidance behaviour (i.e., in social withdrawal) (Ilin and Richter-Levin, 2009; Ruedi-Bettschen et al., 2006; Seigle et al., 2015; Wilson and Koenig, 2014; Wu et al., 2013) and HPA axis blocking prevents the appearance of this behaviour (Lehmann et al., 2013; Wu et al., 2013). Finally, HPA axis activity has been associated with low levels of social approach behaviour both in animals (File and Seth, 2003) and children (Lopez et al., 2004). In this context, it is noteworthy that an impaired hippocampal function may result in a dysregulation of the HPA axis. In fact, in preclinical models, it has been shown that hippocampal lesions can induce social withdrawal (Wilson and Koenig, 2014).

However, as mentioned above, pro-social and aversive processes should not be seen as completely separated and they rely on partially overlapping neural circuitry (Lebow and Chen, 2016; Telzer, 2016) with the bed nucleus of the stria terminalis (BNST) potentially serving as integrating center for limbic network outputs and sensory information. Results from animal work, and to a lesser extent from human data, show that the BNST is involved in sustained fear or anxiety and it is active during imagery of a future threat on the one hand, but also with positive valence preference and motivation for sexual behaviour on the other hand (Dickerson, 2015; Lebow and Chen, 2016). Thus, Lebow et al. proposed that the BNST modulates the salience of information from the environment contexts, which is fundamental for successful social interactions because of the complexity of social situations. However, the BNST contains several sub-nuclei, each one with putative specific functions, and a deeper understanding of their roles is needed for elucidating its role in pathology and designing new therapeutic interventions (Lebow and Chen, 2016). While human research on BNST is still in its infancy (e.g., (Buff et al., 2017)), neuropsychiatric research in animal models started to elucidate the complex role of this brain area (Asok et al., 2018; Duque-Wückens et al., 2018; King et al., 2017; Lebow and Chen, 2016; Newman, 1999). Similarly, a greater complexity has been shown by animal studies also in other brain areas and related neural networks (e.g., (Challis and Berton, 2015; Chaudhury et al., 2013)). Unfortunately, to date, current neuroimaging techniques do not allow a similar degree of detail in human studies. However, with the advance of fMRI techniques (e.g., 7 T fMRI), in the coming years, the neural microcircuits identified by animal studies may be confirmed also in humans.

2.3. Mirroring and mentalizing: building blocks of sociocognitive functioning

Upon the initial processing of stimuli, perception, aversion and affiliation processes undergo further integration. The ability to navigate through our complex social surroundings is in fact largely dependent on further processes that allow the sophisticated interpreting of our own, as well as others’, intentions, emotions, actions, and beliefs (Barrett and Satpute, 2013; Bickart et al., 2014b; Frith and Frith, 2006; Rizzolatti and Sinigaglia, 2016). It is widely believed that these complex socio-cognitive processes are largely accommodated by two interrelated, yet distinct, neurocognitive network assemblies, commonly referred to as the mirroring and mentalizing networks (Barrett and Satpute, 2013; Bickart et al., 2014b; Frith and Frith, 2006; Rizzolatl and Sinigaglia, 2016) (Fig. 1).

The mirroring network comprises a selection of temporal, parietal, and sensory motor brain regions, which employ data on perceived motoric and biological movement (e.g., facial expressions and bodily
gestures) for simulating and interpreting others’ overt actions (Barrett and Satpute, 2013; Rizzolatti and Sinigaglia, 2016; Spunt et al., 2010, 2011; Zaki et al., 2010), as well as their basic emotions (Rizzolatti and Sinigaglia, 2016). Overall, this system allows basic understanding of others’ actions and emotions, by mainly drawing on one’s own sensory, motoric, and visceral representations of what is perceived (Rizzolatti and Sinigaglia, 2016). However, basic understanding of others’ actions and emotions is not sufficient for higher-order inferences on causes and consequences of others’ behavioural repertoires, and this is where the mentalizing network comes into play (Barrett and Satpute, 2013; Frith and Frith, 2006; Rizzolatti and Sinigaglia, 2016).

The mentalizing network (Fig. 1) comprises a more wildly distributed collection of frontoparietal territories, which draw on past experiences and social knowledge for highly enriched and multimodal representation of sociocognitive information (both internally- and externally-oriented) (Barrett and Satpute, 2013; Rizzolatti and Sinigaglia, 2016; Spunt et al., 2010, 2011; Zaki et al., 2010). The original core of this network included the posterior STS, the temporo-parietal junction (TPJ), the anterior temporal poles, the mPFC (Frith and Frith, 2006), posterior cingulate/prefrontal, and inferior frontal gyrus (Schurz et al., 2014). The mentalizing network largely overlaps with the default mode network (DMN) (in particular, some authors identified an overlap between the mentalizing system and three subnetworks of the DMN, although other authors highlighted how only one DNM subnetwork and some DNM hubs overlap with the mentalizing system (Hyatt et al., 2015; Li et al., 2014; Buckner et al., 2008)), to the point that some authors speculated that humans may be predisposed to engage the mentalizing system when not focusing on non-social tasks (Liberman, 2013). However, the DMN becomes more active during rest and after a non-social task is completed (Buckner et al., 2008), suggesting a partial distinction from the mentalizing network (Green et al., 2015; et al., 2015; Hyatt et al., 2015). The mentalizing network is often decomposed into dorsal and ventral subnetworks (Abu-Akel and Shamay-Tsoory, 2011; Kalbe et al., 2016; Lavoie et al., 2016; Poletti et al., 2012; Schlafke et al., 2015; Shayan-Tsoory and Aharon-Peretz, 2007) that perform slightly different functions (Andrews-Hanna et al., 2010; Barrett and Satpute, 2013). The dorsal subnetwork (dorsal ACC, mPFC, precuneus, and temporo-parietal poles) seems more engaged when abstract third-person (exogenous) information is necessary for making sociocognitive inferences (Andrews-Hanna et al., 2010; Barrett and Satpute, 2013). The ventral subnetwork (ventral ACC, mPFC, medial temporal lobe territories) seems more engaged when embodied first-person (endogenous) information is required for sociocognitive inferences (Andrews-Hanna et al., 2010; Barrett and Satpute, 2013). However, this distinction is still preliminary, as more recent data suggest a more complex networks’ structure sustaining mentalizing, with at least three subnetworks sustaining different aspects of mentalizing activity (for detail see (Schurz et al., 2014)). Thus, despite the fact that the core regions of the mentalizing network have been repeatedly identified, the exact functions of its different subnetworks are still to be elucidated in detail, as well as the exact interactions with other brain networks, such as the DNM.

It is also good to mention that mirroring and mentalizing operations often occur automatically with very little effort or explicit deliberation (Frith and Frith, 2006). Mounting evidence suggests that these two networks constantly communicate and interact with each other during sociocognitive processing (Barrett and Satpute, 2013), though the specific nature of these interactions is still under debate (e.g., (Van Overwalle and Vandekerckhove, 2013)). One line of research suggests that these networks seem to act in parallel during the perception of persons, where they either cooperate or compete with each other depending on contextual factors (Barrett and Satpute, 2013; Zaki et al., 2010). Others theorize that these networks are hierarchically related: the process of constructing complex mental state attributions (i.e., mentalizing) is likely preceded by sensorimotor and visceromotor representations of perceived overt behaviours (i.e. mirroring) (Barrett and Satpute, 2013; Spunt et al., 2010, 2011). Taken as a whole, it seems that when minimal sensory input is required, and thus more internally-driven representations are constructed (e.g., introversion and mind-wondering), the mentalizing network might have the overhand, while during externally-driven representations (e.g., object or person perception) the mentalizing and mirroring networks jointly engage to make sense of the social surrounding (Barrett and Bar, 2009; Barrett and Satpute, 2013).

In sum, the mirroring and mentalizing networks allow interpreting our own, as well as others’, intentions, emotions, and actions, and in doing so they enable the uniquely human ability of communicative intent (Frith and Frith, 2006). We are, however, at the early stages of grasping how our brain allows for these complex processes, and perhaps more importantly, how they might go awry and should be normalized in certain psychiatric conditions. Of note, most information on brain regions and circuitry involved in social processes is derived from neuroimaging findings. Thus, the technical limitations (e.g., the spatial resolution in fMRI studies) of this kind of studies do not allow a more detailed analysis of the microcircuits responsible for driving specific behaviours, as previously stated. Animal studies may provide complementary information about the fine organization at the microcircuits level (e.g., (Lebow and Chen, 2016)), hopefully leading to a more complete knowledge of the complex interactions which sustain social behaviours. The PRISM project ((Kas et al., 2017) and Bilderbeck et al. in this issue) aims to increase the current knowledge investigating social behaviours from both clinical and preclinical perspectives in three frequent, severe neuropsychiatric disorders, i.e. SCZ, AD, and MDD. We will describe in a following section how these three disorders, characterized by marked social dysfunction, show perturbations in the abovementioned processes and their putative network assembly (Kennedy and Adolphs, 2012).

3. The social brain: neurotransmitters and social behaviours

In the previous section we revised and integrated the current knowledge about the functional neuroanatomical substrates of social functioning. In this section, the involved neurotransmitters will be detailed with some real-life examples integrating neuroanatomical regions for a physiological understanding of subject’s social behaviour.

The complexity of the processes described above is also reflected by their complexity at the neurotransmitter level. As mentioned above, the selective pressure gradually led to a progressive development of a complex network of neural interactions, which sustains the processing of this class of stimuli. Most of the neurotransmitter systems involved (e.g. dopamine (DA), opioid (OP), and GABA systems (GABA)) probably adapted parts from their general functions to the processing of social stimuli, while only few systems (particularly OXT and vasopressin (AVP), and in a lesser extent also serotonin (5-HT)) show a marked specialization for the processing of these stimuli. These "specialized" neurotransmitter systems seem to orchestrate the neural response to social stimuli, making a secondary use of the other "non-specialized" systems (for a general overview of the neurotransmitter systems discussed see Supplementary material S.1). The amount of evidence achieved so far allows drafting an initial picture of these complex interactions in different aspects of social stimuli processing. Although these interactions are clearly related to the neural activity within the networks described in the previous section, our current knowledge does not enable the precise coupling of neurotransmitters inter-play with activation of specific neural networks. Therefore, in this section we focus on more basic social processes, providing an overview of the involved neurotransmitters inter-play with activation of specific neural networks. Therefore, in this section we will focus on more basic social processes, providing an overview of the involved neurotransmitters inter-play. Where possible, links with the neuroanatomical level are provided. Of note, a large amount of the evidence discussed comes from animal studies. We will specify when findings were confirmed also by human studies. In a following section we will discuss how these processes may be altered by neuropsychiatric disorders, such as AD, SCZ, and MDD, resulting in social dysfunction.
and, eventually, in social withdrawal.

3.1. Social perception

In the previous section, we described the social perception network (Fig. 1) (Bickart et al., 2014b). The functioning of this network requires a complex inter-play of neurotransmitters. Salience is a key attentional mechanism associated with the ability to reorient to (or filter out) salient stimuli (including social ones). This effect has been related to the enhancing of neural responses in the VTA, posterior STS and pre-motor cortex, brain regions related to reward (e.g. nucleus accumbens - NAc, striatum, and OFC), and connectivity among amygdala, insula and caudate (Ma et al., 2016) (i.e. to a modulation of the social perception network and the reward network (Bickart et al., 2014b)). The detection of salient stimuli is centrally regulated by the DA system, which increases phasic activity after salient stimuli detections, promoting attention reorienting and alerting to potentially important sensory cues (Shamay-Tsoory and Abu-Akel, 2016). These alerting signals are sent to salience-coding VTA DA neurons to mesolimbic structures (including central amygdala, BNST, and NAc) to assess their value and valence. The sensitivity of DA neurons to these stimuli depends on basal levels of tonic DA transmission, which in turn are determined by homeostatic biological functions, as well as individual characteristics (Shamay-Tsoory and Abu-Akel, 2016). The OXT system seems to modulate phasic DA activity in response to social stimuli (Groppe et al., 2013), facilitating the salience of this class of stimuli (e.g., (Guastella et al., 2008)), irrespective of their valence, by regulating DA’s salience coding and attention reorienting signals. Interestingly, an up-regulation of the phasic firing of DA neurons within this pathway has been associated with social dysfunction and related to social withdrawal, although only in animals (Campi et al., 2014; Chaudhury et al., 2013). OXT and DA interactions within central amygdala and NAc are thought to have a primary role in the salience stimuli determination (Shamay-Tsoory and Abu-Akel, 2016), while their effects on functional coupling between pre- and post-amygdala and superior colliculi are thought to modulate attention reorienting (likely involving also PFC modulation, as suggested by both animal and humans studies (Rosenfeld et al., 2011)). In turn, preclinical and clinical data suggest that OXT production and release are modulated by the 5-HT system via interaction with 5HT1a, 5HT2b/2c and AVPR1a receptors (Bershad et al., 2016; Kamilari-Britt and Bedi, 2015). This modulation likely reflects the integration of mnemonic information and affective state on the perception of social stimuli (see section 3.3 and 3.4 below) (Svob Strac et al., 2016). This is a first example of how a ‘general’ function such as salience stimulus determinations, shows a sort of “specialization” for social stimuli processing, likely attributable to OXT and 5-HT systems interplay.

Therefore, in a social context such as a friends’ meeting, our perception is strongly enhanced and directed by specific social stimuli (e.g. eye gaze, face expressions, etc.) in order to facilitate their processing. In particular, the 5-HT system modulates OXT system activation, which in turn regulates the DA system. The DA system subsequently promotes the attention through and the salience of social stimuli from the environment. This neurotransmitter interplay results in an activation of the social perception network. Thus, we will pay more attention to social stimuli in general, and particularly to the ones coming from significant others, such as friends, relatives, and partners.

3.2. Social reward and social pain

To understand how our brain evolved to function in a complex social environment, an important question is: why do we perceive social stimuli as rewarding or punishing? Some authors hypothesized that the perception of social stimuli as rewarding is fundamental to develop the other aspects of social brain (Gunaydin and Deisseroth, 2014). Indeed, we form and maintain bonds with conspecifics because of the reward deriving from them, starting from the first interactions (i.e. parental attachment and juvenile social play) to the typical adulthood social interactions (i.e. mating behaviours and aggressive/cooperative behaviours aiming to determine social hierarchies). The feeling of distress caused by social isolation/rejection (i.e. social pain) represents the other side of the coin, which also promotes socialization bond maintenance to avoid these consequences in the future. Clearly, the perception of social stimuli as rewarding and the avoidance of distress due to social separation/rejection have strong evolutionary roots, because of the great advantages deriving from group living compared to solitary living, despite its intrinsically-related costs (e.g., ecological competition and reproductive suppression) (Dunbar and Shultz, 2007). Intuitively, impairments in the processing of social reward and social pain could cause repeated unsuccessful social interactions (e.g., decreasing motivation through social interactions), possibly driving to social withdrawal. In Section 5 we will detail this issue, using as examples MDD, AD, and SCZ.

Several neurotransmitter systems are thought to be involved in the encoding of social stimuli as rewarding or punishing. Intuitively, these processes are likely to be fundamental for the functioning of the social affiliation network and the social aversion network described in the previous section (Fig. 1). Consistently, the areas we will focus on in this section are included in these two networks. As in other forms of rewards, DA plays a primary role. Particularly, DA VTA increases its activity in response to social stimuli, and the degree of downstream DA release is associated with the duration of social interaction (Gunaydin et al., 2014; Scott-Van Zeeland et al., 2010). The main target of VTA DA projection is the NAc (involved both in the social affiliation and social aversion networks, although a spatial differentiation within the ventral striatum has been suggested (Bickart et al., 2014b)), which is thought to encode reward-related signals from the VTA, via an activation of NAc D1-expressing GABA medium spiny neurons (MSNs) (Lobo et al., 2010; Yager et al., 2015). Interestingly, OXT was found to enhance VTA activation in response to cues announcing both social reward and pain stimuli (i.e. anticipation of reward or pain stimuli) (Groppe et al., 2013), but not to non-social stimuli (Dolen et al., 2013). This is another example of how a “specialized” system makes use of a more general one to finely modulate social behaviour. In fact, OXT activation leads to a reinforcement of DA-mediated signals (e.g. strengthening the signal-to-noise ratio in the principal cell circuits through GABA system stimulation (Baribeau and Anagnostou, 2015)) in response to social stimuli only. Moreover, a further step in the interpretation of the complex system interplay, an interaction with 5-HT system is also required for decoding the social stimuli as rewarding or punishing. In detail, the activation of NAc 5-HT1b receptors, which in turn induce long term depression (LTD) in MSNs neurons, is needed to encode social reward (Dolen et al., 2013). But this preliminary picture needs a further step, which is fundamental in processing both social reward and pain stimuli. This is performed by the OP system which contributes together with DA system in mediating the final hedonics aspects of social reward (Blass and Fitzgerald, 1988), as well as the feelings related to social pain (Johnson and Dunbar, 2016), eliciting distress upon separation (through low opioid receptor activity, mainly into NAc shell (Koob and Volkow, 2016)) and comfort upon reunion (through high opioid receptor activity, mainly into NAc shell (Eisenberger, 2012; Koob and Volkow, 2016)). These effects are thought to be related to the direct modulation of the OP system on DA. Indeed, after a reward stimulus, μ-opioid receptor (MOR) activation in NAc directly correlates with DA release (Job et al., 2007). A putative mechanism of action for this modulation is suggested by the expression of OP receptors by MSNs neurons (Yager et al., 2015), which modulates the DA release within NAc, as previously stated. Unfortunately, most data about these fine interplays comes from animal studies, although some consistent evidences have been reported also in humans (Eisenberger, 2012; Koob and Volkow, 2016). The problem is primarily technical with invasive neurochemical techniques (voltammetry, amperometry, microdialysis) with good time resolution unable to be performed in humans while existing non-invasive human-appropriate neurochemical techniques.
(e.g. 1H-MRS) have poor time resolution to link to social behavioural events. At a neural level, these inter-plays might explain how ventral striatum is involved both in the affiliation and aversion networks in humans (Fig. 1) (Bickart et al., 2014b). On the contrary, other brain areas are involved in social pain processing only, such as dACC and anterior insula, as already demonstrated in humans (Eisenberger, 2012). Furthermore, social reward seems to activate OP system within amygdala, left ventral striatum, and anterior insula, while it was de-activated in midline thalamus and sgACC (Hsu et al., 2013; Nummenmaa et al., 2016). In turn, closing the circle, the OP system is modulated by OXT, AVP and other systems, such as the endocannabinoid one (Johnson and Dunbar, 2016), which determine also long-term modulation of the system (in terms of receptors availability in different brain areas), particularly during neurodevelopment in hu-

mans (Nummenmaa et al., 2015). Therefore, at the neurotransmitter level, social pain and social reward processing show a high degree of overlap, since they involve the same actors, although with a different spatial-time activation. However, some neurotransmitter pathways are likely to be involved only in one of these processes, suggesting some degree of specialization also at this level. Unfortunately, current knowledge does not allow to exactly disentangle these specialized neurotransmitter pathways, particularly in humans. Consequently, a large number of findings here discussed comes from animal studies and has still to be replicated in humans, since current human methodologies do not allow the high spatial-time resolution required to elucidate these interactions.

This brief overview allows us to hypothesize how this social reward system acts in a real-life situation. For example, when we meet friends who smile at our arrival (i.e. social reward stimuli), the OXT system is activated, enhancing VTA DA signals through GABA system stimulation. Then, VTA DA neurons generate an OXT-reinforced signal to NAc neurons, which in turn are modulated by 5-HT1b and MOR (mu receptor) heteroreceptors. These modulations determine the amount and duration of DA downstream release, which encodes the subjective feeling of pleasure due to the social stimulus (e.g. friends smile) and increase the desire to maintain the interaction with the source of reward (e.g. the desire to spend more time with our friends). At the neural level, this modulation determines an activation of the social affiliation network. On the other hand, when we separate from our friends, for example to move to another country for a long time, we feel the pain due to the separation. These feelings seem to be mediated by a deac-
tivation of MOR (mu receptor) and probably also of 5-HT1b hetero-
receptors, which result in a decreased NAc DA downstream release, re-
sponsible for the subjective feeling of distress (Eisenberger, 2012). Clearly, these are the final steps of the neurotransmitter signals which processing social stimuli as rewarding or punishing, but they are of fundamental im-
portance because, when altered for example by a psychiatric disease, they might impair the resultant social behaviours.

Obviously, this is only a part of the whole interplay. Social reward and pain stimuli, as other rewarding/punishing stimuli, determine re-
forcement processes through the sources of these stimuli and the mechanics of the anticipation feelings. Again, if these processes are impaired, the motivation through social interactions will most likely decrease, eventually resulting in progressive social withdrawal, as fur-
ther discussed in Section 5. The dorsal raphe nucleus 5-HT projections to NAc seem to play a primary role in processing reinforcement pro-
cesses, such as conditioned learning. It has been shown that 5-HT axons from dorsal raphe nucleus expressed presynaptic OXT and 5-HT1b re-
ceptors at their terminals, which, when stimulated together, lead to a LTD of excitatory synapses onto MSNs in the NAc. These long-term modulations are thought to be involved in reinforcement processes, as suggested by preclinical data (Dolen et al., 2013). Furthermore, also the projection from VTA DA neurons to mPFC and amygdala are involved in these processes, as well as the regulatory feedback provided by the glutamatergic projection from mPFC to NAc. Interestingly, these paths-
ways are involved in the social affiliation network (Fig. 1), the atrophy of which has been associated with higher levels of socio-emotional disengagement and with smaller social network size (Bickart et al., 2014a,b). Other intermediate brain structures such as the medial pre-
optic area (mPOC) (Coria-Avilà et al., 2014) and BNST (Lebow and Chen, 2016) are involved in these processes as well, as observed both in animals and humans. The OXT system seems to increase neural re-
sponses in several reward-related brain areas other than NAc in humans (i.e. insula, precuneus, pgACC, OFC, ventral pallidum, and midbrain (Ma et al., 2016)), overall enhancing the responses to social stimuli. Furthermore, in humans, also the 5-HT system modulates neural ac-
tivity in response to environmental stimuli in limbic and cortical cir-
cuits, including insula, OFC, amygdala, putamen, ventral striatum, hippocampus, VTA and dmPFC (Macoveau, 2014). These effects are thought to be regulated by a direct pathway between 5-HT dorsal raphe neurons and vmPFC (which is included in the affiliation network, Fig. 1), which provides the adaptive cortical control on brainstem cir-
cuits regulating socioemotional decisions and actions, as suggested by both preclinical and clinical data (Challis and Berton, 2015). Cons-
istently, on the basis of preclinical findings, it has been hypothesized that 5-HT dorsal raphe nucleus acts as a hub for current context eval-
uation, encoding how beneficial the current environmental context is for the subject (Luo et al., 2016). The relevance of this pathway on social behaviour has been demonstrated in animals, since a decrease in tonic activity of 5-HT neurons in the dorsal raphe nucleus has been associated with social avoidance behaviour in socially defeated animals (Challis et al., 2013). Again, the inhibitory effects of the 5-HT system on these brain areas are likely to be mediated by the GABA system (Challis et al., 2013; Challis and Berton, 2015). On the other hand, there are other complex interactions between OXT and 5-HT systems, which re-

sult in the necessary fine tuning and feedback of their transmissions. As an example, both in animals and humans, 5-HT activation seems to enhance OXT transmission via activation of 5-HT1A and 5-HT2B/2C receptors (Kamilar-Britt and Bedi, 2015), while OXT activity within raphe nuclei facilitates 5-HT release in this area. Nonetheless, the exact interactions and the hierarchy between 5-HT and OXT systems are still to be elucidated in detail, in particular in humans. However, in animals, OXT was also found to modulate cortical inhibition (Marlin et al., 2015), and these systems likely interact in determining the excitation/inhibition balance in vmPFC, which in turn was associated with value-guided choice process in humans (Jocham et al., 2012) and with social behaviour itself in both animals (Yizhar et al., 2011) and humans (Bickart et al., 2014b; Bicks et al., 2015). Finally, the OP system within the left ventral striatum also plays a fundamental role in motivation through social reward, as suggested by both preclinical and clinical data (Chelnokova et al., 2014; Hsu et al., 2013; Løseth et al., 2014). As previously stated, activation of OP mu receptor is likely needed to in-
duce DA release in this area and, similarly, during anticipation of re-
ward (Koob and Volkow, 2016). From a neural perspective, it is pos-
sible to distinguish two striatum-related networks, a) one sustaining motivation through a future reward (ventral striatum DA circuit, which likely overlaps with the affiliation and aversion networks, Fig. 1 (Bickart et al., 2014b)) and b) one monitoring the outcome of actions to optimize future choices to achieve reward (dorsal striatum DA circuit) (Skuse and Gallagher, 2009), i.e. determining associative learning, likely involving specific nuclei of amygdala, BNST, and some specific areas of VTA (although their involvement has been demonstrated only in primates) (Fudge et al., 2017). In this second network, AVP may play a primary role because of its effect on NAc shell, lateral septal nucleus and other areas of dorsal striatum (Skuse and Gallagher, 2009). Beyond their role in reward processing, motivation and outcome monitoring, these two networks are also responsible for the complex cognitive perception of trust. This complex perception is known to be enhanced by OXT administration, likely through amygdala deactivation and re-
duced amygdala-brainstem regions coupling (Skuse and Gallagher, 2009). Finally, the same neural circuits are also involved in reciprocal altruism, which is likely the result of innate tendency and previous experiences (Skuse and Gallagher, 2009).
3.3. Social learning and bond formation and maintenance

Dynamic situational interplay and anticipation are important, but maintenance is also needed. For example, as we will discuss in Section 5.2, in neurodegenerative disorders (such as AD) the observed social withdrawal (Jost and Grossberg, 1996) is likely to result from a progressive reduction of already established social interactions. Therefore, these effects are important for subsequent social learning and stable bond formation and maintenance. Indeed, repeated reward from a specific social interaction (e.g. repeating meetings with a friend) results in the formation of a stronger social bond, sustained by long-term modifications within these systems. In particular, it is interesting to see how the immediate and delayed phases are linked to different structures. Some authors proposed that, within the rostral shell of NAc, D2 receptors mediate the initial social reward, while D1 receptor activation seems to facilitate the formation of specific bonds (e.g. with a monogamous partner), as suggested by animal studies (Coria-Avila et al., 2014). It has been hypothesized that the initial reward is D2 mediated, while repeated reward from the same source (e.g. a sexual partner) gradually increases the presynaptic D1 receptor expression, facilitating the reward from the same interaction and preventing reward from other interactions, thus facilitating the formation of social bonds (e.g. with monogamous partner or with our infant) (Coria-Avila et al., 2014; Skuse and Gallagher, 2009). However, the exact mechanisms which sustain this long-term modification remain to be fully elucidated, in particular in humans. Some authors suggested that the interaction with OXT and AVP systems are fundamental for social bond formation. Specifically, for bond formation, their effects on salience of social stimuli (which facilitates social recognition) and on reinforcement properties of the DA reward system are thought to be fundamental. At a neural level, the sensory cortices (mainly olfactory, auditory and tactile) project to the medial amygdala and lateral septum, which are critical for social recognition. The medial amygdala and its strictly connected BNST project AVP fibers to the ventral pallidum and lateral septum, whereas OXT fibers in the NAc most likely originate from neurons in the preoptic area (POC) or hypothalamus. As previously mentioned, these areas have been linked with associative learning in primates (Fudge et al., 2017), again suggesting how some systems adapted from their general functions to the processing of social stimuli. Activation of these areas during mating/social interaction may result in local release of these peptides. The ultimate result is the concurrent activation of D2 receptors in the NAc of both sexes, OXT receptors in PFC and NAc of females and AVPV1a receptor in the ventral pallidum of males. As a result, the reinforcing, hedonic properties of social interaction may become coupled with the sensory signatures (e.g. odors and voice) of the partner, resulting in a conditioned partner preference (Young and Wang, 2004). Consistently, these interactions among DA, OXT and AVP systems within the reward-system were found only in monogamous species. However, they were investigated mainly in animal studies (Insel, 2010) and further studies are needed to confirm and better elucidate these mechanisms in humans.

In a real-life situation, the repeated interactions with a conspecific (e.g. friend or partner) will induce specific long term modifications in the DA reward system (D1 receptor) and in its modulator systems (i.e. OXT, 5-HT and OP systems). These long-term modifications facilitate the reward from this specific interaction, which increases motivation and anticipation. Also, associative memories, which link features/sensory signatures of our friend/partner (e.g., odors) to reward expectations, will be formed and recalled during reward anticipation. Thus, when we are waiting to meet our friend, we feel the desire to interact with him and we will experience the excitement due to the expected reward.

3.4. Modulation of affective status

Motivation through social stimuli also depends from our current affective status. This is the first example how an external perturbation may modulate social withdrawal. Indeed, when we feel upset, the desire of social interaction usually decreases (diminished pleasure in activities, including social ones, is one of the criteria for the diagnosis of major depression), while on the contrary, the presence of friends/relatives may help to improve our mood and to cope with distressful events (social therapy is one of the major depression treatments). Consistently, the systems discussed above are also reciprocally involved in the regulation of affects, as suggested by human studies. Particularly, at a neural level, OXT has inhibitory effects on amygdala and amygdala-brainstem functional connectivity (Domes et al., 2007), the ACC, anterior insula, midbrain, OFC and thalamus (Ma et al., 2016) (i.e. OXT inhibits the aversion network, Fig. 1 (Bickart et al., 2014b)). Further, OXT activation reduced neuronal activity in the hypothalamus, hippocampus, and ventrolateral septum, resulting in a reduction of ACTH release and cortisol plasma levels (i.e. to an inhibition of the HPA axis, which acts as a final executor of the social aversion network according to convergent evidence from both preclinical and clinical studies, as previously stated (File and Seth, 2003; Lehmann et al., 2013; Lopez et al., 2004; Wu et al., 2013)). Thus, an activation of OXT system leads to a decrease in anxiety and arousal, promoting a calm status. These effects are also thought to mediate the social buffering effect in humans (Alvare et al., 2010), i.e. the positive effect of social support to cope with distressful events. More in detail, when an aversive/stressful social stimulus is detected, the basolateral amygdala rapidly activates the extensive array of cortico releasing hormone (CRH) neurons located in the central amygdala. In turn, these CRH neurons project to several brain regions involved in emotion modulation, memory processes, and arousal, including BNST and peripheral CRH neurons in the PVN. The activation of the CRH system initiates the series of events that ends in the activation of locus coeruleus (LC) norepinephrine (NE) neurons, which is responsible for producing nonspecific arousal (Aston-Jones et al., 1996), and in the release of cortisol from the adrenal cortex, both in animals and humans (Moore and Depue, 2016; Reppermund et al., 2007; Sandstrom et al., 2011). All these processes result in feeling of anxiety and in hyper-arousal status (for detail see (Aston-Jones et al., 1996; Moore and Depue, 2016; Reppermund et al., 2007; Sandstrom et al., 2011)). OXT seems to mediate the social buffering effect, promoting calmness (i.e. inhibiting defensive behaviours (Tops et al., 2014)) (Baribeau and Anagnostou, 2015), likely through interactions with the CRH system in the paraventricular nucleus (PVN) and BNST (CRH-neurons expressed OXT receptors in these areas, although data in humans are few) (Dabrowska et al., 2011; Heinrichs et al., 2003), which result in an inhibition of the HPA axis (Dabrowska et al., 2011; Karelinia et al., 2011). Consistently, at the neural level, the social buffering effect is thought to be mediated by an activation of PFC (i.e. of the mentalizing and affiliation networks, Fig. 1) and a deactivation of amygdala,
dACC (i.e. of the aversion network, Fig. 1) and frontal cortex, likely through a modulation of GABA system by OXT. The effectiveness of this modulation depends by rearing conditions, which in turn modulate the OXT system development (Hostinar et al., 2014). Obviously, further studies are needed to disentangle more in detail how the OXT system modulates the affective status and mediates the social buffering effect, in particular in humans.

In a real life situation, the presence of friends/relatives, will help us to cope with stressful events, e.g. to face a failure in university examination. On the other hand, after we receive an averse social stimulus (e.g. a failed exam), our averse system will be activated, resulting in anxiety and hyper-arousal feelings with decreased sociality. However, through the support offered by friends/relatives, the magnitude of these feelings will be reduced by the inhibitory effects of OXT system on CRH. Consequently, the activation of the averse system will be moderated and normalized more rapidly, preventing the deleterious effects of a stronger and prolonged HPA axis activation.

3.5. The key role of oxytocin

The OXT system (Supplementary material S.1) has a central role in social behaviours (e.g., (Kanat et al., 2014; Meyer-Lindenberg et al., 2011; Wang et al., 2017)); therefore it seems probable that it is tightly modulated in a fine way with multiple feed-back loops. OXT has excitatory effects on brain areas involved in emotion regulation, such as the mPFC, vIPFC, dIPFC. These effects are in way opposite to the effects of the OXT system discussed above, and are mediated by other neurotransmitter systems, mainly the GABA system (MacDonald et al., 2011; Stan et al., 2014). Generally, OXT increases the GABA interneuron functioning, strengthening the signal-to-noise ratio in the principal cell circuits, although it has been demonstrated only in animals so far (Barbeau and Anagnostou, 2015). The stimulating effect of OXT on GABA system has been suggested also in the dIPFC, where increased GABAergic transmission determines the removal of inhibitory brakes which normally act to suppress the expression of response tendencies that are characteristic of earlier developmental stages, as observed in humans (Mitchell et al., 2015). Moreover, OXT activity within raphe nuclei facilitates 5-HT release in this area, thus promoting an anxiolytic effect and decreasing aggressive behaviours. However, on the basis of preclinical data, it has been also hypothesized that 5-HT may have different effects on social behaviours, by amplifying the preexistent attitude of the subject (de Boer et al., 2009). In turn, 5-HT activation facilitates OXT and AVP secretion in animals (Barbeau and Anagnostou, 2015). On the other hand, social punishment determines the activation of other brain areas, such as dACC and anterior insula, and to a minor extent midline thalamus, ventral striatum, amygdala, primary somatosensory cortex, secondary somatosensory cortex, posterior insula and periaqueductual grey, and sgACC, which are thought to mediate the emotional and cognitive aspects of social pain (Eisenberger, 2012). The activation of these areas is thought to be mediated by the interactions between the OXT and OP systems described above.

In a real-life situation, when we stay with friends, in a positive and friendly environment (e.g., an informal meeting with friends), the activation of the OXT system secondary to social reward promotes both a deactivation of the aversion network (i.e. rostro-dorsal amygdala, caudal ACC, anterior insula, midbrain, OFC, and thalamus, Fig. 1 (Bickart et al., 2014b; Ma et al., 2016)) (i.e. promoting a calm and relaxed status) and a concomitant enhancement of PFC activities, responsible of positive emotional feelings (e.g. happiness). On the contrary, when we meet friends/relatives that are very disappointed by us, the aversion network is activated under OXT and 5-HT system modulation, as well as other areas involved in emotional conflict resolution (i.e. dACC). Thus, we will feel distressed and we will start to think why our friends are disappointed by us and how to resolve this conflict.

3.6. Sex differences

A growing body of evidence suggests that differences between sexes exist concerning these systems, as previously underlined. Intuitively, females and males greatly differ in some aspects of social behaviours, such as parental attachment and mating (Coria-Avila et al., 2014), while in other ones they may be quite similar (e.g., friendship). Consistently, some differences in the neurotransmitter systems discussed above have been found (although mainly in animals, e.g., (Campi et al., 2014)). In particular, in specific areas of the brain (e.g., BNST), differences in the expression of OXT and AVP receptors as well as differences in the expression of these neuropeptides themselves have been reported between the two sexes (for a review see (Dumais and Veenema, 2016)). Nonetheless, these differences are still poorly known, particularly in humans, and further specific investigations are needed to better understand their neurobiological bases.

3.7. Genetic variants

This description of the fine inter-play of neurotransmitters which sustain social functioning underlines the complexity of these processes. As previously mentioned, rearing conditions are thought to explain, at least partially, the variability in social behaviour observed in healthy subjects (Hampton et al., 2016), likely as a result of a gene x environment interaction (e.g. (Calati et al., 2014; Lange et al., 2017b; Singer et al., 2017)). In other words, genetic variants and early life events jointly modulate the resilience/vulnerability of this inter-play in the specific subject (e.g. (Challis et al., 2013; Zhang et al., 2015)). Intuitively, higher vulnerability may lead to higher social dysfunctioning when a pathophysiological process (e.g. AD, SCZ, and MDD) occurs. It is beyond of the scope of the present manuscript to review literature about the genetic findings in social neuroscience (for reviews see (Cole, 2014; Ordonana et al., 2013; Skuse and Gallagher, 2011)). However, in Supplementary Table 1 we reported the genetic variants within the systems described above that have been associated with social functioning and social cognition impairments in different clinical populations, since they might modulate the genetic vulnerability through social dysfunctioning. Further, associations among the same variants and neuropsychiatric disorders such as AD, SCZ, and MDD, were reported as well in order to underline possible convergent pathological mechanisms. However, the available evidence regarding “social genetics” are still largely conflicting and further studies in larger, well-characterized cohorts are needed (e.g., (Wang et al., 2017)).

In Section 5, we will describe in detail how AD, SCZ, and MDD may affect the processes which sustain social functioning, resulting in social dysfunction and, frequently, in social withdrawal.

4. Intermediate cognitive endophenotypes and social functioning

In the previous sections we described the fundamental neural processes underpinning social behavioural patterns and we delineated the underlying neurophysiologic circuitry. However, its complexity is also reflected at a clinical, endophenotype, level, which may independently modulate the final behavioural pattern. For example, we previously discussed how deficits in the attention endophenotype may play a role in the impaired ability of facial emotional recognition observed in patients with amygdala lesions (Adolphs et al., 1994; Vuilleumier et al., 2004). Consistently, other basic cognitive domains have been proved to modulate interpersonal behaviours (e.g., (Bowie et al., 2008; Vlaming et al., 2010)). In line with these data, in the PRISM project (Kas et al., 2017) and Bilderbeck et al. in this same issue) three basic key cognitive domains, namely sensory processing, attention, and working memory, will be investigated in relation to social functioning. One of the main reasons for this selection was that these three specific cognitive domains could be easily investigated also in animal models (e.g. (Wallace et al., 2015)), thus allowing a better dissection of their neurobiological
basis. In this section, we briefly discuss how and to what extent these non-social cognitive domains are involved in social cognition and social functioning. We are well aware that other cognitive domains (e.g., executive functioning and verbal memory (Bell et al., 2009)) may contribute to modulate social behaviours, but we only aim here to highlight how key aspects of human cognition, by modulating social functioning, may determine social dysfunctions - and eventually social withdrawal - when altered, independently from deficits in the social pathway itself. The aim is to dissect as much as possible the behavioural domains in order to facilitate a better understanding of their biological bases.

4.1. Attention, social cognition, and social functioning

Biases in the processing of social stimuli may be driven in part through skewed attentional mechanisms. While the modulation of attention by social factors and vice versa is increasingly acknowledged (Sui and Humphreys, 2016), there is relatively little research on the link between attention and social cognition. As stated above, amygdala lesions have been associated with inability to guide visual attention to the face eye regions, which may explain the associated impairment in emotional face processing (Adolphs et al., 2005; Vuilleumier et al., 2004). We have already discussed the central role of amygdala in the social perception network, as well as in affiliation and aversion networks (Fig. 1). Concerning attention, the amygdala is presumed to play a role in rapid orientation to salient stimuli (Davis and Whalen, 2001) followed by a slower process that engages both the amygdala and cortical regions including the vIPFC (Monk et al., 2006, 2008), which is followed by a slower process that engages both the amygdala and prefrontal regions showing increased activation as a function of increasing social load (Meyer and Lieberman, 2012; Meyer et al., 2012). In particular, the top-down connectivity from prefrontal to posterior regions has been linked to performance in socially demanding tasks (Hillebrandt et al., 2013). Nonetheless, despite this recent evidence, the specific mechanisms behind the possible interaction between social and working memory systems require further exploration. Given the complexity of social interactions, it is likely that individuals with working memory impairments will have more difficult time in considering multiple pieces of social information, thinking through their actions, and referring prior social knowledge (McQuade et al., 2013). Thus, it could be hypothesized that deficits in working memory could lead to a disengagement from social interactions, possibly driving to social withdrawal. Obviously, this hypothesis requires dedicated studies in order to be rigorously investigated.

4.2. Working memory, social cognition, and social functioning

Working memory is another domain which is likely to modulate social behaviour but, similarly to attention, few studies have focused on the evaluation of associations between working memory and social functioning. Nonetheless, studies in SCZ have found that empathic deficits and stronger negative symptoms (including social withdrawal) are linked to lower performance in working memory tasks (Cameron et al., 2002; Pantelis et al., 2003; Smith et al., 2014b). Furthermore, an association between working memory and mentalizing performances has been reported in a neuroimaging paradigm (Spunt and Lieberman, 2013). Moreover, an association between poor working memory performances and broad social dysfunctions has been reported also in healthy children (McQuade et al., 2013). Finally, similarly to attention, working memory was found to modulate interpersonal behaviour (Bowie et al., 2008). Following this line of research, a social working memory network has been proposed, with medial fronto-parietal regions showing increased activation as a function of increasing social load (Meyer and Lieberman, 2012). Additionally, an association between working memory and mentalizing performances has been reported in a neuroimaging paradigm (Spunt and Lieberman, 2013). Moreover, an association between poor working memory performances and broad social dysfunctions has been reported also in healthy children (McQuade et al., 2013). Finally, similarly to attention, working memory was found to modulate interpersonal behaviour (Bowie et al., 2008). Following this line of research, a social working memory network has been proposed, with medial fronto-parietal regions showing increased activation as a function of increasing social load (Meyer and Lieberman, 2012). In particular, the top-down connectivity from prefrontal to posterior regions has been linked to performance in socially demanding tasks (Hillebrandt et al., 2013). Nonetheless, despite this recent evidence, the specific mechanisms behind the possible interaction between social and working memory systems require further exploration. Given the complexity of social interactions, it is likely that individuals with working memory impairments will have more difficult time in considering multiple pieces of social information, thinking through their actions, and referring prior social knowledge (McQuade et al., 2013). Thus, it could be hypothesized that deficits in working memory could lead to a disengagement from social interactions, possibly driving to social withdrawal. Obviously, this hypothesis requires dedicated studies in order to be rigorously investigated.

4.3. Sensory processing, social cognition, and social functioning

More complex aspects of social cognition discussed earlier, such as (facial) emotion recognition, are likely to be adversely affected by abnormalities in more basic sensory processing. For example, early visual deficits, or aberrant processing of pitch changes in speech, are likely to have a deleterious impact on higher level cognitive function. Supporting this idea, there is a growing body of research relating atypical facial expression recognition and resulting social problems observed in ASD to basic deficits in the processing of low spacial frequency information, such as the overall configuration of the face (Vlamings et al., 2010).

In SCZ, some of the most replicated sensory processing deficits involve anomalies in the generation of well-characterised event-related potentials (ERPs), such as the P50 (implicated in sensory gating) and N100 (elicited by an unpredictable stimulus in the absence of task demands) (Turetsky et al., 2009). Both potentials are generated within auditory sensory regions and point to breakdown of processing at early stages of stimulus evaluation (Javitt, 2009). In addition, SCZ is characterised by deficits in mismatch negativity (MMN) generation (Umbricht and Krljes, 2005). MMN is elicited only in response to stimuli that deviate from a predictable sequence, and can apply to auditory attributes such as pitch and duration. As changes in both the pitch and duration of vocalizations can accompany changes in emotions expressed in social interactions, such a deficit may contribute to the abnormal patterns of social behaviour and social judgement typical of the disorder. Consistently, MMN amplitude was found to be highly predictive of SCZ patients’ level of independence in daily life (Light and Braff, 2005). This supports the link between sensory processing and daily function, likely including successful social functioning.

Multisensory events are plentiful in real-life social interactions and
are likely to be critical in supporting adaptation to different social interactions and environments. However, the integration of information obtained from multiple sensory modalities is fundamental in order to correctly understand the contextual environment, in particular the social one, as stated above. This integration was found impaired in SCZ patients, resulting in poorer performances in emotion identification (Calvert et al., 2000; Calvert, 2001; de Gelder et al., 2005) with obvious deleterious consequences on social relationships.

There is also emerging evidence that similar lower-level sensory processing is disrupted in AD and its precursors, such as mild cognitive impairment (MCI). As a matter of fact, MMN response was found altered in auditory (Jiang et al., 2017) and visual tasks (Tales et al., 2008) in early-stage AD patients. The abnormal visual MMN indicates that basic functions associated with the automatic detection of change within the visual environment are abnormal in AD. These effects were strongest in frontal–central areas, suggesting impairment in pre-attentive sensory processes that may be directly relevant to PFC-based socio-cognitive functioning. Thus, some of the high-level social and cognitive deficits observed in AD may be the result, at least in part, of significant deficits in these pre-cursor sensory processing systems, although further evidences supporting this hypothesis are clearly needed.

So far, we have discussed examples of ways in which sensory processing can modulate interpersonal behaviour. Interestingly, however, the relationship has been suggested to be mutual, since loneliness also seems to modulate the neural response to sensory (e.g. threatening and non-threatening) stimuli (Cacioppo et al., 2016). Nonetheless, studies investigating this association are still few and further dedicated investigations are needed to disentangle this issue.

4.4. The relationships among cognition, social cognition, and social functioning

A question which remains open is to what extent social and non-social processes rely on distinct neural processes or share similar neural resources (Adolphs, 2003, 2010). Broadly speaking, a consensus is arising that social cognition is related to but not completely reliant on non-social cognitive processes. Cognitive neuroscience supports this notion. Studies of social and non-social rewards provide a good example of distinct but overlapping networks, since many of the regions involved in non-social valuation processes also support perception of social reward and pain, similarly to the neurotransmitter involved, as previously discussed. However, studies comparing activity patterns elicited by social outcomes with those to monetary rewards reported overlapping as well as differential activity patterns (Izuma et al., 2008; Rademacher et al., 2010; Sescousse et al., 2010; Smith et al., 2010; Spreckelmeyer et al., 2009). Differences may arise due to dependence of social valuation processes on contextual factors, or social values, that are not present or far less relevant to non-social decision-making. For example, social evaluation processes and subsequent decisions have been shown to be influenced by others’ perceived intentions (Gurugu et al., 2010), social comparison (Wright et al., 2011), and in-group vs. out-group membership (Baumgartner et al., 2012). Social principles are likely to influence value-coding, at least in part, through functional connectivity of higher-order cognitive regions involved in social cognition, such as the TPJ and vmPFC (which are involved in the mentalizing and affiliation networks, Fig. 1) (Hare et al., 2010; Janowski et al., 2013; Smith et al., 2014a). These high-order regions may themselves be subdivided into different functional regions, as suggested for example by Scholz et al. (Scholz et al., 2009) who observed that neighbouring but distinct regions of the TPJ were activated by a social (mentalizing) and non-social (orienting) tasks.

Overall, these findings indicate the uniqueness of some aspects of the social brain but also potentially a considerable overlap with brain mechanisms underlying non-social cognition, as previously discussed. Social functioning places higher demands on processing requirements, some of which are non-social in nature. In fact, it has been observed that interpersonal behaviour could be predicted by processing speed, attention and working memory, together with executive functions and depressive and negative symptoms (Bowie et al., 2008). Interestingly, the effects of attention, working memory and processing speed seem to be mediated by their effects on social competence. Therefore, it can be hypothesized that these cognitive deficits induce impairments in the patient’s social competence which eventually results in high social withdrawal, a hypothesis supported by findings in subjects at high-risk of SCZ (Jahshan et al., 2012). In turn, social withdrawal, and the resulting social isolation/loneliness, may cause a worsening of these cognitive deficits (Cacioppo and Hawkley, 2009; Cacioppo et al., 2016; Gow and Mortensen, 2016; Wilson and Koenig, 2014; Hoffman, 2007), resulting in a vicious circle of progressive worsening of the general functioning and the patient’s quality of life. However, further studies investigating in deep social and non social cognitive domains, together with real-life social functioning, are needed to better elucidate these complex interactions.

5. The social brain in different neuropsychiatric disorders (AD, SCZ, and MDD)

We have dissected in the previous sections social functioning from different and complementary perspectives. In this last section we will focus on the topic which has the most clinical and societal impact: how do major neuropsychiatric disorders interact with the social brain? This is a central point given the reciprocal deleterious effects of psychiatric disorders on social functioning and vice versa. A better understanding of how the pathophysiology of psychiatric disorders may influence social brain pathways may pave the way for targeted and innovative treatments. We will therefore review in this perspective three neuropsychiatric disorders (namely AD, SCZ, and MDD) selected among the several ones characterized by social dysfunctions because of their frequencies and heavy burden in Western countries (they account for the 31.5% of disability-adjusted life years - DALYs - associated with neuropsychiatric disorders and substance use disorders (Whiteford et al., 2015), with a global estimated economic burden €312 billion per year) (WHO, 2008; Wittchen et al., 2011) and because social withdrawal often represents one of their first signs, as detailed below. Indeed, these three neuropsychiatric disorders are characterized by a significant change in social behaviour, rather than lifelong social difficulties, such as in ASD. Further, most of these costs are indirect (lack of productivity) and social withdrawal is an important source of indirect costs and it has been identified as one of the main reasons for mental health related disability benefit claims (UK Department for Work and Pensions, 2013). Moreover, these diseases lie heavily on caregivers and impact significantly on their ability to work (Haro et al., 2014). Finally, in these disorders a wide clinical variability in social dysfunctions (both in terms of amount and time course) has been observed, paving the way for the investigation of underlying biological modulators. For these same reasons, these disorders will be investigated in the context of the PRISM project described in this issue ((Kas et al., 2017) and Bilderbeck et al. in this issue). Although we are aware that other neuropsychiatric disorders are characterized by social dysfunctions, as previously stated, in this section we only aim to provide clear examples of how different psychopathological mechanisms may affect the social brain, eventually resulting in the clinical phenotype of social withdrawal.

5.1. The social brain in schizophrenia

Social functioning is highly impaired in SCZ, resulting in a great burden for both patients and relatives (Casqueo-Urizar et al., 2009; Lehman et al., 1982). SCZ patients usually show a reduction in social connections (although a wide variability exists, from almost normal social life to clear social withdrawal (Rossi et al., 2016)), reduced rates of employment, and impaired ability to live an independent life (WHO, 2008). Dysfunctions in social functioning emerge early in life in SCZ,
years before the full onset of the disease. Indeed, large cohort studies showed that subjects who will be affected by SCZ showed higher preference for solitary playing as early as 4 and 6 years of age, lower social confidence at 13 years, and higher social anxiety at 15 years (Jones et al., 1994; Keskinen et al., 2015). Therefore, the progressive pathological process in individuals with SCZ is characterized by some degree of social withdrawal starting in adolescence or even in childhood, in any case well before the full onset of the disease and it represents a core feature of the disease itself (Howes and Murray, 2014). It has been hypothesized that this early social withdrawal may trigger, in vulnerable individuals, the development of a full psychosis through social deprivation (the so-called deafferentation hypothesis (Hoffman, 2007)), resulting in a vicious circle (i.e., subtle social dysfunctions may cause early social withdrawal, which further exacerbate alterations in the brain through deprivation, leading to more severe social withdrawal and facilitating the development/worsening of psychotic symptoms). In parallel, subjects who will develop SCZ showed also delay to reach developmental milestones during childhood and lower scores at educational tests at different ages (Jones et al., 1994; Keskinen et al., 2015), as well as global (Trotta et al., 2015; Woodberry et al., 2008) and specific (Cuesta et al., 2015; Greenland-White et al., 2017; Lenz et al., 2006) cognitive impairments (Sommer et al., 2016). Historically, social dysfunctions have been regarded as ancillary symptoms of a more global cognitive impairment during SCZ premorbid phase, and as part of the negative symptomatology, after the disease onset (Lee et al., 2015a). Nonetheless, in clinical practice, a gap between the observed subtle cognitive impairment (Trotta et al., 2015; Woodberry et al., 2008) and the great social dysfunction in SCZ has been repeatedly reported, suggesting that social dysfunction is not simply a consequence of the SCZ symptomatology. Consistently, a meta-analysis showed that social cognition explains only a surprisingly little variance (15.2%) of functional outcome in SCZ (Fett et al., 2011). It has been suggested that social cognition may fill the gap between “cold” cognition (i.e., the ensemble of information processing that is independent of emotional involvement (Roiser and Sahakian, 2013)) and real functioning. Consistently, a growing amount of evidence demonstrates a greater impairment in social cognition than in “cold” cognition in SCZ (Sora et al., 2017; Pinkham, 2014). Indeed, social cognition is deeply impaired by SCZ, with deficits in 1) emotion processing (Jani and Kasparend, 2017); 2) Theory of Mind (ToM)/mentalizing (Bora et al., 2009; Jani and Kasparend, 2017; Song et al., 2015); 3) social perception and 4) attributional style (Healey et al., 2016; Savla et al., 2013). These deficits have been identified during premorbid status (Eack et al., 2010; van Donkersgoed et al., 2015), they further worsen after illness onset, and persist for the entire lifespan ranging from low to very severe deficits (Rocca et al., 2016), similarly to neurocognitive impairments (Mesgholam-Gately et al., 2009; Trotta et al., 2015). Social cognition is only moderately correlated with “cold” cognition (Rocca et al., 2016; Ventura et al., 2013) and negative symptoms, while it is poorly associated with positive symptoms (excluding disorganization) (Lewandowski et al., 2016; Rocca et al., 2016; Ventura et al., 2013). Therefore, social cognition seems to be, at least partially, an independent domain (Green et al., 2015; et al., 2015; Gur and Gur, 2016), and the partial associations reported above might reflect the impairment of more basic processes involved both in cognition and symptomatology (e.g. emotional experience/expression, decoding and/or integration of environmental information, etc.). Thus, social cognition may play a mediational role between “cold” cognition and social competence (Couture et al., 2011), although it explains only a minor variance of real-world functioning in SCZ (23.3%) (Fett et al., 2011), even when considered together in a unified model with negative, disorganized symptoms, and social competence (from 7 to 25%) (Bowie et al., 2006; Breetke et al., 2005; Couture et al., 2011; Sergi et al., 2006; Vauth et al., 2004). Therefore, it could be possible that other deficits contribute to determine social functioning in SCZ, although real world behaviour is also obviously influenced by factors outside the individual’s control (e.g. level of social support, financial means, personal resources, etc.). In particular, a reduced capacity for anticipating future pleasure resulting from goal-directed action (Simpson et al., 2012), a tendency to maintain more physical distance (i.e. personal space) from others (Holt et al., 2015), and an increased stress response to psychosocial stressors (Lange et al., 2017a) have been already demonstrated in SCZ. Similarly, other processes pertaining to the social functioning are still to be elucidated and further studies investigating these domains together with social cognition, “cold” cognition, and real-world functioning are needed. Based on the available findings we hypothesize that the detection of other individual's determinants of real-world functioning may increase the insight about the disorder and open new therapeutic perspectives in SCZ treatment.

To achieve this, the neurobiological bases of these impairments have to be identified. This will also allow a more neurobiologically based classification of SCZ patients. In the first sections of this review, we have attempted to dissect the neurobiological basis of social dysfunction both at a neuronal and neurotransmitter level. In this section, we will discuss how the pathophysiology of SCZ affects these neurobiological substrates, leading to the impairments exposed above.

Brain structure abnormalities have been demonstrated in SCZ (Amato et al., 2017; van Erp et al., 2016) as well as white matter deficits (Ui et al., 2009). These neuroanatomical abnormalities seem to be present from the first episode of SCZ (Vita et al., 2006), as well as in prodromal/high-risk individuals (Cannon et al., 2015), and they are likely to increase with SCZ duration (Fusar-Poli et al., 2010; Vita et al., 2012). Among these widespread aberrations, some affect key regions within the social brain (Fig. 2). The most part of data pertain to the neural correlates of social cognition impairments. More in detail: 1) emotion control deficits have been associated with an impaired control of the PFC on the amygdala and with aberrant vIPFC activation (Green et al., 2015; et al., 2015); 2) motor resonance (i.e. experience sharing) impairment has been linked to a less-fine-tuned activation of the right inferior parietal lobule and posterior STS (i.e. of areas involved in both the mirror and mentalizing networks, Fig. 1 (Bickart et al., 2014b)) during action observation and imitation (Green et al., 2015); 3) mentalizing impairment has been consistently associated with a decreased/delayed activation of “core” mentalizing network (Fig. 1) during ToM tasks (e.g. the left IFG, vmPFC, orbitofrontal cortex, mPFC, TPJ, and IFG) (Bickart et al., 2014b; Green et al., 2015; Pedersen et al., 2012), together with an aberrant increased activation of areas not usually associated with the specific ToM task (Brune et al., 2008; de Achaval et al., 2012; Frith, 2004; Kronbichler et al., 2017); 4) face perception impairment has been associated with a decreased activation of amygdala, right inferior occipital gyrus, right fusiform gyrus and hippocampal areas, ACC, mPFC, and thalamus (i.e. of the social perception network, Fig. 1(Bickart et al., 2014b)), together with an aberrant activation in the insula, cuneus, parietal lobule, and STG (Green et al., 2015). Similarly, voice perception defects have been associated with decreased activation of key areas of the process, such as the left and mid STG and the bilateral IFG, and aberrant activation in other areas, such as the left middle temporal gyrus and the left insula (Green et al., 2015); 5) attributional style defect has been associated with decreased activations in the IFG, the ventral premotor cortex, primary motor cortex, middle cingulate cortex, amygdala, thalamus, and striatum in happy conditions and to increased activations in the prefrontal/posterior cingulate cortex, and the vmPFC in angry conditions (Park et al., 2009) (i.e. to aberrant activation of aversion and affiliation networks, respectively, Fig. 1 (Bickart et al., 2014b)).

Beyond social cognition, the other processes that can contribute to social dysfunction in SCZ have been less investigated. However, some studies started to elucidate the possible neurobiological correlates of these impairments. For example: 1) deficit in motivation/reward anticipation has been associated with dysfunction in OFC and ventral striatum (or in their connectivity) (Simpson et al., 2012); 2) the aberrant personal space seems to be related to dorsal intraparietal sulcus.
and the ventral premotor cortex connectivity defects (Holt et al., 2014) 3) increased stress response to social stressors has been associated with hippocampal dysfunctions, since HPA axis cortisol peripheral response was conserved (Ciufolini et al., 2014; Lange et al., 2017a). As previously described, these areas are involved in the five networks which are thought to sustain social functioning (Fig. 1). Thus, these preliminary findings may suggest an impairment of these networks, rather than impairments of specific brain areas. Nonetheless, data about these domains are few and further studies are needed to better understand the neurobiological bases of these impairments also at a neural network level and how they modulate social behaviour itself.

At the neurotransmitter level, knowledge is even scarcer, mainly because it is derived from (epi)genetic and pharmacological studies in humans and, largely, from animal studies. Nonetheless, SCZ seems to impact, directly or indirectly, almost all the neurotransmitter systems involved in social functioning. For example, SCZ impacts the DA system (directly or through other system defects during neurodevelopment), leading to an excessive presynaptic DA synthesis and release (Howes and Murray, 2014). This may lead to an inappropriate DA release following different environmental stimuli, which all gained a “hyper-salience” for the SCZ subject (van Os, 2009), independently from their current meaning (a possible source of delusion interpretations (Howes and Murray, 2014)). In the social functioning context, this aberrant DA functioning may disrupt the fine mechanism at the basis of social reward/pain, as well as the delicate inter-play needed for bond formations. Further, also learning processes, which are at the basis of motivation, could be impaired by DA aberrant transmission. In turn, this aberrant transmission may lead to long term alterations in other systems, such as the 5-HT and glutamatergic ones (Howes and Murray, 2014). Moreover, animal studies on SCZ models, suggested that other neurotransmitter systems involved in social behaviours such as the CRH, OXT, and endocannabinoid ones, are altered in their functioning and might result in social withdrawal/dysfunctions (e.g., (Seillier et al., 2013)). Unfortunately, further studies are needed to better understand the fine interactions across the neurotransmitter systems and how they were disrupted by SCZ determining social dysfunctions/withdrawal.

In conclusion of this brief overview, we may consider that SCZ physiopathology directly impairs at multiple sites the brain structures which sustain social functioning (Figs. 1 and 2). This may cause a vicious circle because of the well known negative influence of poor social functioning on SCZ outcome (e.g.,(Tremeau et al., 2016)) and raising the issue of a possible targeted therapeutic intervention on social functioning brain structures.

5.2. The social brain in Alzheimer’s disease

According to several studies the prevalence of non-cognitive psychopathological symptoms in cases of dementia is substantial, often called behavioural and psychological symptoms of dementia (BPSD). Among them, negative-type symptoms (i.e. apathy, avolition, emotional disengagement, lack of initiative and motivation, and social withdrawal) are highly represented (Lyketsos et al., 2002; Saz et al., 2009). Similar to SCZ, a growing body of evidence suggests that these symptoms form a separate cluster of non-cognitive symptoms, which is partially independent from other psychopathological symptoms such as the depressive ones (Reichman et al., 1996; Reichman and Negron, 2001). The relevance of BPSD was demonstrated by the cost for their management (i.e. a third of dementia care cost) (Kales et al., 2015). Moreover, BPSD have been also observed in conditions expected to frequently converge into dementia, such as MCI (Lopez-Anton et al., 2015), and in the prodromal stages of dementia, i.e. in individuals before the manifestation of cognitive or functional decline (Ismail et al.,...
Among BPSD, it is not surprising that social withdrawal is the earliest recognizable psychiatric symptom of AD, occurring on average 33 months before diagnosis (Jost and Grossberg, 1996). Indeed, as previously stated, the enormous amount of brain processes required to initiate and maintain social relationships (Adolphs, 2009) likely reflects an intrinsic vulnerability to pathological insults, which may result in social withdrawal far before the full onset of the disorder. Obviously, social withdrawal is likely preceded by subtle social dysfunctions, which are difficult to detect clinically (Jost and Grossberg, 1996; Kumfor et al., 2014). Interestingly, in the dementia field, it has been demonstrated that objective social isolation and loneliness (i.e. the subjective feeling, and complaint of being alone) have different impacts on outcome (Cacioppo and Hawkley, 2009; Holwerda et al., 2014), supporting the hypothesis of different neurobiological bases for these aspects of social withdrawal. In particular, loneliness (but not objective social isolation) has been associated both with cortical amyloid burden in cognitively normal elderly subjects (Rosenberg, 2016) and with an increased risk of late-life dementia (Donovan et al., 2015; Holwerda et al., 2014). Consistently, the presence of social withdrawal and other negative-type symptoms has been associated with categories of MCI at higher risk of conversion into dementia (Lopez-Anton et al., 2015). Parallel to SCZ, distinguishing social withdrawal from other symptoms, such as apathy and depression in AD and in dementia in general, is also difficult, since partial overlaps and inter-correlations likely exist. For example, apathy (i.e. a loss of motivation and decreased interest in daily activities (Lancot et al., 2017)) may cause decreased reactions to others and to the surrounding environment, leading to social withdrawal (Landes et al., 2001). But, similarly to SCZ (Hoffman, 2007), an inverse or a bidirectional association could be hypothesized as well (Courin and Knapp, 2017; Honda et al., 2013), since social withdrawal itself may decrease motivation through stimuli deprivation in a vicious circle. In the same way, a bidirectional association between social withdrawal and cognitive deficits could be hypothesized. Indeed, specific cognitive deficits, such as word-finding difficulty (Farrell et al., 2014) and face/emotional recognition (Elamin et al., 2012), reduce engagement in social leisure activities leading eventually to social withdrawal. On the other hand, social withdrawal itself seems to determine difficulties in activities of daily living which required cognitive input (Reichman and Negron, 2001).

Irrespective of the causal relationships (i.e. social withdrawal as a prodrome or as a risk factor for AD), from a neurobiological point of view, these associations demonstrate that AD (and dementia) impacts the brain structures which sustain social functioning (Figs. 1 and 2). Current knowledge allows drafting an initial picture of the social brain structures described in the previous sections which are damaged in AD. Similar to SCZ, AD impacts these structures at different levels, resulting in social cognitive deficits and ultimately in social withdrawal (although it is often the first sign of the disorder itself, since subtle cognitive impairments are usually not recognized by patients and relatives). Although social cognition has been less investigated in AD compared to SCZ, a growing body of evidence suggest that it is broadly affected by AD and other neurodegenerative disorders. These impairments usually worsen as the disorder progresses, starting from subtle deficits in the prodromal phases to severe impairments in the advanced phases of the disorders (Elamin et al., 2012; Kumfor et al., 2014). In neurodegenerative disorders, the majority of the studies in this field focused on mentalizing. In AD, deficits in both primary and secondary order belief mentalizing tasks have been reported. Advanced-level mentalizing skills are the first affected by AD, while the more basic mentalizing skills are affected with the disorder progression (Elamin et al., 2012). Similar to SCZ, mentalizing impairments seems to be only partially modulated by general cognition and executive functions in neurodegenerative disorders (Bora et al., 2015; Laisney et al., 2013; Shany-Ur and Rankin, 2011). From a neuroanatomical point of view, in these populations, mentalizing impairment severity has been associated with the degree of atrophy in mentalizing network regions (Fig. 1 (Bickart et al., 2014b)), such as the ventromedial, anterior temporal, and frontoinsular cortices (Adenzato et al., 2010; Bora et al., 2015). Probably, the mentalizing impairments observed in AD are due to damages in these areas, which are usually affected to a lesser extent and in latter phases compared to the behavioural variant of the fronto-temporal dementia (bv-FTD). Consistently, mentalizing impairments in AD are less severe compared to the impairments observed in bv-FTD (Adenzato et al., 2010; Bora et al., 2015). On the contrary, AD affects motivation yet at early stages, resulting in apathy. At the neural level, apathy has been associated with the degree of atrophy in ACC, dIPFC, and striatum (Boulay et al., 2016; Dickerson, 2015), as well as with connection dysfunctions (i.e. white matter damages) between ACC and other brain structures such as the OFC, limbic areas, and basal ganglia (Theleritis et al., 2014). As discussed in the previous sections, the ACC seems of great importance in this context, since it is thought to operate as an interface between emotions (Peng et al., 2012), motor control (MacDonald et al., 2000), cognition (Paus, 2001) and intentional and motivated behavioural responses (Zhou et al., 2014). Consistently, the ACC is involved in the affiliation, aversion, and mentalizing networks mentioned above (Fig. 1), although its different regions are differently involved in these networks (Bickart et al., 2014b). Moreover, impairments in emotion recognition have been hypothesized to reflect the progression of neurodegeneration from the entorhinal cortex and hippocampus towards the lateral temporal neocortex (Elamin et al., 2012), i.e. when neurodegeneration starts to affect the social perception network (Fig. 1). However, neurodegeneration in the entorhinal cortex itself may cause an impairment of the affiliation network as well, resulting in social and emotional detachment from other people (Fig. 1 (Bickart et al., 2014b)). The impairments in abstraction and judgment associated with frontal cortex atrophy may also contribute to social dysfunctions, since these functions are involved in both emotional processing and social cognition (Kennedy and Adolphs, 2012). Moreover, the impaired processing of contextual associations (i.e. the ability to adequately recognize the influence of context in the meaning of events perceived) could determine inadequate behaviours in different social contexts, with predictable deleterious consequences on social life. At a neural level, this impairment has been associated with altered functioning of frontal and temporal areas. In detail, frontal areas, such as the PFC, IPFC, and superior orbital sulcus, are thought to update and associate ongoing contextual information in relation to episodic memory and target-context associations (Ibanez and Manes, 2012), while temporal regions (i.e. amygdala, hippocampus, perirhinal and parahippocampal cortices) index the value learning of target-context associations. Finally, the insular cortex coordinates internal and external milieu in an internal motivational state. In this model, the insula would provide information integration from internal states and social contexts to produce a global feeling state, which in turn facilitates or prevents social interactions. However, as previously mentioned, these impairments likely reflect neural networks dysfunctions, rather than impairments of specific brain areas. Unfortunately, current knowledge does not allow the confirmation of this hypothesis and further studies are needed to link these preliminary findings to impairments at a neural network level.

At the neurotransmitter level, in AD, negative-type symptoms (including social withdrawal) have been associated with impairments in cholinergic, NE, 5-HT, and DA systems (Boulay et al., 2016). In relation to the cholinergic system, its involvement in the pathogenesis of negative-type symptoms such as apathy has been emphasized (Minger et al., 2000; Mori et al., 2014). Specific cholinergic projections from the nucleus basalis of Meynert to the frontal limbic cortical regions play a relevant role for apathy. Moreover, the 5-HT system was found seriously impaired in AD, suggesting its relevance in AD pathophysiology (Lai et al., 2002; Meltzer et al., 1998). In particular, the 5-HT pathways from the raphe nuclei seem to be deeply affected by AD (Hardy et al., 1985; Prakash et al., 2015). As stated in Section 3, the 5-HT transmission throughout these pathways seems to promote a calm status,
decreasing anxiety and aggressive behaviours. Thus, their disruption may lead to agitation and aggressive behaviours, with predictable consequences on social interactions. However, other 5-HT pathways might also contribute to social dysfunction in AD, as suggested by the effects on cognition and social interactions of the 5-HT6 receptor (de Bruin et al., 2016; Prakash et al., 2015), which is mainly expressed in striatum, cerebral cortex, and cerebellum (Parker et al., 2012). On the other hand, the role of DA in the pathogenesis of AD is still not clear (Prakash et al., 2015), although symptoms such as apathy and executive functioning impairment suggest its involvement (Martorana and Koch, 2014). Consistently, meso-striatal DA pathways were found impaired in about half of AD patients, although this finding was reported in 30% of healthy elderly as well (Martorana and Koch, 2014). However, other specific DA alterations were found in AD, such as a reduced expression of D1-like and D2-like receptors in PFC and hippocampus, as well as of D2-like receptors, DAT, and TPH within the NAc (Martorana and Koch, 2014). Interestingly, preclinical data showed that DA system pathology and amyloid deposition are closely related, suggesting a causative role for amyloid on DA dysfunction (Perez et al., 2005). Independent from possible causal relationships, these alterations are likely to account for apathy; this in turn may lead to social disengagement, as previously stated. Interestingly, these DA system alterations have been associated with a worse AD progression, suggesting that it may be possible to distinguish AD patients on the basis of the degree of DA system impairment (Martorana and Koch, 2014). Other neurotransmitter systems involved in social brain are probably affected by AD, such as OXT and AVP (Ishunina and Swaab, 2002), and GABA systems (Martorana and Koch, 2014), thus contributing to the observed social dysfunctions. Unfortunately, these systems were poorly investigated in AD and further studies are needed to better understand their role in AD and its symptomatology, particularly in social dysfunctions. In addition, some of the evidence for a role of neurotransmitters in social behaviour have been derived from other populations, and remain to be confirmed specifically in AD.

In conclusion, AD physiopathology seems to progressively impair social brain at multiple sites, starting from the structures that sustain motivation to the mentalizing network. Similar to SCZ, the resulting poor social functioning accelerates the disease progression in a vicious circle because of the negative influence of social deprivation on cognitive performances. Consistently, a possible targeted therapeutic intervention on social functioning brain structures could be hypothesized, similarly to SCZ.

5.3. The social brain in major depressive disorder

Depressed patients are usually less severely impaired in everyday lives than SCZ or AD patients. Nonetheless, social dysfunction has been recognized as an important sign of depression itself (Hirschfeld et al., 2000). For a long time, it has been attributed to the overall depressive symptomatology. However, similarly to what has been reported for SCZ and AD, nowadays it is recognized as a partially independent domain (Gur and Gur, 2016), which can persist for years after the recovery from the core depressive symptoms (Rhebergen et al., 2010) and correlates with unemployment, disability and decreased work performance (Rizvi et al., 2015). Recently, a comprehensive review on this issue has been published (Kuperberg et al., 2016a), underlining the relevance of social dysfunction in MDD. Providing a complete review of the literature on this issue is beyond the aim of the present section, we encourage the reader to refer to the review by Kuperberg et al. in this same journal for detail (Kuperberg et al., 2016a). In the present section we will link the neurobiological basis of social dysfunction in MDD with the general dissection we detailed in the previous sections.

Social dysfunctions in MDD are pervasive and encompass almost every aspect of one's social capabilities. Unfortunately, data at the neural network level are still lacking and the most part of the studies investigated single brain areas or basic paradigm, such as social reward/pain processes. However, we should keep in mind that social reward/pain paradigms are intrinsically related with the affiliation and aversion networks (Fig. 1), as previously stated. Thus, impairments in these processes probably reflect impairments in the pertinent neural networks, rather than isolated defects. More in detail, in MDD, the main disturbances were found in social reward/pain processes, as well as in processing and mentalizing social signals. These main deficits are likely the causes of other observed impairments, such as the decreased motivation through social interactions, increased sensitivity to peer rejection, diminished cooperativeness, competition avoidance, and alterations in social decision-making (Kupferberg et al., 2016a). The pivotal role of social reward reduction, repeatedly observed in MDD (Derntl et al., 2011; Germine et al., 2011; Nusslock and Alloy, 2017; Rey et al., 2009), has its origin in the NAc. This important structure involved in the social affiliation network (Fig. 1) and encoding reward-related signals is hypo functioning in MDD (Kuperberg et al., 2016a; Laurent and Ablow, 2012). Although the exact causes of NAc hyporesponsivity to social reward stimuli are still not clear, preclinical data suggested an involvement of the DA projections from VTA to NAc (Chaudhury et al., 2013), which in turn are modulated by OXT (Gropp et al., 2013) and probably by the BDNF signalling (Wook Koo et al., 2010). However, as described in the previous sections, NAc activity is further modulated by other brain areas, each one with specific effects on social behaviour. Thus, the reduced social reward observed in MDD is likely related to the fine modulation of NAc (re)activity, which in turn is the result of the different modulations provided by the VTA, thalamus, PFC, and hippocampus projections to the NAc itself (Covington et al., 2010; Kuperberg et al., 2016a; Vialou et al., 2014). On the other hand, MDD patients showed an increased sensitivity to rejection (Elmavil et al., 2014) and a greater magnitude/duration of negative feelings elicited by social rejection (i.e. social pain) (Hsu et al., 2015). In turn, rejection sensitivity was associated with higher rates of internal life stressors (Liu et al., 2014) and might cause higher-level cognitive biases like attributional style and negative expectations, linking rejection sensitivity to depressive symptoms (Liu et al., 2014). Rejection sensitivity eventually leads to maladaptive behavioural responses such as social withdrawal and aggressive response (Kuperberg et al., 2016a). At the neural level, we previously underlined the primary role of amygdala in social pain processing, since its response/magnitude of activation to social rejection was directly associated with emotional pain. However, the feelings of social distress in response to social exclusion were associated mainly with another area, the insula (Masten et al., 2009), which is connected to NAc, thus modulating its activity (Leong et al., 2016). As described in the previous sections, the insula also integrates information from internal states and social contexts, determining a global feeling state. Once more, these preliminary data may reflect an impairment of the social aversion network rather than impairment in specific brain areas, since both amygdala and insula were included in this network (Fig. 1) (Bickart et al., 2014b). Interestingly, amygdala and insula hyperactivity may be attributable to an inadequate control by the dIPFC, which in healthy subjects usually overrides the automatic emotional responses caused by amygdala and insula activation (Hooley et al., 2005). Finally, sgACC activity was associated with negatively biased interpretation of social rejection, which in turn might increase emotional responses to rejection, in a vicious circle which worsens over time leading to maladaptive behavioural responses, such as social withdrawal (Kuperberg et al., 2016a). A possible consequence of this increased sensitivity to social pain is the avoidance of social competition, a characteristic often described in MDD (Kuperberg et al., 2016b; Price et al., 2004). Indeed, MDD patients avoid social competition in order to reduce the damage caused by a potential loss. Unfortunately, the cognitive biases associated with the disorder determine an overestimation of the competition risks and an over self-attrition of competition loses, increasing the belief that they are due to personal undesirable qualities and, thus, the depressive symptoms themselves (Gilbert et al., 2009). At the neural level, this
impairment has been associated with the areas involved in both social reward and pain (and thus partially with the social affiliation and aversion networks, Fig. 1), since competition was associated with both social reward and fear of losing (i.e. anticipation of social pain). However, competition seems to specifically involve other areas as well, such as the bilateral inferior parietal cortex and the habenula (Kupferberg et al., 2016a). Interestingly, habenula activity was found impaired in MDD (Proulx et al., 2014) and preclinical data suggested that it is related to both the regulation and rewarding effects of hierarchy-related behaviours (i.e. competition behaviours) (Chou et al., 2016; Golden, 2014). Thus, this impairment might explain also the reduced motivation to compete observed in MDD (Kupferberg et al., 2016a). Other possible consequences of impaired social reward system were on one hand the increased prosocial preference for fairness at the cost of personal benefits (Destoop et al., 2012; Gradin et al., 2016; Scheele et al., 2013), and on the other hand the difficulty to sustain reciprocal cooperation (Gradin et al., 2016). Indeed, MDD patients showed a decreased evaluation of both personal costs secondary to altruistic punishment and benefits due to cooperation, likely due to a decreased activation of NAc and dorsal caudate nucleus, which are key areas of both the reward system and of the social affiliation network (these systems are likely partially overlapping, as previously stated). Beyond the impairments in social reward/pain processing and their consequences, MDD patients showed also an impaired emotion recognition (Dalili et al., 2015; Gollan et al., 2010) and a negative emotional bias (Bourke et al., 2010; Naranjo et al., 2011), which contribute to deficits in receptive communication (Kupferberg et al., 2016a). Intriguingly, the brain areas involved in the stimuli processing (e.g., for facing processing the amygdala, insula, parahippocampal gyrus, fusiform face area, and putamen (Stuhrmann et al., 2011)) showed a hyperactivity to negative and hypo activity to positive stimuli, representing a putative neural correlate of the negative emotional bias itself. Once more, the aberrant activation of these areas probably reflects an impairment of higher order neural networks, such as the perception, affiliation, and aversion networks (Fig. 1, (Sickart et al., 2014b)). However, an abnormal pattern of activation was also found in ventral PFC and dIPFC during social stimuli (e.g. face processing (Phillips et al., 2003), suggesting that the abnormal activation in the limbic areas and their related neural networks might be due to an inadequate modulation by the PFC, at least partially (Kupferberg et al., 2016a). Additionally, the increased connectivity between the sgACC and the amygdala, which causes a mutual enhancing of abnormal emotion processing during negative stimuli presentation, may represent a neural mechanism for the abnormally increased representation of a social threat (Stuhrmann et al., 2011). Thus, similar to social rejection sensitivity, the general deficit in emotion recognition and the negative bias observed in MDD patients might be due to an impaired top-down control of emotional processing by the dIPFC (Groenewold et al., 2013) and to an abnormal evaluation of social threat provided by the sgACC (Stuhrmann et al., 2011). Clinically, depressed patients fail to correctly recognize emotion in others and tend to attribute negative meaning to social stimuli because of their negative emotional bias (Kupferberg et al., 2016a). Unfortunately, these patients also show impairments in mentalizing and fail to correctly infer others’ mental states (Bora and Berk, 2016) even during remission periods (Inoue et al., 2006). Clearly, the impairments in mentalizing are not comparable to the impairments observed in SCZ (Wolkenstein et al., 2011) and they result in deficits in interpretation of complex tasks, such as understanding humour or paradoxical sarcasm (Kupferberg et al., 2016a). However, the affected brain areas seem to be the same, since abnormal activities in areas of the mentalizing network (Fig. 1), such as the vmPFC, dIPFC, and temporo-parietal regions, have been demonstrated in MDD (Conson et al., 2015; Cusi et al., 2012). Moreover, the increased activity in anterior sgACC has also been associated with deficit in ToM, particularly with an impulsive way of mentalizing (Pincus et al., 2010), likely associated with the negative bias described above. Finally, the mentalizing impairment observed in MDD has also been attributed to connectivity defects across areas involved in ToM, although further studies are needed (Kupferberg et al., 2016a). Another possible consequence of the abovementioned impairments is the reduced empathy observed in MDD. Indeed, although MDD patients may show an even higher level of empathy, their affect-directed (i.e. emotion regulation), automatic causal interpretations of pain (i.e. mentalizing) in others are frequently disturbed, leading to non-conscious assertions of blame, which are usually placed on themselves (Kupferberg et al., 2016a). The neural correlates of empathy in MDD have not been investigated extensively so far, although preliminary data suggested an involvement of the right somatosensory-related cortices and the left-middle-anterior ACC.

The impairments in social functioning summarized above can also be dissected at the neurotransmitter level, although, as stated in the previous sections, the exact interplay and the fine modulation of the different neurotransmitter involved have still to be elucidated in detail. Starting from the impairment in social reward, in MDD an altered endogenous OP activity has been reported, consisting in a reduced activation of MOR in NAc both during social reward stimuli and when recovering from rejection compared to healthy controls (Hsu et al., 2013). However, it is not clear if social reward impairment is due to defects in the opioid system reactivity or, alternatively, if it reflects an impaired mPFC-ventral striatum connectivity (Kupferberg et al., 2016a). Clearly, also the 5-HT system plays a role in social reward dysfunctions in MDD. In detail, a 5-HT system deficiency (mainly in the dorsal raphe nucleus (Takahashi et al., 2012)) has been associated with a decreased DA transmission in the striatum (Navailles and De Deurwaerdere, 2011), leading to a decreased activation after social reward stimuli. Moreover, 5-HT deficiency was also associated to changes in connectivity between mPFC and amygdala, which increased the negative affective bias (Robinson et al., 2013). Data about the molecular basis of hypersensitivity to social rejection are even scarcer. Probably, the OXT system plays a primary role in the regulation of these behaviours, since altered OXT plasma levels were found in MDD patients after social exclusion (Jobst et al., 2015), but further studies are needed to better understand this association (for detail see (McQuaid et al., 2014)). Furthermore, genetic variants within 5-HTT have been associated with amygdala response magnitude to negative social stimuli (Hariri et al., 2002), suggesting an involvement of the 5-HT system also in rejection sensitivity.

In conclusion, even though more studies are needed to better understand the fine interactions across the neurotransmitter systems and how they were disrupted by MDD determining social dysfunctions/withdrawal, from the currently available data we can already infer that in MDD the depressive disorder itself alters many of the core structures which control social behaviour (Fig. 2). Interestingly, many of the dysfunctions we described do, at least partly, reverse after recovery, thus strengthening the view that MDD pathophysiology directly affects adaptive social behaviour mechanism. An impairment which, however, is variable across subjects suggesting that social behaviour control is, at least partly, independent from MDD but they are reciprocally influencing each other.

6. Discussion: social withdrawal as the outcome of social brain impairments

The complexity of the processes that underlie social living emerged clearly in the previous sections. Human brain shows several degrees of specialization for social stimuli processing, which are detectable from the neurotransmitter level to neural network pathways. We provided a global picture of the neural networks and of the neurotransmitters which sustain social functioning (i.e. the social brain), underlining the complex interactions across systems needed for a successful processing of social stimuli in accordance with environmental context and personal previous experiences. Unfortunately, such a high complexity may also be associated with a high susceptibility to several pathogenic stimuli.
Therefore, deficits in any of these processes can result in personal difficulties and interpersonal problems. Consequently, social dysfunctions are frequently observed in a number of neuropsychiatric disorders and often represent the first signs of these diseases. However, subtle social dysfunctions, such as deficits in emotion recognition process or in mentalizing, could be unrecognized in clinical settings (e.g., (Lee et al., 2015b; Pornigo et al., 2015)) until the full onset of the disorder or, more often, the appearance of a clear social dysfunction, such as social withdrawal (NICE, 2014). Social withdrawal in turn contributes to a further worsening of the disorder symptomatology as well as to further deficits in social cognition through social stimuli deprivation in a vicious circle (e.g., (El Haj et al., 2016; Tremeau et al., 2016; Zhong et al., 2017)). Clearly, social functioning as a whole is a complex phenotype (see Van der Wee et al. in this same issue), which is influenced by a variety of socio-demographic features, as well as by basic domain deficits, such as attention, working memory, and sensory processing impairments, as previously discussed. Nonetheless, we previously underlined how social dysfunctions are in some measure independent from other symptoms/deficits as well as from cognitive and even from social cognitive impairments. Therefore, the observed social dysfunctions likely reflect (at least partially) alterations in the social brain itself, which are somehow independent from the other consequences of the affecting disorder.

Following this hypothesis, we discussed how three different neuropsychiatric disorders (namely AD, SCZ, and MDD) all share a final common pathway that affects the social brain (Fig. 2), characterized by social dysfunctions largely similar across the three disorders, although with different degrees of impairment (e.g. mentalizing is greatly impaired in SCZ, while it shows subtle defects in MDD). Nonetheless, in all these three disorders, social dysfunctions often result in the final, deleterious, outcome of social withdrawal, suggesting a final (partially) converging pathway. For example, social withdrawal may be in some measure driven by a systematic de-valuation of social stimuli or positive interactive outcomes, such as cooperation, which are likely linked to the excitatory/inhibitory balance within cortical regions, such as the PFC (Bicks et al., 2015; Yizhar et al., 2011). Reductions in reward associations or increase in associated punishment (or a combination of the two) may contribute to the systematic de-valuation of social stimuli as well. Consistent with this hypothesis, the brain structures and neurotransmitters associated with social deficits are largely the same across the disorders, independent from the pathophysiological mechanism (Fig. 2). Interestingly, despite the fact that only in recent years advances in neuroimaging techniques allow to generate detailed maps of the large-scale structural and functional architecture of human brain networks (e.g., (Braun et al., 2018)), recent evidences suggested that these social dysfunctions reflect defects in the neural networks sustaining social functioning (Fig. 1) (Bickart et al., 2014a, b) rather than single region defects, as previously discussed. Also, the brain regions associated with the social dysfunctions are often part of the same neural networks affected across different neuropsychiatric disorders (Figs. 1 and 2), supporting the hypothesis that social dysfunction, at least partially, may be due to dysfunctions of specific transdiagnostic neural circuits. However, current neuroimaging techniques still fail to identify (patho-) physiological mechanisms that link system-level phenomena to the multiple hierarchies of brain function. Thus, a combined approach applying complementary techniques (e.g., fMRI and electrophysiological measures), as well as preclinical models, is likely needed to increase current knowledge about the neurobiological basis of social functioning/impairment. Clearly, to dissect a complex phenotype such as social dysfunctions (or a particular behavioural outcome such as social withdrawal), the assessment of as much as possible putative contributors is needed. As a matter of fact, we previously underlined how different factors may contribute to cause social dysfunction, and social withdrawal in particular. Nonetheless, not one of these factors can explain sufficiently the observed social dysfunction (i.e., they account only for a minor percentage of the explained variance of the examined behaviour (Bowie et al., 2006; Brekke et al., 2005; Couture et al., 2011; Sergi et al., 2006; Vauth et al., 2004)). However, some authors already tried to elucidate the different modulators of social functioning by applying a detailed assessment of different clinical and cognitive features and advanced statistical analyses (see for example (Bowie et al., 2006, 2008; Fett et al., 2011; Kalin et al., 2015; Pulcu and Elliott, 2015)). These studies focused almost exclusively on SCZ, while data is lacking (or very few) in other neuropsychiatric disorders such as AD and other dementia, as well as MDD. More importantly, taking into account the Research Domain Criteria (RDoc) project launched by the National Institute of Mental Health (NIMH) in the early 2009 (Cuthbert and Insel, 2013; Cuthbert, 2015), to the best of our knowledge there are no studies which investigated with a combined approach social functioning in different neuropsychiatric disorders independent from the diagnosis. However, taking into account the evidence above, social dysfunctions and their most evident expression (i.e. social withdrawal) may represent a transdiagnostic domain being a potentially independent entity in terms of biologic roots. Hence, a better knowledge of these biological factors could inform the development of novel, targeted treatments. We are, however, at the early stages of understanding how our brain allows for these complex processes, and perhaps more importantly, how they might go awry and should be normalized in certain psychiatric conditions.

The PRISM project ((Kas et al., 2017) and Bilderbeck et al. in this issue) investigates for the first time directly these substrates across different neuropsychiatric disorders, aiming to reach a better understanding of the neurobiological basis of social functioning in order to pave the way for new treatments for social dysfunctions. PRISM will focus on social withdrawal in SCZ, AD and MDD, and addresses the treatment needs of these three most prevalent brain disorders in Europe. The economic burden of these three disorders is huge, collectively estimated to cost €312 billion per year (Wittchen et al., 2011). Most of these costs are indirect (lack of productivity). Social withdrawal is an important source of indirect cost and has been identified as one of the main reasons for mental health related disability benefit claims (UK Department for Work and Pensions, 2013). Moreover, these diseases lie heavily on care givers and impact significantly on their ability to work (Haro et al., 2014). For those reasons, also improved clinical assessment tools to classify social withdrawal (and/or underlying domains) objectively have to be developed and validated. These steps are thought to be critical to provide relevant clinical endpoints for future intervention studies. Acquiring knowledge on the neurobiological substrate(s) underlying social withdrawal in SCZ, AD and/or MDD will also depend on the availability of homologous preclinical model systems that allow the functional validation of neural circuits and neurotransmitter systems that have currently been related to social withdrawal. Finally, paving new regulatory paths for transdiagnostic classification tools for social withdrawal requires early engagement of regulators, as well as patient and family organizations during the selection and evaluation of study designs, methods and instruments to facilitate future implementation of these tools.

**Funding**

The PRISM project (www.prism-project.eu) leading to this application has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115916. This Joint Undertaking receives support from the European Union’s Horizon 2020 research and innovation programme and European Federation of Pharmaceutical Industries and Associations (EFPIA).

**Disclaimer**

This publication reflects only the author’s views and neither the IMI 2 JU nor EFPIA nor the European Commission are liable for any use that may be made of the information contained therein.

Dabrowska, J., Hazra, R., Ahern, T.H., Guo, J.D., McDonald, A.J., Mascagni, F., Muller, S. Porcelli et al.


Neuroscience and Biobehavioral Reviews xxx (xxxx) xxxxxx

S. Porcelli et al.

Neuropsychopharmacology 18, 407–430.
Picture of a neuron network.
S. Porcelli et al.

Neuroscience and Biobehavioral Reviews xxx (xxxx) xxx-xxx


