Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Cognitive Neuroscience

Ciclo XXX

Settore Concorsuale di afferenza: 11/E1

Settore Scientifico disciplinare: M-PSI/02

Alexithymia as impairment in constructing the internal representation of emotional stimuli

Presentata da

Francesca Starita

Coordinatore Dottorato

Prof.ssa Monica Rubini

Supervisore

Prof.ssa Elisabetta Làdavas

Esame finale anno 2018

Contents

Abstract

Alexithymia is a personality trait characterized by difficulties in processing emotional stimuli. Here, Experiment 1 shows that alexithymia has a significant impact on the life of individuals, being related to lower emotional intelligence, empathy and wellbeing. Also, Experiment 2 shows that alexithymia is related to the need for more emotional intensity to identify fear in facial expressions. Although these experiments contribute to the literature describing the difficulties of alexithymia, the basic mechanisms underlying such difficulties remain poorly understood. For this reason, the remaining of this thesis focuses on investigating whether differences in emotional learning may characterize alexithymia. Indeed, through emotional learning the internal representation of stimuli is shaped, so that neutral stimuli acquire emotional value. Impairment in this process has been reported in clinical conditions marked by difficulties in emotion processing; nevertheless, this has never been investigated in alexithymia. Given this, Experiment 3 shows that alexithymia is related to impairment in learning the aversive value of stimuli, evidenced by reduced physiological markers of emotional prediction in Pavlovian threat conditioning. On the contrary, Experiment 4 did not find such evidence, when learning appetitive value. Nevertheless, evidence for impairment in learning the appetitive value of stimuli was found in Experiment 5, where electrophysiological markers of prediction and prediction error were assessed during Pavlovian reward conditioning. Finally, Experiment 6 examined the ability to learn the emotional value of actions during instrumental learning, and to use this learned value for adaptive behavior in a new context. Here, alexithymia was related to a difficulty in learning from punishment, marked by longer response time when having to avoid stimuli, which had previously acquired aversive value, encountered in a new context. Together, these results indicate impairment in emotional learning in alexithymia, which may be more severe for aversive than appetitive stimuli. The new insight provided by these results for the understanding of alexithymia is discussed.

Chapter 1. Alexithymia and its difficulties in emotion processing

Alexithymia: Definition, assessment and prevalence

The word alexithymia derives from the Greek *a* = lack*, lexis* = word*, thymos* = emotion. During the 1940s, the construct of alexithymia originated in the clinical context to describe patients, who appeared to have limited awareness of their emotions, to have difficulties in expressing them and to be less focused on their internal world and more preoccupied with physical experiences and details of the external environment (Panaite & Bylsma, 2012). These characteristics are still part of the current definition of alexithymia, which is now considered as a personality trait defined by three aspects: difficulty in identifying feelings, difficulty in describing them to others and a type of thinking focused more on concrete aspects regarding the external environment rather than on introspection on the internal mental life (Sifneos, 1973; Taylor, Bagby, & Parker, 1991). In agreement with this, the most widely used measure to assess alexithymia is the Toronto Alexithymia Scale, a questionnaire with 20 items, answered on a 4 point Likert scale, comprising three subscales corresponding to its three aspects. Importantly, higher score on the measure indicates more difficulties in emotion processing and a cut-off of 61and above has been identified as indicating presence of alexithymia (Bagby, Parker, & Taylor, 1994).

Studies on the general population of different countries suggest a prevalence of the trait ranging from 7% to 18% (Honkalampi, Hintikka, Tanskanen, Lehtonen, & Viinamäki, 2000; Kokkonen et al., 2001; Mason, Tyson, Jones, & Potts, 2005; Salminen, Saarijärvi, Äärelä, Toikka, & Kauhanen, 1999), which increases significantly when examining clinical populations, such as individuals with anxiety (Berthoz, Consoli, Perez-Diaz, & Jouvent, 1999), depression (Hendryx, Haviland, & Shaw, 1991; Honkalampi et al., 2000; S. Li, Zhang, Guo, & Zhang, 2015; Marchesi, Brusamonti, & Maggini, 2000), eating disorders and

substance abuse (Panaite & Bylsma, 2012), leading to the hypothesis that alexithymia might represent a subclinical condition that poses individuals at risk for developing such pathologies. Therefore, given the impact of alexithymia on the wellbeing of the general population, understanding the difficulties in emotion processing that characterize this personality trait and the underling mechanisms, which may give rise to such difficulties, appears of significant importance.

Alexithymia affects multiple aspects of emotion processing

Individuals with alexithymia appear to have difficulties in several domains of emotion processing spanning from the recognition of emotion, to the response to emotional stimuli, the regulation of such response and the appropriate use of emotions to guide decision making.

Extensive research has shown that alexithymia affects the ability to recognize other's emotions. For example, alexithymia appears related to worse performance in recognition of emotional facial expressions (EFEs) (Jessimer & Markham, 1997; Lane et al., 1996), specifically under temporal constraints (Ihme, Sacher, Lichev, Rosenberg, Kugel, Rufer, Grabe, Pampel, Lepsien, Kersting, Villringer, Lane, et al., 2014; Swart, Kortekaas, & Aleman, 2009), which can be restored when stimulus exposure time is extended (Ihme, Sacher, Lichev, Rosenberg, Kugel, Rufer, Grabe, Pampel, Lepsien, Kersting, Villringer, & Suslow, 2014; Pandey & Mandal, 1997; P. D. Parker, Prkachin, & Prkachin, 2005). This suggests that alexithymia is associated to the need for more time to accurately recognize EFEs (Grynberg et al., 2012). Also, alexithymia has been associated with a failure in the modulation of early visual event related potentials (ERPs), recorded with electroencephalography, in response to emotional body postures (Borhani, Borgomaneri, Làdavas, & Bertini, 2016) as well as failure in the modulation of ERPs related to error monitoring in response to EFEs (Maier, Scarpazza, Starita, Filogamo, & Làdavas, 2016). In

addition to visual perception, difficulties have been found in the perception of emotions conveyed through the auditory modality. In fact, alexithymia is associated to decreased performance in the identification of emotional prosody (Goerlich, Aleman, & Martens, 2012) and differences in ERP components related to the processing of emotional prosody (Goerlich et al., 2012, 2012; Goerlich-Dobre et al., 2014) and emotional sounds more in general (Schäfer, Schneider, Tress, & Franz, 2007).

Concerning the response to emotional stimuli, physiological arousal to emotional stimuli has been examined. Although some studies report no difference in autonomic response in alexithymia (Bausch et al., 2011; Easterbrooks, Chaudhuri, & Gestsdottir, 2005; Stone & Nielson, 2001), others found decreased physiological response to emotional stimuli (Bermond, Bierman, Cladder, Moormann, & Vorst, 2010; Franz, Schaefer, & Schneider, 2003; Neumann, Sollers III, Thayer, & Waldstein, 2004; Newton & Contrada, 1994; Pollatos, Schubö, Herbert, Matthias, & Schandry, 2008). In addition, when the information from the physiological arousal is integrated with the subjective reports of emotional response, alexithymia appears to be characterized by a decoupling between the subjective experience and physiological reactivity (Eastabrook, Lanteigne, & Hollenstein, 2013), with subjective reports of greater negative emotional experience relative to physiological arousal (Connelly & Denney, 2007; Eastabrook et al., 2013; Newton & Contrada, 1994; Pollatos et al., 2011). Turning to the response to others' emotions, it is possible that the impairment in emotion perception may affect empathic response. Indeed, alexithymia appears negatively correlated with empathy (Grynberg, Luminet, Corneille, Grèzes, & Berthoz, 2010; Patil & Silani, 2014b), with individuals with alexithymia exhibiting significantly lower scores on empathic concern and the ability to adopt the perspective of the other when compared to those without alexithymia, while having higher scores on the distress experienced during emotionally intense situations (Moriguchi et al., 2006, 2007).

Individuals with alexithymia have also difficulties in regulating their response to emotional stimuli, being more likely to recur to maladaptive cognitive strategies, such as repression, as opposed to adaptive ones, such as reappraisal (Swart et al., 2009). In addition, when specifically asked to implement reappraisal, they fail to report a decrease in emotional arousal and to show a corresponding modulation of brain ERPs, which is instead observed in participants without alexithymia (Pollatos & Gramann, 2012).

Finally, the difficulties of alexithymic individuals in emotion processing affect also decision making. Indeed, emotions play a crucial role in decision making, where somatic changes occurring when pondering a decision guide our behavior to make adaptive choices (Damasio, 2008) and the differences in physiological response to emotional stimuli in alexithymia, might also impair decision making. Specifically, individuals with alexithymia rate utilitarian decisions as more acceptable in moral dilemmas (Patil & Silani, 2014b, 2014a), they fail to consolidate learning about advantageous and disadvantageous choices in gambling tasks (Ferguson et al., 2009) and have stronger preference to choose small rewards immediately available over larger rewards that require a wait time to be obtained (Scarpazza, Sellitto, & di Pellegrino, 2017) compared to individuals without alexithymia.

Neural correlates of alexithymia

To better characterize alexithymia, research has also investigated the brain regions that correlate with the alexithymic trait, when processing emotional stimuli in a broad range of tasks. In particular, a recent meta-analysis has suggested a central role of the amygdala, the insula and the anterior cingulate cortex among the regions underlying difficulties in alexithymia (van der Velde et al., 2013).

The amygdala is an almond shape structure comprising several sub-nuclei and it is located in the medial part of the temporal lobe (LeDoux, 2007). This structure has long been

known to play a crucial role in processing emotional stimuli. Indeed, lesion to the amygdala impairs accurate perception of emotional facial expressions (Adolphs, 2002; Adolphs et al., 1999; Adolphs, Tranel, Damasio, & Damasio, 1994) and memory for emotional events (Adolphs, Denburg, & Tranel, 2001). In alexithymia, several studies have reported decreased activation of the amygdala in response to the presentation of emotional facial expressions (Jongen et al., 2014; Kugel et al., 2008), negative emotional stimuli (Moriguchi & Komaki, 2013; van der Velde et al., 2013; Wingbermühle, Theunissen, Verhoeven, Kessels, & Egger, 2012) and in particular fearful ones (Pouga, Berthoz, de Gelder, & Grèzes, 2010), supporting its role as a structure underlying the difficulties of alexithymic individuals.

The insula is a cortical 'island' located within the lateral sulcus and appears to have a central role in interoception and interoceptive awareness (Craig, 2002, 2003; Critchley, Wiens, Rotshtein, Öhman, & Dolan, 2004). Because of this, it has been argued that the insula supports the representation of physical sensations as affective states (Barrett, 2017b). As such, the insula receives and relays information regarding the internal state of the body (Craig, 2002, 2003) and has found to be responsive in tasks involving viewing others in pain (Bernhardt & Singer, 2012) or disgust (Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012). Given the tendency of individuals with alexithymia to fail to distinguish between bodily arousal and emotional states (Sifneos, 1973), it is not surprising that differences in activity of this region have been correlated with the personality trait. In particular, insular activation appeared to decrease with increasing alexithymia when viewing emotional facial expressions (Reker et al., 2010), while it appeared to increase with increasing alexithymia when viewing others in pain (Bird et al., 2010; Moriguchi et al., 2007). Together these studies support a difference in insular activity in alexithymia, which may vary in direction depending on the experimental conditions.

Finally, the anterior cingulate cortex, located in the medial frontal lobe, appears involved in awareness of core affective state (Lindquist et al., 2012). Although, neuroimaging studies have found both increased (Mériau et al., 2006; Pouga et al., 2010) and decreased (Moriguchi et al., 2007) activation of this region in alexithymia, a meta-analysis suggested that alexithymia may be indeed related to increased activation in this region, reflecting increased cognitive demand to process emotional stimuli (van der Velde et al., 2013). This result would be in keeping with a theoretical account of alexithymia, the so called 'blind-feel' hypothesis, which claims that , alexithymia would be characterized by an intact physiological response and a deficit in emotion concept representation (Lane, Ahern, Schwartz, & Kaszniak, 1997; Lane, Weihs, Herring, Hishaw, & Smith, 2015).

The current issue: from describing alexithymia to understanding alexithymia

Although a large body of research is accumulating showing that alexithymia is related to difficulties in various domains of emotion processing, less attention has been devoted to understanding the basic mechanisms that may underlie such difficulties.

In the wake of the theory of embodied emotion, which argues that perceiving and thinking about emotion involve perceptual, somatovisceral, and motor re-experiencing of the emotion in the self (Niedenthal, 2007; Niedenthal, Winkielman, Mondillon, & Vermeulen, 2009), studies have suggested that alexithymia may be the manifestation of underlying differences in the embodied aspects of emotions. Indeed, the studies investigating the physiological response to emotional stimuli have shown partly impaired interoceptive response to emotional stimuli (Bermond et al., 2010; Franz et al., 2003; Neumann et al., 2004; Newton & Contrada, 1994; Pollatos et al., 2008). In addition, the difficulties in emotion perception has been suggested to result from differences in emotion mimicry, i.e. the activation of the somatosensory system, when viewing another's emotional expression, which

usually simulates the other's emotions in ourselves (Adolphs, 2002). Several studies found impairment in such response alexithymia, which appears to be decreased for fear and enhanced for disgust (Scarpazza, di Pellegrino, & Làdavas, 2014; Scarpazza, Làdavas, & Cattaneo, 2017; Scarpazza, Làdavas, & Pellegrino, 2015). Instead, other studies argue that alexithymia may be related to a broader impairment in interoception, because the ability to perceive physiological signals from the body appears decreased also in non-emotional context (Brewer, Cook, & Bird, 2016; Murphy, Catmur, & Bird, 2017; Shah, Hall, Catmur, & Bird, 2016).

Indeed, differences in emotion embodiment may be one process underlying the difficulties in alexithymia. However, these differences have been investigated with stimuli, such as emotional facial expressions, which elicit an emotional response, not because of their intrinsic properties, but because individuals usually learn to attribute them emotional value (Barrett, 2017a), by actively constructing the internal representation of such stimuli through a process named emotional learning (LeDoux, 1998). Given this, a further step deeper in the understanding of alexithymia can be taken, by investigating whether alexithymia is characterized by differences in the construction of the internal representation of emotional stimuli.

Chapter 2. Understanding the construction of the internal representation of emotional stimuli to understand alexithymia

Emotional stimuli can be defined as stimuli that have relevance for the survival of the organism, such as a predator or food. These stimuli are named unconditioned stimuli (UCS) because they are biologically prepared to elicit an emotional response (LeDoux, 1998), which is marked by interoceptive (e.g. increase arousal), motor (e.g. avoid the threat) and cognitive (e.g. subjective feeling) changes in the organism, occurring in order to promote its survival (Brosch, Pourtois, & Sander, 2010). Nevertheless, in our daily life, countless stimuli can elicit an emotional response, from seeing the frowning expression of a friend to receiving a 'like' for the latest picture we posted on social media. This is because, to increase the chances of survival, we usually expand the range of emotional stimuli, from those that unconditionally elicit an emotional response to those that are often associated with them, actively constructing the internal representation of emotional stimuli. This process is called associative emotional learning.

Associative emotional learning

In associative emotional learning a neutral stimulus or behavior is repeatedly associated to an UCS, such as an aversive or appetitive stimulus. As a consequence, the emotional response that is first observed in reaction to the presentation of the appetitive or aversive stimulus is then transferred on the neutral stimulus or behavior (e.g. Seymour et al., 2004), which is now named conditioned stimulus (CS) or behavior (Pavlov & Anrep, 1927). Ultimately, the internal representation of the neutral stimulus or behavior is changed acquiring emotional value.

Associative emotional learning is a fundamental adaptive mechanism because the CS can be used to predict coming UCS, enabling the organism to appropriately prepare to

respond to the predicted UCS, rather than to simply react to these, once they have occurred (McNally & Westbrook, 2006; Öhman & Mineka, 2001). Nevertheless, for such mechanism to be adaptive, the internal representation of the stimuli need to be accurate and the associated predictions precise. To ensure this, the representation of the stimuli also acts as internal model against which to compare new incoming information. Every time an UCS is predicted, this prediction is compared against the incoming information and, if there is a discrepancy between the two, a prediction error is computed, which drives learning by indicating that the internal representation needs to be updated (Friston, 2010). Given this predictive properties of the internal representation, constructing accurate internal representation of emotional stimuli is not only crucial for effective recognition, response and response regulation to the emotional stimuli per se, but also for anticipating the consequences of these stimuli enabling optimal decision making (Bubic et al., 2010). In sum, accurate predictions and prediction errors during emotional learning are the basis to construct accurate internal representations of emotional stimuli.

Among the various forms of associative emotional learning, two are particularly relevant for this dissertation. The first is Pavlovian conditioning, in which the organism learns contingencies between two stimuli, i.e. a neutral stimulus and a UCS. Here, learning is generally assessed by physiological changes in autonomic nervous system activity, such as in skin conductance response (SCR) (LeDoux, 2014; Phelps & LeDoux, 2005). Before learning, an increase in SCR is observed in response to the UCS. Following learning the sole presentation of the neutral stimulus elicits the increase in SCR, signaling increased expectations regarding the occurrence of UCS following the presentation of the neutral stimulus, which has now become a CS (S. S. Y. Li & McNally, 2014). The second form of associative learning is instrumental learning, in which the organism learns contingencies between a behavior and a UCS. Here, learning is generally assessed by the way in which the

UCS affects behavior. Specifically, a behavior associated to an appetitive UCS should be repeated, while one associated to an aversive UCS should be terminated (Daw & Tobler, 2014).

In addition to changes in physiological response during Pavlovian conditioning and behavior during instrumental learning, changes in subjective experience can also mark emotional learning. According to recent cognitive theories of emotional experience, the subjective experience of emotion is a cognitive construction based on conceptualization of situations (Barrett, 2017a, 2017b; Barrett, Mesquita, Ochsner, & Gross, 2007) or a higher-order cognitive interpretation that emerges from the integration in working memory of lower-order information, coming from within the body and the external environment (LeDoux, 1998; LeDoux & Brown, 2017; LeDoux & Hofmann, 2018). For example, an aversive stimulus, such as physical pain, triggers a defensive response, such as flight response, and interoceptive changes, such as increase in physiological arousal, release of cortisol etc. Nevertheless, the response, the stimulus that caused it and the context require cognitive interpretation to generate an instance of fear. Importantly, the lower-order information is generally used to give rise to a coherent higher-order subjective emotional experience; nevertheless, because of the multiple sources contributing to the lower-order information, the information coming from any individual source may not find direct correspondence in the higher-order subjective emotional experience (LeDoux & Brown, 2017). For example, patients with split brain (Làdavas, Cimatti, Pesce, & Tuozzi, 1993), hemispatial neglect (Tamietto et al., 2015) and affective blindsight (Bertini, Cecere, & Làdavas, 2013; Cecere, Bertini, Maier, & Làdavas, 2014; Tamietto et al., 2009) show intact physiological response to emotional stimuli, despite the absence of awareness for them, but they do not show any fear in their subjective emotional experience.

Neural correlates of associative emotional learning

Several brain regions underlie emotional learning, including the insula, amygdala, anterior cingulate cortex and the striatum. Importantly, although the role of each region has been studied mainly in relation to distinct aspects of learning, they all have reciprocal direct and indirect neural connections, likely constituting a complex intertwined circuit in which activity of each region can be modulated by the others.

The insula and the representation of interoceptive changes associated with unconditioned stimuli

The insula receives information regarding bodily changes elicited by a broad range of appetitive stimuli, such as sensual touch, and aversive stimuli, such as pain (Craig, 2002, 2003; Critchley et al., 2004). In addition, activation of the insula has been found both during the anticipation of aversive (Büchel, Dolan, Armony, & Friston, 1999) and appetitive (Kirsch et al., 2003) stimuli during pavlovian conditioning, which appeared modulated by awareness of the conditioned stimuli, suggesting a role for this structure also in the explicit awareness of an external emotional event (Critchley, Mathias, & Dolan, 2002). So the insula might be involved in processing and predicting interoceptive changes associated with the occurrence of UCS.

Interestingly, a study found stronger activation of the insula during instructed than pavlovian conditioning, where participants learn contingencies between conditioned and unconditioned stimuli through verbal instruction instead of direct experience (Phelps et al., 2001). Authors argued that this difference in activation may suggest a more important role for the insula in the cognitive representation of the aversive properties of UCS and the associated emotional feeling, which may then be relayed to the amygdala. Indeed, the insula has been

suggested as crucial for the representation of physical sensations as affective states (Barrett, 2017b) and the conscious subjective experience of emotion (LeDoux & Brown, 2017).

The amygdala and the acquisition and expression of the association between conditioned and unconditioned stimuli

The amygdala plays a pivotal role in emotional learning, both in pavlovian conditioning (Cardinal, Parkinson, Hall, & Everitt, 2002) and instrumental learning (Averbeck & Costa, 2017), being involved in the acquisition of the association between the neutral stimulus or behavior and appetitive (Bechara, Damasio, Damasio, & Lee, 1999; Murray & Baxter, 2002; Seymour & Dolan, 2008) as well as aversive unconditioned stimuli (LaBar, Gatenby, Gore, LeDoux, & Phelps, 1998; Phelps & LeDoux, 2005). Contrary to the insula, its activation appears independent of awareness of the CS (Critchley et al., 2002), suggesting its importance in the implicit processing of salient stimuli (LeDoux, 2007).

In particular, its role has mostly being studied in Pavlovian conditioning. Here the basolateral complex of the amygdala appears to be the site of association between CS and UCS receiving sensory inputs from multiple modalities and responding to conditioned stimuli associated to aversive as well as appetitive stimuli (Paton, Belova, Morrison, & Salzman, 2006). Its projections then influence the central complex of the amygdala, which seems to modulate physiological changes in the autonomic nervous system (e.g. skin conductance response) through its connections with hypothalamus (Cardinal et al., 2002; McNally & Westbrook, 2006). In addition, its projections to the ventral striatum and prefrontal cortex are thought to influence behavioral responses of approach and avoidance and decision making (Cardinal et al., 2002).

Importantly, single cell recording in animals has shown that some neurons in the basolateral complex of the amygdala may process the valence of UCS. Indeed, while some

neurons respond to both appetitive and aversive UCS, others respond preferentially to one valence over the other. Specifically, neurons projecting to the central complex of the amygdala seem to respond preferentially to aversive stimuli, while those projecting to the nucleus accumbens to appetitive stimuli (Beyeler et al., 2016; Namburi et al., 2015). This evidence suggests that although the amygdala is involved in learning from positive as well as negative emotional stimuli, these may be processed by partly distinct neural circuits. For example, while human lesion studies provide evidence of a causal role of the amygdala in the acquisition of aversive conditioning measured through SCR (Bechara et al., 1995; LaBar, LeDoux, Spencer, & Phelps, 1995), the causal role of the amygdala in the acquisition of reward conditioning is less definitive (Murray & Baxter, 2002; though see Bechara et al., 1999).

The anterior cingulate cortex and the computation of prediction errors

Once the association between CS and UCS has been acquired, the CS becomes predictor of the coming UCS enabling the organism to prepare to respond to it. In this regard, the anterior cingulate cortex (ACC) has been proposed to represent predictions of future events, signaling prediction errors in response to the unexpected non-occurrence of predicted events (Alexander & Brown, 2014). When a predicted event does not occur, a prediction error signals the need to update the prediction and connections between ACC and amygdala can modulate changes in the acquisition of CS-UCS association.

Neuroimaging studies have shown activation of this area during anticipation of pain in aversive conditioning (Büchel, Morris, Dolan, & Friston, 1998; Etkin, Egner, & Kalisch, 2011; Milad et al., 2007) and of monetary reward during appetitive conditioning (Kirsch et al., 2003). In addition, electrophysiological studies provide evidence that activity originating from this area is related to brain components (Nieuwenhuis, Holroyd, Mol, & Coles, 2004),

which appear to code prediction error signals, being characterized by a negative deflection in electrical potential that is enhanced following the violation of predictions (Garofalo, Maier, & di Pellegrino, 2014; Garofalo, Timmermann, Battaglia, Maier, & di Pellegrino, 2016; Gehring & Willoughby, 2002; Hajcak, Moser, Holroyd, & Simons, 2006; Holroyd & Krigolson, 2007; Holroyd, Nieuwenhuis, Yeung, & Cohen, 2003; Talmi, Atkinson, & El-Deredy, 2013).

The striatum and the expression of motivated behavior

The emotional value acquired by the CS can influence behavior. If the CS has acquired appetitive value, the organism will engage in approaching behavior to collect the reward, on the contrary it will engage in avoidance behavior to escape the danger, in case the CS had acquired aversive value. The striatum has been suggested to mediate such motivational behavior (Cardinal et al., 2002) and for this reason its role has been largely investigated in instrumental learning.

While extensive neuroimaging studies suggest that the striatum mediates reward seeking behaviors (Delgado, 2007; Knutson & Cooper, 2005; O'Doherty, 2004), there is accumulating evidence that it is also fundamental for the anticipation of pain (Seymour et al., 2004) and threat avoidance behaviors (Delgado et al., 2009; Reynolds & Berridge, 2008). In particular, prediction error signals, carried by dopaminergic neurons originating in the substantia nigra (Schultz, 1998, 2016), would converge onto the striatum to drive learning. In the dorsal striatum, prediction errors would support acquisition of stimulus-response-reward contingencies, promoting the repetition of actions that are more likely to result in a reward; in the ventral striatum they would update stimulus-reward association more broadly (O'Doherty et al., 2004). In addition, similarly to neurons in the amygdala, single cell recording has shown that largely distinct populations of neurons in the nucleus accumbens, part of the

ventral striatum, encode stimuli that have acquired either aversive or rewarding value, suggesting also in this region partly separate encoding of stimulus valence (Namburi, Al-Hasani, Calhoon, Bruchas, & Tye, 2016).

Rethinking alexithymia: alexithymia as impairment in constructing the internal representation of emotional stimuli

Impairments in emotional learning have been highlighted in a number of psychiatric conditions related to difficulties in emotion processing (Heinz, Schlagenhauf, Beck, & Wackerhagen, 2016), such as depression (Greenberg et al., 2015), anxiety (Lissek et al., 2005, 2014) and psychopathy (Rothemund et al., 2012), suggesting the crucial role of constructing accurate internal representations of emotional stimuli for mental health. Despite this evidence, the relation between emotional learning and subclinical difficulties in emotion processing, and in particular alexithymia, remains largely unexplored.

However, it might be wondered whether alexithymia is also characterized by impairments in emotional learning and more broadly in constructing the accurate internal representation of emotional stimuli. Indeed, because accurate internal representation of emotional stimuli is crucial for recognition, response and response regulation to emotional stimuli and decision making (Bubic et al., 2010), investigating this aspect could provide new insight in the understanding of alexithymia and the mechanisms underlying its difficulties in emotion processing. In addition, given the relation between alexithymia and psychiatric disorders (Panaite & Bylsma, 2012), such investigation could also have clinical implications for the understanding of the relationship between this personality trait and the clinical conditions found to be associated with it.

Chapter 3. Describing alexithymia: difficulties in emotion perception, emotional intelligence, empathy and its impact on psychological wellbeing

Introduction

Alexithymia is a personality trait characterized by difficulty in identifying feelings, difficulty in describing them to others and a type of thinking focused more on concrete aspects regarding the external environment rather than on introspection on the internal mental life (Sifneos, 1973; Taylor et al., 1991). As reviewed in Chapter 1, alexithymia is related to a broad range of difficulties in processing emotional information, which have been described by previous research. In this chapter, two experiments will be presented that extend the current literature on the topic.

Experiment 1 will show the impact of alexithymia on different aspects of the emotional life of individuals. Although some studies have previously investigated this issue, they have never examined a broad range of variables simultaneously in one sample of homogeneous demographic and socio-cultural background. Hence, limiting the conclusion that alexithymia can indeed have a widespread impact on individuals' emotional life. To address this limitation, a series of questionnaires was completed by a group of participants to understand how alexithymia predicts difficulties in emotional intelligence, in empathic response to others' emotions and on psychological wellbeing, focusing on anxiety and depression.

Experiment 2 will focus on emotion perception and the identification of emotional facial expressions (EFEs), in particular. The previous literature has shown that alexithymia is related to the need for more time to accurately recognize intense EFEs (Grynberg et al., 2012) Nevertheless, in everyday life, EFEs are expressed at varying degree of intensity. For this

reason, it seems important to also investigate the impact of emotional intensity in the relationship between alexithymia and EFE identification. Indeed, alexithymia may be related to the need for more emotional intensity to accurately identify EFEs.

Experiment 1: Increasing alexithymia predicts decreasing emotional intelligence, empathy and psychological well being

Introduction

Previous research has found that alexithymia affects various aspects of the emotional life of individuals in the general population, from the ability to manipulate emotional information, to the response to others' emotions, to individuals' psychological wellbeing. Regarding the manipulation of emotional information, the correlation between alexithymia and emotional intelligence (EI) has been investigated. EI is defined as a set of cognitive abilities regarding emotion processing, namely the ability to perceive, understand, use and manage emotions in the self and others (Mayer, Caruso, & Salovey, 1999; Salovey & Mayer, 1990). Previous studies found moderate negative correlation between alexithymia and emotional intelligence (EI) when using both self-report (Austin, Saklofske, & Egan, 2005; Baughman et al., 2011; Grieve & Mahar, 2010; Möira Mikolajczak, Luminet, Leroy, & Roy, 2007; Moïra Mikolajczak, Luminet, & Menil, 2006; J. D. A. Parker, Taylor, & Bagby, 2001; Saklofske, Austin, & Minski, 2003; Webb & McMurran, 2008) and performance measures of EI (Curci, Lanciano, Soleti, Zammuner, & Salovey, 2013; Lumley, Gustavson, Partridge, & Labouvie-Vief, 2005), suggesting that increasing levels of alexithymia are related to decreasing EI.

Regarding the effect of alexithymia on the response to others' emotions, the relation between alexithymia and empathy can be explored, as being able to empathize with others is crucial to understand their feelings and respond to these appropriately (Decety, 2015). In fact, empathy can be conceptualized as an affective state equivalent to the affective state of another person and elicited by imagining or observing the other's affective state, while being aware that our affective state is indeed elicited by the other person's affective state (Bernhardt & Singer, 2012; Singer & Lamm, 2009). Alexithymia has been found to be

negatively correlated to empathy (Grynberg et al., 2010; Patil & Silani, 2014b), with individuals with alexithymia exhibiting significantly lower scores on empathic concern and the ability to adopt the perspective of the other when compared to those without alexithymia, while having higher scores on the distress experienced during emotionally intense situations (Moriguchi et al., 2006, 2007).

Finally, regarding the way in which alexithymia affects psychological wellbeing, two aspects that have been investigated are depression and anxiety. In particular, previous research has found a positive correlation of alexithymia both with depression (Honkalampi et al., 2000; S. Li et al., 2015; Lipsanen, Saarijärvi, & Lauerma, 2004; Mattila, Salminen, Nummi, & Joukamaa, 2006) and anxiety (Berthoz et al., 1999; Marchesi et al., 2000), indicating that as alexithymia increases psychological wellbeing decreases.

One limitation of this literature is that the effect of alexithymia on these different aspects of emotion processing has been investigated in separate populations that differed on demographic and socio-cultural aspects, limiting the conclusion that alexithymia can indeed have a widespread impact on individuals' emotional life. In addition, previous studies have not consistently looked at the contribution of the three different factors of alexithymia in predicting the emotion processing difficulties. For these reasons, the aim of the current experiment was to investigate the effect of alexithymia and its three underlying factors (i.e. difficulty in identifying feelings, difficulty in describing feelings, externally oriented thinking) in predicting EI, empathy, depression and anxiety, in a single sample of young individuals. Based on previous literature, increasing alexithymia was hypothesized to predict decreasing EI, empathy, while increasing depression and anxiety.

Methods

Participants and procedure

A total of 137 participants were recruited for the study. Before completion of the study, participants were informed about the procedure and asked to complete a battery of questionnaires. Five participants were excluded from analysis because they reported having psychiatric or neurological conditions. The final sample included 132 participants (37 males, age: M=23.79, SD=.18 years).

All participants had equivalent educational background and were students at the University of Bologna and gave informed consent to participation. The study was designed and conducted in accordance with the ethical principles of the World Medical Association Declaration of Helsinki and the institutional guidelines of the University of Bologna and was approved by the Ethics Committee of the Department of Psychology.

Independent measure

Toronto Alexithymia Scale (TAS-20; Bagby et al., 1994). It is one of the most commonly used measures of alexithymia. It is a self-report scale, which comprises 20 items rated on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). The scale provides a total score for general alexithymia, which results from the sum of the scores obtained on each question. It also has three subscales: difficulty describing feelings, difficulty identifying feelings and externally-oriented thinking, which measures the tendency of individuals to focus on external events rather than their internal world. It also provides cut-off scoring: total ≤ 51 indicates non-alexithymia, total ≥ 61 indicates alexithymia, $51 <$ total < 61 indicates possible alexithymia.

Dependent measures

Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT; Mayer, Salovey, Caruso, & Sitarenios, 2003). Developed from an intelligence-testing tradition, it is the only available ability-based performance test to measure EI. The test requires participants to solve various emotional tasks and problems and includes 141 items, divided among 8 tasks. These are grouped by two, in order to measure the four abilities of EI: Perceiving Emotions refers to the ability to perceive emotions in oneself and others as well as in objects, art, stories, music; Using Emotions is the ability to generate, use, and feel emotion as necessary to communicate feelings or employ them in other cognitive processes; Understanding Emotions refers to the ability to understand emotional information, to understand how emotions combine and progress through relationship transitions, and to appreciate such emotional meanings; Managing Emotions is the ability to be open to feelings and modulate them in oneself and others to promote personal understanding and growth. So, the test provides a total score and four abilities scores. Standard score based on expert scoring (M=100, SD=15) was calculated for each participant.

Schutte Self-Report Emotional Intelligence Test (Schutte et al., 1998). Based in the ability model of EI by Salovey and Mayer (1990), it is a self-report questionnaire, which includes 33-items to be answered on a 1 (strongly agree) to 5 (strongly disagree) point Likert scale. The questionnaire provides a total score for general EI resulting from the sum of the score obtained on each question, with higher score indicating greater emotional intelligence.

Interpersonal Reactivity Index (IRI; Davis, 1980). It is a self-report questionnaire to measure empathy consisting of 28-items answered on a 5-point Likert scale ranging from 1 (does not describe me well) to 5 (describes me very well). The measure does not provide a total score but has 4 subscales, each made up of 7 different items, and the total score on each

subscales results from the sum of the score obtained on each question. Two subscales refer to the cognitive dimension of empathy, namely perspective taking assesses the tendency to spontaneously adopt the point of view of others and fantasy assesses the tendency to transpose imaginatively into the feelings and actions of fictitious characters in books, movies and plays. The other two subscales refer to the emotional dimension of empathy, namely empathic concern assesses “other-oriented” feelings of sympathy and concern and personal distress assesses "self-oriented" feelings of personal anxiety and unease during tense interpersonal settings.

Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). It is a self-report questionnaire including twenty-one questions assessing symptoms of depression, in addition to a question regarding eating habits, which is not included in the total score. Each question has four possible answers of increasing symptom intensity ranging from 0 to 3. Participants answer questions choosing one of the options, considering the way they felt in the previous couple of weeks. The total score results from summing the score obtained on each question and indicates the severity of symptoms: below 4 indicates possible negation of depression, 5-9 minimal depression, 10-18 mild depression, 19-29 moderate depression, 30-63 severe depression.

State Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). The trait portion of form Y of the STAI was used to assess symptoms of trait anxiety. This is a self-report questionnaire, which comprises 20 items rated on a 4-point Likert scale ranging from ‘Almost never’ to ‘Almost always’. The total score results from the sum of the score obtained on each question and indicates the severity of symptoms, with higher score indicating greater anxiety.

Statistical analyses

A series of linear regressions and multivariate analyses of variance were conducted to investigate how alexithymia predicted scoring on the dependent measures. First, total scores were used. Then, following statistically significant results, the subscales of the measures were used to better understand the scales that contributed to the observed effects. To reduce the possibility of type 1 error due to multiple comparisons, Bonferroni corrected p-level was considered as significance threshold (i.e. p=.005).

Results

Descriptive statistics

Table 1 reports the descriptive statistics for the sample. Note that, on average, the sample fell in the 'non-alexithymia' classification, had minimal/mild symptoms of depression, and average EI level. Also, only 127 participants completed the IRI, so the analyses on this measure will include only those participants.

Table 1 **Descriptive data**

	N	Mean	Std. Deviation	Minimum	Maximum
Toronto Alexithymia Scale	132	44.06	12.95	22.00	73.00
Difficulty in describing feelings	132	12.92	4.96	5.00	24.00
Difficulty in identifying feelings	132	15.19	5.94	7.00	34.00
Externally oriented thinking	132	15.95	4.43	8.00	28.00
Mayer Salovey Caruso Emotional Intelligence Test	132	99.53	11.50	66.90	136.59
Perceiving Emotions	132	113.67	15.32	81.47	145.88
Using Emotions	132	97.38	14.04	69.27	129.70
Understanding Emotions	132	95.49	10.45	69.96	122.14
Managing Emotions	132	92.01	11.14	67.76	124.18
Schutte Self-Report Emotional Intelligence Test	132	122.74	15.32	81.00	157.00
Interpersonal reactivity index					
Fantasy	127	17.71	5.09	5.00	28.00
Personal distress	127	11.92	4.85	3.00	27.00
Empathic concern	127	18.83	3.79	8.00	28.00
Perspective taking	127	18.64	4.44	7.00	28.00
Beck depression inventory	132	9.42	6.18	1.00	42.00
Trait Anxiety (State Trait Anxiety Inventory)	132	43.50	9.70	23.00	69.00

Alexithymia negatively predicts emotional intelligence

Performance EI. A simple linear regression showed that alexithymia was a significant predictor of performance EI (R^2=.31, F(1,130)=13.62, p=.000) and that performance EI decreased with increasing alexithymia (unstandardised β=-.27, t(131)=-3.69, p=.000).

A multivariate analysis of variance (MANOVA) was then conducted to investigate which of the alexithymia subscales predicted performance EI. Although the overall model with the three alexithymia subscales was significant in predicting performance EI (F(4, 125)=5.79, p=.000, partial η^2 = .16; all other ps≥.036), only of externally oriented thinking made a significant contribution to the model (F(1, 125) =21.05, p=.000, partial η^2 = .14; all other

ps≥.022). The ability to manage emotions decreased with increasing externally oriented thinking (unstandardised β =-1.09).

Self-reported EI. A simple linear regression showed that alexithymia was a significant predictor of self-reported EI (R^2=.43, F(1,130)=100.18, p=.000) and that self-reported EI decreased with increasing alexithymia (unstandardised β=-.66, t(131)=-10.01, p=.000).

A multiple linear regression was then conducted to investigate which of the alexithymia subscales predicted self-reported EI. Although the overall model with the three alexithymia subscales was significant in predicting self-reported EI (R^2=.70, F(3,128)=41.00, p=.000), only the difficulty in identifying feelings and externally oriented thinking subscales made a significant contribution to the model (DIF: unstandardised β=-.86, t(131)=-3.44, p=.001; EOT: unstandardised β=-1.66, t(131)=-6.54, p=.000). Self-reported EI decreased with increasing difficulty in identifying feelings and externally oriented thinking.

Alexithymia negatively predicts empathy

A MANOVA was conducted with the total alexithymia score predicting the four empathy scales. Results showed a significant effect of alexithymia on empathy (F(4, 125)=13.56, p=.000, partial η^2 = .31). Specifically, alexithymia predicted perspective taking (F(1, 125) = 23.69, p=.000, partial η^2 = .16) and personal distress (F(1, 125) =20.57, p=.000, partial η^2 = .14) but neither fantasy (F(1, 125)=3.96, p=.049, partial η^2 = .03) nor empathic concern (F(1, 125) = 3.74, p = .055, partial η^2 = .03). In addition, perspective taking decreased with increasing alexithymia (β = -.14), while personal distress increased with increasing alexithymia (β = .14).

The MANOVA with the three alexithymia subscales as predictors of perspective taking and personal distress showed a significant effect only of difficulties in identifying

feelings and externally oriented thinking on the two empathy subscales (DIF: F(2, 122)=9.18, p=.000, partial η^2 = .13; EOT: F(2, 122)=6.77, p=.002, partial η^2 = .10). Specifically, difficulties in identifying feelings predicted personal distress (F(1, 123) =9.80, p=.002, partial η^2 = .07), while externally oriented thinking predicted perspective taking (F(1, 123) = 13.29, p=.000, partial η^2 = .10). Personal distress increased with increasing difficulties in identifying feelings (β = .32), while perspective taking decreased with increasing externally oriented thinking (β = -.34).

Alexithymia positively predicts depression

A simple linear regression showed that alexithymia was a significant predictor of symptoms of depression (R^2=.15, F(1,130)=23.39, p=.000) and that symptoms of depression increased with increasing levels of alexithymia (unstandardised β=.19, t(131)=4.84, p=.000).

A multiple linear regression was then conducted to investigate which of the alexithymia subscales predicted depression. Although the overall model with the three alexithymia subscales was significant in predicting depression (R^2=.17, F(3,128)=8.88, p=.000), only the difficulty in identifying feelings subscale made a significant contribution to the model (unstandardised β=.40, t(131)=3.12, p=.002). Symptoms of depression increased with increasing difficulty in identifying feelings.

Alexithymia positively predicts trait anxiety

A simple linear regression showed that alexithymia was a significant predictor of trait anxiety (R^2=.27, F(1,130)=47.87, p=.000) and that trait anxiety increased with increasing alexithymia (unstandardised β=.39, t(131)=6.92, p=.000).

A multiple linear regression was then conducted to investigate which of the alexithymia subscales predicted anxiety. Although the overall model with the three

alexithymia subscales was significant in predicting anxiety (R^2=.55, F(3,128)=18.72, p=.000), only the difficulty in identifying feelings subscale made a significant contribution to the model (unstandardised β=.77, t(131)=4.20, p=.000). Trait anxiety increased with increasing difficulty in identifying feelings.

Discussion

The aim of the present study was to investigate the role of alexithymia in predicting symptoms of emotional intelligence (EI), empathy, depression and anxiety in a group of young individuals. The data revealed that alexithymia predicted all these constructs. In particular, increasing levels of alexithymia were related to decreasing EI and empathy, while increasing depression and anxiety. In addition, different aspects of alexithymia were differentially related to these dependent measures. Specifically, increasing difficulty in identifying feelings was related to decreasing self-reported EI, increasing personal distress during emotionally intense situations, as measured on an empathy scale, as well as increasing depression and anxiety. Also, increasing externally oriented thinking was related to decreasing ability to manage emotions, as measured on a performance measure of EI, self-reported EI and decreasing ability to adopt the point of view of another person, as measured on an empathy scale. Instead, difficulty in describing feelings did not appear to predict any of the dependent measures, suggesting that the ability to correctly identify and introspect on the emotions a person is feeling may play a more significant role than the ability to verbally discuss these emotions with somebody else, in predicting the capacity to manipulate emotional information, manage emotional experience and respond to other people's emotional experiences, in addition to individual psychological wellbeing.

Extending the previous literature that used measures of trait EI (Baughman et al., 2011; J. D. A. Parker et al., 2001; Saklofske et al., 2003; Webb & McMurran, 2008),

increasing alexithymia predicted decreasing self-reported ability EI, corroborating Grieve and Mahar (2010) and deepening their results, showing that difficulty in identifying feelings and externally oriented thinking are the factors that affect self-reported EI. In addition, of particular significance is the result on the negative relationship between alexithymia and performance EI. The only two existing studies on the topic corroborate the main finding that alexithymia is also negatively related to performance EI (Curci et al., 2013; Lumley et al., 2005), suggesting that as individuals report having increasing difficulties in identifying and describing emotions, they indeed show worse performance in tasks involving emotional information. When looking at the different subscales, only externally oriented thinking predicted difficulty in managing emotions, which indicates that lack of introspection may render more challenging to effectively manage emotional situations. This is in contrast to the previous studies, which found that all subscales of alexithymia predicted scoring on the different EI branches (Lumley et al., 2005) or that difficulty in identifying and describing feelings predicted difficulty in perception and understanding of emotions (Curci et al., 2013). These differences may in part result from the lack of correction for multiple comparisons, which may have inflated the possibility of type I error in previous studies, and in part result from methodological differences, such as demographics of the participants. Additional research should be conducted to clarify how the different factors of alexithymia affect the different EI abilities.

The relationship between alexithymia and empathy was mostly in line with previous literature (Grynberg et al., 2010; Moriguchi et al., 2006, 2007; Patil & Silani, 2014b). Indeed, alexithymia negatively predicted the tendency to adopt the point of view of the other, while positively predicted the distress elicited by the perception of others' distress. In addition, no relationship was found between alexithymia and the tendency to transpose imaginatively into the feelings and actions of fictitious characters in books, movies and plays. Contrary to the

previous literature, no relationship between alexithymia and empathic concern was found. This may have resulted by the fact that here all the variables were simultaneously analyzed in one statistical model, which reduces the possibility of type I error that may have occurred in previous studies, which used separate analyses to test the relationship between alexithymia and each aspect of empathy. Focusing on the alexithymia subscales, difficulty in identifying feelings was positively related to personal distress during emotionally intense situations, while externally oriented thinking was negatively related to the ability to adopt the point of view of another person. This result suggests a coherent relationship between the cognitive and emotional aspects of each construct, with an emotional aspect of alexithymia impacting an emotional component of empathy, while a cognitive aspect of alexithymia impacting a cognitive component of empathy, in line with Grynberg et al. (2010).

The positive correlation between alexithymia and depression and the relation between difficulty in identifying feelings and depression are in line with the previous literature (Berthoz et al., 1999; S. Li et al., 2015; Marchesi et al., 2000). However, a recent meta-analysis also highlighted, respectively, moderate and a weak correlations between difficulty in describing feelings and externally oriented thinking and depression (S. Li et al., 2015). Nevertheless, it also points out that the relationship between these two factors and depression seems less consistent, suggesting they may have a more marginal role in the relationship between alexithymia and depression than the difficulty in identifying feelings, as suggested by the present results.

The result that alexithymia, and specifically difficulty in identifying feelings, positively predicted anxiety is also in line with the only three studies that investigated the relation between alexithymia and anxiety in the general population (Berthoz et al., 1999; Grynberg et al., 2010; Marchesi et al., 2000), replicating the results of Grynberg et al. (2010)

and extending the results of Berthoz et al., (1999), whose sample comprised only women and used an older version of the TAS, and the results of Marchesi et al. (2000), who had recruited their sample from individuals admitted to the emergency room, representing a possible confounding factor to the generalisability of results.

To conclude, the current experiment shows that alexithymia can affect different domains of emotion processing, from the ability to manipulate emotional information, to the response to others' emotions, to individuals' psychological wellbeing. These results also highlight the importance of the study of alexithymia, which can have a significant impact on individuals' emotional life, despite being a subclinical trait.

Experiment 2: individuals with alexithymia require higher emotional intensity to recognize fearful facial expressions

Introduction

Previous research found alexithymia to be related to worse performance in the identification of emotional facial expressions (EFEs) (Jessimer & Markham, 1997; Lane et al., 1996). Specifically, under temporal constraints alexithymia has been related to lower accuracy in EFE recognition (Ihme, Sacher, Lichev, Rosenberg, Kugel, Rufer, Grabe, Pampel, Lepsien, Kersting, Villringer, Lane, et al., 2014; Swart et al., 2009), which seems to be overcome when stimulus exposure time is extended (Ihme, Sacher, Lichev, Rosenberg, Kugel, Rufer, Grabe, Pampel, Lepsien, Kersting, Villringer, & Suslow, 2014; Pandey & Mandal, 1997; P. D. Parker et al., 2005). This suggests that alexithymia is associated to the need for more time to accurately recognize EFEs (Grynberg et al., 2012).

Despite this evidence, previous research has focused on the response to intense EFEs. Nevertheless, these are rarely encountered in everyday life, where individuals are faced with the challenge of identifying emotional expression displayed at varying degrees of intensity. In fact, alexithymia may be hypothesized to be related not only to the need for more time but also for more perceptual information to identify EFEs. Therefore, manipulating the intensity of EFEs could offer the opportunity to extend current literature on the impact of alexithymia on EFE identification.

In this regard, two studies exist that have used morphed faces to understand the role of emotional intensity in the relationship between alexithymia and EFE identification. Nevertheless, they present contrasting findings and have the limitation of focusing mainly on alexithymia within the autistic population. Specifically, the first study found alexithymia to be related to less precision, expressed as higher attribution threshold, in the identification of

morphed EFEs both in the autistic and control group (Cook, Brewer, Shah, & Bird, 2013). On the contrary, the second study found alexithymia to be related to reduced accuracy in identifying EFEs at low emotional intensity in the autistic but not the control group (Ketelaars, In't Velt, Mol, Swaab, & van Rijn, 2016), raising the possibility that autism per se may represent a confounding factor contributing to the results. Therefore, it appears that further research is needed in order to understand the role of emotional intensity in the relationship between alexithymia and EFE identification.

The aim of the present study was to investigate the impact of emotional intensity in the relationship between alexithymia and EFE identification. To this end, participants with low (LA) and high (HA) levels of alexithymia were tested in their ability to identify static morphed EFEs, which ranged from neutral to intense emotional expression. Individuals with high, as compared to low, levels of alexithymia were hypothesized to require more emotional intensity to be as efficient as LA in identifying the presence of the emotion in the face, hence showing decreased performance in this task.

Methods

Participants

Three-hundred university students completed the 20-item Toronto Alexithymia Scale (TAS-20; Taylor, Bagby, & Parker, 2003). Depending on the score, students were classified as LA (TAS-20 $\leq$ 36) or HA (TAS-20 $\geq$ 61) (Franz, Schaefer, Schneider, Sitte, & Bachor, 2004) and were then randomly contacted to participate in the study. Once in the laboratory, the alexithymia module of the structured interview for the Diagnostic Criteria for Psychosomatic Research (DCPR; Mangelli, Semprini, Sirri, Fava, & Sonino, 2006) was administered to increase reliability of screening and confirm TAS-20 classification. This measure has 6 questions to evaluate presence or absence of alexithymic characteristics. At least 3

characteristics are required for a classification of alexithymia. Participants with discordant classification on the two measures did not complete the task (n = 1). Due to the high co-occurrence of alexithymia and depression (S. Li et al., 2015), participants completed the Beck Depression Inventory (Beck et al., 1961) and did not complete the experimental task in case their score was higher than the moderate depression cut-off (i.e. 19, n = 1).

Forty volunteers with no history of major medical, neurological or psychiatric disorders completed the study: 20 LA (6 males; TAS-20 M = 30.25, SD = 4.12; age M = 24.55, SD = 2.98 years); 20 HA (6 males; TAS-20 M = 63.37, SD = 2.25; age M = 23.03, SD = 2.32 years). A priori targets for sample size and data collection stopping rule were based on sample and effect sizes reported in the literature on alexithymia and EFE identification (sample size of an average of 38 participants in total as indicated in a recent review (Grynberg et al., 2012)).

All participants had equivalent educational backgrounds and were students at the University of Bologna. The study was designed and conducted in accordance with the ethical principles of the World Medical Association Declaration of Helsinki and the institutional guidelines of the University of Bologna and was approved by the Ethics Committee of the Department of Psychology. All participants gave informed written consent to participation after being informed about the procedure of the study.

Independent measure

Stimuli consisted of black and white photographs of 20 actors (10 males) with each actor depicting 3 EFEs, respectively of happiness, disgust and fear. Half of the pictures were taken from the Karolinska Directed Emotional Faces database (Lundqvist, Flykt, & Öhman, 1998) and half from the Pictures of Facial Affect database (Ekman & Friesen, 1976). Pictures were trimmed to fit an ellipse in order to uniform them and remove distracting features from the

face, such as hair or ears and non-facial contours. Each emotional facial expression was then morphed with the neutral facial expression of the corresponding identity using *Abrosoft FantaMorph*, (2009) in order to create stimuli of 20% increments of emotional intensity ranging from 0% to 100% emotional intensity. This resulted in a total of 360 stimuli (20x13cm size), i.e. 20 actors expressing 3 emotions with 6 degrees of intensity (0%, 20%, 40%, 60%, 80%, 100%; Fig. 1).

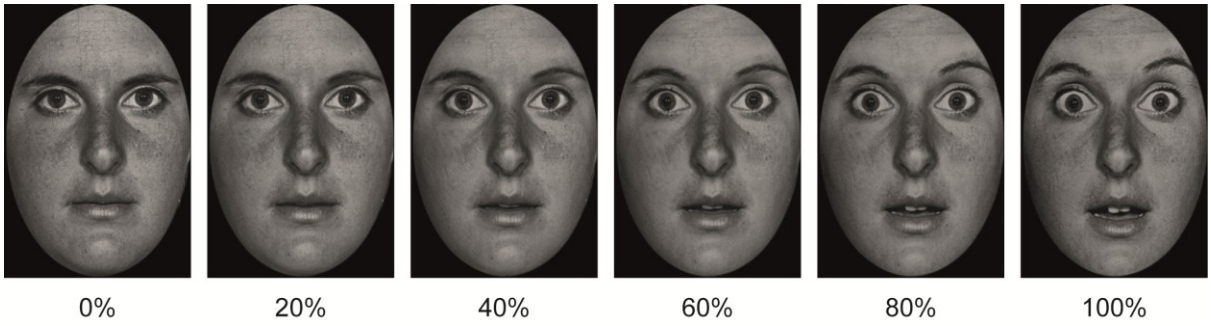

Figure 1. Example of morphed pictures of fearful facial expressions used as EFEs ranging from 0% to 100% emotional intensity.

Each trial started with the presentation of a fixation cross (500ms) in the centre of the screen followed by the stimulus (100ms) and subsequently a black screen (3000ms) during which participants could provide the answer by pressing one of four keys with their index and middle finger of either hand. Keys were labeled "N" for neutral (i.e. Italian = "neutro"), "F" for happiness (i.e. Italian = "felicità"), "P" for fear (i.e. Italian = "paura") and "D" for disgust (i.e. Italian = "disgusto"). The experiment consisted of 360 randomized trials divided in two blocks of 180 trials so that participants could rest if desired. The order of keys was counterbalanced between participants.

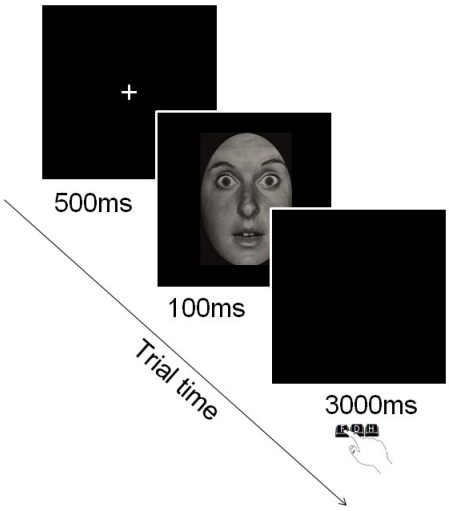

Figure 2. Illustration of experimental trial. The stimulus is represented by a fearful face, as an example.

Procedure

The experiment took place in a sound attenuated room with dimmed light. Participants sat in a relaxed position on a comfortable chair in front of a computer monitor (17'', 60Hz refresh rate) used for stimuli presentation at 57cm distance. Participants were instructed that at each trial a face would briefly appear on the screen and their task would be to identify the emotion expressed by the face by pressing one of four keys. Before beginning the task, participants familiarized with the position of keys by having the experimenter calling out loud in random order the keys and participants pressing them until they felt confident they could press them correctly while fixating the screen.

Dependent measure

For each emotional facial expression, the expression identification rate at each intensity level was computed. Then, for each subject, expression identification rates for each emotional facial expression were fit to a psychometric function using a generalized linear model with a binomial distribution in MATLAB software (MathWorks, Natick, MA, USA) (Nakajima, Minami, & Nakauchi, 2017). The point of subjective equality (PSE) was then calculated and

used for statistical analysis. This represented the percentage of emotional intensity at which subjects had equal probability to identify the facial expression as neutral or emotional.

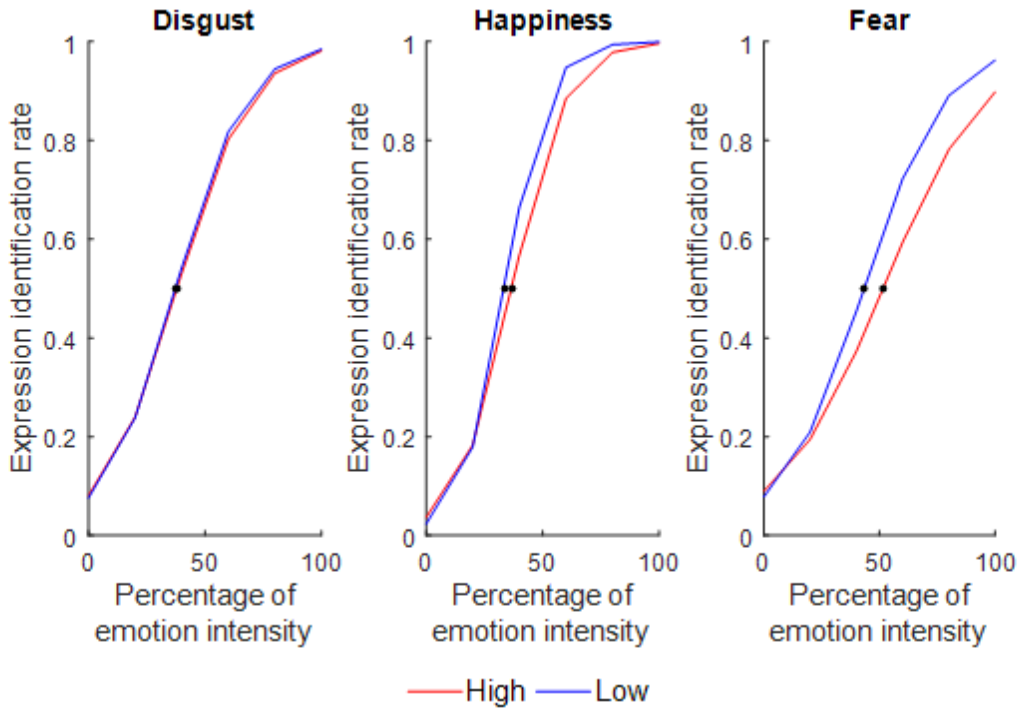

Figure 3. Average psychometric function for each emotional expression as a function of group. The dots represent the point of subjective equivalence (PSE).

Results

A 3x2 repeated measures analysis of variance (emotion: happiness, disgust, fear; group: low alexithymia, high alexithymia) on PSE scores showed a significant group by emotion interaction ($F(2,76)=4.69$, $p=0.012$, partial $\eta^2 = 0.11$). Bonferroni post-hoc test shows that HA had higher PSE compared to LA only for the fearful emotional facial expression ($p=0.004$; $M_{HA}=54.05$, $M_{LA}=43.71$). Therefore, HA need more emotional intensity to identify fearful facial expressions compared to LA.

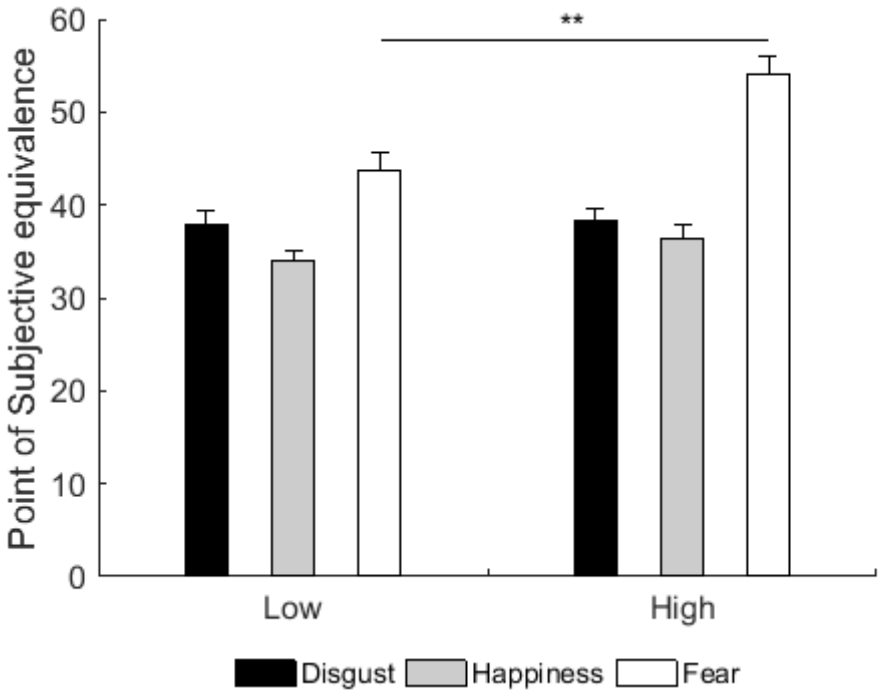

Figure 4. Mean point of subjective equivalence (PSE) for each emotional facial expression as a function of group. Participants with high alexithymia have higher PSE than those with low alexithymia for the fearful facial expression, indicating they need more emotional intensity to identify fearful facial expressions. Error bars represent standard errors. Significant differences are indicated as follows: **p< .01.

Discussion

The aim of the present study was to investigate the role of alexithymia in processing EFEs. Specifically, alexithymia was hypothesized to be related to the need for more emotional intensity to identify static morphed EFEs ranging from neutral to intense emotional expression. Results showed that HA had higher point of subjective equivalence than LA when identifying fearful EFE, indicating the need of more emotional intensity to identify the presence of that expression in the face. Crucially, while previous studies showed that HA need more time to identify EFEs as efficiently as LA (Grynberg et al., 2012), the present study extends the current literature suggesting that HA also need more perceptual information, specifically to identify fearful EFEs.

The difficulty in processing fearful EFE is in line with previous literature. Indeed, compared to LA, HA rate the expression of fearful, but not other EFEs, as less intense

(Prkachin, Casey, & Prkachin, 2009), they fail to show enhanced remapping of fear on their own somatosensory system (Scarpazza et al., 2014, 2015) and enhanced electrophysiological response when encoding fearful body postures (Borhani et al., 2016). These results have been interpreted in light of the decreased activation of the amygdala observed in alexithymia in response to the presentation of EFEs (Jongen et al., 2014; Kugel et al., 2008), negative emotional stimuli (Moriguchi & Komaki, 2013; van der Velde et al., 2013; Wingbermühle et al., 2012) and in particular fearful ones (Pouga et al., 2010). Although involved in processing EFEs in general (Fusar-Poli et al., 2009), the amygdala appears a crucial structure in processing fearful EFEs. Indeed, lesion of this structure has been found to impair the recognition of fearful EFEs (Adolphs et al., 1999, 1994). Furthermore, in addition to processing fear, the amygdala appears to have a role in processing emotional intensity. The increase of emotional intensity enhances amygdala activation, as shown by increased amygdala response during the presentation of high as compared to low intensity EFEs (Winston, O'Doherty, & Dolan, 2003). Therefore, it is possible that a reduced response in the amygdala in HA may underlie the present results.

In contrast to the difference found in response to fearful EFEs, no difference between the groups was found when identifying happy or disgusted facial expressions. In this regard, previous behavioral studies have reported mixed results. Alexithymia was either related to a global deficit to recognize EFEs, including happiness and disgust (Ihme, Sacher, Lichev, Rosenberg, Kugel, Rufer, Grabe, Pampel, Lepsien, Kersting, Villringer, Lane, et al., 2014; Jessimer & Markham, 1997), or, in line with the current results, the difficulty of alexithymia was limited to the identification of EFEs other than happiness and disgust (Ihme, Sacher, Lichev, Rosenberg, Kugel, Rufer, Grabe, Pampel, Lepsien, Kersting, Villringer, & Suslow, 2014; Prkachin et al., 2009). In addition, with regards to happiness, the majority of neuroimaging studies failed to report a relationship between alexithymia and differences in

brain activity in response to happy EFEs (Duan, Dai, Gong, & Chen, 2010; Eichmann, Kugel, & Suslow, 2008; Ihme, Sacher, Lichev, Rosenberg, Kugel, Rufer, Grabe, Pampel, Lepsien, Kersting, Villringer, Lane, et al., 2014; Ihme, Sacher, Lichev, Rosenberg, Kugel, Rufer, Grabe, Pampel, Lepsien, Kersting, Villringer, & Suslow, 2014, 2014; Kugel et al., 2008; but see Reker et al., 2010 for contrasting results). Regarding disgusted EFEs instead, no neuroimaging studies on alexithymia and EFEs recognition could be found that included such facial expression among their stimuli, leaving this issue open for future research.

Being this the first study to investigate the impact of emotional intensity in the relationship between alexithymia and EFE identification on healthy participants, it is crucial to point its limitation in order to guide future research. To limit the duration of the task, not all basic emotions were included among the stimuli, restricting the implications of the results. Therefore, future studies should investigate response also to the emotions not included here. Also, the time of stimuli presentation was quite limited (100ms), because previous studies showed that exposure time plays a significant role in the ability of HA to identify EFEs (Grynberg et al., 2012), future studies could investigate whether time plays a role also in the current results. Exposure time could be extended to clarify whether the difficulty in the identification of fear would remain or not when this is increased.

To conclude, the present study shows that alexithymia requires more perceptual information to identify fearful EFEs, possibly suggesting insufficient amygdala activation in response to these stimuli to enable efficient recognition of such emotional expression.

Conclusion

The experiments presented in this chapter contribute to the literature describing the difficulties of individuals with alexithymia in processing emotional information. Specifically, Experiment 1 shows that alexithymia has a widespread effect on the emotional life of individuals being related to decreasing emotional intelligence, empathy and psychological wellbeing, highlighting the importance of studying this personality trait. In addition, Experiment 2 shows that alexithymia is related to the need for more emotional intensity to identify fear in emotional facial expressions, extending the previous literature that had shown alexithymia is related to the need for more time to identify emotions in facial expressions.

Chapter 4. Understanding alexithymia: impairments in emotional learning

Introduction

Although the literature has dedicated extensive research to describe the difficulties in emotion processing related to alexithymia, significantly less studies have been dedicated to the understanding of the basic mechanisms that may underlie such difficulties. Nevertheless, this type of investigation could provide new insight in the understanding and conceptualization of this personality trait.

Although in our daily life countless stimuli can elicit an emotional response, only a restricted range of stimuli is biologically programmed to do so, which includes appetitive and aversive unconditioned stimuli (LeDoux, 1998). However, to increase the chances of survival, we usually expand the range of emotional stimuli, from those that unconditionally trigger an emotional response to those that are often associated with them, thanks to a process called associative emotional learning. In associative emotional learning a neutral stimulus or behavior and an unconditioned stimulus are repeatedly associated. As a consequence, the emotional response that is first triggered by the unconditioned stimulus is transferred on the neutral stimulus or behavior (e.g. Seymour et al., 2004), which acquires emotional value becoming predictor of the unconditioned stimulus. As such, these internal representations are used to predict coming emotional stimuli, preparing the organism to respond to the predicted emotional stimuli (McNally & Westbrook, 2006; Öhman & Mineka, 2001). Nevertheless, for such mechanism to be adaptive, these predictions need to be precise. To ensure this, every time an emotional stimulus is predicted, this prediction is compared against the incoming information and if there is a discrepancy between the two, a prediction error is computed to change the erroneous prediction driving learning (Friston, 2010). Therefore, precise

computation of predictions and prediction errors are necessary for successful emotional learning and ultimately accurate construction of emotions.

Given this premise, Chapter 2 raised the question that the difficulties in emotion processing of alexithymia could be a manifestation of an underlying impairment in emotional learning. In order to test this question, four experiments will be presented in the following chapter. Experiment 3 and 4 will examine the way in which alexithymia impacts the construction of emotional predictions, assessing the psychophysiological correlates of emotional predictions during Pavlovian aversive and appetitive conditioning respectively. Experiment 5 will not only examine the impact of alexithymia on emotional predictions but also prediction errors, which are equally crucial to construct accurate internal representations of emotional stimuli. In particular, it will assess the electrophysiological correlates of prediction and prediction error during appetitive conditioning. In all experiments, alexithymia is hypothesized to be related to decreased ability to compute predictions and prediction errors during emotional learning. Finally, Experiment 6 will examine how alexithymia affects behavior both during learning as well as after learning, when the information learned has to be used in a new context to make effective decisions, here focusing on instrumental rather than Pavlovian learning. In this case, alexithymia is hypothesized to be related to a difficulty in learning from rewards and punishments and a difficulty in then using such learning effectively to drive behavior in a new context.

Experiment 3: reduced psychophysiological correlates of prediction of aversive stimuli during threat conditioning in alexithymia

Introduction

Pavlovian threat conditioning is a crucial adaptive mechanism that enables the individual to learn to predict the occurrence of aversive events in order to efficiently avoid them (Maren, 2001). In pavlovian threat conditioning, a neutral stimulus is paired with an aversive unconditioned stimulus (UCS), which elicits innate emotional responses, named unconditioned response (UR). After repeated pairing of the two stimuli, the individual learns to predict the occurrence of UCS at the presentation of the neutral stimulus, which becomes a conditioned stimulus (CS). In the end, the sole presentation of CS elicits an anticipatory response in preparation to the occurrence of UCS, called conditioned response (CR). Both UR and CR are marked by physiological changes in autonomic nervous system activity, such as increased skin conductance response (SCR) (LeDoux, 2014; Phelps & LeDoux, 2005). Higher SCR, in response to CS, signals increased expectations regarding the occurrence of UCS following the presentation of CS, indicating that the association between CS and UCS has been learnt (S. S. Y. Li & McNally, 2014). Additionally, changes in subjective affective experience accompany the physiological changes and higher arousal and lower pleasantness are generally reported by participants at the presentation of CS compared to a neutral stimulus (Tabbert et al., 2011).

The aim of this experiment was to investigate the differences between LA, HA and individuals with medium levels of alexithymia (MA) in learning to predict an aversive emotional event during pavlovian threat conditioning. Sixty HA, LA and MA, as measured by the 20-item Toronto Alexithymia Scale (TAS-20)(Taylor, Bagby, & Parker, 2003), completed a classical threat conditioning task with partial reinforcement (Garofalo et al.,

2014; Schiller, Levy, Niv, LeDoux, & Phelps, 2008). On each trial, one of two colored squares was presented. During habituation none of the stimuli was reinforced to ensure there were no baseline differences in response to the stimuli. During acquisition one stimulus was reinforced with a mild electric stimulation (UCS) on 80% of trials (CS+) while the other was never reinforced (CS-). During extinction no stimulation was administered. Changes in SCR were recorded continuously during the experiment as a psychophysiological indicator of the degree of prediction of the UCS. In addition, to assess subjective experience, participants reported the level of anxiety and fear experienced at the presentation of each CS. HA were hypothesized to have lower prediction of the aversive emotional event (i.e. UCS), exhibiting lower SCR in response to the CS+ compared to MA and LA, while no differences between MA and LA were expected.

Methods

Participants

Three-hundred university students completed the 20-item Toronto Alexithymia Scale (TAS-20)(Taylor et al., 2003). Depending on the score, students were classified as LA (TAS-20 $\leq$ 36), MA (36 < TAS-20 < 61) or HA (TAS-20 $\geq$ 61) (Franz et al., 2004). Individuals from the three groups were randomly contacted and asked to participate in the study. Due to the high co-occurrence of alexithymia and depression(S. Li et al., 2015), participants completed the Beck Depression Inventory (Beck et al., 1961), and were excluded from the in case their score was higher than the moderate depression cut-off (i.e. 19; n = 5). Sixty-two university students with no history of neurological or psychiatric disorders completed the study. After the experimental task, explicit awareness of the contingency between CS and UCS was assessed. Two participants were removed from analysis due to failure in reporting the correct association between stimuli. The final sample included in the analysis consisted of

60 participants (22 males; age $M = 24.03$, $SD = 2.38$ years old) divided in three groups: 20 LA (7 males; TAS-20 $M = 30.42$, $SD = 3.79$; age $M = 24.67$, $SD = 2.83$ years old); 20 MA (8 males; TAS-20 $M = 46.10$, $SD = 6.18$; age $M = 24.06$, $SD = 1.80$ years old); 20 HA (7 males; TAS-20 $M = 63.63$, $SD = 2.39$; age $M = 23.35$, $SD = 2.32$). A priori targets for sample size and data collection stopping rule were based on sample and effect sizes reported in the literature on classical fear conditioning (sample size around 17-19 participants per group as indicated by a recent meta-analysis (Fullana et al., 2015)).

Because anxiety is known to affect SCR in classical threat conditioning (Duits et al., 2015), levels of anxiety were measured with the State-Trait Anxiety Inventory (Spielberger et al., 1983). Levels of anxiety in the three groups differed significantly both for state ($F(2, 57) = 5.86$, $p = .005$) and trait anxiety ($F(2, 57) = 21.54$, $p < .001$). Post-hoc Newman-Keuls test showed that for state anxiety LA ($M = 33.84$) had significantly lower levels of anxiety compared to MA ($M = 40.22$; $p = .009$) and HA ($M = 39.79$; $p = .006$), while for trait anxiety all groups differed significantly from each other, with LA ($M = 36.42$) showing lower levels of trait anxiety compared to MA ($M = 44.32$; $p = .001$) and HA ($M = 51.53$; $p < .001$) and MA showing lower levels of trait anxiety compared to HA ($p = .003$). Correlations between levels of anxiety and differential SCR response were explored to exclude an effect of anxiety on results.

The study was designed and conducted in accordance with the ethical principles of the World Medical Association Declaration of Helsinki and the institutional guidelines of the University of Bologna and was approved by the Ethics Committee of the Department of Psychology. All participants gave informed written consent to participation after being informed about the procedure of the study.

Independent measure

Pavlovian threat conditioning. The task consisted in a classical differential threat conditioning paradigm with partial reinforcement (Garofalo et al., 2014; Schiller et al., 2008). Two isoluminant colored squares represented the CS. The UCS consisted of a mild electric stimulation of 200ms in duration generated by a Digitimer Stimulator (Model DS7, Digitimer Ltd., UK) administered to the inner wrist of the right hand, to which two electrodes were attached. The intensity of the stimulation was set with a standard workup procedure. It was initially set at 0.5mA and increased of 1mA until participants reported it as being highly uncomfortable but not painful.

Each trial consisted in the presentation of one CS in the centre of a computer screen (17'', refresh rate 60Hz) for 6 seconds followed by an inter trial interval of 12 seconds during which a fixation cross was presented (Fig. 5). The task included 40 trials (20 for each CS) divided in three blocks: habituation, acquisition and extinction.

At the beginning of habituation, instructions appeared on the screen stating that two different images would be presented one at the time in the centre of the screen, no stimulation would be administered and the task of the participant would be to carefully observe the images. Habituation included 4 trials (2 for each CS) to ensure the absence of any baseline differences within and between groups in response to the images. At the beginning of acquisition, similar instructions stated that the same two images would appear one at the time in the centre of the screen and that one of them might be paired with the stimulation. The task of participants remained to carefully observe the images. No information was given about contingencies between images and stimulation. Acquisition included 16 trials (8 for each CS). CS+ was reinforced in 80% of the trials (n = 6), while CS- was never reinforced. Extinction

followed acquisition without any instructions. It included 20 trials (10 for each CS) and no stimulation was administered.

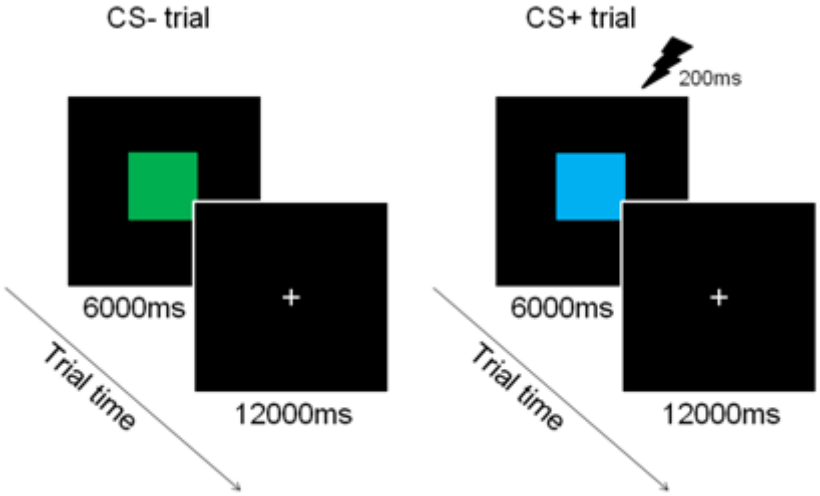

Figure 5. Illustration of a CS- trial and a CS+ trial, in which the presentation of the CS+ was reinforced by the shock.

Stimuli were presented in pseudo-randomized order, no more than two presentations of the same stimulus occurred in a row (Schiller et al., 2008). The first two trials of acquisition always included one CS- and one CS+ presented randomly. The color of the square associated to the CS+ and CS- was counterbalanced across participants.

Dependent measures

Skin conductance response recording and processing. The SCR was recorded through two Ag/AgCl electrodes (TSD203 Model; Biopac Systems, USA), filled with isotonic hyposaturated conductant attached to the distal phalanges of the second and third finger of participants' left hand and held with Velcro straps. The SCR signal was continuously recorded at 200Hz and amplified using a DC amplifier (Biopac GSR100; Biopac Systems, USA) with 5µS/V gain factor and 10Hz low pass filter. The analogue signal was digitalized using the MP-150 digital converter (Biopac Systems, USA) and fed into AcqKnowledge 3.9 software (Biopac Systems, USA).

SCR data were analyzed using MATLAB (The MathWorks, Inc., USA) custom-made scripts (Garofalo et al., 2014). SCR was calculated as the peak-to-peak amplitude difference of the largest deflection in the 0.5-4.5sec latency window after stimulus onset. Regarding SCR to UCS, stimulus onset was represented by the administration of shock, while regarding SCR to CS, stimulus onset referred to the time of CS appearance. The SCR was transformed into microsiemens (μS) and calculated for each trial. Minimum response criterion was 0.02μS and smaller responses were encoded as zero. Square root transformation was conducted on raw SCR to normalize the data distribution and SCR were scaled to each subject's maximal UCS response to account for interindividual variability (Schiller et al., 2008).

Both SCR to UCS and CS- were analyzed to ensure that groups did not differ in their physiological response to the stimulation or in the anticipatory response to CS- but only to a conditioned stimulus that predicts an aversive event (i.e. CS+). Regarding SCR to UCS, both peak response and average response were analyzed. For each participant, peak response represented the highest SCR in response to six shocks while mean response was the average of the SCRs to the six shocks. Regarding the response to CSs, SCRs during the three phases of conditioning were analyzed separately. Concerning habituation, all trials were included in the analysis. With regards to acquisition, the first two trials were not included in the analysis because participants learned the association between UCS and CS+ after its first pairing with the shock. Regarding extinction, all trials were included in the analysis but this phase was divided in two blocks (early and late extinction). Then, mean SCR of each participant was computed to produce four average scores representing the SCR of each subject during habituation, acquisition, early and late extinction. These were then averaged to obtain the SCR during the different phases for each alexithymia group.

Assessment of subjective anxiety, fear and contingency awareness. Participants were asked to report the level of anxiety and fear experienced at the presentation of each CS on separate visual analogue scales ranging from 0 (not at all) to 100 (extreme). The order of questions was counterbalanced across participants.

Participants were also asked to indicate which of the two stimuli was associated with the stimulation to ensure explicit awareness of pairing between CS+ and UCS. Participants who failed to report the correct association were removed from analysis.

Procedure

The experiment took place in a sound attenuated room with dimmed light. Participants were seated in a chair in front of a computer monitor at ~70 cm distance. Once seated, the experimental procedure was explained and written informed consent was obtained from participants. Then SCR electrodes were attached and correct recording of the signal was ensured. Participants were then told the task consisted of two phases and that instructions would appear on the screen for each phase. Following completion of the task, subjective reports were completed.

Results

No difference in UCS intensity and peak and mean SCR to UCS among groups

Univariate ANOVAs were used to evaluate differences in UCS intensity and mean and peak SCR to the UCS among groups. Results showed no significant differences among the three groups in either UCS intensity ($F(2, 57) = 1.06$, $p = .353$, partial $\eta^2 = .04$), peak SCR in response to UCS ($F(2, 57) = .31$, $p = .736$, partial $\eta^2 = .01$) or mean SCR in response to UCS ($F(2, 57) = .80$, $p = .456$; partial $\eta^2 = .03$). On average, the intensity of the stimulation

received by participants and the physiological response to it did not differ significantly among groups.

No difference in SCR among groups or stimuli during habituation

A 3x2 RM ANOVA (group: LA, MA, HA; stimulus: CS+, CS-) was carried out to analyze SCR during habituation. Analysis on SCR showed no significant main effect of group ($F(2, 57) = .71$, $p = .494$, partial $\eta^2 = .02$), stimulus ($F(1, 57) = .84$, $p = .361$, partial $\eta^2 = .01$) or interaction ($F(2, 57) = .38$, $p = .681$, partial $\eta^2 = .01$), confirming that at baseline there were neither within group nor between group differences in response to the two conditioned stimuli.

SCR after habituation

A 3x2x3 RM ANOVA (group: LA, MA, HA; stimulus: CS+, CS-; phase: acquisition, early extinction, late extinction) was carried out to analyze SCR in the phases following habituation. Analysis on SCR showed significant stimulus by phase by group interaction ($F(4,114) = 3.64$, $p = .008$, partial $\eta^2 = .11$). This interaction was further explored conducting separate ANOVAs for acquisition and extinction.

Reduced SCR in anticipation of shock during acquisition in high alexithymia

The 3x2 RM ANOVA (group: LA, MA, HA; stimulus: CS+, CS-) on SCR during acquisition showed a significant main effect of group ($F(2, 57) = 3.26$, $p = .046$, partial $\eta^2 = 0.10$) and stimulus ($F(1, 57) = 84.74$, $p < .001$, partial $\eta^2 = 0.60$). Crucially, this was secondary to a stimulus by group interaction ($F(2, 57) = 3.26$, $p = .046$, partial $\eta^2 = .10$) was found, indicating that groups differed in SCR to the two conditioned stimuli. Newman-Keuls test showed that despite all groups showing significant difference in response to CS+ as compared to CS- ($M_{LACS-} = .12$ µS, $M_{LACS+} = .30$ µS; $p < .001$; $M_{MACS-} = .16$ µS, $M_{MACS+} = .31$ µS; $p < .001$; $M_{HACS-} = .09$ µS, $M_{HACS+} = .18$ µS; $p = .007$), there was a significant

difference between groups in response to CS+. Specifically, HA had significantly lower SCR compared to LA ($p = .007$) and MA ($p = .015$). No difference was found in SCR to CS+ between LA and MA or in response to CS- among any of the groups (all $p > .262$; Fig. 6). Therefore, all groups showed differential increase in SCR to CS+ compared to CS-. However, SCR to CS+ exhibited by HA was significantly lower than SCR exhibited by the other two groups. On the contrary, responses to CS- were comparable among groups.

Given that difference in SCR between CS+ and CS- may be influenced by the levels of anxiety, correlations between these two variables were explored to exclude a significant contribution of anxiety to the results. Neither trait nor state anxiety correlated significantly with difference in SCR to the two conditioned stimuli (all $p > .292$).

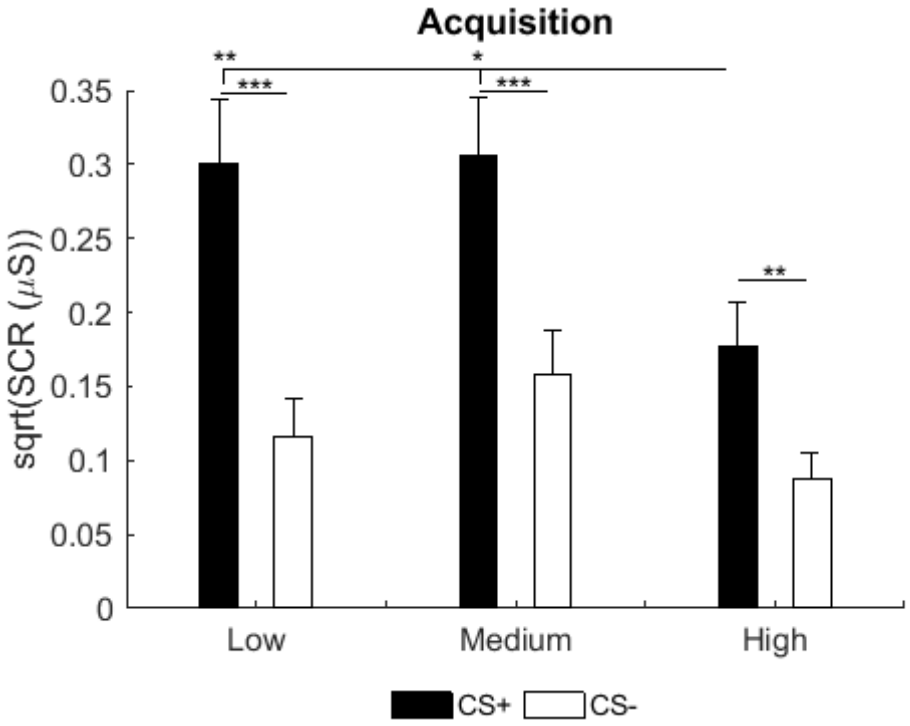

Figure 6. Mean square rooted skin conductance response (SCR) to CS+ and CS- during acquisition as a function of alexithymia group. Although all groups show higher SCR to CS+ than CS-, participants with high alexithymia have show significantly less SCR to CS+ compared to those with medium and low alexithymia. Error bars represent standard errors. Significant differences are indicated as follows: *p< .05; **p< .01; ***p<.001.

Enhanced extinction in high alexithymia

A 3x2x2 RM ANOVA (group: LA, MA, HA; stimulus: CS+, CS-; phase: early extinction, late extinction) on SCR during extinction showed neither a significant main effect nor interaction of time with the other factors was found (all $p > .183$). A main effect of stimulus was present ($F(1, 57) = 24.30$, $p < .001$, partial $\eta^2 = 0.30$), although secondary to the stimulus by group interaction. Crucially, there was a significant stimulus by group interaction ($F(2, 57) = 5.53$, $p = .007$, partial $\eta^2 = 0.16$). Newman-Keuls test showed that LA and MA maintained a significantly higher SCR to CS+ compared to CS- ($M_{LACS-} = 0.10$ µS, $M_{LACS+} = 0.15$ µS; $p = .037$; $M_{MACS-} = 0.10$ µS, $M_{MACS+} = 0.24$ µS; $p < .001$). On the contrary, HA extinguished the differential SCR response to the CSs ($M_{HACS-} = 0.07$ µS, $M_{HACS+} = 0.09$ µS; $p = .355$; Fig. 7).

Also in this phase neither trait nor state anxiety correlated significantly with the difference in SCR between CS+ and CS- (all $p > .616$).

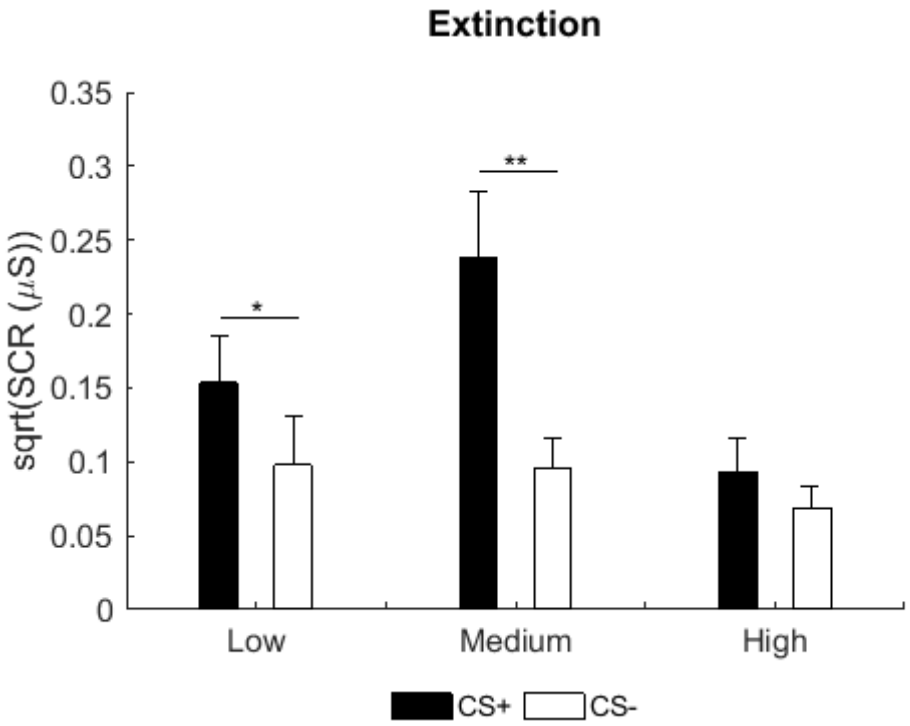

Figure 7. Mean square rooted skin conductance response (SCR) to CS+ and CS- during extinction as a function of alexithymia group. While participants with low and medium alexithymia maintained higher SCR to CS+ than CS-, participants with high alexithymia extinguished the differential SCR once the CS+ was no more reinforced. Error bars represent standard errors. Significant differences are indicated as follows: *p< .05; **p< .01.

No difference in subjective reports of anxiety and fear among groups

3x2 RM ANOVAs (group: LA, MA, HA; stimulus: CS+, CS-) were conducted on subjective reports of fear and anxiety experienced at the presentation of the conditioned stimuli. Both the subjective report on anxiety and fear showed a main effect of stimulus (respectively: $F(1, 57) = 170.41$, $p < .001$, partial $\eta^2 = .75$; $F(1, 57) = 133.34$, $p < .001$, partial $\eta^2 = .70$). The reported anxiety to CS+ ($M_{CS+} = 59.84\%$) was higher than to CS- ($M_{CS-} = 23.56\%$; $p < .001$) as well as the reported fear to CS+ ($M_{CS+} = 46.13\%$) was higher than to CS- ($M_{CS-} = 13.41\%$; $p < .001$). In contrast, no significant main effect of group or interaction was found either for anxiety or fear (all ps≥.117).

Discussion

This experiment investigated whether HA presented reduced prediction of aversive emotional stimuli. To this end, changes in SCR and subjective emotional experience were

assessed during pavlovian threat conditioning to assess differences among LA, MA and HA in predicting the occurrence of a negative emotional event by learning patterns of association between CS+ and UCS. Participants correctly associated CS+ and UCS, suggesting that they explicitly identified the stimulus that anticipated the negative emotional event. In addition, groups did not differ in the intensity of UCS received, SCR to it and emotional experience reported in response to presentation of CS+. On the contrary, results showed significant differences among HA and MA and LA in SCR during acquisition that exacerbated during extinction.

Specifically, during acquisition all three groups learned the anticipatory value of CS+ in predicting UCS, as indicated by higher SCR to CS+ compared to CS-. However, the degree of physiological response elicited by the anticipation of UCS in HA was lower compared to MA and LA, as shown by significantly lower SCR to CS+. This reduced response intensified during extinction, when the differential SCR to CS+ extinguished in HA while it was maintained in MA and LA. This suggested that the response elicited by the anticipation of UCS disappeared as soon as the predictive value of CS+ was no more reinforced by the administration of UCS. Crucially, this result did not appear to be dependent solely on the reduced SCR to CS+ during acquisition. These differences between the groups were attributable neither to differences in the intensity of UCS, because all groups received comparable intensities of stimulation, nor to reactivity to UCS itself, because groups did not differ in mean or peak SCR amplitude to UCS. In addition, groups did not differ in their SCR during habituation, acquisition or extinction to CS-, indicating comparable physiological response to neutral stimuli as well. Therefore, although HA seem to learn to differentiate a neutral from a conditioned stimulus, they appear less responsive in predicting the negative consequences of a conditioned stimulus compared to LA and MA. This becomes particularly evident once the conditioned stimulus ceased to be reinforced revealing a difficulty in

maintaining the prediction learned over time. As soon as the conditioned stimulus was no more reinforced by the aversive stimulus, the emotional value that HA had learnt to attribute to the conditioned stimulus disappeared.

Physiological changes in the anticipation of negative emotional events have been proposed to be a crucial component of emotional experience (Seth, Suzuki, & Critchley, 2012) and they have the adaptive function of guiding attention towards the source of the events preparing the organism to effectively identify, respond and regulate the response to such event (LeDoux, 2014; Öhman & Mineka, 2001). Therefore, the anticipation of emotional events might be crucial for effective emotion processing. Results suggest that HA are less able to predict the coming emotional event and possibly its consequences, which would be crucial to allow rapid identification, response and regulation of the response to such event. This difference may represent a shared underlying mechanism contributing to the difficulties of this group in emotion processing, which are particularly evident in ambiguous contexts, such as the recognition of emotional stimuli during limited time constraints, decision making in moral dilemmas and emotional response regulation (Grynberg et al., 2012; Ihme, Sacher, Lichev, Rosenberg, Kugel, Rufer, Grabe, Pampel, Lepsien, Kersting, Villringer, Lane, et al., 2014; Patil & Silani, 2014b; Pollatos & Gramann, 2012).

Predicting the emotional future seems to involve more complex mechanisms than just learning about the contingency between CS+ and UCS (S. S. Y. Li & McNally, 2014). In fact, the individual is required to learn the causal relationship between the CS+ and UCS. At each learning trial, UCS acts as teaching signal strengthening the response to CS+. The strength of this teaching signal is modulated by predictions regarding the occurrence of UCS following the presentation of CS+ (McNally, Johansen, & Blair, 2011). A brain circuit seems to be responsible for such process involving the periaqueductal gray, relaying the UCS

teaching signal to the amygdala through indirect pathways via the thalamus, which then project to the medial prefrontal cortex (mPFC), orbitofrontal cortex (OFC) and anterior cingulate cortex (ACC). Once in the amygdala, the UCS teaching signal then modulates plasticity at CS+ input synapses strengthening the response to CS+ (McNally et al., 2011). The amygdala then sends an output to the regions that regulate activity in the autonomic nervous system, to generate changes in SCR (Knight, Nguyen, & Bandettini, 2005; Olsson & Phelps, 2007; Pape & Pare, 2010). Indeed, previous research has shown decreased activation of mPFC, ACC and amygdala during processing of negative emotional stimuli in HA (Reker et al., 2010; van der Velde et al., 2014). Similarly, this circuit might be less active also during to the anticipation of negative emotional events.

The lower physiological response in anticipation to UCS in HA was not reflected in lower subjective reports of fear and anxiety. HA reported comparable levels of anxiety and fear experienced at the presentation of CS+ to LA and MA. These data would support a decoupling between the subjective experience and physiological response to emotional stimuli in alexithymia (Eastabrook et al., 2013) and more broadly corroborate theories that argue for a distinction between processes generating the physiological response in threat conditioning and those that give rise to conscious feelings of fear and anxiety (LeDoux, 2014; Öhman & Mineka, 2001). In fact, while the amygdala is crucial structure in generating SCR to CS+, cortical areas seem to be involved in attributing meaning to interoceptive inputs to construct the experience of an emotion (Lindquist et al., 2012). Such dissociation has been observed in a number of other conditions. For example, patients with lesions to the amygdala have shown diminished (LaBar et al., 1995) or absent (Bechara et al., 1995) SCR to an aversively conditioned stimulus, despite intact unconditioned response and awareness about the association between conditioned and unconditioned stimulus. On the contrary, patients with split brain (Làdavas et al., 1993), hemispatial neglect (Tamietto et al., 2015) and

affective blindsight (Bertini et al., 2013; Cecere et al., 2014; Tamietto et al., 2009) have shown intact physiological response in the absence of awareness for emotional stimuli. Nevertheless, although physiological responses and awareness for emotion can be separated, somatic and interoceptive information regarding one's own body is generally incorporated with semantic and contextual knowledge to generate an integrated representation of affective state (Barrett et al., 2007; Critchley et al., 2004) and this might be the case for LA and MA. However, in HA the physiological and cognitive aspect of emotional experience may remain decoupled possibly contributing to their difficulties. Despite comparable cognitive aspects of emotional experience, lower physiological response in anticipation of emotional events might hinder HA from effectively preparing to respond to emotional events.

To conclude, the present experiment shows that HA exhibit decreased prediction of an aversive emotional event compared to LA and MA, indicating a disruption in HA in learning to attribute a corresponding emotional value to previously neutral stimuli and use them as cues to predict the emotional future. The next experiment will investigate whether this difficulty is present also in the case of appetitive stimuli.

Experiment 4: no difference in psychophysiological correlates of prediction of appetitive stimuli during reward conditioning in alexithymia

Introduction

The previous experiment showed that HA presented reduced psychophysiological response in anticipation of aversive emotional events during threat conditioning. Nevertheless, it remains to be studied whether such impairment in emotional predictions is also evident for appetitive events during reward conditioning.

Similarly to pavlovian threat conditioning, pavlovian reward conditioning enables the individual to learn to predict the occurrence of appetitive events fundamental for survival and to prepare to efficiently approach them. In this form of conditioning, a neutral stimulus is repeatedly paired with an appetitive UCS, such as the delivery of monetary reward, so that the neutral stimulus becomes a CS anticipating the delivery of the reward. Again, increase in SCR and changes in subjective affective experience to the CS are observed indicating learning of the association between CS and UCS (Delgado, Gillis, & Phelps, 2008).

The aim of this experiment was to investigate the differences between LA, HA and MA in learning to predict an appetitive emotional event during pavlovian reward conditioning. Sixty LA, MA and HA, as measured by the 20-item Toronto Alexithymia Scale (TAS-20) (Taylor et al., 2003), completed a classical reward conditioning task with partial reinforcement. On each trial, one of two colored squares was presented. During habituation none of the stimuli was reinforced to ensure there were no baseline differences in response to the stimuli. During acquisition one stimulus was reinforced with the sound of a coin dropping in a piggy-bank (UCS) on 80% of trials (CS+) while the other was never reinforced (CS-). During extinction no sound was played. Changes in SCR were recorded continuously during the experiment as a psychophysiological indicator of the degree of prediction of the UCS. In

addition, to assess subjective experience, participants reported the level of liking, excitement and happiness experienced at the presentation of each CS. HA were hypothesized to show decreased anticipation of the UCS following the presentation of CS+, hence exhibit lower SCR to CS+ compared to MA and LA. No differences between MA and LA were expected.

Methods

Participants

Three-hundred university students completed the 20-item Toronto Alexithymia Scale (TAS-20)(Taylor et al., 2003). Depending on the score, students were classified as LA (TAS-20 $\leq$ 36), MA (36 < TAS-20 < 61) or HA (TAS-20 $\geq$ 61) (Franz et al., 2004). Individuals from the three groups were randomly contacted and asked to participate in the study. Due to the high co-occurrence of alexithymia and depression (S. Li et al., 2015), participants completed the Beck Depression Inventory (Beck et al., 1961), and were excluded from the in case their score was higher than the moderate/severe depression cut-off (i.e. 19; n = 3).

Sixty university students with no history of neurological or psychiatric disorders completed the study and were included in the analysis (21 males; age M = 21.94, SD = 1.69 years) divided in three groups: 20 LA participants (7 males; TAS-20 M = 31.89, SD = 2.51; age M = 21.44, SD = 1.60 years); 20 MA participants (7 males; TAS-20 M = 44.21, SD = 6.09; age M = 22.5, SD = 1.83 years); 20 HA participants (7 males; TAS-20 M = 63.73, SD = 2.31; age M = 21.88, SD = 1.54 years). A priori targets for sample size and data collection stopping rule were based on sample and effect sizes reported in the literature on classical conditioning (sample size around 17-19 participants per group as indicated by a recent meta-analysis (Fullana et al., 2015)).

Levels of anxiety were measured with the State-Trait Anxiety Inventory (Spielberger et al., 1983). Levels of anxiety in the three groups differed significantly both for state ($F(2, 57) = 11.27$, $p<.001$) and trait anxiety ($F(2, 57) = 21.40$, $p < .001$). Post-hoc Newman-Keuls test showed that for state anxiety LA ($M = 34.26$) had significantly lower levels of anxiety compared to HA ($M = 41.42$; $p < .001$) but not to MA ($M = 35.68$; $p = .377$), while for trait anxiety all groups differed significantly from each other, with LA ($M = 38.16$) showing lower levels of trait anxiety compared to MA ($M = 44.32$; $p = .008$) and HA ($M = 52.84$; $p < .001$) and MA showing lower levels of trait anxiety compared to HA ($p < .001$). Correlations between levels of anxiety and differential SCR response were explored to exclude an effect of anxiety on results. Neither trait nor state anxiety correlated significantly with difference in SCR to the two conditioned stimuli (all $p \geq .194$).

The study was designed and conducted in accordance with the ethical principles of the World Medical Association Declaration of Helsinki and the institutional guidelines of the University of Bologna and was approved by the Ethics Committee of the Department of Psychology. All participants gave informed written consent to participation after being informed about the procedure of the study.

Independent measure

Pavlovian reward conditioning. The task consisted in a classical differential reward conditioning paradigm with partial reinforcement. Two isoluminant colored squares represented the CS. The UCS consisted of a sound of a coin dropping into a piggy bank of 1000ms in duration, which indicated participants were earning 1€. The volume of the sound was kept constant across participants.

Each trial consisted in the presentation of one CS in the centre of a computer screen (17'', refresh rate 60Hz) for 6 seconds followed by an inter trial interval of 12 seconds during

which a fixation cross was presented (Fig. 8). The task included 44 trials (22 for each CS) divided in three blocks: habituation, acquisition and extinction.

At the beginning of habituation, instructions appeared on the screen stating that two different images would be presented one at the time in the centre of the screen, no stimulation would be administered and the task of the participant would be to carefully observe the images. Habituation included 4 trials (2 for each CS) to ensure the absence of any baseline differences within and between groups in response to the images. At the beginning of acquisition similar instructions stated that the same two images would appear one at the time in the centre of the screen and that one of them might be paired with the sound of the coin and the task of participants was to carefully observe the images. No information was given about the contingencies between images and the sound. Acquisition included 20 trials (10 for each CS). CS+ was reinforced in 80% of the trials (n = 8), while CS- was never reinforced. Extinction followed acquisition without any instructions. It included 20 trials (10 for each CS) and no sound was played.

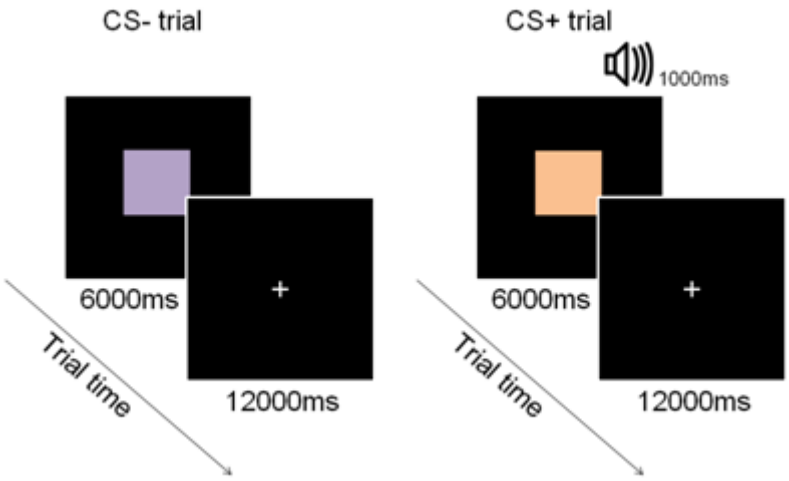

Figure 8. Illustration of a CS- trial and a CS+ trial, in which the presentation of the CS+ was reinforced by the sound of the coin.

Stimuli were presented in pseudo randomized order, no more than two presentations of the same stimulus occurred in a row (Schiller et al., 2008). The first two trials of acquisition always included one CS- and one CS+ presented randomly. The color of the square associated to the CS+ and CS- was counterbalanced across participants.

Dependent measures

Skin conductance response recording and processing. The SCR was recorded through two Ag/AgCl electrodes (TSD203 Model; Biopac Systems, USA), filled with isotonic hyposaturated conductant attached to the distal phalanges of the second and third finger of participants' left hand and held with Velcro straps. The SCR signal was continuously recorded at 200Hz and amplified using a DC amplifier (Biopac GSR100; Biopac Systems, USA) with 5μS/V gain factor and 10Hz low pass filter. The analogue signal was digitalized using the MP-150 digital converter (Biopac Systems, USA) and fed into AcqKnowledge 3.9 software (Biopac Systems, USA).

SCR data were analyzed using MATLAB (The MathWorks, Inc., USA) custom-made scripts (Garofalo et al., 2014). SCR was calculated as the peak-to-peak amplitude difference of the largest deflection in the 0.5-4.5sec latency window after stimulus onset. Regarding SCR to UCS, stimulus onset was represented by the beginning of the sound playing, while regarding SCR to CS, stimulus onset referred to the time of CS appearance. The SCR was transformed into microsiemens (μS) and calculated for each trial. Minimum response criterion was 0.02μS and smaller responses were encoded as zero. Square root transformation was conducted on raw SCR to normalize the data distribution and SCR were scaled to each subject's maximal UCS response to account for interindividual variability (Schiller et al., 2008).

Both SCR to UCS and CS- were analyzed to ensure that groups did not differ in their physiological response to the sound or in the anticipatory response to CS- but only to a conditioned stimulus that predicts an appetitive event (i.e. CS+). Regarding SCR to UCS, both peak response and average response were analyzed. For each participant, peak response represented the highest SCR in response to the eight coins while mean response was the average of the SCRs to the eight coins. Regarding the response to CSs, SCRs during the three phases of conditioning were analyzed separately. Concerning habituation, all trials were included in the analysis. With regards to acquisition, the first two trials were not included in the analysis because participants learned the association between UCS and CS+ after its first pairing with the coin sound. Regarding extinction, all trials were included in the analysis but this phase was divided in two blocks (early and late extinction). Then, mean SCR of each participant was computed to produce four average scores representing the SCR of each subject during habituation, acquisition, early and late extinction. These were then averaged to obtain the SCR during the different phases for each alexithymia group.

Assessment of subjective liking, excitement, happiness and contingency awareness. Participants were asked to report how much they liked each CS and the level of excitement and happiness experienced at the presentation of each CS on separate visual analogue scales ranging from 0 (not at all) to 100 (extreme). The order of questions was counterbalanced across participants.

Participants were also asked to indicate which of the two stimuli was associated with the coin to ensure explicit awareness of pairing between CS+ and UCS. All participants reported the correct association.

Procedure

The experiment took place in a sound attenuated room with dimmed light. Participants were seated in a chair in front of a computer monitor at ~70 cm distance. Once seated, the experimental procedure was explained and written informed consent was obtained from participants. Then SCR electrodes were attached and correct recording of the signal was ensured. Participants were then told the task consisted of two phases and that instructions would appear on the screen for each phase. Also, participants were familiarized with the sound of the coin and told it might play in the second phase. They were also reassured that the coins earned were real money they would be given at the end of the session. Following completion of the task, subjective reports were completed and the reward was given.

Results

No difference in peak and mean SCR to UCS among groups

Univariate ANOVAs were used to evaluate differences in UCS intensity and mean and peak SCR to the UCS. Results showed no significant differences among the three groups in either peak SCR in response to UCS (M_{LA} = 0.73 μS, M_{MA} = 0.83 μS, M_{HA} = 0.92 μS; $F(2, 57) = 2.28$, $p = .111$, partial $\eta^2 = .07$) or mean SCR in response to UCS (M_{LA} = 0.40 μS, M_{MA} = 0.43 μS, M_{HA} = 0.49 μS; $F(2, 57) = 0.72$, $p = .489$, partial $\eta^2 = .02$). On average, the physiological response to the UCS did not differ significantly among groups.

No difference in SCR among groups or stimuli during habituation

A 3x2 RM ANOVA (group: LA, MA, HA; stimulus: CS+, CS-) was carried out to analyze SCR during habituation. Analysis on SCR showed no significant main effect of group ($F(2, 57) = 1.69$, $p = .194$, partial $\eta^2 = .05$), stimulus ($F(1, 57) = .71$, $p = .400$, partial $\eta^2 = .01$) or interaction ($F(2, 57) = 1.87$, $p = .163$, partial $\eta^2 = .06$), confirming that at baseline

there were neither within group nor between group differences in response to the two conditioned stimuli.

No difference between groups in the acquisition and extinction of conditioning

A 3x2x3 RM ANOVA (group: LA, MA, HA; stimulus: CS+, CS-; phase: acquisition, early extinction, late extinction) was carried out to analyze SCR in the phases following habituation. Analysis on SCR showed significant main effect of phase ($F(2,114) = 7.03$, $p = .001$, partial $\eta^2 = .11$). However, this was secondary to a significant stimulus by phase interaction ($F(2,114) = 3.44$, $p = .035$, partial $\eta^2 = .06$). Post-hoc Newman Keuls test indicated that during acquisition participants had higher SCR to the CS+ than CS- ($p=.004$; $M_{CS+}=.12$, $M_{CS-}=.16$; Fig 9). On the contrary there was no difference in SCR to the conditioned stimuli during extinction (all $ps \geq .855$; Fig 10). Also, no main effect or interaction with the factor group was found (all $ps \geq .232$). The results indicate that groups showed comparable acquisition and extinction of appetitive conditioning.

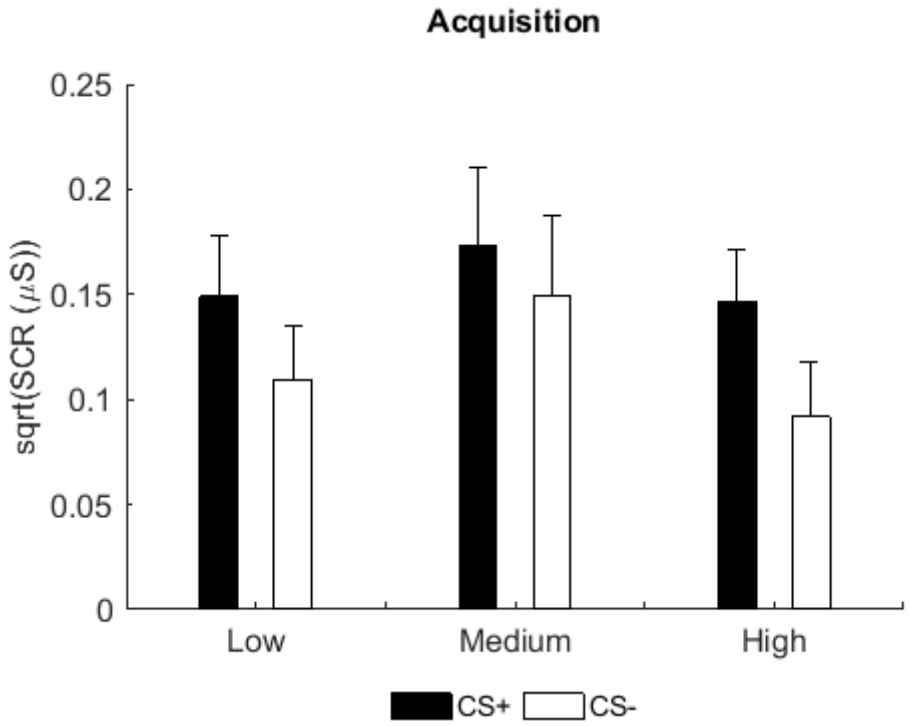

Figure 9. Mean square rooted skin conductance response (SCR) to CS+ and CS- during acquisition as a function of alexithymia group. In all groups, SCR was higher to the CS+ than CS-. This differential response appears comparable across groups, as suggested by a main effect of stimulus shown by the statistical analysis, indicating significantly higher SCR to CS+ than CS-. Error bars represent standard errors.

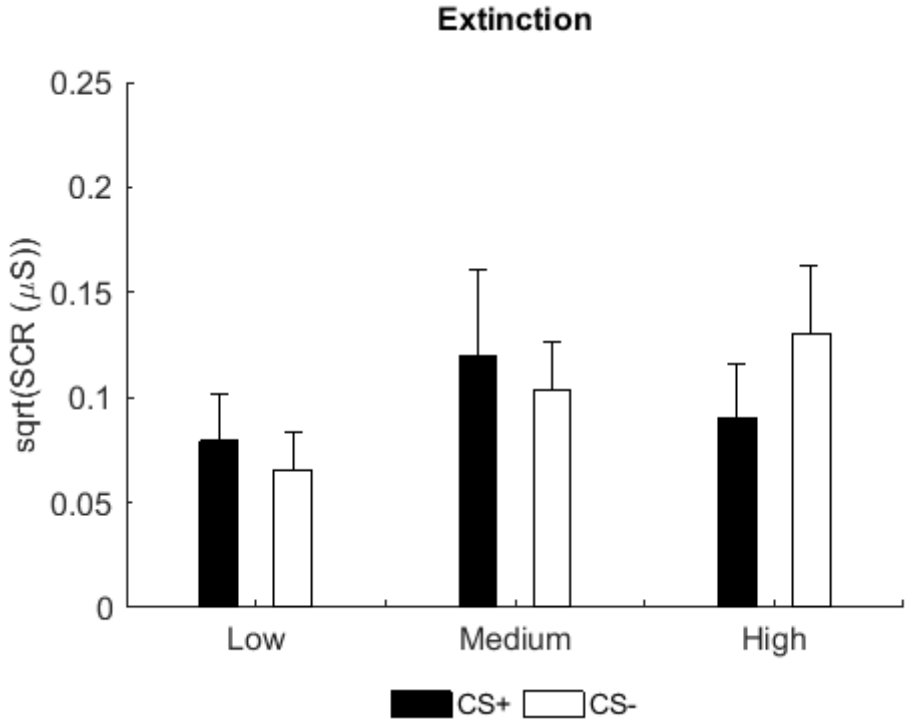

Figure 10. Mean square rooted skin conductance response (SCR) to CS+ and CS- during extinction as a function of alexithymia group. All groups extinguished differential SCR between CS+ and CS-, as suggested by the statistical analysis. Error bars represent standard errors.

No difference in subjective reports of liking, excitement and happiness among groups

3x2 RM ANOVAs (group: LA, MA, HA; stimulus: CS+, CS-) were conducted on subjective reports of liking, excitement and happiness experienced at the presentation of the CSs. The subjective report on liking showed a significant main effect of stimulus ($F(1, 57) = 37.48$, $p < .001$, partial $\eta^2 = .40$). Participants liked the CS+ more than the CS- ($p < .001$, $M_{CS+} = 60.64\%$, $M_{CS-} = 35.38\%$). No significant main effect of group or interaction was found (all ps≥.294).

The subjective report on excitement showed a main effect of stimulus ($F(1, 57) = 69.69$, $p < .001$, partial $\eta^2 = .55$) and a stimulus by group interaction ($F(1, 57) = 3.66$, $p = .032$, partial $\eta^2 = .11$). Nevertheless, this appeared to be driven by a spurious comparison, because within each group participants were more excited at the presentation of the CS+ than the CS- (all ps≤.002) and between groups there were no differences in the degree of excitement to the CS+ (all ps≥.122) or CS- (all ps≥.113) .

The subjective report on happiness showed a significant main effect of stimulus ($F(1, 57) = 58.03$, $p < .001$, partial $\eta^2 = .50$). Participants were happier at the presentation of the CS+ than the CS- ($p < .001$, $M_{CS+} = 57.17\%$, $M_{CS-} = 28.13\%$). No significant main effect of group or interaction was found (all ps≥.062).

Discussion

This experiment investigated whether alexithymia presented reduced prediction of appetitive emotional events. To this end, changes in SCR were recorded during pavlovian reward conditioning to assess differences among LA, MA and HA in predicting the occurrence of a positive emotional event by learning patterns of association between CS+ and UCS.

Similarly to the previous experiment, behavioral results showed that participants correctly associated CS+ and UCS, suggesting that they explicitly identified the stimulus that anticipated the appetitive event. In addition, groups did not differ in the SCR to the appetitive stimulus and emotional experience reported in response to presentation of CS+. Specifically, all groups reported comparable arousal in response to the appetitive stimulus. In addition, they also reported higher levels of subjective liking, excitement and happiness in response to the stimulus predicting the appetitive UCS than the neutral one. However, contrary to the results found during threat conditioning, no significant difference was found in physiological arousal in anticipation of the appetitive stimulus among the three alexithymia groups in any of the phases of conditioning. Indeed, all groups appeared to learn to discriminate a neutral stimulus from a stimulus predicting an appetitive event. In addition, HA appeared to acquire and extinguish the prediction of a coming appetitive stimulus to a degree comparable to LA and MA.

The difference in results between the two experiments may be driven by differences in the neural circuit underling threat and reward conditioning and the dependent measure used to assess prediction of the UCSs. Although changes in SCR are a standard dependent measure to assess reward conditioning, these might be differentially modulated in threat as opposed to reward learning. Indeed, changes in SCR have been shown to be driven by variation in neural activity in the connections from the central complex of the amygdala to the hypothalamus (Cardinal et al., 2002; McNally & Westbrook, 2006). In this regard, a recent study, performing cell activity recording in mice, showed different neural populations responded to the anticipation of threat and reward. Specifically, the anticipation of aversive stimuli seems to mainly excite neurons projecting from the basolateral amygdala to the central amygdala, while, that of appetitive stimuli, neurons projecting from the basolateral amygdala to the nucleus accumbens (Beyeler et al., 2016). Therefore, the amygdala might

have a more prominent role in modulating SCR in threat than reward conditioning. From the data, it can be noted that the overall level of SCR is higher during threat conditioning than reward conditioning. Indeed, while there is evidence of its causal role in the acquisition of threat conditioning measured through SCR in humans (Bechara et al., 1995; LaBar et al., 1995), the causal role of the amygdala in the acquisition of reward conditioning is less definitive (Murray & Baxter, 2002; though see Bechara et al., 1999). Possibly, in this experiment, changes in SCR might have been sensitive in discriminating between CS- and CS+, but might not be sensitive enough in uncover subtle group differences in emotional learning.

To conclude, the current experiment did not find any significant difference in learning to predict appetitive stimuli in alexithymia. The present results, together with those of the previous experiment, may suggest that the difficulties of HA in predicting emotional events may be more severe for aversive than appetitive emotional events. Nevertheless, it is possible that this negative result may be related to the dependent measure used to assess reward conditioning. This leaves open the possibility that HA may still have difficulties in predicting appetitive emotional stimuli, which may become evident with measures of neural rather than physiological activity.

Experiment 5: reduced electrophysiological correlates of prediction and prediction error to appetitive stimuli during pavlovian reward conditioning in alexithymia

Introduction

While Experiment 3 showed reduced prediction of aversive stimuli in alexithymia during threat conditioning, Experiment 4 did not show a similar impairment in the prediction of appetitive stimuli during reward conditioning. Nevertheless, the result of Experiment 2 may partly be related to the dependent measure used in the experiment, leaving open the possibility of alexithymia being indeed related to impairment in predicting coming appetitive stimuli, which may become evident with the use of a dependent measure that reflects a more direct measure of neural activity related to emotional prediction. Therefore, recording of electrophysiological activity through electroencephalogram (EEG) and specifically event related potentials (ERP) representing emotional prediction signals could be used. In this regard, the CUE-P300 component has been previously shown to be sensitive to the anticipation of reward (Pfabigan et al., 2014; Sutton, Braren, Zubin, & John, 1965). This is a positive deflection in electrical potential observed at centroparietal electrodes peaking 300-600ms after CS presentation, which appears to have a role in the evaluation of motivationally salient stimuli (Pornpattananangkul & Nusslock, 2015), being larger for stimuli predicting reward compared to no-reward (Broyd et al., 2012; Goldstein et al., 2006) and could be used to assess the prediction of appetitive stimuli during pavlovian reward conditioning.

In addition to prediction, prediction error plays an equally fundamental role in constructing accurate internal representation of emotional stimuli and learning in general. In fact, during associative learning, on each trial the organism compares the reinforcer delivered (or not) on that trial (e.g. monetary reward) with its own prediction about the delivery of that

reinforcer, which had been learned during previous trials (Daw & Tobler, 2014). In case of a mismatch between the prediction and the actual outcome, a prediction error occurs, which signals the need for more learning to update the prediction (Friston, 2010). The EEG can be used to also record electrophysiological correlates of such prediction error, recording an ERP named feedback related negativity (FRN). This is a negative deflection in electrical potential observed at frontocentral electrodes between 200-350ms after a stimulus indicating whether reward or no reward is delivered (Sambrook & Goslin, 2015). As such, it is calculated as the difference wave resulting from subtracting the electrical potential following reward omission from that following reward delivery (Sambrook & Goslin, 2015). Importantly, this component is modulated by expectations, such that the difference wave is more negative for unexpected compared to expected events, corroborating its role in encoding a prediction error signal (e.g. Hajihosseini & Holroyd, 2013; Talmi et al., 2013; Walsh & Anderson, 2012).

The aim of the current experiment is to assess differences in electrophysiological correlates of predictions as well as prediction errors in HA and LA for appetitive stimuli. To this end, EEG was recorded continuously from HA and LA participants during a pavlovian reward conditioning task, in order to assess CUE-P300 during the anticipation of reward and FRN in response to the unexpected reward-related feedback. During the task, two CS (CS1 & CS2) were presented. Each CS was followed by a feedback indicating delivery of monetary reward or of no-reward. In 80% of trials, CS1 was associated to the delivery of reward, while CS2 to the delivery of no-reward. As the task progresses, LA should learn such associations and construct corresponding predictions of CS1 anticipating reward while CS2 no-reward. Therefore LA are hypothesized to have enhanced CUE-P300 in response to CS1 than CS2. On the contrary, HA are hypothesized to have difficulties in learning such predictions and present a reduction in the modulation of the CUE-P300 compared to LA. Importantly, the task also included two events that violate the learned predictions. Specifically, in the

remaining 20% of trials, the stimulus-feedback association was inverted, resulting in an unexpected reward-related feedback. In LA, this is hypothesized to lead to a prediction error, manifested as enhanced FRN in response to the unexpected compared to the expected reward-related feedback. On the contrary, HA are hypothesized to have reduced prediction error, because of the reduced prediction. Therefore, HA should have reduced modulation of the FRN compared to LA. In addition to EEG data, behavioral measures of participants' degree of learning and changes in subjective experience in response to the CS were also collected, although, given the results of previous experiments, no difference between groups was expected on these measures.

Methods

Participants

Three-hundred individuals completed the 20-item Toronto Alexithymia Scale (TAS-20; Taylor, Bagby, & Parker, 2003). Depending on the score, students were classified as LA (TAS-20 ≤ 36) or HA (TAS-20 ≥ 61) (Franz et al., 2004) and were then randomly contacted to participate in the study. Once in the laboratory, the alexithymia module of the structured interview for the Diagnostic Criteria for Psychosomatic Research (DCPR; (Mangelli et al., 2006) was administered to increase reliability of screening and confirm TAS-20 classification. Participants with discordant classification on the two measures did not complete the task (n = 1). Due to the high co-occurrence of alexithymia and depression (S. Li et al., 2015), participants completed the Beck Depression Inventory (Beck et al., 1961) and did not complete the experimental task in case their score was higher than the moderate/severe depression cut-off (i.e. 19, n = 1). Levels of anxiety were measured with the State-Trait Anxiety Inventory (Spielberger et al., 1983).

All participants had equivalent educational backgrounds and were students at the University of Bologna. The study was designed and conducted in accordance with the ethical principles of the World Medical Association Declaration of Helsinki and the institutional guidelines of the University of Bologna and was approved by the Ethics Committee of the Department of Psychology. All participants gave informed written consent to participation after being informed about the procedure of the study.

Forty-two volunteers with no history of major medical, neurological or psychiatric disorders completed the study (22 LA, 21HA). Data from two LA participants and 1 HA participant were removed from further analysis due to excessive noise in the EEG signal. In total 40 participants were included in the analysis: 20 LA (6 males; age $M = 21.97$, $SD = 1.57$ years; TAS-20 $M = 31.80$, $SD = 2.82$; STAI state $M = 35.16$, $SD = 5.08$; STAI trait $M = 40.65$, $SD = 6.38$); 20 HA (6 males; age $M = 21.97$, $SD = 2.27$ years; TAS-20 $M = 64.00$, $SD = 4.30$; STAI state $M = 37.65$, $SD = 6.09$; STAI trait $M = 48.47$, $SD = 8.80$). HA had significantly higher trait anxiety than LA ($t(38)=3.19$, $p=.003$); nevertheless, correlations between anxiety and the dependent measures were not significant (all $ps \geq .081$).

A priori targets for sample size and data collection stopping rule were based on sample and effect sizes reported in a previous EEG study on alexithymia and error related negativity (ERN) (Maier et al., 2016), a mediofrontal component that is also considered a prediction error signal following performance error (Alexander & Brown, 2011; Gehring, Goss, Coles, Meyer, & Donchin, 1993).

Independent measures

Pavlovian appetitive conditioning. Two neutral stimuli representing the conditioned stimuli (CS) were presented followed by a feedback indicating the delivery of reward or of no-reward. Each CS consisted in a 3cm white square with a Japanese hiragana on it, reward

consisted in the writing "1€" and no-reward in the writing "0€". In order to manipulate reward expectations, the percentage of reward delivery was varied for each CS. Specifically, CS1 was followed by the delivery of reward in 80% (i.e. expected reward condition) of the trials and no-reward in 20% of trials (i.e. unexpected no-reward condition). On the contrary, CS2 followed by to the delivery of no-reward in 80% of the trials (i.e. expected no-reward condition) and reward in 20% of trials (i.e. unexpected reward condition) (Table 2). Stimuli were displayed on a 17-inch color monitor (refresh rate 60Hz) with a black background, at a viewing distance of 80 cm. A PC running E-prime 2.0 (Psychology Software Tools, Pittsburgh, PA) controlled stimulus presentation and behavioral response recording. Stimuli were balanced for luminance, complexity and color saturation.

Table 2. Illustration of experimental conditions with an example CS1 and CS2 stimuli.

Stimulus	Feedback Type	Feedback Expectancy	Experimental condition
CS1 	1€	80%	Expected Reward
	0€	20%	Unexpected No-reward
CS2 	0€	80%	Expected No-reward
	1€	20%	Unexpected Reward

Each trial consisted in the presentation of a fixation cross in the center of the screen for 500ms, followed by the presentation of CS to the right or left of the fixation cross for 1500ms, followed by the feedback for 1000ms, followed by an inter trial interval of 1000-1500ms during which a blank screen was presented (Fig. 11). To ensure the attention of participants was maintained throughout the task, 48 catch trials were included during which, a scrambled image of the CS was presented, followed by the presentation of the blank screen. The task included 848 trials in total, divided in 8 blocks of 108 trials each. The order of trials was randomized and in each block the two CSs were presented the same number of times.

After the first four blocks, both CS were changed and participants had to learn the new stimulus-feedback associations. The assignment of each CS to a particular condition was counterbalanced among participants.

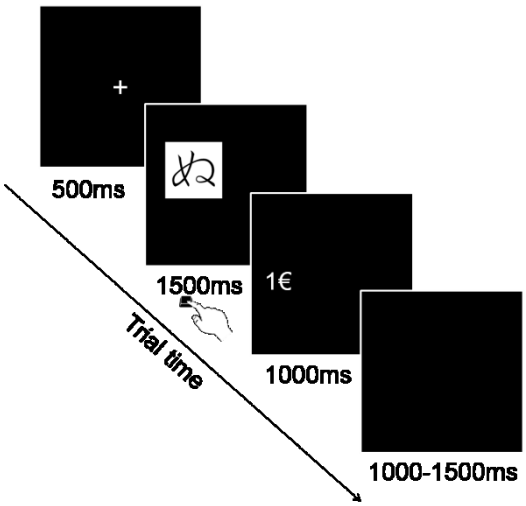

Figure 11. Illustration of experimental trial.

Dependent measures

Explicit assessment of learning. To evaluate understanding of the task and explicit learning of stimulus-feedback association, participants were asked to report how many stimuli they saw and to briefly describe them. All participants correctly reported to having seen two stimuli and were able to describe them. In addition, participants were asked what they noticed about the stimulus-feedback association. A score of 1 was given for correctly reporting which stimulus was mostly associated with which feedback, while a score of 0.5 for reporting that one stimulus was mostly associated with reward while the other with no-reward but failing to identify which one, all other responses were given a score of 0.

Self-report liking. To assess changes in subjective experience following conditioning, participants were asked to report how much they liked each CS on a visual analogue scale ranging from 0 (not at all) to 100 (extremely).

EEG recording and pre-processing. The electroencephalogram (EEG) was recorded with Ag/AgCl electrodes (Fast n Easy Electrodes; Easycap, Herrsching, Germany) from 59 electrode sites (Fp1, Fp2, AF3, AF4, AF7, AF8, F1, F2, F3, F4, F7, F8, FC1, FC2, FC3, FC4, FC5, FC6, FT7, FT8, C1, C2, C3, C4, C5, C6, T7, T8, CP1, CP2, CP3, CP4, CP5, CP6, TP7, TP8, P1, P2, P3, P4, P5, P6, P7, P8, PO3, PO4, PO7, PO8, O1, O2, FPz, AFz, Fz, FCz, Cz, CPz, Pz, POz, Oz) and from the right mastoid. The reference electrode was placed on the left mastoid and the ground electrode on the right cheek. Signal impedance was maintained below 10 KΩ. The electro-oculogram (EOG) was recorded from above and below the left eye and from the outer canthi of both eyes. EEG and EOG were recorded with a band-pass filter of 0.01–100 Hz, amplified by a BrainAmp DC amplifier (Brain Products, Gilching, Germany) and digitized at a sampling rate of 1000 Hz.

The EEG data were pre-processed using EEGLAB free toolbox, version 14.1.0b (Delorme & Makeig, 2004), and custom routines written in MATLAB R2013a (The MathWorks, Natick, MA). The ERP data were re-referenced offline to the right mastoid (Luck, 2014) and filtered with a 1- to 30-Hz band-pass filter. To extract the signal in response to the CSs, stimulus-locked epochs from –100 to 700ms relative to the appearance of the CS were extracted from the continuous EEG. To extract the signal in response to the feedbacks, stimulus-locked epochs from –100 to 500ms relative to the appearance of the feedback were extracted from the continuous EEG. Epochs were baseline-corrected using the average voltage during the 100ms pre-stimulus window. Epochs were excluded in case the voltage on a single channel exceeded 400μV to remove trials with large voltage fluctuations and in case they contained data deviating more than 5SDs from the mean of the joint probability distribution to remove improbable data. To correct remaining artifacts, the data were subjected to a temporal independent component analysis (Jutten & Herault, 1991; Makeig, Bell, Jung, & Sejnowski, 1996) using the infomax algorithm (Bell & Sejnowski, 1995).The

resulting component matrix was screened for independent components (ICs) representing stereotyped artifact activity, such as horizontal (saccades) and vertical (blinks) eye movements, and muscle artifacts. This was done using a multistep correlational template-matching process, implemented in CORRMAP, version 1.03 (Viola et al., 2009). Topographies of ICs labeled as artifacts by the CORRMAP procedure were visually inspected and then calculated out of the data using inverse matrix multiplication. Finally, to avoid a possible influence of differing numbers of trial on the average ERP results, the trial number was matched for the four experimental conditions through data re-sampling (Garofalo et al., 2016). For each participant, the condition with the smallest number of trials was identified and the corresponding number of trials was randomly drawn from each of the remaining conditions for 1000 iterations. Then, an average ERP was calculated for each iteration separately for each condition. The resulting components were then averaged over all iterations separately for each condition producing four ERPs per participant used for data analysis.

ERP quantification

Cue-P300 in response to the stimulus. Scalp topographies of the mean voltage in the interval 300-500ms following CS appearance showed maximum activation at parietal electrodes in both groups (Fig. 12). The grand average waveforms at electrode Pz, showed a positive deflection peaking around 350ms after CS onset in both groups, resembling the CUE-P300 component (Fig. 13). Statistical analysis was performed on individual waveforms representing the mean amplitude recorded in the 300-500ms interval following CS appearance.

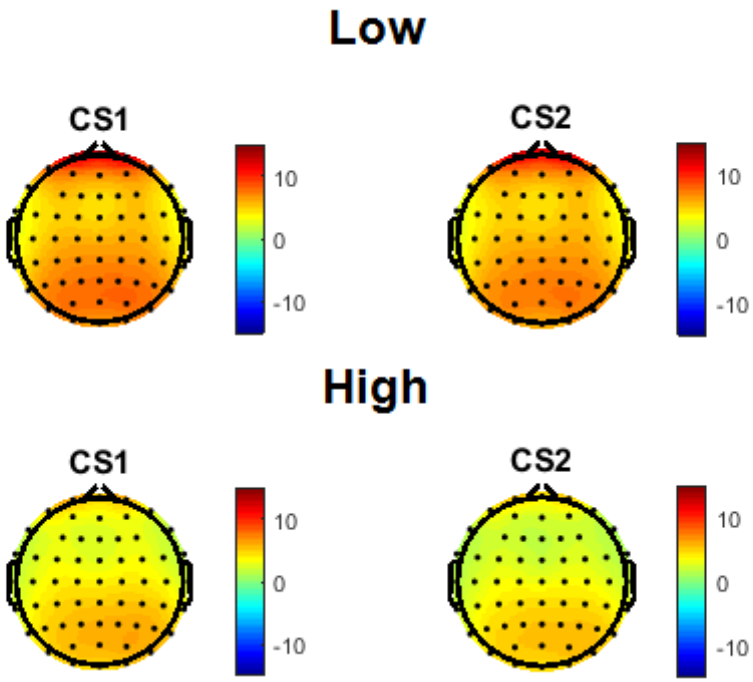

Figure 12. Scalp topographies of the mean voltage in the interval 300-500ms following CS1 and CS2 appearance for the low and high alexithymia group showing maximum activation at parietal electrodes in both groups

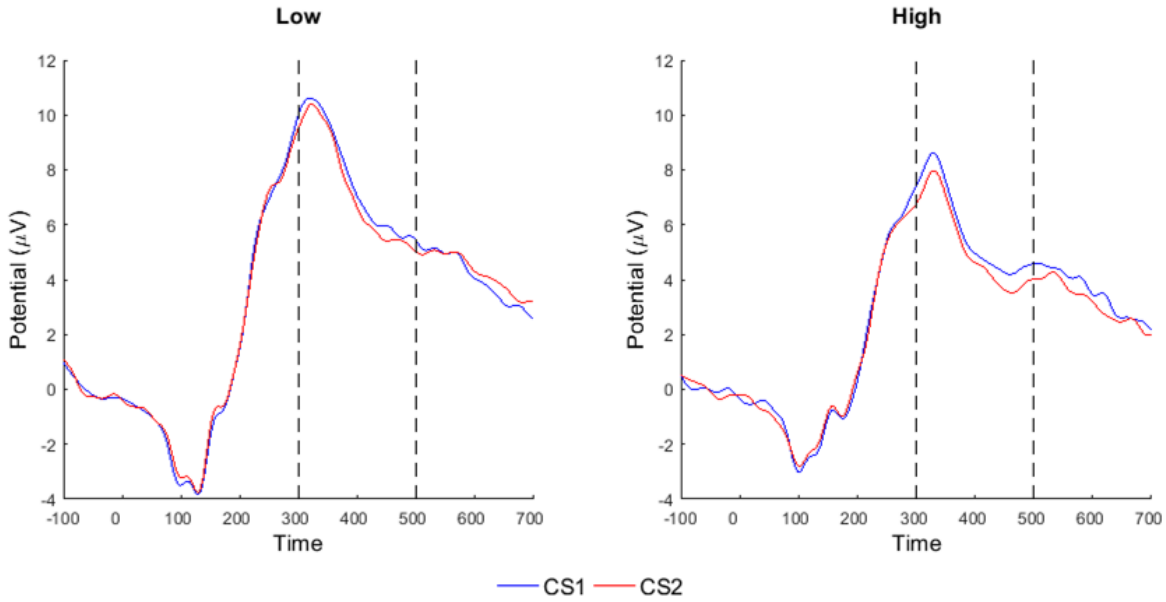

Figure 13. Grand average waveforms at electrode Pz for CS1 and CS2 for low and high alexithymia group. 0 on the x axis indicates time point of stimulus appearance; dashed lines indicate time interval for analysis.

FRN in response to the feedback. Scalp topographies of the mean voltage difference between reward and no-reward feedback in the interval 200-300ms following feedback

showed maximum difference in activation at central electrodes in both groups (Fig. 14). For each condition, the grand average waveforms at electrode Cz, where FRN has been previously reported (Sambrook & Goslin, 2015), showed a negative deflection peaking around 250ms after feedback onset in both groups (Fig 15). Separate FRNs were computed for the expected and unexpected conditions, calculating the difference wave between the no-reward and reward conditions for the expected and unexpected condition for each group. Statistical analysis was performed on the mean amplitude of the FRN recorded in the 200-300ms following feedback appearance.

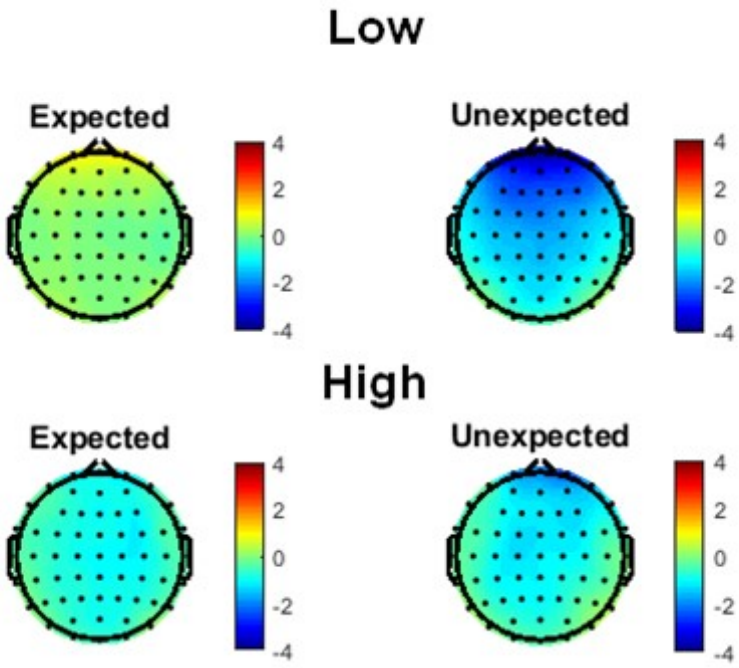

Figure 14. Scalp topographies of the mean voltage difference between no-reward and reward feedback in the interval 200-300ms following feedback for the expected and unexpected conditions in the low and high alexithymia group showing maximum difference in activation at central electrodes in both groups

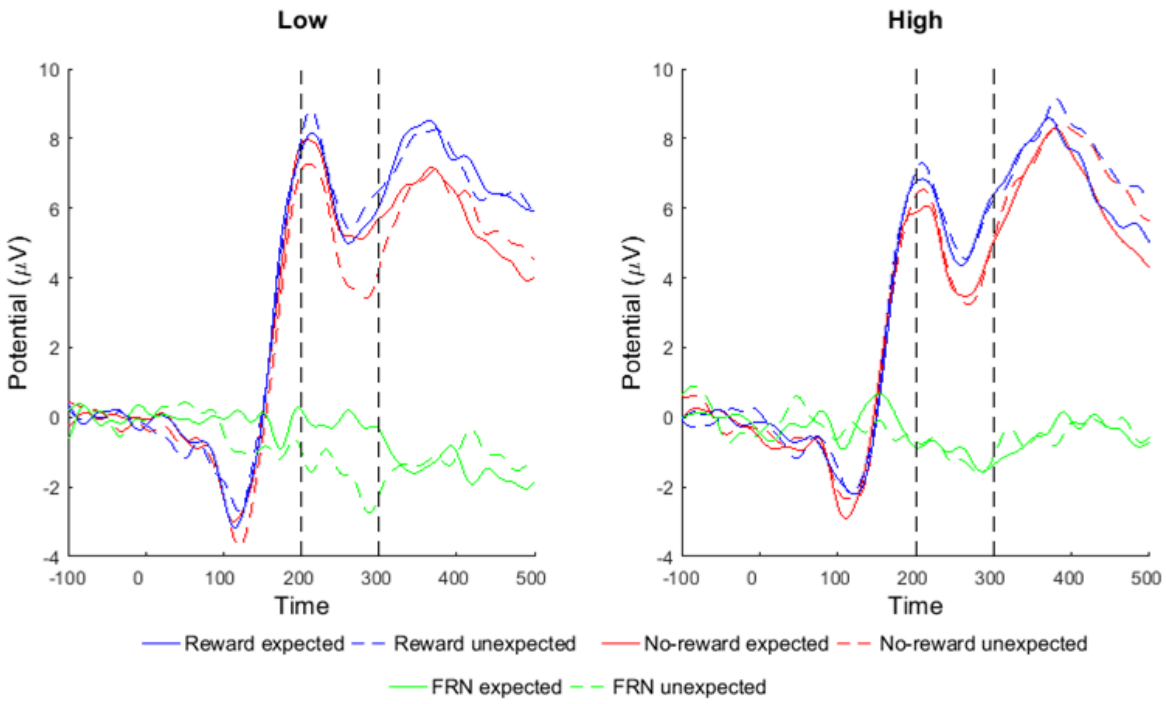

Figure 15. Grand average waveforms at electrode Cz for reward and no-reward in the expected and unexpected conditions, as well as FRN in the low and high alexithymia group. 0 on the x axis indicates time point of stimulus appearance; dashed lines indicate time interval for analysis.

Procedure

The task took place in a sound attenuated room with dimmed light. After the EEG cap was mounted, participants were asked to seat comfortably on the chair. They were instructed that on each trial they would see a stimulus appearing on the left or right of the screen followed by the feedback "1€" or "0€" that indicated the reward for that trial. Their task was to pay attention to the association between the stimulus and feedback because they would earn part of the reward at the end of the task. In addition, participants were instructed to press one of two keys on the keyboard corresponding to the side of presentation of the symbol, as soon as the stimulus appeared on the screen. They were asked to be as fast and as accurate as they could in their response, but were informed that neither speed nor accuracy of response would affect the feedback appearance to avoid the possibility they associated the reward with their actions. In addition, they were told that sometimes, a scrambled picture would appear

instead of the stimulus and they should not press any key. Participants could take a break to rest at the end of each block.

At the end of the fourth block and of the task, participants were asked to report how much they liked each CS. Also, participants were asked to report how many stimuli they saw and to briefly describe them. All participants correctly reported to having seen two stimuli and were able to describe them. In addition, participants were asked what they noticed about the stimulus-feedback association.

At the end of the session, participants received 10€ as a reward.

Results

Reduced prediction in anticipation of reward-related feedback in alexithymia

The 2x2x2 RM ANOVA (CS type: CS1, CS2; group: LA, HA) conducted on the mean amplitude of the CUE-P300 revealed a main effect of group. CUE-P300 amplitude in response to the CSs was larger for LA than HA ($F(1, 38)=5.25$, $p=.028$, partial $\eta^2 = 0.12$; $M_{LA}=7.53$, $M_{HA}=5.52$; Fig. 16). All other factors were not significant (all $ps \geq .112$). The result suggests HA had reduced anticipation of reward-related outcome compared to LA.

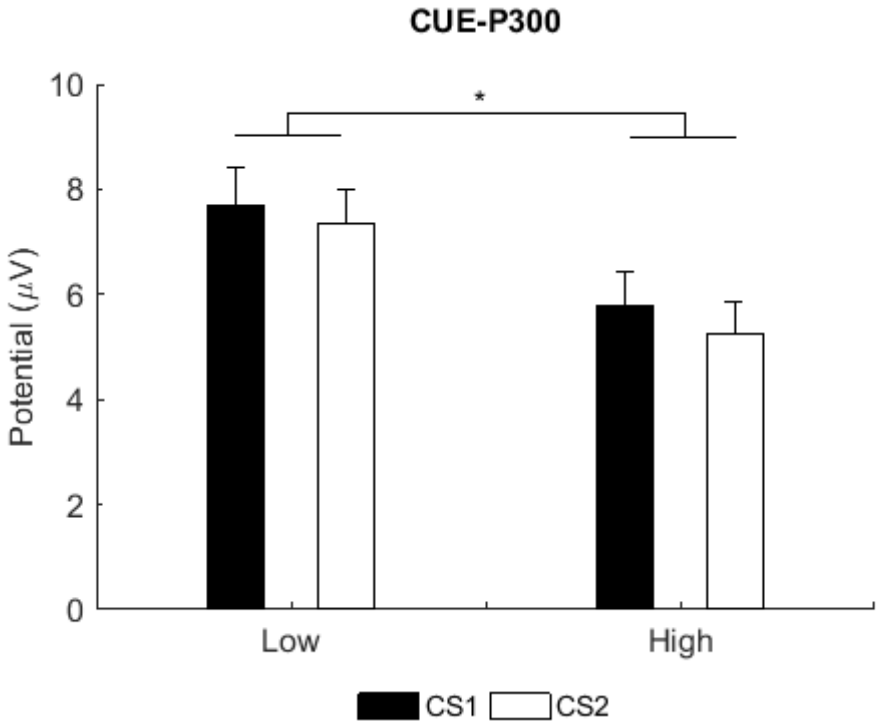

Figure 16. CUE-P300 mean voltage for CS1 and CS2 as a function of alexithymia group. Higher voltage was found for individuals with low than high alexithymia. Error bars represent standard errors. Significant differences are indicated as follows: *p< .05.

Absent prediction error to unexpected reward-related feedback in alexithymia

The 2x2 RM ANOVA (feedback expectation: expected, unexpected; group: LA, HA) conducted on the mean amplitude of the FRN revealed a main effect of feedback expectation ($F(1, 38)=4.39$, $p=.043$, partial $\eta^2 = 0.104$), which was secondary to the group by feedback expectation interaction ($F(1, 38)=4.39$, $p=.043$, partial $\eta^2 = 0.104$; Fig. 17). Bonferroni post-hoc test showed that LA had larger FRN for the unexpected compared to the expected condition ($p=.020$, $M_{unexpected}=-1.70$, $M_{expected}=-.143$). On the contrary HA, did not show any modulation of the FRN ($p=1.000$). All other comparisons were not significant ($p\geq.619$). These results suggest that HA did not compute a prediction error in response to unexpected reward-related outcome compared to LA.

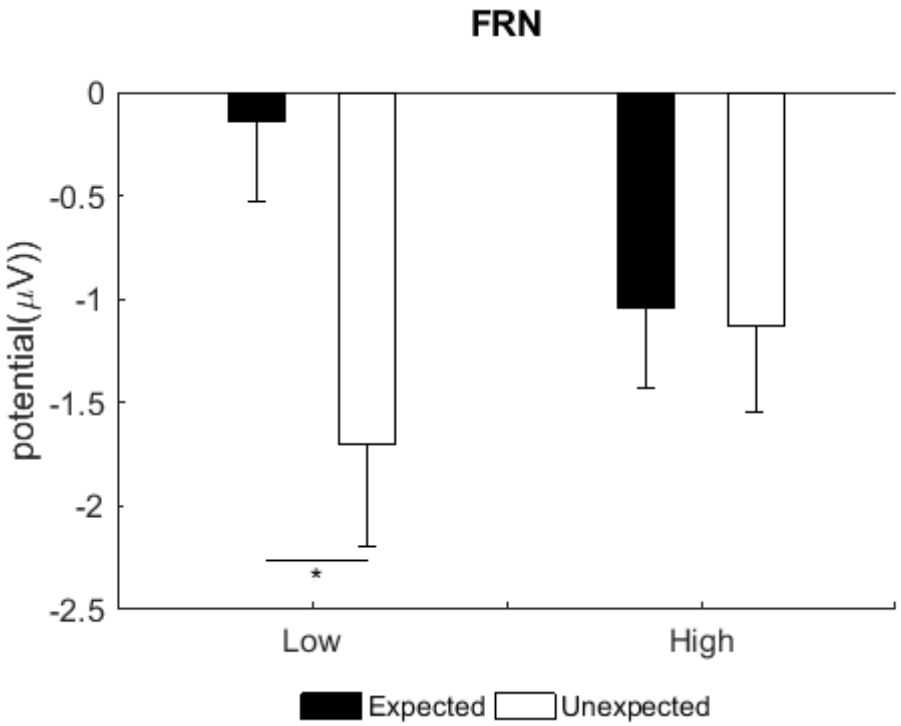

Figure 17. FRN mean voltage for expected and unexpected condition as a function of alexithymia group. A significant difference between the two conditions was found only in the low alexithymia group. Error bars represent standard errors. Significant differences are indicated as follows: *p< .05.

Reduced explicit learning of CS-feedback contingencies in alexithymia

The 2x2x2 RM ANOVA (CS type: CS1, CS2; block: 1-4, 5-8; group: LA, HA) showed a main effect of group ($F(1, 38)=4.22$, $p=.047$, partial $\eta^2 = 0.10$; Fig. 18). HA ($M=0.86$) showed less understanding of the CS-feedback contingencies compared to LA ($M=0.97$). All other factors were not significant ($p \geq .090$).

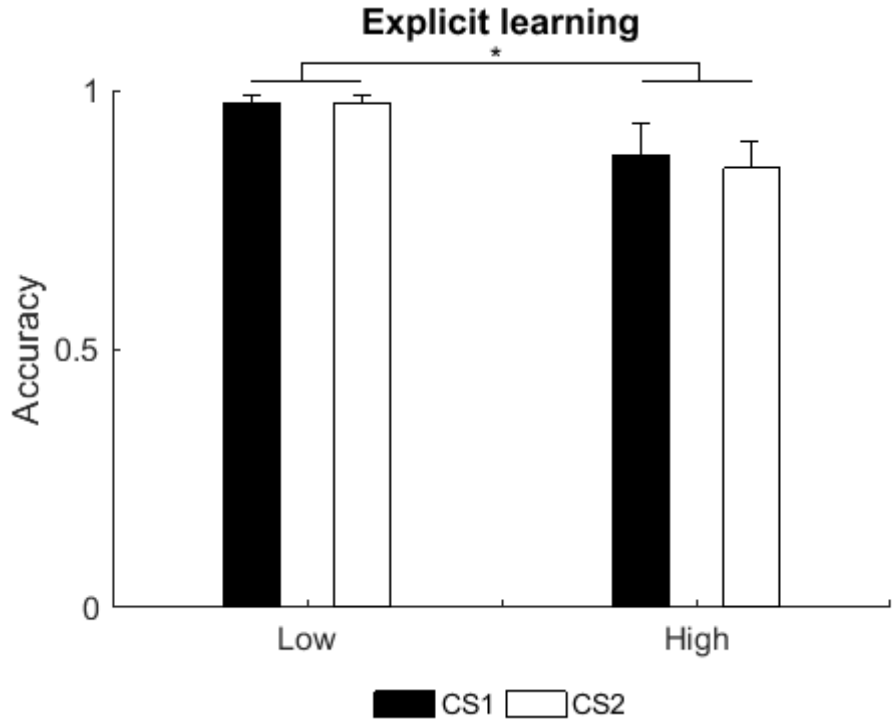

Figure 18. Mean accuracy of stimulus feedback association for CS1 and CS2 as a function of alexithymia group. The low alexithymia group had better accuracy in reporting the stimulus-feedback association. Error bars represent standard errors. Significant differences are indicated as follows: *p< .05.

No difference in subjective liking of CSs between groups

The 2x2x2 RM ANOVA (CS type: rewarding, non-rewarding; block: 1-4, 5-8; group: LA, HA) showed a main effect of CS type ($F(1, 37)=10.129$, $p=.003$, partial $\eta^2 = 0.21$). In both groups, participants liked the rewarding CS more than the non-rewarding CS ($M_{rewarding}=72.03$, $M_{non\text{-}rewarding}=55.91$). All other factors were not significant ($p\geq.193$). Note that data for one LA participant was missing because the software failed to record the responses.

Discussion

The current experiment tested the electrophysiological correlates of prediction and prediction error during reward conditioning in LA and HA. Electrophysiological results found that HA had reduced CUE-P300 during the anticipation of monetary reward-related feedback compared to LA. In addition, HA did not show an enhancement of FRN in response

to unexpected reward-related feedback compared to LA. Behavioral results showed that HA also had less understanding of the association between stimuli and reward-related feedback. These results indicate that HA have reduced prediction of coming rewards and prediction error in response to unexpected reward-related feedback suggesting decreased ability to construct accurate internal representations of coming appetitive stimuli.

The CUE-P300 has been previously identified as an electrophysiological marker of reward prediction, being larger not only for reward compared to no-reward but also for larger compared to smaller rewards (Broyd et al., 2012; Goldstein et al., 2006; Pornpattananangkul & Nusslock, 2015). Following this, larger CUE-P300 was hypothesized in LA in response to the stimulus predicting reward than the one predicting no-reward. Contrary to this hypothesis and the previous literature, there was no evidence of such modulation of the CUE-P300 in LA. This result may be related to the fact that in the current experiment, both stimuli predicted reward delivery, even though with different probabilities and although the CS2 predicted reward delivery only in 20% of trials, leading to overall less earning compared to CS1, the magnitude of reward was the same for both stimuli. Indeed, previous literature has shown a modulation of the cue-P300 related to reward magnitude (Goldstein et al., 2006). However, whether this component is modulated also by reward likelihood has not been clarified. Here the results suggest that reward likelihood does not seem to affect the amplitude of CUE-P300.

The crucial result, however, regards the difference in amplitude of the between LA and HA, which was reduced in HA compared to LA, suggesting reduced prediction of coming appetitive stimuli in alexithymia. This result is in line with the current literature on alexithymia that has shown reduced P300 in response to emotional stimuli (Bermond, Righart, Ridderinkhof, & Moormann, 2008; Pollatos et al., 2011; Pollatos & Gramann, 2012).

In addition, the broader literature on emotion processing shows enhancement of the P300 in response to motivationally salient stimuli, such as emotional pictures, gain or loss of money and target stimuli in attention tasks (Nieuwenhuis, Aston-Jones, & Cohen, 2005; Olofsson, Nordin, Sequeira, & Polich, 2008; Polich, 2007). This effect has been interpreted as an increase in attentional resources devoted towards such stimuli. In keeping with this, HA may attribute less motivational salience to the stimuli that anticipate coming rewards compared to LA and pay less attention to the information conveyed by such stimuli, ultimately impairing the accurate prediction of coming reward.

This deduction is further supported by the other main result of the current experiment, namely the lack of modulation of the FRN in response to unexpected reward-related feedback in HA compared to LA. The FRN has been argued to represent an electrophysiological marker of prediction error, signaling the non-occurrence of a predicted event, enabling the organism to correct its prediction in order to construct accurate expectations regarding coming events (Alexander & Brown, 2011). In addition, according to the Pearce and Hall's model of learning, the size of the prediction error would modulate the amount of attention driven to the cues predicting the reward-related feedback, with larger prediction error increasing attention hence promoting learning (Roesch, Esber, Li, Daw, & Schoenbaum, 2012). Here, the results indicate that HA do not compute a prediction error in response to unexpected reward-related feedback, hindering accurate emotional learning and exacerbating the difficulty in constructing accurate representations of coming appetitive stimuli.

There is evidence that the P300 and FRN may be related to the dopamine system, which plays a fundamental role in emotional learning and in encoding prediction errors (Schultz, 1998, 2016). Specifically, P300 amplitude showed positive correlation with blood oxygen level dependent response in the ventral striatum during reward anticipation (Pfabigan

et al., 2014), appears positively related to availability of D2 receptors in the striatum (Pogarell et al., 2011), has been related with polymorphism in genes encoding dopamine D2 (Berman et al., 2006) and D3 (Mulert et al., 2006) receptors and increases following administration of amphetamine (McKetin, Ward, Catts, Mattick, & Bell, 1999). Regarding the FRN, studies suggest this component originates from activity in the ACC (Holroyd et al., 2004; Warren, Hyman, Seamans, & Holroyd, 2015), which may be related to phasic changes in dopamine activity resulting from dopaminergic projections from midbrain structures to this area (Nieuwenhuis et al., 2004; Sambrook & Goslin, 2015; Walsh & Anderson, 2012). It might be possible that the differences observed in HA in the electrophysiological components may be related to differences in the dopaminergic system. In addition, differences in activity in the ACC have also been shown in alexithymia (van der Velde et al., 2014), possibly suggesting an overlapping neural mechanism for the difficulty in emotional learning in alexithymia and its broader difficulties in processing emotional stimuli.

Contrary to the previous experiments, the difficulty in reward learning in HA, evidenced by electrophysiological, data was corroborated by the behavioral data showing less accurate verbal report regarding the stimulus-reward association in HA compared to LA. One difference between the current experiment and the previous ones, is that, rather than having one CS associated to reward while another never associated to reward, both CSs were associated to reward, differing only in their reinforcement rate. It is possible, that in these circumstances, more effort may be required to learn the accurate value of each stimulus and this may be especially challenging for HA, manifested as worse behavioral performance in reporting stimulus-reward contingency. Despite this, groups did not show a difference in the degree of liking of the CSs. Both groups liked the CS mostly followed by reward more than the one mostly followed by no-reward to comparable degree. Similarly to what reported in Experiment 3, these results suggest dissociation between processes underlying different

aspects of emotional learning and that HA may present difficulties in some but not all of these aspects.

In summary, the results on the CUE-P300 and FRN suggest a broken feedback loop in HA between prediction and prediction error signals during pavlovian reward conditioning. HA show reduced prediction of coming appetitive emotional stimuli. This inaccurate prediction may then lead to absent prediction error, which impairs the correction of the inaccurate prediction, further contributing to the reduced prediction and to a global impairment in constructing accurate internal representations of appetitive emotional events.

Experiment 6: Alexithymia requires more time to accurately avoid stimuli that have acquired aversive value following instrumental learning

Introduction

The experiments presented so far suggest that alexithymia is related to difficulties in constructing accurate predictions of coming aversive and appetitive stimuli during Pavlovian conditioning. One limitation of this type of learning is that the organism learns the contingency between a stimulus (CS) and an outcome (UCS). Nevertheless, in everyday life, the organism is an active agent in its surrounding environment, changing its behavior based on the outcome it might lead to. Therefore, in addition to make predictions about stimuli, the organism also needs to learn predictions about the outcomes following its actions, in order to select the actions that can increase survival. This process is named instrumental learning. During instrumental learning, the organism performs different actions, which can be rewarded with the delivery of appetitive stimuli or punished with the delivery of aversive stimuli. Following the repetition of the actions, the organism learns to predict which actions will be rewarded and which actions will be punished. As a consequence, rewarded actions will be repeated while punished actions will be terminated, making it another form of learning crucial for survival (Daw & Tobler, 2014). In addition, the environment is not a static, rather it is ever-changing, so that the same stimuli are rarely encountered in the same context twice. Therefore, the organism is also required to flexibly use the information learned about the value of actions associated to the stimuli encountered in one context, in order to ensure adaptive behavior, when the same stimuli are encountered in a novel context.

Given the above information, the aim of the current experiment is to investigate whether alexithymia also affects instrumental learning and the ability to effectively use the information acquired during this type of learning to ensure adaptive behavior in a novel

context. To this end, LA and HA completed the Probabilistic Selection Task (Frank, Seeberger, & O'Reilly, 2004; Frank, Woroch, & Curran, 2005). This includes two phases: learning and testing. In the learning phase, participants are faced with three pairs of characters (AB, CD, and EF). Within each pair, choosing one character is more likely to lead to reward (and less likely to lead to punishment) than choosing the other. Importantly, the probability of reward and punishment differ for each character (Fig. 20), so that each character, and the choice associated to it, acquire a more or less positive or negative value compared to the remaining characters. On each trial, participants choose one character of the pair and reward (positive feedback) or punishment (negative feedback) following the choice is provided. By trial and error, participants are required to learn the character in each pair more likely to lead to reward. Then, during testing, participants are again faced with pairs of characters; however, all possible combinations of the characters encountered during learning are presented. Participants' task remains to choose the character in each pair more likely to lead to reward; nevertheless, no feedback is provided about the choice. The testing phase enables to test whether participants can use the information learned about the value of each character, and the choice associated to it, to make effective choices when the old characters are presented within new pairs. Based on the results of the previous experiments, HA were hypothesized to be less able to learn accurate predictions from reward and punishment during instrumental learning, showing worse performance than LA in the learning phase. In addition, they were also hypothesized to be less able to use what was learned flexibly in a new context, hence showing worse performance compared to LA in choice behavior also during testing.

Methods

Participants

Three-hundred individuals completed the 20-item Toronto Alexithymia Scale (TAS-20; Taylor, Bagby, & Parker, 2003). Depending on the score, students were classified as LA (TAS-20 ≤ 36) or HA (TAS-20 ≥ 61) (Franz et al., 2004) and were then randomly contacted to participate in the study. Once in the laboratory, the alexithymia module of the structured interview for the Diagnostic Criteria for Psychosomatic Research (DCPR; Mangelli et al., 2006) was administered to increase reliability of screening and confirm TAS-20 classification. Participants with discordant classification on the two measures did not complete the task (n = 8). Due to the high co-occurrence of alexithymia and depression (S. Li et al., 2015), participants completed the Beck Depression Inventory (Beck et al., 1961) and did not complete the experimental task in case their score was higher than the moderate/severe depression cut-off (i.e. 19, n = 3). Levels of anxiety were measured with the State-Trait Anxiety Inventory (Spielberger et al., 1983)

All participants had equivalent educational backgrounds and were students at the University of Bologna. The study was designed and conducted in accordance with the ethical principles of the World Medical Association Declaration of Helsinki and the institutional guidelines of the University of Bologna and was approved by the Ethics Committee of the Department of Psychology. All participants gave informed written consent to participation after being informed about the procedure of the study.

Forty-two volunteers with no history of major medical, neurological or psychiatric disorders completed the study (20 LA, 22HA). Data from two HA were removed from analysis because they failed to achieve the performance criterion required for successful completion of the learning phase of the task. In total 40 participants were included in the

analysis: 20 LA (6 males; age *M* = 21.44, *SD* = 1.65 years; TAS-20 *M* = 31.89, *SD* = 2.58; STAI state *M* = 34.26, *SD* = 3.43; STAI trait *M* = 38.16, *SD* = 50.55); 20 HA (6 males; age *M* = 21.83 *SD* = 1.85 years; TAS-20 *M* = 64.70, *SD* = 4.59; STAI state *M* = 40.15, *SD* = 6.67; STAI trait *M* = 50.55, *SD* = 8.92). HA had significantly higher state ($t(38)=3.53$, $p=.001$) and trait anxiety ($t(38)=5.35$, $p<.001$) than LA; nevertheless there was no significant correlation between anxiety and the dependent measures (all $ps\geq.076$).

Independent measures

The experimental task consisted in the Probabilistic Selection Task by (Frank et al., 2004, 2005). This includes two phases: learning and testing.

Learning. This phase was a reinforcement learning procedure. On each trial a pair of stimuli consisting of hiragana characters appeared on the screen. Every time a pair appeared, the participant chose one of the two characters pressing a key on the keyboard. Following the choice, feedback appeared on the screen indicating whether the choice was correct (reward) or incorrect (punishment). These consisted of a hand with a thumb up or down respectively. In total, there were three pairs of stimuli (AB, CD, and EF). In each pair, each character had a predetermined probability of being followed by the correct feedback. Specifically, for the AB pair, choosing A led to correct feedback (reward) 80% of the time and incorrect (punishment) in the remaining 20% of the time, whereas B led to correct feedback (reward) only 20% of the time. For the CD pair, choosing C led to correct feedback (reward) 70% of the time, whereas D led to correct feedback (reward) only 30% of the time. For the EF pair, choosing E led to correct feedback (reward) 60% of the time, whereas F led to correct feedback (reward) only 40% of the time (Fig. 20). Participants' task was to learn to choose the character in each pair that leads to correct feedback the majority of trials.

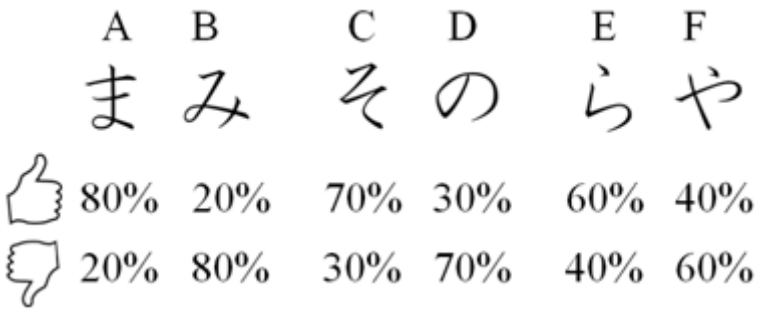

Figure 19. Illustration of examples of characters and the feedback probability associated with each character for each type of feedback.

A performance criterion was introduced for each pair to ensure participants achieved comparable level of learning before moving to the testing phase. Specifically, this was 65% of A for AB, 60% of C in CD and 50% of E in EF. Learning was evaluated at the end of each training block consisting of 60 trials (20 per stimulus pair) for a maximum of 4 blocks. Participants who did not achieve the criterion after 4 blocks were excluded from further analysis (n=1 HA). After achieving the criterion, participants proceeded to the testing phase.

Each trial consisted in the presentation of a fixation cross in the center of the screen for 500ms, followed by the presentation of the pair of characters during which participants could provide their choice by pressing the corresponding key. Key press terminated stimulus presentation and participants had a maximum of 3000ms to provide their answer. This was followed by the feedback for 1000ms, followed by an inter trial interval of 1000-1500ms during which a blank screen was presented (Fig. 21). The order of presentation of stimuli was randomized across trials. The type of characters constituting each pair was counterbalanced across participants.

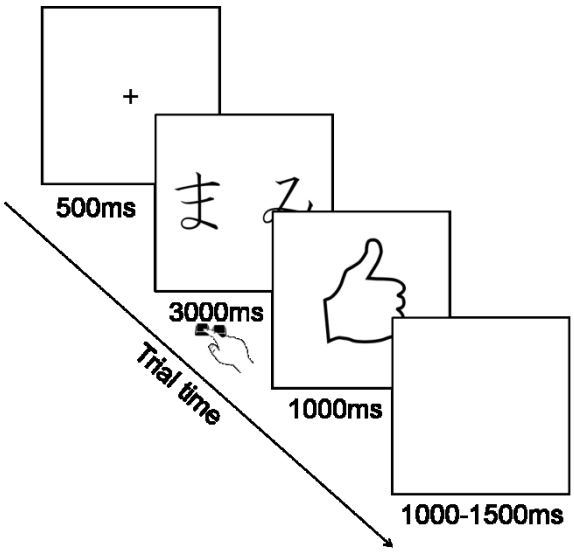

Figure 20. Illustration of experimental trial for the learning phase.

Testing. This phase enabled to evaluate how the acquired learning affected choice behavior when the same characters are presented in a new context. So, the old pairs of characters were presented in addition to new pair of characters resulting from all the possible combinations of pairs of characters.

On each trial a pair appeared on the screen and participants chose one of the two characters. No feedback was given about the choice. Participants' task was to choose the character in each pair they thought was the correct one. Participants were also told to follow their instinct when they were not sure about which character to choose.

Each trial consisted in the presentation of a fixation cross in the center of the screen for 500ms, followed by the presentation of the pair of characters during which participants could provide their choice by pressing the corresponding key. Key press terminated stimulus presentation and participants had a maximum of 3000ms to provide their answer. This was followed by an inter trial interval of 1000-1500ms during which a blank screen was presented

(Fig. 22). The order of presentation of stimuli was randomized across trials. There were 90 trials in total (6 per pair).

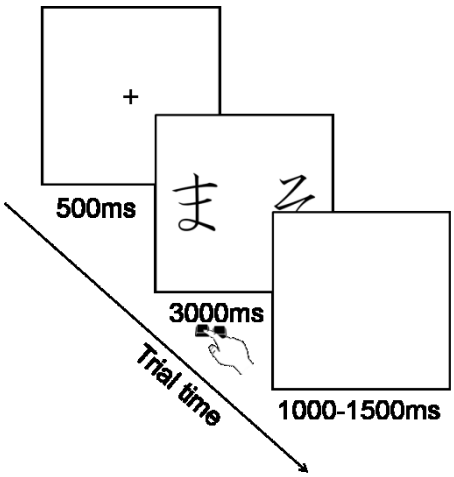

Figure 21. Illustration of experimental trial for the testing phase.

Dependent measures

Number of blocks completed during learning. The number of blocks completed in order to achieve the performance criterion was counted for each participant to then test whether there were any group differences.

Early learning. The percentage of accurate response and average response times for accurate responses for the first block were evaluated to test differences between groups in early acquisition of learning (Waltz, Frank, Robinson, & Gold, 2007).

Degree of exploration during early learning. The probability of changing response following either positive or negative feedback was calculated during the first block of learning in order to test group differences.

Retaining of learning in a novel context. We verified that subjects retained the performance criterion for successful learning also during the testing phase, to ensure learning was retained. As a consequence, participants whose accuracy in choosing the correct character when faced

with old pairs (AB, CD or EF) did not equal the performance criterion were excluded from further analysis because their data were not interpretable. Then differences between groups in accuracy and response times for the old pairs were tested.

Using what was learned, in a new context. On each trial, participants were faced with one out of four possible types of choice (Table 3). First, they could be faced by a pair consisting of one correct character and one incorrect character (conflict of choice: low conflict) and where the probability of the correct stimulus of having been rewarded was greater than the probability of the incorrect stimulus of having been punished (type of choice: choose positive). This included AD, AF and CF pairs. Second, they could be faced by a pair consisting of one correct character and one incorrect character (conflict of choice: low conflict) and where the probability of the incorrect stimulus of having been punished was greater than the probability of the correct stimulus of having been rewarded (type of choice: avoid negative). This included BC, BE and DE pairs. Third, they could be faced by a pair consisting of two correct characters (conflict of choice: high conflict) and where one had higher probability of having been previously rewarded compared to the other (type of choice: choose positive). This included AC, AE and CE pairs. Four, they could be faced by a pair consisting of two incorrect characters (conflict of choice: high conflict) and where one had higher probability of having been previously punished compared to the other (type of choice: avoid negative). This included BD, BF and DF pairs. The percentage of accurate response and the average response time for accurate choices for each participant and for each type of choice were calculated to test differences in performance between groups during testing.

Table 3. Illustration of the different types of choice during the testing phase

Stimulus	Value of stimuli	Conflict of choice	Type of choice
AC	A: 80% positive C: 70% positive	High	Choose positive
AD	A: 80% positive D: 70% negative	Low	Choose positive
AE	A: 80% positive E: 60% positive	High	Choose positive
AF	A: 80% positive F: 60% negative	Low	Choose positive
BC	B: 80% negative C: 70% positive	Low	Avoid negative
BD	B: 80% negative D: 70% negative	High	Avoid negative
BE	B: 80% negative E: 60% positive	Low	Avoid negative
BF	B: 80% negative F: 60% negative	High	Avoid negative
CE	C: 70% positive E: 60% positive	High	Choose positive
CF	C: 70% positive F: 60% negative	Low	Choose positive
DE	D: 70% negative E: 60% positive	Low	Avoid negative
DF	D: 70% negative F: 60% negative	High	Avoid negative

Results

No difference in the number of blocks required to complete learning

An independent sample t-test showed no significant difference between the two groups in the average number of blocks completed to achieve the performance criterion ($t(38)=1.09$, $p=.281$; $M_{low}=1.90$, $M_{high}=1.60$). The groups required a comparable number of blocks to achieve the performance criterion.

No difference in the degree of exploration during early learning

A 3x2x2 RM ANOVA (type of pair: AB, CD, EF; type of feedback: correct, incorrect; group: LA, HA) showed a significant main effect of type of feedback ($F(1, 38)=29.26$, $p<.001$,

partial η^2=.43). Newman-Keuls post-hoc comparison indicated that on any given trial, participants were more likely to switch choice towards the other character in the pair, if they had received incorrect feedback in the previous trial for the same pair than if they had received correct feedback (p<.001, $M_{positive}$=.18, $M_{negative}$=.31). In addition, there was no main effect or interaction with the factor group (all ps≥.270) indicating that groups had comparable degree of exploration while learning the correct character in each pair.

HA tend to be slower in choosing the correct character in the EF than CD pair

A 3x2 RM ANOVA on the accuracy (type of pair: AB, CD, EF; group: LA, HA) showed a main effect of the type of pair (F(2, 76)=5.38, p=.006, partial η^2=.12). Newman-Keuls post-hoc comparison indicated that at the end of the first block participants achieved lower response accuracy to the EF (M=.640) pair compared to AB (M=.787; p=.008) and CD (M=.762; p=.013), while there was no significant difference in the accuracy between the response to AB and CD (p=.605). In addition, there was no main effect or interaction with the factor group (all ps≥.818) indicating that at the end of the first block groups had comparable acquisition of learning.

Next, we tested differences in response times for correct trials. Note that only subjects with valid response times were included (i.e. participants with at least one accurate response on any pair). A 3x2 RM ANOVA (type of pair: AB, CD, EF; group: LA, HA) showed a trend for a pair by group interaction (F(2,72)=3.06, p=.053, partial η^2=.08). Newman-Keuls post-hoc comparison indicated that, while LA had no difference in response times between the three pairs (all ps≥.218), there was a tendency in HA to be slower in choosing the correct character in the EF pair (M=1299.1ms) compared to the CD pair (M=1167.7ms, p=.075; Fig. 23). All other within-group comparisons were not significant (p≥.162).

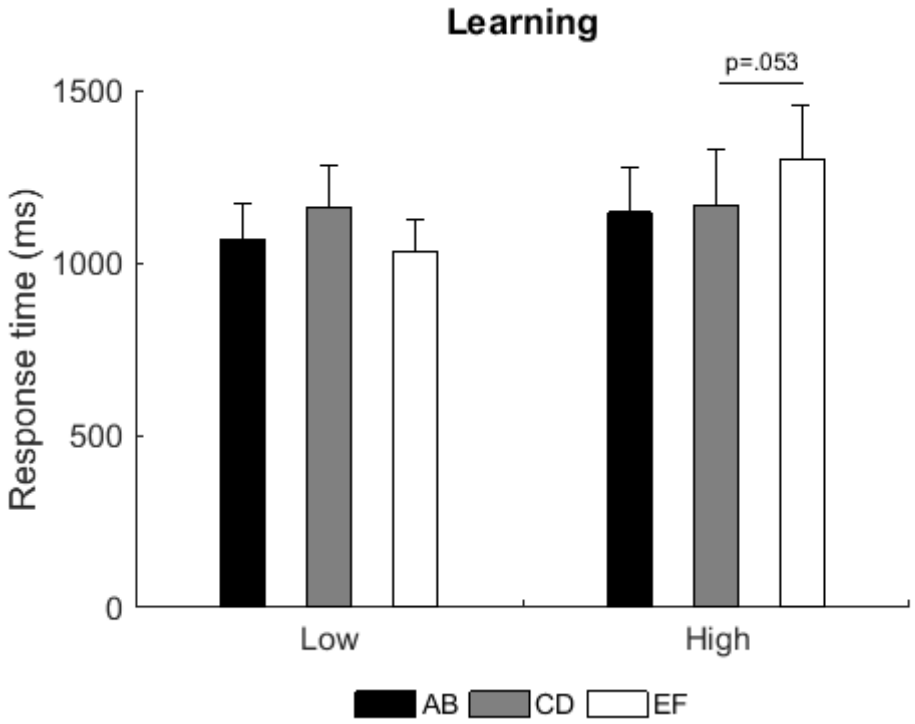

Figure 22. Mean response time for each stimulus pair as a function of alexithymia group. In the high alexithymia group, participants were slower in choosing the correct character in the EF than in the CD pair. Error bars represent standard errors.

No difference in retaining of learning in a novel context

First, 1 LA and 3 HA were not able to retain the acquired learning and were excluded from further analysis.

A 3x2 RM ANOVA on the accuracy (type of pair: AB, CD, EF; group: LA, HA) showed a main effect of the type of pair ($F(2, 68)=4.46$, $p=.015$, partial $\eta^2=.11$). Newman-Keuls post hoc comparisons showed that participants were less accurate in responding to the EF (M=.89) pair compared to AB (M=.97, p=.013) and CD (M=.95, p=.039), while there was no difference in accuracy between AB and CD (p=.420). There was no main effect or interaction with the factor group (all ps≥.376), indicating that HA and LA retained learning to a comparable level.

The 3x2 RM ANOVA on the response times, also showed no difference between groups as well as no difference between stimulus pairs (all ps≥.090).

HA require more time to avoid a negative stimulus than choosing a positive one, when in a novel context

The 2x2x2 RM ANOVA (type of learning: choose positive, avoid negative; type of conflict: low conflict, high conflict; group: LA, HA) on mean accuracy showed a significant main effect of conflict ($F(1,34)=97.09$, $p<.001$, partial $\eta^2=.74$). Newman-Keuls post hoc comparisons showed that participants were less accurate when facing pairs with high conflict choice ($M=.91$) than low conflict choice ($M=.55$; $p<.001$).

The 2x2x2 RM ANOVA (type of learning: choose positive, avoid negative; type of conflict: low conflict, high conflict; group: LA, HA) on the reaction times showed a significant valence by conflict interaction ($F(1,34)=14.56$, $p<.001$, partial $\eta^2=.30$). Newman-Keuls post hoc comparisons showed that participants were slower in a high conflict choice, specifically when accurately avoiding the most negative character in a pair of two negative characters than when having to make any other choice (all $ps<.001$). Crucially, there was a significant valence by group interaction ($F(1,34)=4.38$, $p=.044$, partial $\eta^2=.11$). Newman-Keuls post hoc comparisons showed that HA were slower when accurately avoiding a negative stimulus ($M=1396.8$ms) than when choosing a positive one ($M=1095.3$, $p<.001$), while LA had no significant difference when faced by these two choices ($M_{negative}=1017.5$, $M_{positive}=113.9$, $p=.357$; Fig. 24). All other comparisons were not significant (all $ps\geq.098$). This result suggests that HA require more time to avoid a negative stimulus compared to choosing a positive one, when in a novel context. On the contrary, LA require comparable amount of time to make either type of choice.

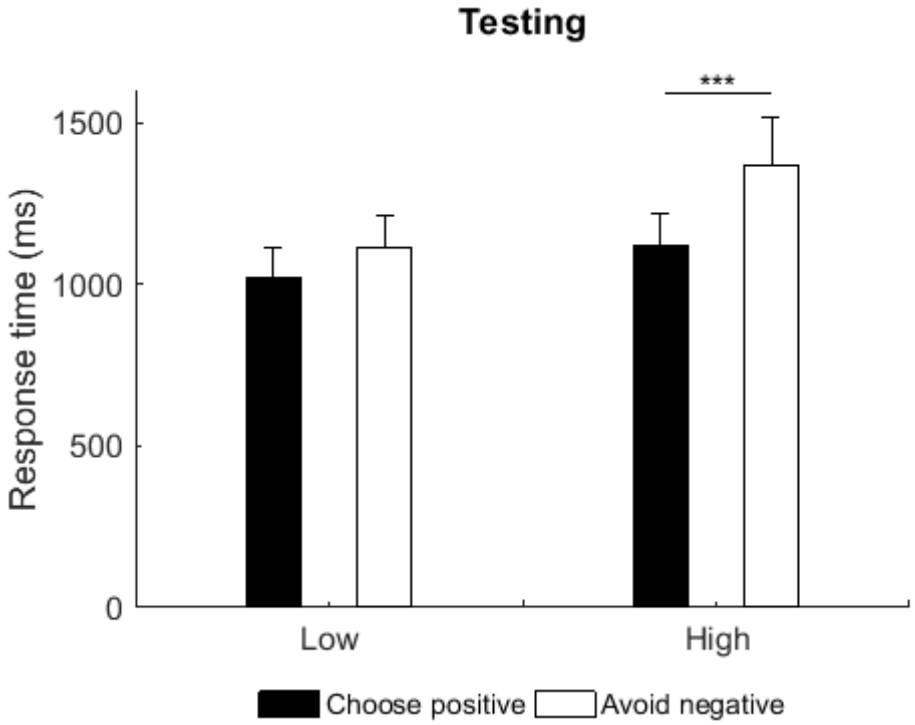

Figure 23. Mean response time for the 'choose positive' and 'avoid negative' conditions as a function of alexithymia group. In the high alexithymia group, participants were slower when avoiding the negative than when choosing the positive stimulus. Error bars represent standard errors. Significant differences are indicated as follows: ***p< .001.

Discussion

The current experiment tested the effect of alexithymia on participants' ability to learn from reward and punishment in an instrumental learning task, the ability to retain the acquired learning and the ability to use the information learned about the value of stimuli flexibly to make adaptive choices, when such stimuli are encountered in a novel context.

First of all, the main results were in line with previous literature. During learning, participants were more likely to change choice of character in a pair if their previous choice received incorrect than if it received correct feedback (Frank, Moustafa, Haughey, Curran, & Hutchison, 2007), indicating that the two types of feedback were effective as punishment and reward. Indeed, participants changed their behavior according to the feedback received repeating rewarded choices and diminishing punished ones. During testing, the analysis on accuracy showed that participants were equally accurate in choosing positive and avoid

negative stimuli, replicating Frank et al. (2007, 2005) but less accurate in making the correct choice in high conflict than in low conflict situation, when faced with the new pairs. Indeed, having to choose between one positive and one negative stimulus seems easier that having to choose between two positive or two negative stimuli, which differ only in their reinforcement rate. In addition, the analysis on reaction times showed that participants were slower when having to avoid the most negative stimulus among two negative ones than when making any other choice. Together these data suggest that making the correct choice in low conflict situation is easier than in high conflict situation and, in particular, when faced by two negative stimuli, participants require more time to maintain accuracy comparable to when choosing between two positive stimuli, suggesting the former may represent the overall most difficult choice.

Regarding group differences, initial hypotheses were only partially supported. During learning, the two groups did not differ significantly in the number of blocks completed to learn the value of stimuli, the degree of exploration following reward or punishment, and the accuracy in identifying the correct character in each pair. Nevertheless, analysis on reaction time, showed that HA, but not LA, had a tendency to be slower, during the early phase of learning, when accurately identifying the correct character in the EF than in the CD pair. Importantly, while in the CD pair, the percentage difference in reinforcement rate between characters was 40%, in the EF pair, the percentage difference in reinforcement rate between characters was only 20%. Therefore, as difference in reinforcement rate between two stimuli decreases, HA may find increasingly difficult to learn the value of individual stimuli and associated actions, requiring more time to maintain choice accuracy. During testing, three HA but only one LA did not retain the learned value of the characters, possibly suggesting more HA may have difficulties in retaining the internal representation of the value of stimuli and actions, once these are no more reinforced. Despite this, HA and LA who retained learning

did not differ in choice behavior to old pairs of characters. However, when faced by new pairs of characters, although groups showed comparable accuracy in choice behavior, the analysis on response times indicated a difficulty of HA in efficiently avoiding characters, which had acquired negative value, encountered in a new context. Indeed, while LA were equally efficient in avoiding negative or choosing positive characters, HA were slower when accurately avoiding negative characters than when choosing positive ones, regardless of the level of conflict involved in the choice. Therefore, HA may find more difficult to effectively avoid previously learned negative stimuli encountered in a new context than choosing positive ones. This may also suggest that, although during the learning phase HA were able to learn to differentiate the positive from the negative stimulus in each pair of character, the quality of such learning may have differed between the two stimuli, being relatively weaker for the negative stimuli, becoming evident during the testing phase.

Previous studies suggest that the neurotransmitter dopamine plays a significant role in relative performance between choosing positive/avoid negative stimuli in the current task. Indeed, while healthy controls show no difference in performance between the two conditions, patients with Parkinson's show a change in performance depending on their current medication status. When on medication, they are relatively better at choosing positive than avoiding negative stimuli in the new context, while they show the opposite pattern when off medication (Frank et al., 2004). This is because during learning, on medication, the neurotransmitter would be sufficient to enable dopamine bursts following positive feedback, promoting learning the value of positive stimuli, but it would also block the effect of dopamine dips following negative feedback, because of the prolonged occupancy of dopamine receptors, hindering learning the value of negative stimuli. Off medication, instead, such dopamine bursts cannot occur resulting in the opposite learning pattern (Maia & Frank, 2011). In addition, the ability to avoid negative stimuli, in particular, seems to be related to

differences in genotype associated to density of postsynaptic D2 receptors, which are crucial for learning from low dopamine levels, as it is the case in response to negative feedback (Frank et al., 2004). Indeed, performance in avoiding negative stimuli increasing with increasing density of D2 receptors (Frank et al., 2007), and decreases in individuals carrying an allele of a genetic polymorphism associated with a reduction in D2 receptor density by up to 30% (Klein et al., 2007). Therefore, it might be possible that alexithymia may be related to differences in the dopamine system and in particular in those aspects supporting learning from negative feedback.

To conclude, the results of the current experiment suggest that alexithymia may be related to a difficulty in learning the individual value of two stimuli and the actions associated to them as difference in reinforcement rate between them decreases. In addition, individuals with alexithymia also have a difficulty in learning from punishment, which becomes evident when having to avoid stimuli, which had previously acquired aversive emotional value, encountered in a novel context.

Conclusion

In summary, the four experiments presented in this chapter show that alexithymia is related to a difficulty in emotional learning. Individuals with alexithymia appear less able to learn the contingencies between neutral stimuli and actions and their emotional consequences, ultimately failing constructing accurate internal representations of emotional stimuli. Disruption in this process may represent a unifying mechanism underling the difficulties in emotion processing observed in alexithymia. Indeed, having accurate internal representations of emotional stimuli enables their prediction and preparation to respond to them (McNally & Westbrook, 2006; Öhman & Mineka, 2001). This is not only crucial for effective recognition, response and response regulation to the emotional stimulus per se, but also for anticipating the consequences of the emotional event enabling optimal decision making (Bubic et al., 2010).

Chapter 5. General discussion: Rethinking alexithymia

Alexithymia is a personality trait characterized by difficulty in identifying feelings, in describing them to others and a type of thinking focused more on concrete aspects regarding the external environment than on introspection (Sifneos, 1973; Taylor et al., 1991). As reviewed in Chapter 1, alexithymia is related to a broad range of difficulties in processing emotional stimuli and in Chapter 3 two experiments contributed to this literature. In particular, Experiment 1 showed that alexithymia has a widespread impact on the emotional life of individuals, being related to lower emotional intelligence, empathy and psychological wellbeing, supporting the significance of studying this personality trait. In addition, Experiment 2 focused on the perception of emotional facial expressions, showing that alexithymia is related to the need for more emotional intensity to identify fear in emotional facial expressions, extending the previous literature that had shown alexithymia is related to the need for more time to identify emotions in facial expressions (Grynberg et al., 2012). Although this evidence significantly extends the literature dedicated to the description of the difficulties in emotion processing of alexithymia, the basic mechanisms that may underlie such difficulties remain poorly understood. Nevertheless, this type of investigation could provide new insight in the conceptualization of this personality trait.

The internal representations of emotional stimuli in alexithymia

A basic mechanism that is crucial to process emotional stimuli successfully is emotional learning. As discussed in Chapter 2, in our daily life countless stimuli can elicit an emotional response, nevertheless only a restricted range of stimuli is biologically programmed to do so (LeDoux, 1998). However, to increase the chances of survival, through emotional learning, we can shape the internal representations of stimuli in order to attribute them emotional value because of their association with appetitive and aversive stimuli. By doing so, emotional

learning supports the construction of an internal model of emotional stimuli, which can be used to predict the emotional future, anticipating coming appetitive and aversive stimuli and preparing the organism to respond to them effectively (Bubic et al., 2010; McNally & Westbrook, 2006; Öhman & Mineka, 2001). For this reason, the remaining experiments investigated the possible differences in the construction of internal representations of emotional events, examining the process of emotional learning in alexithymia.

Unstable and imprecise internal representations of emotional stimuli shown by psycho-, electro-physiological signals and behaviors

The experiments in Chapter 4 showed that alexithymia is related to impairments in different aspects of emotional learning. Specifically, Experiment 3 focused on learning from aversive stimuli, showing that alexithymia is related to a reduction of physiological markers of emotional prediction when anticipating aversive stimuli during Pavlovian conditioning. Importantly, here alexithymia was not related to a difference in the physiological changes elicited by the aversive stimulus per se, rather individuals with alexithymia had difficulty in learning to transfer the emotional response triggered by the aversive stimulus on a neutral stimulus associated with it, suggesting impairment in constructing accurate internal representations of aversive stimuli. Experiment 4 used a similar paradigm to look at differences in physiological markers of emotional prediction, when anticipating appetitive stimuli. Nevertheless, here no difference was found in relation to alexithymia, suggesting that the physiological markers of emotional prediction may be spared when facing appetitive stimuli. Nevertheless, in Experiment 5, when electrophysiological markers of emotional predictions were investigated during reward conditioning, alexithymia showed decreased anticipation of coming appetitive stimuli. This was accompanied also by absent markers of prediction error in response to unexpected feedback related to the appetitive stimuli. These results suggest a broken feedback loop, in which individuals with alexithymia may attribute

less motivational salience and devote less attention to the stimuli that predict appetitive emotional stimuli, showing reduced prediction of the latter. This may then lead to absent prediction error and failure in correcting the inaccurate prediction, enhancing the reduced prediction. Furthermore, this was accompanied by less accurate verbal report regarding the stimulus-reinforcer. Finally, Experiment 6 examined the ability to learn the emotional value of actions, in addition to that of stimuli, by having participants complete an instrumental learning task. Furthermore, this experiment also investigated the ability to use the internal representation of the emotional value of stimuli and actions acquired in one context, to ensure adaptive behavior when the same stimuli are encountered in a novel context. Here, the results showed that alexithymia was mainly related to two difficulties. The first was a difficulty in learning the individual value of stimuli and the actions associated to them as the difference in reinforcement rate between the stimuli decreased. The second was a difficulty in learning from punishment, which became evident when having to avoid stimuli, which had previously acquired aversive emotional value, encountered in a novel context. Indeed, individuals with alexithymia were slower when accurately avoiding stimuli that had acquired negative value than when choosing those that had acquired positive value. On the contrary, such difference in performance was not found in individuals without alexithymia, who were equally efficient in avoiding negative or choosing positive stimuli.

Together, the data suggest that alexithymia is related to an inability to construct accurate internal representations of emotional events and actions, failing to adaptively expand the representations of emotional stimuli by attributing emotional value not only to appetitive and aversive unconditioned stimuli but also to the stimuli that predict them. In particular, the internal representations appeared unstable and imprecise. The former was suggested by the enhanced extinction of threat conditioning, once the value of the conditioned stimulus was no more reinforced by the aversive event, and by a higher rate of participants with alexithymia

being unable to retain the emotional value of stimuli and actions, learned during instrumental learning, in a subsequent testing phase. The latter was suggested by decreased predictions during aversive and appetitive conditioning, the absence of prediction error during appetitive conditioning and the difficulty in learning the value of individual stimuli, as differences in reinforcement rate between stimuli decreased.

The internal representations of aversive stimuli may be more severely affected than for appetitive stimuli

In addition, the current data on emotional learning seem to suggest that, in alexithymia, the internal representations of emotional stimuli may be differentially affected depending on their valence. Specifically, the representation of aversive emotional stimuli may be more severely affected than that of appetitive stimuli. Indeed, the data from Experiment 3 and 4 showed that the physiological marker of emotional prediction was decreased in alexithymia for aversive but not for appetitive stimuli. Also, in Experiment 6, alexithymia was related to a difficulty in learning from punishment but not from reward, during instrumental leaning.

This difference in results may be explained by the fact that the neural circuits involved in emotional learning present substantial overlap but also significant differences when it comes to encoding stimulus valence. For example, although some neurons in the basolateral complex of the amygdala respond to both appetitive and aversive stimuli, others respond preferentially to one valence over the other and these project to different output regions (Beyeler et al., 2016; Namburi et al., 2015). In particular, only those preferentially responding to aversive stimuli project to the central amygdala, whose connections to the hypothalamus appear to drive the physiological arousal in anticipation of threat (Cardinal et al., 2002; McNally & Westbrook, 2006). For this reason, it is possible that psychophysiological changes in anticipation of emotional stimuli may be differentially

affected by alexithymia depending on the valence of the emotional stimulus, as suggested by the reduced skin conductance response in anticipation of shock, in Experiment 3, but not of monetary reward, in Experiment 4. In addition, with regards to behavioral changes triggered by emotional stimuli, although dopamine may drive both approach of appetitive stimuli and avoidance of aversive ones, it seems to do so acting on different dopamine receptors, D1 receptors would mediate approach behavior while D2 avoidance behavior (Frank et al., 2007). Therefore, it is also possible for alexithymia to impair avoidance behavior of aversive stimuli but not approach behavior for positive ones, as suggested by the slower response time when avoiding negative stimuli compared to when choosing positive ones in Experiment 6.

Such discrepancy, between the construction of internal representations of aversive and appetitive emotional stimuli, appears in line with the broader literature on alexithymia, which has suggested that alexithymia may be associated with more difficulties in processing aversive as opposed to appetitive emotional stimuli. For example, individuals with alexithymia rate the expression of fearful faces, but not happy ones, as less intense (Prkachin et al., 2009), and fail to show enhanced remapping of fear on their own somatosensory system compared to those with no alexithymia, while the remapping of happiness remains comparable between the two groups (Scarpazza et al., 2014). Also in the current thesis, Experiment 2 showed that participants with alexithymia needed more emotional intensity to identify fear in emotional facial expressions, but not happiness. Moreover, studies investigating the neural response to emotional stimuli, found decreased activation of the amygdala in alexithymia, which seems to be specific to aversive stimuli, including fearful bodies and observation of pain in others (Kugel et al., 2008; Pouga et al., 2010; Reker et al., 2010; van der Velde et al., 2014).

Decoupling between psycho- and electro-physiological signals and subjective reports

Contrary to what might have been expected, no evidence was found that the subjective report of the emotional experience during the presentation of conditioned stimuli was affected in alexithymia. This was true both when participants were provided with an emotional word (e.g. fear, happiness) and they had to rate the intensity of the corresponding emotional experience (Experiment 3, 4), as well as when participants were asked to indicate the value attributed to the stimuli, rating how much they liked them (Experiment 5). This difference in results deserves further discussion because it is informative not only for the understanding of alexithymia, but also to extend the literature on the construction of internal representations of emotional stimuli.

First of all, this difference suggests that the processes giving rise to psychophysiological and electrophysiological signals, assessed here, in response to the conditioned stimuli, are partly dissociated from those giving rise to the subjective emotional experience associated with them. Indeed, the neural circuits eliciting the physiological changes in response to emotional stimuli have been proposed to mainly involve subcortical regions, while those eliciting the conscious subjective emotional experience to mainly involve cortical regions (LeDoux, 1998; LeDoux & Brown, 2017; LeDoux & Hofmann, 2018). For example, neuroimaging evidence shows that, while the amygdala is active also in lack of awareness for aversive stimuli, the insula is only active when participants are aware of the threat (Critchley et al., 2002). Therefore, while the amygdala would be more relevant for the implicit processing of emotional stimuli and elicitation of psychophysiological changes in response to them, the insula may be more relevant for their cognitive representation of the subjective emotional experience (Critchley et al., 2002; LeDoux, 2007). In particular, with regard to the shock administered in Experiment 3, the nociceptive information conveyed by the shock is processed by multiple neural pathways. On one side, the nociceptive information

would reach the basolateral complex of the amygdala (Paton et al., 2006), whose projections would then influence the central complex of the amygdala and then, through its connections with hypothalamus, elicit the physiological changes in response to the conditioned stimuli (Cardinal et al., 2002; McNally & Westbrook, 2006). On the other side, the nociceptive information would also reach the insula (Craig, 2002, 2003; Critchley et al., 2004). The insula then may modulate the physiological changes elicited by the amygdala because of its connections with this structure (Phelps et al., 2001), but, more importantly, appears involved in the cognitive representation of physical sensations as affective states (Barrett, 2017b) and the conscious subjective experience of emotion (LeDoux & Brown, 2017), as suggested by stronger response of this region in instructed compared to pavlovian threat learning, where the unconditioned stimulus is expected but never experienced (Phelps et al., 2001). In addition, the subjective emotional experience, would not only arise from the cognitive interpretation of this nociceptive information, but also from the integration in working memory of this information together with additional interoceptive information as well as information coming from other sources, such as the external environment or semantic knowledge (LeDoux, 1998; LeDoux & Brown, 2017; LeDoux & Hofmann, 2018). Therefore, focusing on Experiment 3, the cognitive interpretation of the nociceptive information, processed by the insula, may also have been integrated with information from these other sources, to give rise to a subjective emotional experience, which did not arise directly from the physiological response elicited by the emotional stimuli. In sum, the current data support dissociation between the physiological changes and the subjective emotional experience to emotional stimuli.

Turning to the understanding of alexithymia, the data may suggest that alexithymia affects the physiological response to conditioned stimuli sparing the subjective emotional experience, contrasting with the ‘blind-feel’ hypothesis of alexithymia, which argues that

alexithymia would be characterized by an intact physiological response to emotional stimuli and a deficit in emotion concept representation (Lane et al., 1997). However, the picture seems more complex than this, because the current data should be considered in light of the previous literature, which found both comparable (Bausch et al., 2011; Easterbrooks et al., 2005; Stone & Nielson, 2001) and decreased (Bermond et al., 2010; Franz et al., 2003; Neumann et al., 2004; Newton & Contrada, 1994; Pollatos et al., 2008) physiological response to emotional stimuli together with no difference (Franz et al., 2003), increased (Eastabrook et al., 2013; Newton & Contrada, 1994; Pollatos et al., 2011) or decreased (Stone & Nielson, 2001) subjective reports of emotional experience. Therefore, alexithymia may affect one level, the other or both depending on the experimental conditions. In the present experiments, it may indeed be possible that the subjective emotional experience may have not been affected by alexithymia. Nevertheless, given the contribution of multiple sources of information to the subjective emotional experience, it is also possible that participants with alexithymia may have relied on information from the external context, when providing the subjective reports of emotional experience, enabling them to obtain a report comparable to the individuals without alexithymia. In fact, two additional aspects should be taken into consideration when discussing these results. First, participants were provided with specific emotional words or dimensions and only needed to rate their intensity. It is possible, that if participants were faced by more open questions asking to describe what they experienced during the presentation of the stimuli, differences between groups may have been revealed, as it was the case for the assessment of explicit learning regarding the stimulus-feedback contingency in Experiment 5. In addition, Experiment 6 did not include a subjective report of emotional value attributed to the stimuli, which might also have revealed differences between groups. In fact, computing the value of the single stimuli in the experiment may have been more difficult compared to the other experiments, because of the higher number of stimuli

and their more subtle differences in reinforcement rates. Therefore, differences in subjective emotional experience in alexithymia may become evident, once individuals face conditions in which relying on information from the external environment becomes more challenging and they are required to rely primarily on information coming from bodily signals to construct their subjective emotional experience. This seems a plausible hypothesis, which could be tested by future studies, given the evidence that alexithymia is related to a general impairment in interoception, even in non-emotional context (Brewer et al., 2016; Murphy et al., 2017; Shah et al., 2016).

Implications of the current results for the understanding of alexithymia

The definition of alexithymia

The current definition of alexithymia includes three aspects: difficulty in identifying feelings, difficulty in describing them to others and a type of thinking focused more on concrete aspects regarding the external environment rather than on introspection on the internal mental life (Sifneos, 1973; Taylor et al., 1991).

Regarding the identification and description of emotions, it is possible that the difficulties in this domain are not always apparent, as suggested by the results on the subjective emotional experience during emotional learning. However, these may become evident when conditions for processing emotional stimuli become challenging. Indeed, this was also suggested by the first two experiments. Specifically, in Experiment 1 there was no evidence that alexithymia was related to impairment in the identification of emotions expressed by facial expressions, when assessed on the performance measure of emotional intelligence; nevertheless, in Experiment 2, individuals with alexithymia needed more emotional intensity to perceive fear in facial expressions. One difference between the two tasks is that in Experiment 1 participants could take as much time as needed to observe the

face and choose their response, on the contrary in Experiment 2 the presentation of the face and the time for response was limited, possibly increasing the difficulty of the task. Indeed, previous literature has shown that alexithymia appears related to worse performance in recognition of emotional facial expressions under temporal constraints (Ihme, Sacher, Lichev, Rosenberg, Kugel, Rufer, Grabe, Pampel, Lepsien, Kersting, Villringer, Lane, et al., 2014; Swart et al., 2009), which can be restored when stimulus exposure time is extended (Ihme, Sacher, Lichev, Rosenberg, Kugel, Rufer, Grabe, Pampel, Lepsien, Kersting, Villringer, & Suslow, 2014; Pandey & Mandal, 1997; P. D. Parker et al., 2005), corroborating the idea that the difficulty in identifying feelings may be manifested only under specific conditions.

More interesting appears the third factor of alexithymia, namely externally oriented thinking, which is described as a cognitive style focused more on concrete events rather than on introspection and has been associated with an impoverished fantasy life and imaginative capacity (Sifneos, 1973; Taylor et al., 1991). In this regard, the ability to construct internal representations of stimuli in the environment through learning is not only crucial for predicting the future, but also for imagination and mental imagery more in general (Bubic et al., 2010; Hassabis & Maguire, 2007; Moulton & Kosslyn, 2009; Weber et al., 2017). So it is possible that this third factor, which describes alexithymia, may be a manifestation of the impairment in constructing accurate internal representations of emotional stimuli. To test whether alexithymia is indeed related to poor imaginative capacity, individuals with and without alexithymia could be assessed on a mental imagery task (e.g. Bertossi, Aleo, Braghittoni, & Ciaramelli, 2016; Hassabis, Kumaran, & Maguire, 2007). In particular, because only the internal representation of emotional stimuli should be impaired in alexithymia, alexithymia could be hypothesized to be related to decreased richness of emotions included in imaginary scenarios but comparable richness of concrete aspects of the scenarios, such as the presence of objects or entities.

The difficulties in emotion processing

Having accurate internal representations of emotional stimuli is crucial to predict the emotional future, so that organisms can appropriately prepare to respond to emotional stimuli, rather than simply reacting to them once they have occurred (McNally & Westbrook, 2006; Öhman & Mineka, 2001). These predictive representations enable effective recognition, response and response regulation to the emotional stimulus and optimal decision making (Bubic et al., 2010). Therefore, the unstable and imprecise internal representations, shown by psycho-, electro-physiological and behavioral markers, in alexithymia may represent a unifying mechanism underling the difficulties observed in multiple aspects of emotion processing. Indeed, results from the first two experiments corroborate this.

Regarding the recognition of emotional stimuli, in Experiment 2, alexithymia was related to the need for more emotional intensity to perceive fear in emotional facial expressions. Because emotional faces are learned emotional stimuli (Barrett, 2017a), this result may have been a manifestation of an underlying imprecise internal representation of fearful faces. A detailed representation would be required to be able to perceive the somatic features, which define such emotional expression, also when not displayed at full intensity.

Regarding the regulation of emotional response, in Experiment 1, alexithymia was negatively correlated with the ability to manage emotions, measured on a performance measure of emotional intelligence. In order to successfully complete this task, participants are required to abstract their knowledge about emotional stimuli and apply it to a hypothetical emotional situation, in order to find the best strategy to manage it. This abstraction can be achieved only with an accurate internal representation of emotional stimuli, which act as a general model to be applied also to theoretical scenarios in the lack of experiential information. In addition, alexithymia was positively correlated with personal distress

experienced during emotionally charged situations measured on an empathy questionnaire. The higher the alexithymia, the more the person was overwhelmed by the emotional distress of others'. This may also be related to the lack of an accurate predictive internal model, which can facilitate the preparation to deal with emotionally charged situations adaptively.

Regarding decision making, the previous literature on alexithymia, has shown a relationship between this personality trait and differences in decision making. For example, in the context of risky decisions during a gambling task, participants with alexithymia show lower performance than those without alexithymia. In particular, participants were tested on the Iowa Gambling Task (Bechara, Damasio, & Damasio, 2003), in which they are presented with four decks of cards and have to learn to choose the decks that will maximize their profit. The decks have different reinforcement rates, with two decks leading primarily to loss of money while two to win of money. Usually, following an initial exploratory phase, participants learn to distinguish the different decks, directing their choice behavior preferentially towards the advantageous decks. Indeed, while this occurred in participants with low alexithymia, those with high alexithymia showed a different pattern of behavior. After the exploratory phase, they did shift their choice towards advantageous decks; nevertheless, this lasted only for a limited number of trials, showing another shift in choice behavior, returning to choose disadvantageous decks (Ferguson et al., 2009). The authors argued that the results suggest an inability of individuals with high alexithymia to consolidate learning due to a reduced sensitivity to monetary losses, which appears in line with what was found here in Experiment 6, with alexithymia being related to reduced ability to learn from punishment and avoid negative stimuli. Successful performance on the Iowa Gambling Task has also been related to the ability to produce appropriate physiological changes in anticipation of choice behavior, which would be necessary to accurately predict the outcome following the choice and guide behavior towards adaptive choices (Damasio, 2008). In

particular, during the task healthy participants generate changes in skin conductance response, following feedback about their decision indicating win or loss of money. As the task progresses, such changes can also be observed prior the decision, as participants think to the choice to make. Absence of this anticipatory response, following damage to the amygdala or ventromedial prefrontal cortex, is related to impaired performance in the task, with patients being unable to learn which the most adaptive choices are (Bechara et al., 1999; Bechara, Tranel, Damasio, & Damasio, 1996). This evidence supports the importance of emotional learning and of psychophysiological markers of prediction of future outcomes for successful decision making. Therefore, the reduced performance in the Iowa Gabling Task in alexithymia, found by Ferguson et al. (2009), may be a manifestation of an underlying impairment in the production of appropriate psychophysiological markers of prediction of future negative outcomes. Indeed, this was shown by Experiment 3, where participants with alexithymia exhibited reduced skin conductance response in anticipation of an aversive stimulus and enhanced extinction of this response indicating impairment in maintaining the prediction of the aversive stimulus even in absence of emotional reinforcement.

The underlying brain mechanisms

From the literature reviewed in the first two chapters, an overlap between the brain areas that play a role in alexithymia and emotional learning appeared evident, in particular, the amygdala, the anterior cingulate cortex and the insula, in keeping with the idea that the areas involved in the construction of internal representations of emotional stimuli are also more broadly involved in processing such stimuli. Thanks to the dependent measures assessed here, some considerations can be made also regarding the possible brain mechanisms involved in alexithymia, though they remain more speculative given that no measure of regional brain activity was collected.

With regard to the amygdala, given its crucial role in the generation of skin conductance response in anticipation of threat (Bechara et al., 1995; LaBar et al., 1998, 1995; Phelps & LeDoux, 2005), the decreased psychophysiological arousal found in alexithymia in anticipation of shock, in Experiment 3, may indicate reduced activity in this structure. This would be in line with the previous literature on alexithymia, which has reported decreased activation of the amygdala in response to the presentation of emotional facial expressions (Jongen et al., 2014; Kugel et al., 2008), negative emotional stimuli (Moriguchi & Komaki, 2013; van der Velde et al., 2013; Wingbermühle et al., 2012) and in particular fearful ones (Pouga et al., 2010).

Regarding the anterior cingulate cortex, the absence of modulation of feedback related negativity evidenced by the electrophysiological recording in Experiment 5, would suggest differences in activity in this region, as this component is generally related to activity in the anterior cingulate cortex (Holroyd et al., 2004; Warren et al., 2015). This would also be in line with previous evidence from the broader literature on alexithymia (van der Velde et al., 2013), although the direction of such differences remains unclear, as neuroimaging studies have found both increased (Mériau et al., 2006; Pouga et al., 2010) and decreased (Moriguchi et al., 2007) activation of this region in alexithymia.

Regarding the insula, its role in emotional learning appears to be broader than the other two regions, since it may not only be involved in processing he sensory information conveyed by unconditioned stimuli, such as nociceptive information conveyed by shock (Craig, 2002, 2003; Critchley et al., 2004), but it also may be involved in the cognitive representation of subjective emotional experience (Barrett, 2017b; LeDoux & Brown, 2017). Given that, here, in Experiment 2 no evidence for differences in the intensity of the shock received and the subjective reports of emotional experience was found in alexithymia, it is

possible that the functionality of the insula may remain spared in these tasks in alexithymia. Nevertheless, the literature on alexithymia has previously found changes in insula activation, though both increased (Bird et al., 2010; Moriguchi et al., 2007; Wiebking & Northoff, 2015) and decreased insula activation (Reker et al., 2010) have been reported. Therefore, the role of this area in the emotional learning in alexithymia should be clarified.

In addition, the results of the present experiments seem to suggest possible differences in the dopamine system in alexithymia. Indeed, dopamine plays a crucial role in learning both from appetitive and aversive stimuli, so that when such stimuli occur unpredicted, phasic changes in dopamine release act as a teaching signal, coding a prediction error that enables the construction of accurate internal representations of emotional events (Schultz, 2016; Wenzel, Rauscher, Cheer, & Oleson, 2015). In particular, the amplitude of the CUE-P300 and FRN, assessed in Experiment 5, and the efficiency in avoiding negative stimuli, assessed in Experiment 6, have been previously put in relationship with the dopaminergic system and the density of D2 dopamine receptors (Berman et al., 2006; Frank et al., 2007, 2004; Maia & Frank, 2011; McKetin et al., 1999; Mulert et al., 2006; Nieuwenhuis et al., 2004; Pfabigan et al., 2014; Pogarell et al., 2011; Sambrook & Goslin, 2015; Walsh & Anderson, 2012). So, the decreased amplitude of the CUE-P300 and FRN and lower performance in avoiding negative stimuli found in alexithymia may suggest decreased density of D2 dopamine receptors, which may then impair the functionality of the dopamine teaching signal. In this regard, one study found that carriers of an allele associated with a reduction in D2 dopamine receptor, together with an allele associated with lower activity-dependent secretion of brain-derived neurotrophic factor, had significantly higher scores of alexithymia, compared to participants with other allelic variations (Klein et al., 2007). Therefore, there is preliminary evidence of the involvement of the dopamine system in alexithymia and future studies could further explore the relationship between differences in the dopamine system and alexithymia.

Final remarks

In sum, the present thesis contributed to extend the current literature on the description and understanding of alexithymia, a personality trait defined by difficulty in identifying feelings, in describing them and a type of thinking focused on concrete aspects rather than on introspection (Sifneos, 1973; Taylor et al., 1991). As shown by Experiment 1, alexithymia has a significant impact on the life of individuals, being related to lower emotional intelligence, empathy and psychological wellbeing, highlighting the importance of studying this personality trait. In addition, when focusing on the perception of emotions, alexithymia appears related, not only to the need for more time, as previously shown (Grynberg et al., 2012), but also to the need for more emotional intensity to identify fear in facial expressions, as shown by Experiment 2. Although these experiments extend the current literature describing the difficulties of alexithymia in processing emotional stimuli, the basic mechanisms underlying such difficulties remain poorly understood.

For this reason, the remaining of the thesis focused on investigating one mechanism that may underlie such difficulties, namely emotional learning. Indeed, through emotional learning, the internal representation of stimuli is shaped, so that neutral stimuli acquire emotional value. As such, emotional learning supports the construction of an internal model of emotional stimuli, which is used to predict the emotional future, in order to respond effectively to coming emotional stimuli (Bubic et al., 2010; McNally & Westbrook, 2006; Öhman & Mineka, 2001). Therefore, impairment in emotional learning may then compromise processing of emotional stimuli. In fact, impairment in this process has been reported in clinical conditions marked by difficulties in emotion processing, such as depression (Greenberg et al., 2015) or anxiety (Lissek et al., 2005, 2014); nevertheless, this has never

been investigated in alexithymia before. Here, four experiments were dedicated to this investigation.

Experiment 3 showed that alexithymia is related to impairment in learning the aversive value of stimuli, evidenced by reduced physiological markers of emotional prediction in Pavlovian threat conditioning. On the contrary, such evidence was not found in Experiment 4 when physiological markers of emotional prediction were assessed during Pavlovian reward conditioning. Despite this, evidence for impairment in learning appetitive value of stimuli, in alexithymia, was found in Experiment 5, where electrophysiological markers of emotional prediction and prediction error were assessed during Pavlovian reward conditioning. Finally, Experiment 6 examined the ability to learn the emotional value of actions during instrumental learning, and to use this learned value for adaptive behavior in a new context. Alexithymia was related to a difficulty in learning from punishment, marked by longer response time when having to avoid stimuli, which had previously acquired aversive value, encountered in a new context. Considered together, these results indicate impairment in emotional learning in alexithymia. In other words, alexithymia appears related to impairment in attributing emotional value to previously neutral stimuli and aberrant internal representations of such stimuli. In addition, this impairment may be more severe when learning the value of aversive than appetitive stimuli, in line with the broader literature on alexithymia, which reported altered processing of negative stimuli, while processing of positive stimuli may be partly spared (e.g. Kugel et al., 2008; Pouga et al., 2010; Prkachin et al., 2009; Reker et al., 2010; Scarpazza et al., 2014; van der Velde et al., 2014).

Contrary to the differences found in the psychophysiological and electrophysiological markers of emotional learning, no evidence was found that the subjective emotional experience during emotional learning was affected in alexithymia. The contrasting result,

between these different types of measures of emotional response, corroborates recent cognitive theories of emotion processing, which argue that the processes giving rise to psychophysiological and electrophysiological signals in response to emotional stimuli, are partly dissociated from those giving rise to the subjective emotional experience associated with them (Barrett, 2017b; Barrett et al., 2007; LeDoux, 1998; LeDoux & Brown, 2017; LeDoux & Hofmann, 2018). In addition, turning to the understanding of alexithymia, the difference between the results of psychophysiological and electrophysiological markers and subjective emotional experience, may suggest that alexithymia may affect the former while sparing the latter aspect of processing of emotional stimuli. Nevertheless, it should also be remembered that the subjective emotional experience is constructed by the integration of information coming from multiple sources, both from the internal (i.e. bodily signals) and external environment (LeDoux, 1998; LeDoux & Brown, 2017; LeDoux & Hofmann, 2018). Therefore, it is also possible that, despite the aberrant psychophysiological and electrophysiological markers of emotional learning, participants with alexithymia may have used contextual information to construct their subjective emotional experience, obtaining a report comparable to the individuals without alexithymia. Future studies could clarify whether differences in subjective emotional experience in alexithymia become evident, when individuals are required to rely primarily on bodily signals to construct their subjective emotional experience, which may indeed be possible, given that reduced interoception was previously found in alexithymia (Brewer et al., 2016; Murphy et al., 2017; Shah et al., 2016).

Finally, the differences found in emotional learning in alexithymia also offer new insight into the understanding of alexithymia. Indeed, the difficulties in perceiving, responding and regulating the response to emotional stimuli as well as using them to make adaptive decisions found in the previous literature on alexithymia, may be the manifestation of underlying inaccurate internal representations of emotional stimuli. This seems also

supported by the common neural mechanisms underlying the difficulties in processing emotional stimuli and in emotional learning characteristic of alexithymia. In particular, the present results suggest possible differences in activity of the amygdala and anterior cingulate cortex in emotional learning in alexithymia, areas that have been previously pointed out among the neural correlates of alexithymia (van der Velde et al., 2013). In addition, the results of the thesis suggest that differences in the dopamine system may underlie alexithymia, in line with preliminary evidence from a previous behavioral genetic study (Klein et al., 2007). Future research should further explore these aspects, in order to extend the understanding of the differences in the internal representations of emotional stimuli in alexithymia and their underlying neural mechanisms.

Reference list

Abrosoft FantaMorph. (2009). (Version 4.1). Abrosoft. Retrieved from www.fantamorph.com

Adolphs, R. (2002). Recognizing Emotion from Facial Expressions: Psychological and Neurological Mechanisms. *Behavioral and Cognitive Neuroscience Reviews*, *1*(1), 21–62. https://doi.org/10.1177/1534582302001001003

Adolphs, R., Denburg, N. L., & Tranel, D. (2001). The amygdala's role in long-term declarative memory for gist and detail. *Behavioral Neuroscience*, *115*(5), 983. https://doi.org/10.1037/0735-7044.115.5.983

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*(6507), 669–672. https://doi.org/10.1038/372669a0

Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, A. J., Phelps, E. A., … Damasio, A. R. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia*, *37*(10), 1111–1117. https://doi.org/10.1016/S0028-3932(99)00039-1

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience*, *14*(10), 1338–1344. https://doi.org/10.1038/nn.2921

Alexander, W. H., & Brown, J. W. (2014). A general role for medial prefrontal cortex in event prediction. *Frontiers in Computational Neuroscience*, *8*. https://doi.org/10.3389/fncom.2014.00069

Austin, E. J., Saklofske, D. H., & Egan, V. (2005). Personality, well-being and health correlates of trait emotional intelligence. *Personality and Individual Differences*, *38*(3), 547–558. https://doi.org/10.1016/j.paid.2004.05.009

Averbeck, B. B., & Costa, V. D. (2017). Motivational neural circuits underlying reinforcement learning. *Nature Neuroscience*, *20*(4), 505–512. https://doi.org/10.1038/nn.4506

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia scale-I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, *38*(1), 23–32. https://doi.org/10.1016/0022-3999(94)90005-1

Barrett, L. F. (2017a). *How Emotions Are Made: The Secret Life of the Brain*. Houghton Mifflin Harcourt.

Barrett, L. F. (2017b). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23. https://doi.org/10.1093/scan/nsw154

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The Experience of Emotion. *Annual Review of Psychology*, *58*, 373–403. https://doi.org/10.1146/annurev.psych.58.110405.085709

Baughman, H. M., Schwartz, S., Schermer, J. A., Veselka, L., Petrides, K. V., & Vernon, P. A. (2011). A behavioral-genetic study of alexithymia and its relationships with trait emotional intelligence. *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies*, *14*(6), 539–543.

Bausch, S., Stingl, M., Hartmann, L. C., Leibing, E., Leichsenring, F., Kruse, J., … Leweke, F. (2011). Alexithymia and script-driven emotional imagery in healthy female subjects: no support for deficiencies in imagination. *Scandinavian Journal of Psychology*, *52*(2), 179–184. https://doi.org/10.1111/j.1467-9450.2010.00847.x

Bechara, A., Damasio, H., & Damasio, A. R. (2003). Role of the Amygdala in Decision-Making. *Annals of the New York Academy of Sciences*, *985*(1), 356–369. https://doi.org/10.1111/j.1749-6632.2003.tb07094.x

Bechara, A., Damasio, H., Damasio, A. R., & Lee, G. P. (1999). Different Contributions of the Human Amygdala and Ventromedial Prefrontal Cortex to Decision-Making. *Journal of Neuroscience*, *19*(13), 5473–5481.

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science (New York, N.Y.)*, *269*(5227), 1115–1118.

Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1996). Failure to Respond Autonomically to Anticipated Future Outcomes Following Damage to Prefrontal Cortex. *Cerebral Cortex*, *6*(2), 215–225. https://doi.org/10.1093/cercor/6.2.215

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159.

Berman, S. M., Noble, E. P., Antolin, T., Sheen, C., Conner, B. T., & Ritchie, T. (2006). P300 development during adolescence: effects of DRD2 genotype. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *117*(3), 649–659. https://doi.org/10.1016/j.clinph.2005.11.012

Bermond, B., Bierman, D. J., Cladder, M. A., Moormann, P. P., & Vorst, H. C. M. (2010). The cognitive and affective alexithymia dimensions in the regulation of sympathetic responses. *International Journal of Psychophysiology*, *75*(3), 227–233. https://doi.org/10.1016/j.ijpsycho.2009.11.004

Bermond, B., Righart, R., Ridderinkhof, K. R., & Moormann, P. P. (2008). Alexithymia and the brain potential P300. *Netherlands Journal of Psychology*, *64*(2), 65–77. https://doi.org/10.1007/BF03076408

Bernhardt, B. C., & Singer, T. (2012). The Neural Basis of Empathy. *Annual Review of Neuroscience*, *35*(1), 1–23. https://doi.org/10.1146/annurev-neuro-062111-150536

Berthoz, S., Consoli, S., Perez-Diaz, F., & Jouvent, R. (1999). Alexithymia and anxiety: compounded relationships? A psychometric study. *European Psychiatry*, *14*(7), 372–378. https://doi.org/10.1016/S0924-9338(99)00233-3

Bertini, C., Cecere, R., & Làdavas, E. (2013). I am blind, but I "see" fear. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *49*(4), 985–993. https://doi.org/10.1016/j.cortex.2012.02.006

Bertossi, E., Aleo, F., Braghittoni, D., & Ciaramelli, E. (2016). Stuck in the here and now: Construction of fictitious and future experiences following ventromedial prefrontal damage. *Neuropsychologia*, *81*(Supplement C), 107–116. https://doi.org/10.1016/j.neuropsychologia.2015.12.015

Beyeler, A., Namburi, P., Glober, G. F., Simonnet, C., Calhoon, G. G., Conyers, G. F., … Tye, K. M. (2016). Divergent Routing of Positive and Negative Information from the Amygdala during Memory Retrieval. *Neuron*, *90*(2), 348–361. https://doi.org/10.1016/j.neuron.2016.03.004

Bird, G., Silani, G., Brindley, R., White, S., Frith, U., & Singer, T. (2010). Empathic brain responses in insula are modulated by levels of alexithymia but not autism. *Brain*, *133*(5), 1515–1525. https://doi.org/10.1093/brain/awq060

Borhani, K., Borgomaneri, S., Làdavas, E., & Bertini, C. (2016). The effect of alexithymia on early visual processing of emotional body postures. *Biological Psychology*, *115*, 1–8. https://doi.org/10.1016/j.biopsycho.2015.12.010

Brewer, R., Cook, R., & Bird, G. (2016). Alexithymia: a general deficit of interoception. *Royal Society Open Science*, *3*(10). https://doi.org/10.1098/rsos.150664

Brosch, T., Pourtois, G., & Sander, D. (2010). The perception and categorisation of emotional stimuli: A review. *Cognition and Emotion*, *24*(3), 377–400. https://doi.org/10.1080/02699930902975754

Broyd, S. J., Richards, H. J., Helps, S. K., Chronaki, G., Bamford, S., & Sonuga-Barke, E. J. S. (2012). An electrophysiological monetary incentive delay (e-MID) task: A way to decompose the different components of neural response to positive and negative monetary reinforcement. *Journal of Neuroscience Methods*, *209*(1), 40–49. https://doi.org/10.1016/j.jneumeth.2012.05.015

Bubic, A., Von Cramon, D. Y., Schubotz, R. I., Bubic, A., Cramon, D. Y. von, & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, *4*, 25. https://doi.org/10.3389/fnhum.2010.00025

Büchel, C., Dolan, R. J., Armony, J. L., & Friston, K. J. (1999). Amygdala-hippocampal involvement in human aversive trace conditioning revealed through event-related functional magnetic resonance imaging. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *19*(24), 10869–10876.

Büchel, C., Morris, J., Dolan, R. J., & Friston, K. J. (1998). Brain Systems Mediating Aversive Conditioning: an Event-Related fMRI Study. *Neuron*, *20*(5), 947–957. https://doi.org/10.1016/S0896-6273(00)80476-6

Cardinal, R. N., Parkinson, J. A., Hall, J., & Everitt, B. J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience & Biobehavioral Reviews*, *26*(3), 321–352. https://doi.org/10.1016/S0149-7634(02)00007-6

Cecere, R., Bertini, C., Maier, M. E., & Làdavas, E. (2014). Unseen Fearful Faces Influence Face Encoding: Evidence from ERPs in Hemianopic Patients. *Journal of Cognitive Neuroscience*, *26*(11), 2564–2577. https://doi.org/10.1162/jocn_a_00671

Connelly, M., & Denney, D. R. (2007). Regulation of emotions during experimental stress in alexithymia. *Journal of Psychosomatic Research*, *62*(6), 649–656. https://doi.org/10.1016/j.jpsychores.2006.12.008

Cook, R., Brewer, R., Shah, P., & Bird, G. (2013). Alexithymia, Not Autism, Predicts Poor Recognition of Emotional Facial ExpressionsPsychological Science. *Psychological Science*, *24*(5), 723–732.

Craig, A. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews Neuroscience*, *3*(8), 655–666. https://doi.org/10.1038/nrn894

Craig, A. (2003). Interoception: the sense of the physiological condition of the body. *Current Opinion in Neurobiology*, *13*(4), 500–505. https://doi.org/10.1016/S0959-4388(03)00090-4

Critchley, H. D., Mathias, C. J., & Dolan, R. J. (2002). Fear Conditioning in Humans:The Influence of Awareness and Autonomic Arousal on Functional Neuroanatomy. *Neuron*, *33*(4), 653–663. https://doi.org/10.1016/S0896-6273(02)00588-3

Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience*, *7*(2), 189–195. https://doi.org/10.1038/nn1176

Curci, A., Lanciano, T., Soleti, E., Zammuner, V. L., & Salovey, P. (2013). Construct Validity of the Italian Version of the Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT) v2.0. *Journal of Personality Assessment*, *95*(5), 486–494. https://doi.org/10.1080/00223891.2013.778272

Damasio, A. (2008). *Descartes' Error: Emotion, Reason and the Human Brain*. Random House.

Davis, M. H. (1980). Individual differences in empathy: A multidimensional approach. *APA PsycNET*. Retrieved from http://ezproxy.library.nyu.edu:2101/index.cfm?fa=search.displayRecord&id=6F711544-B62F-2D27-7781-CA40B6A7175F&resultID=24&page=1&dbTab=all&search=true

Daw, N. D., & Tobler, P. N. (2014). Chapter 15 - Value Learning through Reinforcement: The Basics of Dopamine and Reinforcement Learning. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics (Second Edition)* (pp. 283–298). San Diego: Academic Press. https://doi.org/10.1016/B978-0-12-416008-8.00015-2

Decety, J. (2015). The neural pathways, development and functions of empathy. *Current Opinion in Behavioral Sciences*, *3*, 1–6. https://doi.org/10.1016/j.cobeha.2014.12.001

Delgado, M. R. (2007). Reward-Related Responses in the Human Striatum. *Annals of the New York Academy of Sciences*, *1104*(1), 70–88. https://doi.org/10.1196/annals.1390.002

Delgado, M. R., Gillis, M. M., & Phelps, E. A. (2008). Regulating the expectation of reward via cognitive strategies. *Nature Neuroscience*, *11*(8), 880–881. https://doi.org/10.1038/nn.2141

Delgado, M. R., Jou, R. L., LeDoux, J., Phelps, L., Delgado, M. R., Jou, R. L., … Phelps, E. A. (2009). Avoiding negative outcomes: tracking the mechanisms of avoidance learning in humans during fear conditioning. *Frontiers in Behavioral Neuroscience*, *3*, 33. https://doi.org/10.3389/neuro.08.033.2009

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

Duan, X., Dai, Q., Gong, Q., & Chen, H. (2010). Neural mechanism of unconscious perception of surprised facial expression. *NeuroImage*, *52*(1), 401–407. https://doi.org/10.1016/j.neuroimage.2010.04.021

Duits, P., Cath, D. C., Lissek, S., Hox, J. J., Hamm, A. O., Engelhard, I. M., … Baas, J. M. P. (2015). Updated Meta-Analysis of Classical Fear Conditioning in the Anxiety Disorders. *Depression and Anxiety*, *32*(4), 239–253. https://doi.org/10.1002/da.22353

Eastabrook, J. M., Lanteigne, D. M., & Hollenstein, T. (2013). Decoupling between physiological, self-reported, and expressed emotional responses in alexithymia. *Personality and Individual Differences*, *55*(8), 978–982. https://doi.org/10.1016/j.paid.2013.08.001

Easterbrooks, M. A., Chaudhuri, J. H., & Gestsdottir, S. (2005). Patterns of emotional availability among young mothers and their infants: A dydaic, contextual analysis. *Infant Mental Health Journal*, *26*(4), 309–326. https://doi.org/10.1002/imhj.20057

Eichmann, M., Kugel, H., & Suslow, T. (2008). Difficulty identifying feelings and automatic activation in the fusiform gyrus in response to facial emotion. *Perceptual and Motor Skills*, *107*(3), 915–922. https://doi.org/10.2466/pms.107.3.915-922

Ekman, P., & Friesen, W. (1976). *Picture of facial affect*. Palo Alto: Consulting Psychologists Press.

Etkin, A., Egner, T., & Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in Cognitive Sciences*, *15*(2), 85–93. https://doi.org/10.1016/j.tics.2010.11.004

Ferguson, E., Bibby, P. A., Rosamond, S., O'Grady, C., Parcell, A., Amos, C., … O'Carroll, R. (2009). Alexithymia, Cumulative Feedback, and Differential Response Patterns on the Iowa Gambling Task. *Journal of Personality*, *77*(3), 883–902. https://doi.org/10.1111/j.1467-6494.2009.00568.x

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences*, *104*(41), 16311–16316. https://doi.org/10.1073/pnas.0706111104

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By Carrot or by Stick: Cognitive Reinforcement Learning in Parkinsonism. *Science*, *306*(5703), 1940–1943. https://doi.org/10.1126/science.1102941

Frank, M. J., Woroch, B. S., & Curran, T. (2005). Error-Related Negativity Predicts Reinforcement Learning and Conflict Biases. *Neuron*, *47*(4), 495–501. https://doi.org/10.1016/j.neuron.2005.06.020

Franz, M., Schaefer, R., & Schneider, C. (2003). Psychophysiological Response Patterns of High and Low Alexithymics Under Mental and Emotional Load Conditions. *Journal of Psychophysiology*, *17*(4), 203–213. https://doi.org/10.1027/0269-8803.17.4.203

Franz, M., Schaefer, R., Schneider, C., Sitte, W., & Bachor, J. (2004). Visual Event-Related Potentials in Subjects With Alexithymia: Modified Processing of Emotional Aversive Information? *American Journal of Psychiatry*, *161*(4), 728–735. https://doi.org/10.1176/appi.ajp.161.4.728

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

Fullana, M. A., Harrison, B. J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., & Radua, J. (2015). Neural signatures of human fear conditioning: an updated and

extended meta-analysis of fMRI studies. *Molecular Psychiatry*. https://doi.org/10.1038/mp.2015.88

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., … Politi, P. (2009). Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *Journal of Psychiatry & Neuroscience : JPN*, *34*(6), 418–32.

Garofalo, S., Maier, M. E., & di Pellegrino, G. (2014). Mediofrontal negativity signals unexpected omission of aversive events. *Scientific Reports*, *4*. https://doi.org/10.1038/srep04816

Garofalo, S., Timmermann, C., Battaglia, S., Maier, M. E., & di Pellegrino, G. (2016). Mediofrontal Negativity Signals Unexpected Timing of Salient Outcomes. *Journal of Cognitive Neuroscience*, 1–10. https://doi.org/10.1162/jocn_a_01074

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A Neural System for Error Detection and Compensation. *Psychological Science*, *4*(6), 385–390. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x

Gehring, W. J., & Willoughby, A. R. (2002). The Medial Frontal Cortex and the Rapid Processing of Monetary Gains and Losses. *Science*, *295*(5563), 2279–2282. https://doi.org/10.1126/science.1066893

Goerlich, K. S., Aleman, A., & Martens, S. (2012). The Sound of Feelings: Electrophysiological Responses to Emotional Speech in Alexithymia. *PLOS ONE*, *7*(5), e36951. https://doi.org/10.1371/journal.pone.0036951

Goerlich-Dobre, K. S., Witteman, J., Schiller, N. O., van Heuven, V. J. P., Aleman, A., & Martens, S. (2014). Blunted feelings: Alexithymia is associated with a diminished neural response to speech prosody. *Social Cognitive and Affective Neuroscience*, *9*(8), 1108–1117. https://doi.org/10.1093/scan/nst075

Goldstein, R. Z., Cottone, L. A., Jia, Z., Maloney, T., Volkow, N. D., & Squires, N. K. (2006). The effect of graded monetary reward on cognitive event-related potentials and behavior in young healthy adults. *International Journal of Psychophysiology*, *62*(2), 272–279. https://doi.org/10.1016/j.ijpsycho.2006.05.006

Greenberg, T., Chase, H. W., Almeida, J. R., Stiffler, R., Zevallos, C. R., Aslam, H. A., … Phillips, M. L. (2015). Moderation of the Relationship Between Reward Expectancy and Prediction Error-Related Ventral Striatal Reactivity by Anhedonia in Unmedicated Major Depressive Disorder: Findings From the EMBARC Study. *The American Journal of Psychiatry*, *172*(9), 881–891. https://doi.org/10.1176/appi.ajp.2015.14050594

Grieve, R., & Mahar, D. (2010). The emotional manipulation–psychopathy nexus: Relationships with emotional intelligence, alexithymia and ethical position. *Personality and Individual Differences*, *48*(8), 945–950. https://doi.org/10.1016/j.paid.2010.02.028

Grynberg, D., Chang, B., Corneille, O., Maurage, P., Vermeulen, N., Berthoz, S., & Luminet, O. (2012). Alexithymia and the Processing of Emotional Facial Expressions (EFEs): Systematic Review, Unanswered Questions and Further Perspectives. *PLoS ONE*, *7*(8), e42429. https://doi.org/10.1371/journal.pone.0042429

Grynberg, D., Luminet, O., Corneille, O., Grèzes, J., & Berthoz, S. (2010). Alexithymia in the interpersonal domain: A general deficit of empathy? *Personality and Individual Differences*, *49*(8), 845–850. https://doi.org/10.1016/j.paid.2010.07.013

Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2006). The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biological Psychology*, *71*(2), 148–154. https://doi.org/10.1016/j.biopsycho.2005.04.001

Hajihosseini, A., & Holroyd, C. B. (2013). Frontal midline theta and N200 amplitude reflect complementary information about expectancy and outcome evaluation. *Psychophysiology*, *50*(6), 550–562. https://doi.org/10.1111/psyp.12040

Hassabis, D., Kumaran, D., & Maguire, E. A. (2007). Using Imagination to Understand the Neural Basis of Episodic Memory. *Journal of Neuroscience*, *27*(52), 14365–14374.

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*, *11*(7), 299–306. https://doi.org/10.1016/j.tics.2007.05.001

Heinz, A., Schlagenhauf, F., Beck, A., & Wackerhagen, C. (2016). Dimensional psychiatry: mental disorders as dysfunctions of basic learning mechanisms. *Journal of Neural Transmission*, *123*(8), 809–821. https://doi.org/10.1007/s00702-016-1561-2

Hendryx, M. S., Haviland, M. G., & Shaw, D. G. (1991). Dimensions of Alexithymia and Their Relationships to Anxiety and Depression. *Journal of Personality Assessment*, *56*(2), 227–237. https://doi.org/10.1207/s15327752jpa5602_4

Holroyd, C. B., & Krigolson, O. E. (2007). Reward prediction error signals associated with a modified time estimation task. *Psychophysiology*, *44*(6), 913–917. https://doi.org/10.1111/j.1469-8986.2007.00561.x

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., & Cohen, J. D. (2003). Errors in reward prediction are reflected in the event-related brain potential. *Neuroreport*, *14*(18), 2481–2484.

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., Nystrom, L., Mars, R. B., Coles, M. G. H., & Cohen, J. D. (2004). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nature Neuroscience*, *7*(5), nn1238. https://doi.org/10.1038/nn1238

Honkalampi, K., Hintikka, J., Tanskanen, A., Lehtonen, J., & Viinamäki, H. (2000). Depression is strongly associated with alexithymia in the general population. *Journal of Psychosomatic Research*, *48*(1), 99–104. https://doi.org/10.1016/S0022-3999(99)00083-5

Ihme, K., Sacher, J., Lichev, V., Rosenberg, N., Kugel, H., Rufer, M., … Suslow, T. (2014). Alexithymia and the labeling of facial emotions: response slowing and increased motor and somatosensory processing. *BMC Neuroscience*, *15*(1), 40. https://doi.org/10.1186/1471-2202-15-40

Ihme, K., Sacher, J., Lichev, V., Rosenberg, N., Kugel, H., Rufer, M., … Suslow, T. (2014). Alexithymic features and the labeling of brief emotional facial expressions – An fMRI study. *Neuropsychologia*, *64*, 289–299. https://doi.org/10.1016/j.neuropsychologia.2014.09.044

Jessimer, M., & Markham, R. (1997). Alexithymia: A Right Hemisphere Dysfunction Specific to Recognition of Certain Facial Expressions? *Brain and Cognition*, *34*(2), 246–258. https://doi.org/10.1006/brcg.1997.0900

Jongen, S., Axmacher, N., Kremers, N. A. W., Hoffmann, H., Limbrecht-Ecklundt, K., Traue, H. C., & Kessler, H. (2014). An investigation of facial emotion recognition impairments in alexithymia and its neural correlates. *Behavioural Brain Research*, *271*, 129–139. https://doi.org/10.1016/j.bbr.2014.05.069

Jutten, C., & Herault, J. (1991). Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, *24*(1), 1–10. https://doi.org/10.1016/0165-1684(91)90079-X

Ketelaars, M. P., In't Velt, A., Mol, A., Swaab, H., & van Rijn, S. (2016). Emotion recognition and alexithymia in high functioning females with autism spectrum

disorder. *Research in Autism Spectrum Disorders*, *21*, 51–60. https://doi.org/10.1016/j.rasd.2015.09.006

Kirsch, P., Schienle, A., Stark, R., Sammer, G., Blecker, C., Walter, B., … Vaitl, D. (2003). Anticipation of reward in a nonaversive differential conditioning paradigm and the brain reward system:: an event-related fMRI study. *NeuroImage*, *20*(2), 1086–1095. https://doi.org/10.1016/S1053-8119(03)00381-1

Klein, T. A., Neumann, J., Reuter, M., Hennig, J., Cramon, D. Y. von, & Ullsperger, M. (2007). Genetically Determined Differences in Learning from Errors. *Science*, *318*(5856), 1642–1645. https://doi.org/10.1126/science.1145044

Knight, D. C., Nguyen, H. T., & Bandettini, P. A. (2005). The role of the human amygdala in the production of conditioned fear responses. *NeuroImage*, *26*(4), 1193–1200. https://doi.org/10.1016/j.neuroimage.2005.03.020

Knutson, B., & Cooper, J. C. (2005). Functional magnetic resonance imaging of reward prediction. *Current Opinion in Neurology*, *18*(4), 411–417.

Kokkonen, P., Karvonen, J. T., Veijola, J., Läksy, K., Jokelainen, J., Järvelin, M.-R., & Joukamaa, M. (2001). Prevalence and sociodemographic correlates of alexithymia in a population sample of young adults. *Comprehensive Psychiatry*, *42*(6), 471–476. https://doi.org/10.1053/comp.2001.27892

Kugel, H., Eichmann, M., Dannlowski, U., Ohrmann, P., Bauer, J., Arolt, V., … Suslow, T. (2008). Alexithymic features and automatic amygdala reactivity to facial emotion. *Neuroscience Letters*, *435*(1), 40–44. https://doi.org/10.1016/j.neulet.2008.02.005

LaBar, K. S., Gatenby, J. C., Gore, J. C., LeDoux, J. E., & Phelps, E. A. (1998). Human Amygdala Activation during Conditioned Fear Acquisition and Extinction: a Mixed-Trial fMRI Study. *Neuron*, *20*(5), 937–945. https://doi.org/10.1016/S0896-6273(00)80475-4

LaBar, K. S., LeDoux, J. E., Spencer, D. D., & Phelps, E. A. (1995). Impaired fear conditioning following unilateral temporal lobectomy in humans. *The Journal of Neuroscience*, *15*(10), 6846–6855.

Làdavas, E., Cimatti, D., Pesce, M. D., & Tuozzi, G. (1993). Emotional evaluation with and without conscious stimulus identification: evidence from a split-brain patient. *Cognition and Emotion*, *7*(1), 95–114. https://doi.org/10.1080/02699939308409179

Lane, R. D., Ahern, G. L., Schwartz, G. E., & Kaszniak, A. W. (1997). Is Alexithymia the Emotional Equivalent of Blindsight? *Biological Psychiatry*, *42*(9), 834–844. https://doi.org/10.1016/S0006-3223(97)00050-4

Lane, R. D., Lee, S., Reidel, R., Weldon, V. B., Kaszniak, A., & Schwartz, G. E. (1996). Impaired Verbal and Nonverbal Emotion Recognition in Alexithymia. *Psychosomatic Medicine*, *58*(3), 203–210.

Lane, R. D., Weihs, K. L., Herring, A., Hishaw, A., & Smith, R. (2015). Affective agnosia: Expansion of the alexithymia construct and a new opportunity to integrate and extend Freud's legacy. *Neuroscience & Biobehavioral Reviews*, *55*, 594–611. https://doi.org/10.1016/j.neubiorev.2015.06.007

LeDoux, J. (1998). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. Simon and Schuster.

LeDoux, J. (2007). The amygdala. *Current Biology*, *17*(20), R868–R874. https://doi.org/10.1016/j.cub.2007.08.005

LeDoux, J. (2014). Coming to terms with fear. *Proceedings of the National Academy of Sciences*, *111*(8), 2871–2878. https://doi.org/10.1073/pnas.1400335111

LeDoux, J., & Brown, R. (2017). A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Sciences*, *114*(10), E2016–E2025. https://doi.org/10.1073/pnas.1619316114

LeDoux, J., & Hofmann, S. G. (2018). The subjective experience of emotion: a fearful view. *Current Opinion in Behavioral Sciences*, *19*(Supplement C), 67–72. https://doi.org/10.1016/j.cobeha.2017.09.011

Li, S. S. Y., & McNally, G. P. (2014). The conditions that promote fear learning: Prediction error and Pavlovian fear conditioning. *Neurobiology of Learning and Memory*, *108*, 14–21. https://doi.org/10.1016/j.nlm.2013.05.002

Li, S., Zhang, B., Guo, Y., & Zhang, J. (2015). The association between alexithymia as assessed by the 20-item Toronto Alexithymia Scale and depression: A meta-analysis. *Psychiatry Research*, *227*(1), 1–9. https://doi.org/10.1016/j.psychres.2015.02.006

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *The Behavioral and Brain Sciences*, *35*(3), 121–143. https://doi.org/10.1017/S0140525X11000446

Lipsanen, T., Saarijärvi, S., & Lauerma, H. (2004). Exploring the Relations between Depression, Somatization, Dissociation and Alexithymia – Overlapping or Independent Constructs? *Psychopathology*, *37*(4), 200–206. https://doi.org/10.1159/000080132

Lissek, S., Bradford, D. E., Alvarez, R. P., Burton, P., Espensen-Sturges, T., Reynolds, R. C., & Grillon, C. (2014). Neural substrates of classically conditioned fear-generalization in humans: a parametric fMRI study. *Social Cognitive and Affective Neuroscience*, *9*(8), 1134–1142. https://doi.org/10.1093/scan/nst096

Lissek, S., Powers, A. S., McClure, E. B., Phelps, E. A., Woldehawariat, G., Grillon, C., & Pine, D. S. (2005). Classical fear conditioning in the anxiety disorders: a meta-analysis. *Behaviour Research and Therapy*, *43*(11), 1391–1424. https://doi.org/10.1016/j.brat.2004.10.007

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. MIT Press.

Lumley, M. A., Gustavson, B. J., Partridge, R. T., & Labouvie-Vief, G. (2005). Assessing Alexithymia and Related Emotional Ability Constructs Using Multiple Methods: Interrelationships Among Measures. *Emotion September 2005*, *5*(3), 329–342.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska directed emotional faces (KDEF)*. Psychology section, Karolinska Institutet.

Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience*, *14*(2), 154–162. https://doi.org/10.1038/nn.2723

Maier, M. E., Scarpazza, C., Starita, F., Filogamo, R., & Làdavas, E. (2016). Error monitoring is related to processing internal affective states. *Cognitive, Affective, & Behavioral Neuroscience*, 1–13. https://doi.org/10.3758/s13415-016-0452-1

Makeig, S., Bell, A. J., Jung, T.-P., & Sejnowski, T. J. (1996). Independent Component Analysis of Electroencephalographic Data. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), *Advances in Neural Information Processing Systems 8* (pp. 145–151). MIT Press. Retrieved from http://papers.nips.cc/paper/1091-independent-component-analysis-of-electroencephalographic-data.pdf

Mangelli, L., Semprini, F., Sirri, L., Fava, G. A., & Sonino, N. (2006). Use of the Diagnostic Criteria for Psychosomatic Research (DCPR) in a Community Sample. *Psychosomatics*, *47*(2), 143–146. https://doi.org/10.1176/appi.psy.47.2.143

Marchesi, C., Brusamonti, E., & Maggini, C. (2000). Are alexithymia, depression, and anxiety distinct constructs in affective disorders? *Journal of Psychosomatic Research*, *49*(1), 43–49. https://doi.org/10.1016/S0022-3999(00)00084-2

Maren, S. (2001). Neurobiology of Pavlovian Fear Conditioning. *Annual Review of Neuroscience*, *24*(1), 897–931. https://doi.org/10.1146/annurev.neuro.24.1.897

Mason, O., Tyson, M., Jones, C., & Potts, S. (2005). Alexithymia: Its prevalence and correlates in a British undergraduate sample. *Psychology and Psychotherapy: Theory, Research and Practice*, *78*(1), 113–125. https://doi.org/10.1348/147608304X21374

Mattila, A. K., Salminen, J. K., Nummi, T., & Joukamaa, M. (2006). Age is strongly associated with alexithymia in the general population. *Journal of Psychosomatic Research*, *61*(5), 629–635. https://doi.org/10.1016/j.jpsychores.2006.04.013

Mayer, J. D., Caruso, D. R., & Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence*, *27*(4), 267–298. https://doi.org/10.1016/S0160-2896(99)00016-1

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion (Washington, D.C.)*, *3*(1), 97–105.

McKetin, R., Ward, P. B., Catts, S. V., Mattick, R. P., & Bell, J. R. (1999). Changes in Auditory Selective Attention and Event-Related Potentials Following Oral Administration of D-amphetamine in Humans. *Neuropsychopharmacology*, *21*(3), 1395341. https://doi.org/10.1016/S0893-133X(99)00017-2

McNally, G. P., Johansen, J. P., & Blair, H. T. (2011). Placing prediction into the fear circuit. *Trends in Neurosciences*, *34*(6), 283–292. https://doi.org/10.1016/j.tins.2011.03.005

McNally, G. P., & Westbrook, R. F. (2006). Predicting danger: The nature, consequences, and neural mechanisms of predictive fear learning. *Learning & Memory*, *13*(3), 245–253. https://doi.org/10.1101/lm.196606

Mériau, K., Wartenburger, I., Kazzer, P., Prehn, K., Lammers, C.-H., van der Meer, E., … Heekeren, H. R. (2006). A neural network reflecting individual differences in cognitive processing of emotions during perceptual decision making. *NeuroImage*, *33*(3), 1016–1027. https://doi.org/10.1016/j.neuroimage.2006.07.031

Mikolajczak, M., Luminet, O., Leroy, C., & Roy, E. (2007). Psychometric Properties of the Trait Emotional Intelligence Questionnaire: Factor Structure, Reliability, Construct, and Incremental Validity in a French-Speaking Population. *Journal of Personality Assessment*, *88*(3), 338–353. https://doi.org/10.1080/00223890701333431

Mikolajczak, M., Luminet, O., & Menil, C. (2006). Predicting resistance to stress: incremental validity of trait emotional intelligence over alexithymia and optimism. Retrieved January 11, 2017, from http://www.redalyc.org/articulo.oa?id=72709512

Milad, M. R., Quirk, G. J., Pitman, R. K., Orr, S. P., Fischl, B., & Rauch, S. L. (2007). A Role for the Human Dorsal Anterior Cingulate Cortex in Fear Expression. *Biological Psychiatry*, *62*(10), 1191–1194. https://doi.org/10.1016/j.biopsych.2007.04.032

Moriguchi, Y., Decety, J., Ohnishi, T., Maeda, M., Mori, T., Nemoto, K., … Komaki, G. (2007). Empathy and Judging Other's Pain: An fMRI Study of Alexithymia. *Cerebral Cortex*, *17*(9), 2223–2234. https://doi.org/10.1093/cercor/bhl130

Moriguchi, Y., & Komaki, G. (2013). Neuroimaging studies of alexithymia: physical, affective, and social perspectives. *BioPsychoSocial Medicine*, *7*(1), 8. https://doi.org/10.1186/1751-0759-7-8

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

Moulton, S. T., & Kosslyn, S. M. (2009). Imagining predictions: mental imagery as mental emulation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1521), 1273–1280. https://doi.org/10.1098/rstb.2008.0314

Mulert, C., Juckel, G., Giegling, I., Pogarell, O., Leicht, G., Karch, S., … Rujescu, D. (2006). A Ser9Gly Polymorphism in the Dopamine D3 Receptor Gene (DRD3) and Event-

Related P300 Potentials. *Neuropsychopharmacology*, *31*(6), 1300984. https://doi.org/10.1038/sj.npp.1300984

Murphy, J., Catmur, C., & Bird, G. (2017). Alexithymia Is Associated With a Multidomain, Multidimensional Failure of Interoception: Evidence From Novel Tests. *Journal of Experimental Psychology. General*. https://doi.org/10.1037/xge0000366

Murray, E. A., & Baxter, M. G. (2002). The amygdala and reward. *Nature Reviews Neuroscience*, *3*(7), 563. https://doi.org/10.1038/nrn875

Nakajima, K., Minami, T., & Nakauchi, S. (2017). Interaction between facial expression and color. *Scientific Reports*, *7*. https://doi.org/10.1038/srep41019

Namburi, P., Al-Hasani, R., Calhoon, G. G., Bruchas, M. R., & Tye, K. M. (2016). Architectural Representation of Valence in the Limbic System. *Neuropsychopharmacology*, *41*(7), 1697–1715. https://doi.org/10.1038/npp.2015.358

Namburi, P., Beyeler, A., Yorozu, S., Calhoon, G. G., Halbert, S. A., Wichmann, R., … Tye, K. M. (2015). A circuit mechanism for differentiating positive and negative associations. *Nature*, *520*(7549), 675–678. https://doi.org/10.1038/nature14366

Neumann, S. A., Sollers III, J. J., Thayer, J. F., & Waldstein, S. R. (2004). Alexithymia predicts attenuated autonomic reactivity, but prolonged recovery to anger recall in young women. *International Journal of Psychophysiology*, *53*(3), 183–195. https://doi.org/10.1016/j.ijpsycho.2004.03.008

Newton, T. L., & Contrada, R. J. (1994). Alexithymia and repression: contrasting emotion-focused coping styles. *Psychosomatic Medicine*, *56*(5), 457–462.

Niedenthal, P. M. (2007). Embodying Emotion. *Science*, *316*(5827), 1002–1005. https://doi.org/10.1126/science.1136930

Niedenthal, P. M., Winkielman, P., Mondillon, L., & Vermeulen, N. (2009). Embodiment of emotion concepts. *Journal of Personality and Social Psychology*, *96*(6), 1120–1136. https://doi.org/10.1037/a0015574

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, *131*(4), 510–532. https://doi.org/10.1037/0033-2909.131.4.510

Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. H. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neuroscience & Biobehavioral Reviews*, *28*(4), 441–448. https://doi.org/10.1016/j.neubiorev.2004.05.003

O'Doherty, J. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Current Opinion in Neurobiology*, *14*(6), 769–776. https://doi.org/10.1016/j.conb.2004.10.016

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. *Science*, *304*(5669), 452–454. https://doi.org/10.1126/science.1094285

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483–522. https://doi.org/10.1037/0033-295X.108.3.483

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

Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature Neuroscience*, *10*(9), 1095–1102. https://doi.org/10.1038/nn1968

Panaite, V., & Bylsma, L. M. (2012). Alexithymia. In V. S. Ramachandran (Ed.), *Encyclopedia of Human Behavior (Second Edition)* (pp. 92–99). San Diego: Academic Press. Retrieved from http://www.sciencedirect.com/science/article/pii/B9780123750006000185

Pandey, R., & Mandal, M. K. (1997). Processing of facial expressions of emotion and alexithymia. *British Journal of Clinical Psychology*, *36*(4), 631–633. https://doi.org/10.1111/j.2044-8260.1997.tb01269.x

Pape, H.-C., & Pare, D. (2010). Plastic Synaptic Networks of the Amygdala for the Acquisition, Expression, and Extinction of Conditioned Fear. *Physiological Reviews*, *90*(2), 419–463. https://doi.org/10.1152/physrev.00037.2009

Parker, J. D. A., Taylor, G. J., & Bagby, R. M. (2001). The relationship between emotional intelligence and alexithymia. *Personality and Individual Differences*, *30*(1), 107–115. https://doi.org/10.1016/S0191-8869(00)00014-3

Parker, P. D., Prkachin, K. M., & Prkachin, G. C. (2005). Processing of Facial Expressions of Negative Emotion in Alexithymia: The Influence of Temporal Constraint. *Journal of Personality*, *73*(4), 1087–1107. https://doi.org/10.1111/j.1467-6494.2005.00339.x

Patil, I., & Silani, G. (2014a). Alexithymia increases moral acceptability of accidental harms. *Journal of Cognitive Psychology*, *26*(5), 597–614. https://doi.org/10.1080/20445911.2014.929137

Patil, I., & Silani, G. (2014b). Reduced empathic concern leads to utilitarian moral judgments in trait alexithymia. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00501

Paton, J. J., Belova, M. A., Morrison, S. E., & Salzman, C. D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature*, *439*(7078), 865–870. https://doi.org/10.1038/nature04490

Pavlov, I. P., & Anrep, G. V. (1927). *Conditioned reflexes; an investigation of the physiological activity of the cerebral cortex.* London: Oxford Univ. Press.

Pfabigan, D. M., Seidel, E.-M., Sladky, R., Hahn, A., Paul, K., Grahl, A., … Lamm, C. (2014). P300 amplitude variation is related to ventral striatum BOLD response during gain and loss anticipation: An EEG and fMRI experiment. *NeuroImage*, *96*(Supplement C), 12–21. https://doi.org/10.1016/j.neuroimage.2014.03.077

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the Amygdala to Emotion Processing: From Animal Models to Human Behavior. *Neuron*, *48*(2), 175–187. https://doi.org/10.1016/j.neuron.2005.09.025

Phelps, E. A., O'Connor, K. J., Gatenby, J. C., Gore, J. C., Grillon, C., & Davis, M. (2001). Activation of the left amygdala to a cognitive representation of fear. *Nature Neuroscience*, *4*(4), 437–441. https://doi.org/10.1038/86110

Pogarell, O., Padberg, F., Karch, S., Segmiller, F., Juckel, G., Mulert, C., … Koch, W. (2011). Dopaminergic mechanisms of target detection — P300 event related potential and striatal dopamine. *Psychiatry Research: Neuroimaging*, *194*(3), 212–218. https://doi.org/10.1016/j.pscychresns.2011.02.002

Polich, J. (2007). Updating P300: An Integrative Theory of P3a and P3b. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Pollatos, O., & Gramann, K. (2012). Attenuated modulation of brain activity accompanies emotion regulation deficits in alexithymia. *Psychophysiology*, *49*(5), 651–658. https://doi.org/10.1111/j.1469-8986.2011.01348.x

Pollatos, O., Schubö, A., Herbert, B. M., Matthias, E., & Schandry, R. (2008). Deficits in early emotional reactivity in alexithymia. *Psychophysiology*, *45*(5), 839–846. https://doi.org/10.1111/j.1469-8986.2008.00674.x

Pollatos, O., Werner, N. S., Duschek, S., Schandry, R., Matthias, E., Traut-Mattausch, E., & Herbert, B. M. (2011). Differential effects of alexithymia subscales on autonomic reactivity and anxiety during social stress. *Journal of Psychosomatic Research*, *70*(6), 525–533. https://doi.org/10.1016/j.jpsychores.2010.12.003

Pornpattananangkul, N., & Nusslock, R. (2015). Motivated to win: Relationship between anticipatory and outcome reward-related neural activity. *Brain and Cognition*, *100*(Supplement C), 21–40. https://doi.org/10.1016/j.bandc.2015.09.002

Pouga, L., Berthoz, S., de Gelder, B., & Grèzes, J. (2010). Individual differences in socioaffective skills influence the neural bases of fear processing: The case of alexithymia. *Human Brain Mapping*, *31*(10), 1469–1481. https://doi.org/10.1002/hbm.20953

Prkachin, G. C., Casey, C., & Prkachin, K. M. (2009). Alexithymia and perception of facial expressions of emotion. *Personality and Individual Differences*, *46*(4), 412–417. https://doi.org/10.1016/j.paid.2008.11.010

Reker, M., Ohrmann, P., Rauch, A. V., Kugel, H., Bauer, J., Dannlowski, U., … Suslow, T. (2010). Individual differences in alexithymia and brain response to masked emotion faces. *Cortex*, *46*(5), 658–667. https://doi.org/10.1016/j.cortex.2009.05.008

Reynolds, S. M., & Berridge, K. C. (2008). Emotional environments retune the valence of appetitive versus fearful functions in nucleus accumbens. *Nature Neuroscience*, *11*(4), nn2061. https://doi.org/10.1038/nn2061

Roesch, M. R., Esber, G. R., Li, J., Daw, N. D., & Schoenbaum, G. (2012). Surprise! Neural Correlates of Pearce-Hall and Rescorla-Wagner Coexist within the Brain. *The European Journal of Neuroscience*, *35*(7), 1190–1200. https://doi.org/10.1111/j.1460-9568.2011.07986.x

Rothemund, Y., Ziegler, S., Hermann, C., Gruesser, S. M., Foell, J., Patrick, C. J., & Flor, H. (2012). Fear conditioning in psychopaths: Event-related potentials and peripheral measures. *Biological Psychology*, *90*(1), 50–59. https://doi.org/10.1016/j.biopsycho.2012.02.011

Saklofske, D. H., Austin, E. J., & Minski, P. S. (2003). Factor structure and validity of a trait emotional intelligence measure. *Personality and Individual Differences*, *34*(4), 707–721. https://doi.org/10.1016/S0191-8869(02)00056-9

Salminen, J. K., Saarijärvi, S., Äärelä, E., Toikka, T., & Kauhanen, J. (1999). Prevalence of alexithymia and its association with sociodemographic variables in the general population of Finland. *Journal of Psychosomatic Research*, *46*(1), 75–82.

Salovey, P., & Mayer, J. D. (1990). Emotional Intelligence. *Imagination, Cognition and Personality*, *9*(3), 185–211. https://doi.org/10.2190/DUGG-P24E-52WK-6CDG

Sambrook, T. D., & Goslin, J. (2015). A neural reward prediction error revealed by a meta-analysis of ERPs using great grand averages. *Psychological Bulletin*, *141*(1), 213–235. https://doi.org/10.1037/bul0000006

Scarpazza, C., di Pellegrino, G., & Làdavas, E. (2014). Emotional modulation of touch in alexithymia. *Emotion (Washington, D.C.)*, *14*(3), 602–610. https://doi.org/10.1037/a0035888

Scarpazza, C., Làdavas, E., & Cattaneo, L. (2017). Invisible side of emotions: somato-motor responses to affective facial displays in alexithymia. *Experimental Brain Research*, 1–12. https://doi.org/10.1007/s00221-017-5118-x

Scarpazza, C., Làdavas, E., & Pellegrino, G. di. (2015). Dissociation between Emotional Remapping of Fear and Disgust in Alexithymia. *PLOS ONE*, *10*(10), e0140229. https://doi.org/10.1371/journal.pone.0140229

Scarpazza, C., Sellitto, M., & di Pellegrino, G. (2017). Now or not-now? The influence of alexithymia on intertemporal decision-making. *Brain and Cognition*, *114*, 20–28. https://doi.org/10.1016/j.bandc.2017.03.001

Schäfer, R., Schneider, C., Tress, W., & Franz, M. (2007). Cortical augmenting in alexithymic subjects after unpleasant acoustic stimulation. *Journal of Psychosomatic Research*, *63*(4), 357–364. https://doi.org/10.1016/j.jpsychores.2007.03.015

Schiller, D., Levy, I., Niv, Y., LeDoux, J. E., & Phelps, E. A. (2008). From Fear to Safety and Back: Reversal of Fear in the Human Brain. *The Journal of Neuroscience*, *28*(45), 11517–11525. https://doi.org/10.1523/JNEUROSCI.2265-08.2008

Schultz, W. (1998). Predictive Reward Signal of Dopamine Neurons. *Journal of Neurophysiology*, *80*(1), 1–27.

Schultz, W. (2016). Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience*, *18*(1), 23–32.

Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., & Dornheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences*, *25*(2), 167–177. https://doi.org/10.1016/S0191-8869(98)00001-4

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Consciousness Research*, *2*, 395. https://doi.org/10.3389/fpsyg.2011.00395

Seymour, B., & Dolan, R. (2008). Emotion, Decision Making, and the Amygdala. *Neuron*, *58*(5), 662–671. https://doi.org/10.1016/j.neuron.2008.05.020

Seymour, B., O'Doherty, J. P., Dayan, P., Koltzenburg, M., Jones, A. K., Dolan, R. J., … Frackowiak, R. S. (2004). Temporal difference models describe higher-order learning in humans. *Nature*, *429*(6992), 664–667. https://doi.org/10.1038/nature02581

Shah, P., Hall, R., Catmur, C., & Bird, G. (2016). Alexithymia, not autism, is associated with impaired interoception. *Cortex*. https://doi.org/10.1016/j.cortex.2016.03.021

Sifneos, P. E. (1973). The prevalence of "alexithymic" characteristics in psychosomatic patients. *Psychotherapy and Psychosomatics*, *22*(2), 255–262.

Singer, T., & Lamm, C. (2009). The Social Neuroscience of Empathy. *Annals of the New York Academy of Sciences*, *1156*(1), 81–96. https://doi.org/10.1111/j.1749-6632.2009.04418.x

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). Manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press.

Stone, L. A., & Nielson, K. A. (2001). Intact Physiological Response to Arousal with Impaired Emotional Recognition in Alexithymia. *Psychotherapy and Psychosomatics*, *70*(2), 92–102. https://doi.org/10.1159/000056232

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science (New York, N.Y.)*, *150*(3700), 1187–1188.

Swart, M., Kortekaas, R., & Aleman, A. (2009). Dealing with Feelings: Characterization of Trait Alexithymia on Emotion Regulation Strategies and Cognitive-Emotional Processing. *PLOS ONE*, *4*(6), e5751. https://doi.org/10.1371/journal.pone.0005751

Tabbert, K., Merz, C. J., Klucken, T., Schweckendiek, J., Vaitl, D., Wolf, O. T., & Stark, R. (2011). Influence of contingency awareness on neural, electrodermal and evaluative responses during fear conditioning. *Social Cognitive and Affective Neuroscience*, *6*(4), 495–506. https://doi.org/10.1093/scan/nsq070

Talmi, D., Atkinson, R., & El-Deredy, W. (2013). The Feedback-Related Negativity Signals Salience Prediction Errors, Not Reward Prediction Errors. *The Journal of Neuroscience*, *33*(19), 8264–8269. https://doi.org/10.1523/JNEUROSCI.5695-12.2013

Tamietto, M., Castelli, L., Vighetti, S., Perozzo, P., Geminiani, G., Weiskrantz, L., & Gelder, B. de. (2009). Unseen facial and bodily expressions trigger fast emotional reactions. *Proceedings of the National Academy of Sciences*, *106*(42), 17661–17666. https://doi.org/10.1073/pnas.0908994106

Tamietto, M., Cauda, F., Celeghin, A., Diano, M., Costa, T., Cossa, F. M., … de Gelder, B. (2015). Once you feel it, you see it: Insula and sensory-motor contribution to visual awareness for fearful bodies in parietal neglect. *Cortex*, *62*, 56–72. https://doi.org/10.1016/j.cortex.2014.10.009

Taylor, G. J., Bagby, R. M., & Parker, J. D. A. (1991). The Alexithymia Construct: A Potential Paradigm for Psychosomatic Medicine. *Psychosomatics*, *32*(2), 153–164. https://doi.org/10.1016/S0033-3182(91)72086-0

Taylor, G. J., Bagby, R. M., & Parker, J. D. a. (2003). The 20-Item Toronto Alexithymia Scale: IV. Reliability and factorial validity in different languages and cultures. *Journal of Psychosomatic Research*, *55*(3), 277–283. https://doi.org/10.1016/S0022-3999(02)00601-3

van der Velde, J., Gromann, P., Swart, M., Wiersma, D., Haan, L. de, Bruggeman, R., … Aleman, A. (2014). Alexithymia influences brain activation during emotion perception but not regulation. *Social Cognitive and Affective Neuroscience*, nsu056. https://doi.org/10.1093/scan/nsu056

van der Velde, J., Servaas, M. N., Goerlich, K. S., Bruggeman, R., Horton, P., Costafreda, S. G., & Aleman, A. (2013). Neural correlates of alexithymia: A meta-analysis of emotion processing studies. *Neuroscience & Biobehavioral Reviews*, *37*(8), 1774–1785. https://doi.org/10.1016/j.neubiorev.2013.07.008

Viola, F. C., Thorne, J., Edmonds, B., Schneider, T., Eichele, T., & Debener, S. (2009). Semi-automatic identification of independent components representing EEG artifact.

Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, *120*(5), 868–877. https://doi.org/10.1016/j.clinph.2009.01.015

Walsh, M. M., & Anderson, J. R. (2012). Learning from experience: Event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neuroscience and Biobehavioral Reviews*, *36*(8), 1870–1884. https://doi.org/10.1016/j.neubiorev.2012.05.008

Waltz, J. A., Frank, M. J., Robinson, B. M., & Gold, J. M. (2007). Selective reinforcement learning deficits in schizophrenia support predictions from computational models of striatal-cortical dysfunction. *Biological Psychiatry*, *62*(7), 756–764. https://doi.org/10.1016/j.biopsych.2006.09.042

Warren, C. M., Hyman, J. M., Seamans, J. K., & Holroyd, C. B. (2015). Feedback-related negativity observed in rodent anterior cingulate cortex. *Journal of Physiology-Paris*, *109*(1), 87–94. https://doi.org/10.1016/j.jphysparis.2014.08.008

Webb, D., & McMurran, M. (2008). Emotional intelligence, alexithymia and borderline personality disorder traits in young adults. *Personality and Mental Health*, *2*(4), 265–273. https://doi.org/10.1002/pmh.48

Weber, T., Racanière, S., Reichert, D. P., Buesing, L., Guez, A., Rezende, D. J., … Wierstra, D. (2017). Imagination-Augmented Agents for Deep Reinforcement Learning. *ArXiv:1707.06203 [Cs, Stat]*. Retrieved from http://arxiv.org/abs/1707.06203

Wenzel, J. M., Rauscher, N. A., Cheer, J. F., & Oleson, E. B. (2015). A Role for Phasic Dopamine Release within the Nucleus Accumbens in Encoding Aversion: A Review of the Neurochemical Literature. *ACS Chemical Neuroscience*, *6*(1), 16–26. https://doi.org/10.1021/cn500255p

Wiebking, C., & Northoff, G. (2015). Neural activity during interoceptive awareness and its associations with alexithymia—An fMRI study in major depressive disorder and non-

psychiatric controls. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00589

Wingbermühle, E., Theunissen, H., Verhoeven, W. M. A., Kessels, R. P. C., & Egger, J. I. M. (2012). The neurocognition of alexithymia: evidence from neuropsychological and neuroimaging studies. *Acta Neuropsychiatrica*, *24*(2), 67–80. https://doi.org/10.1111/j.1601-5215.2011.00613.x

Winston, J. S., O'Doherty, J., & Dolan, R. J. (2003). Common and distinct neural responses during direct and incidental processing of multiple facial emotions. *NeuroImage*, *20*(1), 84–97. https://doi.org/10.1016/S1053-8119(03)00303-3