Atypical interoception as a common risk factor for psychopathology: A review

Rebecca Brewer¹, Jennifer Murphy¹* & Geoffrey Bird²,³

Atypical interoception as a common risk factor for psychopathology: A review

¹Department of Psychology, Royal Holloway, University of London
²Department of Experimental Psychology, University of Oxford
³Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King’s College London

*Corresponding author
Department of Psychology
Royal Holloway, University of London
Egham, Surrey
TW20 0EX

Jennifer.murphy@rhul.ac.uk
Abstract

The inadequacy of a categorial approach to mental health diagnosis is now well-recognised, with many authors, diagnostic manuals and funding bodies advocating a dimensional, trans-diagnostic approach to mental health research. Variance in interoception, the ability to perceive one’s internal bodily state, is reported across diagnostic boundaries, and is associated with atypical functioning across symptom categories. Drawing on behavioural and neuroscientific evidence, we outline current research on the contribution of interoception to numerous cognitive and affective abilities (in both typical and clinical populations), and describe the interoceptive atypicalities seen in a range of psychiatric conditions. We discuss the role that interoception may play in the development and maintenance of psychopathology, as well as the ways in which interoception may differ across clinical presentations. A number of important areas for further research on the role of interoception in psychopathology are highlighted.

Keywords: Interoception, interoceptive accuracy, psychopathology, comorbidity
Highlights

- Reviews evidence that atypical interoception is observed across multiple disorders
- Interoception is associated with performance in fundamental domains of functioning
- Interoception is associated with cross-disorder symptomology
- Argues that interoception may be a common vulnerability for psychopathology
- Outstanding questions for future research are outlined
Table of Contents

1. Structure of the Paper ................................................................. 4
2. Introduction to Interoception ...................................................... 7
   2.1. What is Interoception? ......................................................... 7
   2.2. How is Interoception Measured? ......................................... 10
   2.3. The Physiological and Neurological Basis of Interoception ....... 20
3. The relevance of Interoception to Mental Health ......................... 25
   3.1. The Role of Interoception in Typical Functioning: Emotional Processing .... 25
       3.1.1. The Role of Interoception in Emotional processing: Evidence from alexithymia ......................................................... 34
       3.2. The Role of Interoception in Typical Functioning: Learning and Decision-Making .......................................................... 40
4. Increased Prevalence of Interoceptive Impairments in Clinical Conditions ...... 47
   4.1. Evidence of interoceptive impairment in psychiatric and neurological conditions ................................................................................. 47
       4.1.1. Interoceptive impairment in psychiatric and neurological conditions: Evidence from alexithymia ......................................................... 51
5. The Link Between Interoceptive Impairment and Clinical Symptoms Across Disorders ............................................................... 53
   5.1. Interoceptive impairment and affective deficits ......................... 54
   5.2. Interoceptive impairment and learning, reward and decision-making deficits ............................................................... 57
   5.3. Other domains potentially affected by interoceptive impairment .... 63
       5.3.1 Sensory processing ......................................................... 63
       5.3.2. Pain sensitivity ............................................................... 65
       5.3.3. Repetitive behaviours and compulsions .............................. 66
       5.3.4. Self-other distinction ..................................................... 67
       5.3.5. Further social impairments ............................................ 70
       5.3.6. Eating behaviours ....................................................... 72
       5.3.7. Physical health ............................................................ 74
       5.3.8. Sleep ........................................................................ 75
6. What gives rise to poor interoception? The Development of Typical Interoception and How it May Go Awry ...................................................... 76
7. Outstanding Questions .................................................................. 81
   7.1. Is interoception a unitary ability? ........................................... 81
   7.2. If interoception accounts for symptom commonalities across disorders, what are differences between clinical groups attributable to? ................................. 89
   7.3. What is the precise nature of the relationship between interoception and psychopathology? ................................................................. 91
   7.4. Is intact interoception required online, or only during development? .... 92
   7.5. How can interoceptive abilities be improved? ............................ 94
8. Conclusion ............................................................................... 97
References .................................................................................. 99
1. Structure of the Paper

Debate as to the utility of a psychiatric nosology, such as that contained in the Diagnostic and Statistical Manual (5th edition; DSM-5; American Psychiatric Association, 2013), for research and health care delivery, is far from resolved. While diagnostic manuals such as DSM-5 now acknowledge that a continuous diagnostic approach may be favourable, substantial diversion from the categorical approach to diagnosis has not yet been achieved. Opponents of attempts to develop a classification of mental disorders point to the cross-disorder nature of many symptoms such as working memory impairments or emotion regulation difficulties (Cuthbert & Insel, 2013), and to the high rates of comorbidity and co-occurrence (Hasin & Kilcoyne, 2012; Kessler, Chui, Demler, & Walters, 2005) within psychiatry, such that half of those who meet diagnostic criteria for one disorder also meet diagnostic criteria for another disorder. Responses to this challenge tend to adopt one of two approaches. First, in the approach adopted by the National Institute of Health’s Research Domain Criteria (RDoC), a collection of clinicians and basic scientists agree on a number of areas of psychological functioning, for example within domains such as positive and negative valence systems, cognitive systems, social processes, and arousal and regulatory mechanisms. Within this approach, traditional diagnostic categories are ignored and each area of functioning is itself the unit of analysis (e.g. a study may seek to determine the genetic basis of trait anxiety, regardless of whether this anxiety is displayed by individuals with Generalised Anxiety Disorder or Autism Spectrum Disorder). Similarly, the Hierarchical Taxonomy of Psychopathology (HiTOP) provides a dimensional classification system with which to diagnose psychopathology, again identifying a range of continua representing variation within the typical and clinical populations, which may be used to classify patients (Kotov et al., 2017). In a related approach (e.g. Krueger, Caspi, Moffitt, & Silva, 1998), factor analysis has been used to uncover a higher-order structure among diagnostic symptoms. Original work suggested a two-factor structure, made up of an Internalising factor (predisposing towards mood and anxiety disorders) and an Externalising factor (predisposing towards substance abuse and antisocial disorders), which was later supplemented by a third Thought
Disorder factor (predisposing towards psychotic symptoms; Kotov, Chang, et al., 2011; Kotov, Ruggero, et al., 2011). Several recent studies have extended this conceptualisation, however, demonstrating the existence of a single higher-order factor, namely the ‘P Factor’, representing lesser-to-greater severity of psychopathology with associated disruption in neural circuitry (Caspi, Houts, Belsky, & Goldman-Mellor, 2014; Lahey et al., 2012; Selzam, Coleman, Caspi, Moffitt, & Plomin, 2018). Although differing in their methodological approach, all of these conceptualisations suggest that there are similarities in symptoms across psychopathology, and that impairments in a specific psychological process may lead to a range of symptoms that are shared across traditional diagnostic categories.

One such psychological process that has clear relevance within these frameworks is interoception, the perception of, interpretation of, or attention to, one’s internal bodily state (note that henceforth ‘internal state’ refers to bodily states). A number of authors have previously highlighted the role that atypical interoception may play in a range of mental health conditions (Barrett & Simmons, 2015; Bonaz et al., 2021; Khalsa & Lapidus, 2016; Khalsa et al., 2018; Murphy, Brewer, Catmur, & Bird, 2017; Paulus & Stein, 2010; Quadt, Critchley, & Garfinkel, 2018; Tsakiris & Critchley, 2016), and we have previously proposed that interoceptive atypicalities may represent a common vulnerability factor for psychopathology (Brewer & Bird, 2019; Brewer, Cook, & Bird, 2016a; Murphy, Geary, Millgate, Catmur, & Bird, 2017). Despite consensus in the literature regarding the relevance of interoception for psychopathology, and the recent publication of a number of reviews relevant to interoception, no papers to date have provided a detailed review of the evidence surrounding the relationship between interoceptive abilities and mental health across a broad range of clinical conditions. In this paper, we review the evidence that atypical interoception may represent a key risk factor for the development of psychopathology, across a range of traditional diagnostic categories.

As an introduction to the topic, we first define interoception, and discuss its measurement and neural basis. Subsequently, in order to show that interoception is a sufficiently basic process for
atypicalities to affect multiple symptom domains, we describe the contribution of interoception to physical and mental health. In particular, we highlight the contribution of interoception to multiple aspects of emotional (and socio-emotional) processing, and for multiple aspects of learning and decision-making. We then review evidence for the presence of interoceptive impairment in multiple clinical conditions. Drawing these two sections together, we critically assess the available evidence suggesting that interoceptive impairments at least contribute to, or are responsible for, a range of symptoms across clinical conditions. This latter point includes explanation of how interoceptive deficits can explain heterogeneity within conditions and commonalities across clinical conditions. We argue that many symptoms are a product of atypical interoception, rather than consequences of individual disorders themselves. For example, it might be the case that a particular impairment associated with interoceptive deficits (e.g., emotion processing difficulties) seen in many, but not all, individuals diagnosed with eating disorders is due to the fact that a significant proportion of this population has atypical interoception (which may not be a product of the eating disorder per se). If true, the presence or absence of atypical interoception would explain the variance within the eating disorders population with respect to emotion processing. Furthermore, the fact that atypical interoception is seen in several other clinical conditions means that emotion processing difficulties should also be seen in other conditions besides eating disorders.

If our conjecture is correct, and atypical interoception confers a general susceptibility to psychopathology, then it is useful to address what may cause atypical interoception, particularly in the context of how individuals develop typical interoception. Furthermore, it is useful to consider the potential for intervention when interoception becomes problematic. These questions are addressed before we consider some outstanding questions fundamental to the idea that atypical interoception contributes to symptoms across conditions and suggest ways in which these might be addressed. Note that, in line with dimensional approaches to psychopathology, the current paper discusses interoception within the context of a wide range of conditions, including neurodevelopmental disorders such as autism, and neurodegenerative disorders, such as dementia.
2. Introduction to Interoception

2.1. What is Interoception?

On a surface level it is easy to define interoception; it is the ability to perceive the internal state of one’s own body. However, such a seemingly simple definition masks uncertainty in specifying exactly how to define ‘internal’, and raises issues relating to the distinction between perception and recognition of such states. Whilst early definitions included visceral (internal) sensations only, more recent definitions include signals that are not visceral (e.g., affective touch) but which are processed using similar neural pathways as other interoceptive signals (see Khalsa & Lapidus, 2016; Chen et al., 2021). Thus, contemporary definitions of interoception include signals sent either via lamina I of the spinal and trigeminal dorsal horn to the anterior insula and anterior cingulate cortex (Craig, 2002) or via cranial nerves to the nucleus of the solitary tract (Critchley & Harrison, 2013); see Section 2.3 for a detailed description of the neural basis of interoception). Accordingly, signals relating to hunger, satiety, itch, thirst, muscular effort, bladder, gastrointestinal, respiratory, cardiac, temperature, blood (PH, glucose level), vasomotor flush, air hunger, shudder, sensual touch, genital sensation, bruising, headache, broken bones and many more visceral sensations are typically considered to be interoceptive signals (Khalsa & Lapidus, 2016), though this remains debated (Ceunen, Vlaeyen, & Van Diest, 2016; Critchley & Garfinkel, 2017). Anterior insula has also been implicated in the processing of smell (Plailly et al., 2007), suggesting that olfaction could also be considered interoceptive under broader definitions. Similarly, taste appears to be represented in both the anterior dorsal insula, as well as more posteriorly (Small, 2010). Debate also exists as to whether proprioception (the sense of the position of one’s body, particularly the position of the limbs in relation to the trunk) should also be considered interoceptive (Cameron, 2001; Vaitl, 1996); although information originates from inside the body, it is processed by distinct neurological systems, in particular the posterior insula (Bottini et al., 2001; Fasold, Heinau, Trenner, Villringer,
Further debate concerns the level of the cognitive representation required in order for interoception to have occurred (e.g., Khalsa et al., 2018). At one level, any instance of homeostatic control involves interoception. For example, the release of insulin in response to hyperglycemia relies on the detection of excess glucose in the blood. Such detection fulfils the definition of interoception even though this process may occur without the individual’s conscious awareness, and various computational models have been put forward to explain the role of interoception in maintaining homeostasis (Petzschner, Garfinkel, Paulus, Koch, & Khalsa, 2021). A higher degree of conscious awareness is associated with the perception of interoceptive signals. At this level, interoception has occurred when the individual is consciously aware that their interoceptive state has changed. Crucially, it is possible to distinguish this from a still higher level in which the individual is able to recognise the interoceptive signal explicitly. An individual with conscious perception, but without explicit recognition, of interoceptive signals may be aware that their internal state has changed, but be unable to label that state. For example, they may be aware that they are experiencing an unusual internal state, but be unaware of which specific interoceptive change they are experiencing. At the highest level of recognition, an individual would be able to detect that their state has changed, and to recognise and label their new interoceptive state (e.g., as heat, hunger, etc.). This characterisation is not without controversy, however, especially when applied to emotional internal states, when the specific theory of emotion one subscribes to plays a role. For example, under Schacter & Singer’s (1962) model of emotion, determining whether one is hungry or hot may be achieved on the basis of interoceptive signals alone, but distinguishing between the emotions of surprise and fear, for example, may rely on the integration of interoceptive cues and the cognitive evaluation of situational cues. Under such a model, it may be possible for an individual with intact interoceptive awareness to fail to recognise their own emotional internal states beyond a
coarse valence level if they lack the experience or ability necessary to evaluate situational cues from the environment.

Beyond these (already problematic) definitions of interoception, separate components of interoception have been defined. Garfinkel & Critchley (2013) proposed a tripartite model of interoception comprising of three facets of interoceptive ability: 1) interoceptive sensitivity, one’s accuracy of perceiving one’s internal state (for example, how accurately an individual can count their heartbeats when instructed to do so), 2) interoceptive sensibility, an individual’s self-reported beliefs regarding their attention to and accuracy of perceiving internal signals and 3) interoceptive awareness, a metacognitive measure reflecting the level of correspondence between an individual’s true accuracy (interoceptive sensitivity) and their self-rated interoceptive sensibility. Following this initial description a number of variants of this model have been proposed (Khalsa et al., 2018; Murphy, Catmur, & Bird, 2018; Murphy, Catmur, & Bird, 2019); for example, one recent model highlights a need to distinguish between one’s interoceptive accuracy and one’s propensity to attend to interoceptive information, along both subjectively reported and objectively assessed axes (Murphy, Catmur, & Bird, 2017; Murphy, Catmur, & Bird, 2019). Previously available measures of interoception (see Section 2.2) led interoceptive accuracy to be measured most frequently using objective, performance-based tasks, and interoceptive attention to be measured using self-report measures. However, assuming both interoceptive attention and interoceptive accuracy can be measured using subjective and objective measures, Murphy et al. (2017; 2019) proposed a 2 x 2 dimensional structure of interoception, reflecting both what is measured (accuracy versus attention) as well as how it is measured (self-report versus objective performance) (see Figure 1). Within this 2 x 2 structure, interoceptive awareness, as defined by Garfinkel and Critchley (2013), can be seen as the degree to which self-report and accuracy measures correlate with each other, either along the accuracy or the attention dimension. The propensity to use internal signals in one’s daily life may also be an additional interoceptive dimension (Murphy et al., 2019), and this may be a function of interoceptive accuracy, attention and metacognitive interoceptive awareness combined.
Notably, as described above, there are multiple levels of representation at which interoception can occur (e.g. implicit homeostasis, conscious perception of a signal without recognition of the specific signal, recognition without the requirement for a verbal label, and verbal labelling of the signal), and tasks tend to tap into different levels of this hierarchy. For example, threshold tasks simply require participants to detect the presence or absence of a sensation, interoceptive load tasks require participants to determine whether there has been a change in signal strength, while questionnaire measures require individuals to reflect on distinct verbally labelled interoceptive states. The difficulties identifying relationships between different interoceptive tasks may, therefore, be exacerbated by tasks tapping into different levels of representation as well as different dimensions of interoception (attention, accuracy and metacognitive abilities) and different measurement types (objective performance or beliefs).

2.2. How is Interoception Measured?

The measurement of interoception is generally acknowledged to be the biggest difficulty within the field (for a review, see Quigley, Kanoski, Grill, Barrett, & Tsakiris, 2021). Studies of interoceptive accuracy have relied almost exclusively upon measures of heartbeat counting or discrimination (also referred to as ‘heartbeat detection’). In heartbeat counting procedures, participants are asked to count their heartbeats without physically measuring their heartbeat over a series of intervals (typically between 3 and 6 time intervals in the range of ~25 to 100 seconds) (Dale & Anderson,
1978; Schandry, 1981). The difference between an objective measurement and the participants’ estimate is taken as a measure of interoceptive sensitivity. Variants of this task also exist where counting procedures are replaced with tapping in time with one’s heartbeat (Ludwick-Rosenthal & Neufeld, 1985). In heartbeat discrimination procedures, participants are asked to determine whether visual or auditory stimuli are presented synchronously or asynchronously with their heartbeat. Whilst cardiac sensitivity tasks are extensively used, and in some respects reliable, (Brener & Kluvitse, 1988; Jones, 1994; Wildman & Jones, 1982), their suitability for research has been questioned. First, approximately 40% of typical individuals are not able to consciously perceive their heartbeat (Khalsa, Rudrauf, Sandesara, Olshansky, & Tranel, 2009) making these tasks unsuitable for quantifying interoceptive sensitivity in individuals with poor interoception. While this issue has been addressed using infusions of isoproterenol (a beta-adrenergic agonist) to increase intensity of cardiac (and respiratory) signals (e.g. Hassanpour et al., 2018; Khalsa et al., 2009), this technique is invasive and so not always feasible to use. Second, heartbeat may be perceived via (exteroceptive) touch receptors due to the vibration of the chest wall, with factors such as the individual’s percentage of body fat (Rouse, Jones, & Jones, 1988), systolic blood pressure (O’Brien, Reid, & Jones, 1998) resting heart rate, and heart rate variability (Knapp-Kline & Kline, 2005) potentially affecting the extent to which heartbeat is perceived via this route. Indeed, whilst such influence of physiology may not be a limitation per se (as presumably such factors act upon one’s ability to perceive cardiac signals across all situations), physiological factors may impact on the above tasks of interoceptive sensitivity specifically; for example, individual differences in resting heartrate may alter the number of hits and misses in the heartbeat counting task. Likewise, elevated blood pressure may influence the time at which external signals are perceived as synchronous with one’s heartbeat in the heartbeat discrimination task (O’Brien et al., 1998). At present the mechanism by which these physiological factors influence performance on tasks of cardiac interoceptive accuracy remains unknown.
Whilst the factors described above may affect performance in any cardiac-based measure of interoceptive accuracy, there are important considerations for each task specifically. The heartbeat counting task is also strongly influenced by an individual’s beliefs regarding their own or the average resting heart rate (Brener & Ring, 2016; Murphy, Millgate, et al., 2018; Ring & Brener, 1996; Ring, Brener, Knapp, & Mailloux, 2015; Windmann, Schonecke, Fröhlig, & Maldener, 1999). Indeed, beliefs about one’s heartbeat can alter performance on a range of tasks; false feedback about heart rate, for example, has been found to alter perception of one’s effort during exercise (Iodice, Porciello, Bufalari, Barca, & Pezzulo, 2019). The extent to which beliefs about heart rate affect interoceptive accuracy estimates in the heartbeat counting task may depend on the task instructions given; whilst early instructions encouraged participants to ‘estimate’ if they could not feel their heartbeat (Brener & Ring, 2016; Schandry, 1981) more recent approaches advocate a response of zero in this instance (Murphy, Brewer, Hobson, & Catmur, 2018) given evidence that the instructions given to participants may alter the pattern of results obtained (Desmedt et al., 2020; Desmedt, Luminet, & Corneille, 2018; Ehlers, Breuer, Dohn, & Fiegenbaum, 1995). Equally problematic, is the fact that the task is often administered in the absence of a control task, meaning that an individual’s performance may be influenced by other factors (e.g., attention, motivation) that are not interoceptive. Whilst newer research has advocated the use of a timing control task (Ainley, Brass, & Tsakiris, 2014), it is likely that this time estimation is not an adequate control for the factors affecting performance on counting based measures of interoception (Desmedt et al., 2020). Notably, it has also been argued that the psychometric properties of the HCT are less than ideal; for example, performance on the HCT appears to differ across different time intervals and a weak correlation is often observed between the participants’ actual and reported heartbeats (Zamariola, Maurage, et al., 2018). Such evidence has been taken to suggest that individuals generally display poor perception of cardiac signals (Zamariola, Maurage, Luminet, & Corneille, 2018). Importantly, Zamariola et al. (2018) also found that the correlation between reported and actual number of heartbeats did not differ across individuals in the first and second quintile of HCT performers,
suggesting that the task may also not be sensitive to differences in cardiac accuracy in better
performers (but see Ainley, Tsakiris, Pollatos, Schulz, & Herbert (2020)). These factors, combined
with the general factors affecting cardiac-based measures of interoceptive accuracy, lead to
corns about the reliability and validity of this task. Indeed, the few studies examining the test-
retest reliability of the heartbeat counting task over longer durations suggest fairly poor temporal
stability in children and adults (Ferentzi, Drew, Tihanyi, & Köteles, 2018; Koch & Pollatos, 2014;
Murphy, Cheesman, et al., 2019). Whilst this may suggest that differing degrees of interoceptive
accuracy are a product of an individual’s state, rather than a stable trait (Wittkamp, Bertsch, Vögele,
& Schulz, 2018), it is possible that temporal changes in this task may be due to changes across time
in physiological or psychological factors that are not interoceptive.

In contrast to the heartbeat counting task, research suggests that the heartbeat discrimination
task is less influenced by beliefs (Phillips, Jones, Rieger, & Snell, 1999), making this a preferred
method for quantifying interoceptive accuracy. However, this task is not without limitations; for
example, a high number of trials are required to gain precise estimates of ability (Kleckner,
Baumann Wormwood, Simmons, Barrett, & Quigley, 2015), and concerns have been raised over the
difficulty of this task, meaning that it may not be suitable for examining interoceptive accuracy at
the lower range of ability (Brener & Ring, 2016). Further, although Ring and Brener (2018) argue
that humans are able to judge very precisely (within 20ms) whether two stimuli in different
modalities are simultaneous (Zampini, Guest, Shore, & Spence, 2005), a sizeable portion of
variance in this task appears to be accounted for by the ability to determine synchronicity of (non-
interoceptive) cross-modal signals (Knapp, Ring, & Brener, 1997). This problem is exacerbated by
the fact that individuals perceive heartbeats at different locations within the body, each of which is
associated with a different temporal delay (Brener & Kluvitse, 1988; Christopher Ring & Brener,
1992). Control tasks assessing these abilities are rarely employed (although for an exception see
Garfinkel et al., 2016). Variations also exist as to the exact methodology employed for the heartbeat
discrimination task; whilst two-alternative forced choice procedures are used most frequently, it has
been argued that such methods do not account for individual differences in the delay at which individuals perceive an external signal to be synchronous with their heartbeat, and that lengthier procedures (for example the method of constant stimuli) may be more appropriate (Brener & Ring, 2016). A recently developed task addresses these issues; in the Phase Adjustment Task (Plans et al., 2020), participants are required to adjust the phase relationship between a tone and their heartbeat in order to achieve synchrony. This task is quicker to complete than improved versions of the heartbeat discrimination task (such as the method of constant stimuli), and allows participants to identify any delay between the tone and their heartbeat as ‘synchronous’; meaning it both accounts for the individual differences in delay length perceived as synchronous that cause problems for standard versions of the heartbeat discrimination task, and makes it especially useful in clinical populations who may struggle with longer procedures or who require a high level of precision in their responses.

In addition, in contrast to the HCT, non-interoceptive participants cannot perform well on the task simply by knowing their resting heart rate. The relationship with psychopathology is yet to be investigated, however, making future research utilising this task in clinical groups a priority.

Surprisingly, scores on the heartbeat counting and discrimination tasks often correlate only modestly, if at all (Forkmann et al., 2016; Hickman, Seyedsalehi, Cook, Bird, & Murphy, 2020; Kandasamy et al., 2016; Knoll & Hodapp, 1992; Phillips et al., 1999; Ring & Brener, 2018; Schulz, Lass-Hennemann, Sütterlin, Schächinger, & Vögele, 2013; Weisz, Balazs, & Adam, 1988). Whilst this may reflect the unreliability of the measures, it is possible that different abilities are quantified by each task (Khalsa & Lapidus, 2016). Such a conjecture is supported by evidence that different neural activation has been associated with each task (Schulz, 2016) and that certain perturbations (e.g., stress) have differential effects on performance across these tasks (Schulz et al., 2013). Such differences may largely reflect task demands; while the heartbeat counting task (at least in theory) quantifies an individual’s ability to perceive their heartbeat, the discrimination procedure requires the individual to integrate an internal signal with an external signal (Couto et al., 2015). Given questions over the validity and specificity of certain measures of cardiac interoception, recent
efforts have focused on the development of new measures of cardiac interoceptive accuracy (e.g., Plans et al., 2020).

Beyond tasks measuring interoception from heartbeat perception, a small number of studies have used a measure of gastric distension; Whitehead and colleagues, for example, assessed participants’ ability to detect whether stomach contractions coincided with an exteroceptive light stimulus, as well as their ability to control their gastric motility, and observed a positive relationship between perception of cardiac and gastric signals (Whitehead & Drescher, 1980). Similarly, detection rates and intensity ratings of colon distension have been investigated (Hölzl, Erasmus, & Möltner, 1996) and Zaman et al. (2016) measured ratings of esophageal distension intensity, identifying thresholds for initial perception, discomfort and pain. The invasive nature of these tests, however, means that they have not been used extensively. An alternative assessment of gastric interoception is the water load test, which involves participants consuming water until the point of perceived fullness (e.g. Herbert, Muth, Pollatos, & Herbert, 2012; Koch, Hong, & Xu, 2000). An adapted version of this task (the Two Step Water Load Test; WLT-II) uses a two-stage procedure, whereby participants first drink until perceived satiation, and then to maximum fullness, in an attempt to control for stomach capacity (Van Dyck et al., 2016). It is assumed that interoceptive ability is negatively associated with volume of water consumed (or the proportion of water consumed in step one relative to step two in WLT-II). As such, participants’ ability to perceive their heartbeat was negatively correlated with amount of water consumed in the original water load task (Herbert, Muth, et al., 2012). FMRI has also been used to assess interoceptive cortex activity while participants attend to sensations from the stomach (Simmons et al., 2013), experience proximal and distal distension of the oesophagus (Aziz et al., 2000), or attempt to regulate their gastric responses to a virtual rollercoaster (Li, Zucker, Kragel, Covington, & Labar, 2017). These studies implicated interoceptive cortex, but do not provide an objective measure of interoceptive accuracy. Electrogastrography has also been used to assess gastric activity non-invasively, with activity being linked to neural responses in interoceptive regions (Rebollo, Devauchelle, Béranger, & Tallon-
Baudry, 2018) and cardiac perception (Herbert, Muth, et al., 2012), for example. See Wolpert, Rebollo, & Tallon-Baudry (2020) for a review of electrogastrogram use.

Tests of respiratory effort have also been used to assess interoceptive ability. A respiratory resistance threshold task, for example, involves participants breathing through a circuit with varying levels of resistance, induced by filters, and determining when resistance is present (Garfinkel et al., 2016; Harver, Katkin, & Bloch, 1993). Participants’ ability to detect resistance is taken as a measure of respiratory interoceptive accuracy. Respiratory tasks have also been used that require participants to breathe air that varies in CO₂ concentration, and rate the frequency (‘faster breathing’) and volume (‘deeper breathing’) of their respiration using a tickbox to indicate the presence of each perturbation (Bogaerts et al., 2005; van den Bergh et al., 2004). Interoceptive accuracy is inferred from the strength of the relationship between subjective ratings and objective respiratory frequency and volume.

While the cardiac, respiratory and gastric tests described above are the most commonly used measures used to explicitly assess interoceptive ability, as a wide variety of signals are considered to be interoceptive in nature, we would suggest that a range of existing perceptual sensitivity measures are in fact also measures of interoceptive ability. These would include tests of taste (Murphy et al., 2017; Stevens, Cruz, Hoffman, & Patterson, 1995), hunger, thirst and satiety (Harshaw, 2008), oral temperature (Guest et al., 2007), muscular effort (Herbert, Ulbrich, & Schandry, 2007; Murphy et al., 2017), pain (e.g. de Zwaan, Biener, Bach, Wiesnagrotzki, & Stacher, 1996; Pollatos, Dietel, Gündel, & Duschek, 2015), and (depending on one’s definition of interoception; Khalsa & Lapidus, 2016) interoception may also be involved in the existing measures of perception of proprioceptive signals (e.g. Moberg, 1983; Ponzo, Kirsch, Fotopoulou, & Jenkinson, 2018, 2019), and balance (e.g. Evkaya, Karadag-saygi, Karali, & Giray, 2019). Notably, established explicit measures of interoception, such as heartbeat perception, correlate with some of these tasks. Participants with better perception of their heartbeats, for example, appear to be more sensitive to their cardiac signals, and exert less physical effort, in a free cycling situation than those
with poorer heartbeat perception (Herbert et al., 2007). Individuals with higher cardiac accuracy are also more sensitive to, and less tolerant of, pain (Pollatos, Füstös, & Critchley, 2012).

The addition of these tests to the existing limited battery of tests of interoceptive ability would greatly reduce the over-reliance on tests of heartbeat perception, and serve to overcome the limitations of the methods described above. Similarly, this would allow comparison of abilities across interoceptive domains, and a more reliable measure of overall interoceptive ability. There is an obvious need to assess interoceptive abilities across a range of domains, as it is thus far unclear whether or not interoception is a unitary construct, or one that consists of multiple interoceptive domains (Vaitl, 1996). Some studies have found moderate correlations between perception of different internal states, such as heartbeat and gastric distension (Herbert, Muth, et al., 2012; Van Dyck et al., 2016; Whitehead & Drescher, 1980), or between rectal distention and both cold perception and heat-induced pain (Horing, Kugel, Brenner, & Zipfel, 2013). Increasing hunger through fasting also appears to increase cardiac accuracy, suggesting potential overlap in the processing of these interoceptive signals (Herbert, Herbert, et al., 2012). Conversely, many studies have found non-significant relationships between performance across different interoceptive domains. Evidence suggests, for example, that cardiac and pain perception (Werner, Duschek, Mattern, & Schandry, 2009) and cardiac and respiratory abilities (Garfinkel et al., 2016; Harver et al., 1993; Nicholson, Williams, Carpenter, & Kallitsounaki, 2019; Pollatos, Herbert, Mai, & Kammer, 2016; although see Steptoe & Noll, 1997) may dissociate. Early evidence also found that detection of high blood pressure, sweaty hands, and shortness of breath were not associated (Steptoe & Vögele, 1992). The most comprehensive investigation of multiple interoceptive domains within the same participants found no relationships between performance across the domains of cardiac, gastric, taste, pain, proprioceptive, and vestibular perception (Ferentzi, Bogdány, Szabolcs, & Csala, 2018). Taken together, these findings indicate that one’s performance on a single interoceptive task may not reflect one’s ability to perceive or recognise other internal signals. This is consistent with evidence that intracranial stimulation of different insula regions is associated with distinct internal
sensations (Stephani, Fernandez-Baca Vaca, Maclunas, Koubeissi, & Lüders, 2011). It is also important to assess abilities across interoceptive signals as some bodily signals may simply be easier to detect than others, potentially because interoceptive signals vary highly in terms of their timescales and amplitudes (Khalsa et al., 2018). Perception of heartbeats for example, has been suggested to be easier than perception of other visceral signals (Kollenbaum, Dahme, & Kirchner, 1996).

As described above, each of the existing tasks assessing interoceptive accuracy has its own limitations. In addition to these task-specific limitations is a task-general problem regarding variation in signal strength. While individuals can vary in their ability to perceive a given signal accurately, signals arising within the body are also not consistent across individuals or across time within the same individual. One may perform poorly at tasks of interoceptive accuracy owing to poor perception and/or interpretation of a signal, or owing to this signal being particularly weak. Indeed, cardiac perception accuracy has been found to increase following an increase in cardiac signal strength, for example that induced by isoproterenol infusions or exercise (Jones & Hollandsworth, 1981; Khalsa et al., 2009). In contrast to tasks assessing exteroceptive perception, it is very difficult to control the stimulus to be perceived in interoceptive tasks. Even in tasks where an interoceptive signal is manipulated, the experimenter can only control the stimulus inducing the interceptive change, rather than this signal change itself. Interoceptive tasks therefore tend to rely on measuring (rather than controlling) interoceptive signals objectively where possible, but the extent to which it is possible to measure internal signals varies across interoceptive domain; it is far easier to measure cardiac signals accurately than hunger or satiety, for example. This difficulty leaves the conclusions of many studies on individual differences in interoception open to interpretation, as it is possible that atypical performance is due to differences in signal strength, rather than in interoceptive accuracy. Where psychopathology is concerned, this may be a particular issue as many clinical conditions are associated with co-occurring physical atypicalities, making it feasible that interoceptive signal strength varies with mental health. Where relationships between
psychopathology and interoceptive task performance are observed, therefore, these may be driven by atypical processing of the interoceptive signal, or differences in the signal itself. Future work should aim to manipulate and control across participants, or at least measure and account for, interoceptive signal strength where possible.

Beyond measures of interoceptive accuracy (sensitivity under Garfinkel and Critchely’s (2013) model), a number of self-report measures have been used to assess self-reported interoception (interoceptive sensibility under Garfinkel and Critchely’s (2013) model). These include the Body Perception Questionnaire (BPQ; Porges, 1993), the Multidimensional Assessment of Interoceptive Awareness (MAIA; Mehling et al., 2012), the Body Consciousness Questionnaire (Miller, Murphy, & Buss, 1981), the interoception subscale of the Eating Disorder Inventory (Garner, Olmsted, & Polivy, 1983), the Body Awareness Questionnaire (Shields, Mallory, & Simon, 1989), the Interoception Sensory Questionnaire (Fiene, Ireland, & Brownlow, 2018), and the Self Awareness Questionnaire (Longarzo et al., 2015). These measures are limited, however, by the fact that subjective perception of interoceptive attention is often confounded with the extent to which interoceptive signals are present within the individual (BPQ), or by assessing multiple facets of interoception, for example interoceptive accuracy and attention (MAIA and BCQ). Note that confidence ratings are often used to quantify self-reported interoception in tasks of interoceptive accuracy (e.g., the heartbeat counting or discrimination procedures) but these are often uncorrelated with self-report questionnaires of interoception (Garfinkel, Seth, Barrett, Suzuki, & Critchley, 2015; Murphy et al., 2019). While efforts have been made to assess the relationship between self-report questionnaire measures and objective measures of interoception, findings have varied across studies.

Notably, these findings are likely to vary depending on whether self-report measures assess interoceptive accuracy or attention. While the majority of interoceptive questionnaire measures focus on interoceptive attention alone, or conflate interoceptive accuracy and attention (e.g. Fiene, Ireland, & Brownlow, 2018; Mehling et al., 2012; Shields, Mallory, & Simon, 1989), self-report measures of interoceptive accuracy have recently been developed, namely the Interoception
Confusion Questionnaire (Brewer et al., 2016a) and the Interoceptive Accuracy Scale (Murphy, Brewer, Plans, Khalsa, & Catmur, 2019). While these different self-report interoceptive accuracy measures have been found to correlate with each other, neither appears to be associated with self-reported interoceptive attention (Murphy et al., 2019). Further, while weak or non-significant correlations have been found between self-reported interoceptive attention and objectively measured interoceptive accuracy (Ferentzi et al., 2018; Garfinkel et al., 2016, 2015; Meessen et al., 2016; Murphy et al., 2019; Whitehead, Drescher, Heiman, & Blackwell, 1977), self-reported interoceptive accuracy appears to correlate with performance on the heartbeat counting task (Murphy et al., 2019), lending some support to the importance of distinguishing between interoceptive accuracy and attention and a 2 x 2 dimensional structure of interoception. We therefore recommend that future research includes both measures of interoceptive attention (e.g., the IATS; Gabriele, Spooner, Brewer, & Murphy, 2020) and accuracy (e.g., the IAS) in order to distinguish between these two dimensions of self-reported interoception. Similarly, it is necessary that future work aims to develop not only improved objective measures of interoceptive accuracy, but also objective measures of interoceptive attention (for example experience sampling procedures; Murphy, Catmur, et al., 2019).

While questions remain concerning the unitary or fractionated nature of interoception across internal signals (e.g., cardiac, gastric, and respiratory channels), it is difficult to determine whether clinical symptoms are associated with a general interoceptive impairment, or atypicalities in specific interoceptive domains. While it may be the case that interoception is a unitary ability, and that observed dissociations are due to differences in task demands, it is also possible that different interoceptive channels, or clusters of channels, exist, and that these show different associations with distinct clinical profiles. Further work is required in order to differentiate between these possibilities.

2.3. The Physiological and Neurological Basis of Interoception
As discussed briefly above, recent definitions of interoception tend to class bodily signals as interoceptive if they are sent via lamina 1 to the AI or ACC (Craig, 2002), or via cranial nerves (vagus and glossopharyngeal) to the nucleus of the solitary tract (Critchley & Harrison, 2013). These pathways are discussed in more detail in this section. It should be noted that multiple humoral pathways also exist that convey certain interoceptive signals (thirst, blood sugar) though it is beyond the scope of this paper to discuss these in detail (see Critchley & Harrison, 2013). For a recent, more thorough, review of the neural circuitry of interoception, see Berntson & Khalsa (2021).

Small diameter Aδ and C fibres innervate all tissues in the body, and project sympathetic afferents monosynaptically to lamina 1 of the spinal and trigeminal dorsal horn (the most dorsal section of the spinal cord) (Panneton, 1991; Woolf & Fitzgerald, 1983). Lamina 1 neurons are divided into physiologically, morphologically and chemically distinct classes, selective to modality, and responsive to various physical sensations (Andrew & Craig, 2001; Craig, Krout, & Andrew, 2001; Han, Zhang, & Craig, 1998; Light & Willcockson, 1999; Yu et al., 1999). From here, projections exist to the parabrachial nucleus (directly and via A1 and A2 catecholamine cell groups), and the hypothalamus via the A1 group (Craig, 2002). Parasympathetic afferents concerned with the heart and digestive tract (vagus nerve) and taste, cardiac, and general visceral sensory information (glossopharyngeal nerve), on the other hand, project to the nucleus of the solitary tract (NTS), which also projects to the parabrachial nucleus, periaqueductal grey (PAG) and hypothalamus. The parabrachial nucleus is the major brain region involved in integration of the homeostatic afferents from the wide range of interoceptive cues (Craig, 1995), and projects to the medial and basal ventral medial nuclei of the thalamus (VMb) (Krut & Loewy, 2000; Krukoff, Harris, & Jhamandas, 1993; Saper, 2002). Parabrachial projections also connect to the PAG, known as the mesencephalic homeostatic motor centre of the brain. The PAG is involved in guiding goal-directed activity (autonomic, neuroendocrine and behavioural) in order to maintain homeostasis (Canteras & Swanson, 1992; Saper, 2002). The hypothalamus, which is the diencephalic homeostatic motor centre, plays a similar role (Canteras & Swanson, 1992). The NTS also projects
to the VMb, and a phylogenetically new pathway (only present in primates) projects from Lamina 1 neurons to thalamic nuclei; the ventral caudal portion of the medial dorsal nucleus (MDvc), the posterior ventral medial nucleus (VMpo) and the VMb. VMpo and VMb pathways project to the dorsal section of the insula (the limbic sensory cortex), which is bi-directionally connected to the ACC, OFC (which is centrally involved in processing emotion and reward for decision making (Bechara, Damasio, & Damasio, 2000)), hypothalamus and amygdala. The MDvc, on the other hand, projects directly to the ACC. Both the ACC and the insula project descending connections, in order to control the homeostatic integration sites in the brainstem, as well as projecting to the AI. These projections to the AI are again phylogenetically new, being specific to primates.

Whilst the posterior dorsal insula appears to provide a cortical representation of interoceptive sensations, and is activated in response to a wide range of internal signals (e.g., temperature, pain, itch, affective touch; Craig, 2002), this information is then relayed to the mid insula. The mid insula also receives information from other regions, for example the amygdala, hypothalamus and the secondary somatosensory cortex, the latter conveying non-homeostatic sensory information, and is thought to be where information is integrated (Ceunen, Vlaeyen, & Van Diest, 2016; Craig, 2008). Like the posterior insula, the mid insula often shows activation in response to several internal states (Craig, 2002). However, it is in the AI that bodily states (generally parasympathetic and sympathetic in the left and right hemispheres respectively (Craig, 2005, 2002, 2009)) are ‘re-represented’, allowing for interoception (subjective experience of these states) (Craig, 2005, 2011, 2002, 2009). Projections to the ACC then allow for appropriate motor responses to the subjective feelings represented in AI to be selected and completed, in line with proposals that the AI and ACC are complimentary limbic sensory (subjective feeling) and motor (response) (Craig, 2002), or input and output (Medford & Critchley, 2010) areas, respectively. The two cortical areas tend to be activated simultaneously (e.g. Medford & Critchley, 2010), consistent with their strong anatomical (Augustine, 1996; Moisset et al., 2010) and functional (Cauda et al., 2011; Sridharan, Levitin, & Menon, 2008; Taylor, Seminowicz, & Davis, 2009) connectivity. The
AI and ACC are implicated in a wide range of interoceptive states (Ibañez, Gleichgerrcht, & Manes, 2010), including hunger and satiety (Del Parigi et al., 2002), heat (Craig, Chen, Bandy, & Reiman, 2000; Olausson et al., 2005), thirst (de Araujo, Kringelbach, Rolls, & McGlone, 2003; Farrell et al., 2006), itch (Ikoma, Steinhoff, Ständer, Yosipovitch, & Schmelz, 2006; Mochizuki et al., 2007), affective touch (Francis et al., 1999; Gordon et al., 2013), gastric distension of the stomach (Stephan et al., 2003; Vandenbergh et al., 2005), distension of the bladder (Jarrahi et al., 2015), oesophageal and rectal stimulation (Coen & Gregory, 2007; Eickhoff et al., 2006; Hobday et al., 2001), taste (Kinomura et al., 1994; Small, 2010), pain (Derbyshire, 2003; Peyron et al., 2000), and fatigue (Caseras et al., 2008), as well as changes in heart rate (Critchley, Corfield, Chandler, Mathias, & Dolan, 2000), blood pressure (Harper, Bandler, Spriggs, & Alger, 2000), respiration (Liotti et al., 2001; Pattinson et al., 2009), and glucose levels (Allport et al., 2004). Insula and ACC activation have also been associated with suppression of natural urges involving interoceptive processes, for example suppression of breathing and voiding (Banzett et al., 2000; Kuhtz-Buschbeck et al., 2005; Seseke et al., 2006). The insula is also involved in processing body movement and proprioception, which are classed as interoceptive states under some definitions (e.g. Cameron, 2001; Vaitl, 1996), although the posterior insula is more commonly implicated than the anterior insula (Bottini et al., 2001; Farrer et al., 2003; Farrer & Frith, 2002; Fasold et al., 2008; Ferrè et al., 2012; Karnath & Baier, 2010; Mazzola, Faillenot, Barral, Mauguière, & Peyron, 2012; Petit & Beauchamp, 2003; Zu Eulenburg et al., 2013). Crucially, the re-representation of states into a meta-representation in AI is associated with subjective awareness of these states. While the objective temperatures of stimuli (non-painful and painful) are represented in the dorsal posterior insula, for example, the subjective ratings of these stimuli correlate best with AI activity (Craig et al., 2000; Kong et al., 2006).

Similarly, AI activation has been found to correlate with subjective fullness (relating to gastric distension) (Stephan et al., 2003), subjective orgasm quality in females (Ortigue, Grafton, & Bianchi-Demicheli, 2007), subjective unpleasantness of dyspnea (von Leupoldt et al., 2008), and bladder distension (Jarrahi et al., 2015), and AI activity (and volume) is associated with individuals’
ability to perceive their own heart beats (Critchley, Wiens, Rotshtein, Ohman, & Dolan, 2004). This study also found involvement of the AI and ACC in interoceptive attention (when participants were attending to timing of heart beats, relative to attending to pitch of auditory tones). AI activity has also been associated with subjectively reported fatigue following inflammation (Harrison et al., 2009).

Studies of patients with insula lesions have generally observed interoceptive difficulties (see Ibañez et al., 2010 for a review). The specific location of the lesion appears to play a role in the resulting impairment, both in terms of the interoceptive domains affected, and whether deficits are observed at the level of perception or recognition (Jones, Ward, & Critchley, 2010) Further, stimulation of the insula in humans has been found to elicit unpleasant sensations in the throat, mouth and nose (Krolak-Salmon et al., 2003), changes in heart rate (Oppenheimer, Gelb, Girvin, & Hachinski, 1992), inhibition of respiration, gastrointestinal motility, abdominal sensations and nausea (Penfield & Faulk, 1955). rTMS stimulation of the anterior insula has also been found to cause reduced interoceptive accuracy in both the cardiac and respiratory domains (Pollatos et al., 2016, but see Coll, Penton, & Hobson, 2017). While much evidence therefore suggests that the insula is integral to consciously representing feeling states, it is worth noting that some evidence indicates that interoceptive states, such as itch, tickle, pain and temperature, as well as emotions, may be experienced even following bilateral insula lesion (Damasio, Damasio, & Tranel, 2013). Damasio and colleagues therefore emphasise the role of brain stem, thalamic, hypothalamic, and somatosensory regions (alongside interoceptive cortices) in representing these states. Similarly, ACC and AI lesions do not appear to impair the perception of heart beats (Khalsa, Rudrauf, Feinstein, & Tranel, 2009), although it is possible that exteroceptive signals (for example due to vibration of the chest wall) may contribute to this ability. Studies in typical populations also question the role of the AI; for example, a recent meta-analysis observed that the posterior insula, and other regions not including the AI, were most commonly associated with attention to cardiac signals (Schulz, 2016). Such evidence suggests that further work is required to elucidate the exact
mapping of neural activity to facets of interoception across interoceptive domains. Overall, it appears that the AI and ACC are involved in interoception, but subcortical and somatosensory regions also contribute to the representation of internal states, and that different areas may support different aspects of interoceptive ability (see Section 2.1). Indeed, whilst the above model appears to propose a sequential hierarchical process from spinal, vagal and glossopharyngeal afferents to the AI (with the AI supporting the perception of internal states) and ACC it should be noted that there is a large degree of interaction (or ‘crosstalk’) between levels, from higher to lower areas, and between channels conveying interoceptive and exteroceptive information (Critchley & Garfinkel, 2017; Critchley & Harrison, 2013). As such, it appears that multiple structures contribute towards interoception, with the anterior insula and anterior cingulate cortex playing a crucial role in the perception and conscious awareness of these signals.

3. The relevance of Interoception to Mental Health

Interoception plays a role in a wide range of psychological functions (Quigley et al., 2021). Our central thesis is that atypical interoception is associated with impairments in a number of domains of functioning, and that such impairments characterise a broad range of psychiatric and neurological conditions. Such a conjecture has been raised previously (Barrett & Simmons, 2015; Khalsa & Lapidus, 2016; Khalsa et al., 2018; Murphy, Brewer, Catmur, & Bird, 2017; Paulus & Stein, 2010; Quadt, Critchley, & Garfinkel, 2018; Tsakiris & Critchley, 2016) either implicating interoception in specific disorders, or psychiatry more generally. Building on these, here we provide an in-depth overview regarding the proposed role of interoception in a range of abilities that relate to mental health. The following sections outline the role of interoception in two major domains of functioning of relevance to mental health; emotional processing and learning and decision-making.

3.1. The Role of Interoception in Typical Functioning: Emotional Processing

Interoception has been relatively understudied within psychology and cognitive neuroscience, certainly in comparison to exteroceptive senses such as vision and audition. The one domain in which interoception has a long history of study, however, is that of emotional processing (Pace-
Schott et al., 2019). Historical debates between the James-Lange (James, 1894; Lange, 1885) and Cannon-Bard (Bard, 1928; Cannon, 1929) positions centred on whether awareness of physiological arousal (i.e. interoception of arousal) in and of itself constituted an emotion, or whether cognitive appraisal of the individual’s situation was also necessary. Most modern theories are a variant on the Schacter-Singer (Schachter & Singer, 1962) model (see also Cantril & Hunt, 1932), which suggests that emotions are a combination of awareness of physiological arousal and cognitive appraisal of contextual cues. As such, interoception of arousal plays a central role in the experience of emotion, with interoception necessary for accurate detection of emotional signals, and judgements of emotional intensity, which are both necessary to identify one’s emotional state (e.g. Bechara & Naqvi, 2004). The necessity of intact interoceptive ability for typical emotional functioning is further highlighted when one considers second-order effects of interoception on affective processes such as emotional memory (see Section 3.2), emotion regulation and moral reasoning. Empirical evidence supports the necessity of interoception for all aspects of emotional processing. Much of this evidence utilises an individual differences approach to demonstrate that, across individuals, interoceptive accuracy is correlated with emotional lability (Rainer Schandry, 1981), emotion regulation (Füstös, Gramann, Herbert, & Pollatos, 2012; Kever, Pollatos, Vermeulen, & Grynberg, 2015), arousal focus (Barrett, Quigley, Bilss-Moreau, & Aronson, 2004), and emotional intensity (Füstös et al., 2012; Herbert, Herbert, & Pollatos, 2011; Pollatos, Herbert, Matthias, & Schandry, 2007; Pollatos, Traut-Mattausch, Schroeder, & Schandry, 2007; Wiens, Mezzacappa, & Katkin, 2000).

In support of a role for interoception in emotion, much evidence suggests that better perception of internal signals leads to more intense experience of one’s own emotions (e.g. Ferguson & Katkin, 1996; Pollatos, Gramann, & Schandry, 2007; Pollatos, Traut-Mattausch, et al., 2007; Wiens et al., 2000; although see Zamariola, Luminet, Mierop, & Corneille, 2019) and greater arousal focus when recounting one’s emotional experiences (Barrett et al., 2004). Higher interoceptive accuracy (rather than autonomic reactivity) also predicts decreased tolerance of (and
therefore increased sensitivity to pain (Pollatos et al., 2012), although the separate sensory and affective components of pain (Fernandez & Turk, 1992) complicate this finding. Depression, associated with less intense emotions, has also consistently been associated with reduced physiological reactivity to positive stimuli (Bylsma, Morris, & Rottenberg, 2008; Sloan & Sandt, 2010), which presumably makes perception of this physiological reaction more difficult. Further, specific internal states are associated with particular emotions; for example, associations exist between disgust and cardiac and gastric activity (Harrison, Gray, Gianaros, & Critchley, 2010), anger and increased heart rate and temperature (Ekman, Levenson, & Friesen, 1983; Ray, Wilhelm, & Gross, 2008), fear and increased heart rate and blood pressure (Ekman et al., 1983; Schwartz, Weinberger, & Singer, 1981), and surprise and increased skin conductance and decreased blood volume pulse (change in blood volume per heart beat) (Jang, Park, Park, Kim, & Sohn, 2015). In line with this, individuals tend to perceive specific emotional states as similar to specific non-emotional (interoceptive) internal states (Brewer et al., 2016a), and there is consistency among individuals in where within the body specific emotions (Nummenmaa, Glerean, Hari, & Hietanen, 2014) and non-emotional interoceptive signals (Nummenmaa, Hari, Hietanen, & Glerean, 2018) are experienced. Where interoceptive accuracy itself is concerned, as well as leading to more intense emotional experiences, evidence suggests that those with increased cardiac accuracy are more likely to physically express emotion (Ferguson & Katkin, 1996). Overall, it seems that perception of non-affective interoceptive states is associated with perfection of affective states, supporting the hypothesis that interoception and emotion are intrinsically linked.

With respect to regulation of one’s own emotions, it has been argued that interoceptive accuracy is of specific benefit when individuals use a reappraisal strategy to regulate emotion. Reappraisal is an especially effective and adaptive form of emotion regulation in which negative situations are cognitively re-appraised such that they are viewed as either neutral or positive in order to limit the negative affect they induce (Gross & John, 2003). Füstös et al. (2012) reasoned that individuals with better interoception should be better able to regulate their emotions through
reappraisal, owing to earlier identification of negative states (enabling earlier intervention to prevent the escalation and consolidation of negative states), and more precise individuation of emotional states (enabling more targeted, and therefore effective, intervention (Barrett, Gross, Christensen, & Benvenuto, 2001)). These authors provided evidence for this hypothesis in two studies. They first demonstrated that both the magnitude of the P3 ERP component, which reflects the intensity of emotional experience induced by an external stimulus (Pollatos, Gramann, et al., 2007), and participants’ subjective experience of emotion, were predicted by interoceptive accuracy. Their second study demonstrated that interoceptive accuracy predicted the effectiveness of emotional reappraisal in reducing the P3 magnitude, physiological arousal, and subjective experience in response to negative stimuli. Interestingly, reduced interoceptive accuracy may increase the likelihood of suppression, a less adaptive form of emotion regulation (in which emotion-consistent behaviours and/or thoughts are inhibited), owing to the fact that less perceptible interoceptive signals are likely to be easier to inhibit. The increased use of suppression strategies over reappraisal strategies is thought to lead to a variety of negative physical and mental health outcomes (Cutuli, 2014; Gross & John, 2003; John & Gross, 2004).

Beyond neurological measures, research has tended to utilise self-report measures of emotion regulation. Reports of difficulties decreasing one’s negative emotions in interviews analysed using Interpretative Phenomenological Analysis, for example, were associated with self-reported interoceptive difficulties (Zamariola, Frost, Van Oost, Corneille, & Luminet, 2019). Further, individuals with better cardiac interoceptive accuracy report using both cognitive reappraisal and suppression emotion regulation strategies more frequently than those with poorer cardiac accuracy (Kever et al., 2015; Pollatos, Matthias, & Keller, 2015). In line with this, Weiss and colleagues found that, in a sample of typical individuals and those with multisomatoform disorders, higher interoceptive accuracy was associated with better self-reported emotion regulation in the ‘frustration tolerance’ and ‘affect differentiation and affect tolerance’ subscales of a self-regulation scale (Weiss, Sack, Henningsen, & Pollatos, 2014). A similar relationship between
interoception and emotion regulation has been observed in developmental samples. In a sample of 9-16 year olds, heartbeat counting task performance was negatively correlated with self-reported maladaptive emotion regulation, but not associated with self-reported adaptive emotion regulation (De Witte, Sütterlin, Braet, & Mueller, 2016). In 4-6 year old children, performance on an emotion regulation vignette task was also related to interoceptive accuracy scores on a cardiac perception task, although it should be noted that this child-friendly interoceptive measure reflected the degree of over- or under-estimation of change in heart rate across two conditions, so is not comparable to the more commonly used heartbeat counting task (Schaan et al., 2019). In line with these findings are reports of emotion regulation and interoceptive difficulties within a sample of obese individuals relative to those with lower weight (Willem, Gandolphe, Roussel, Verkindt, & Pattou, 2019), and poorer metacognitive interoceptive awareness (although higher interoceptive accuracy) being associated with more emotional eating (Young et al., 2017). Taken together, these results suggest that interoception may play an important role in the regulation of one’s own emotions. Inconsistent with this, however, is the finding that neither self-reported interoception (BAQ) nor cardiac interoceptive accuracy was associated with mood following negative mood induction, or self-reported emotion regulation (Zamariola, Luminet, et al., 2019). Notably, however, mood *per se* does not directly assess emotion regulation; individuals who experienced an emotion more intensely may also be more effective emotion regulators, thereby reducing the emotion to a greater extent. Future work should therefore aim to assess emotion regulation more directly, for example through change in emotional response, and its relationship with interoceptive measures.

Interoception is not only relevant to self-focussed emotional processing. Evidence also links interoceptive abilities to one’s responsiveness to others’ emotions. For example, Terasawa et al. (2014) demonstrated that those with better interoceptive accuracy were more likely to report experiencing emotional responses to images of others’ emotional expressions. Similarly, both accuracy of expression recognition and facial mimicry of avatars’ emotional expressions, which may prompt empathic responding (Bird & Viding, 2014; Coll et al., 2017), appear to be higher in
those with better cardiac interoceptive accuracy (Chick, Rounds, Hill, & Anderson, 2019). In adolescents, those with good cardiac interoceptive accuracy also appear to recognise others’ facial expressions of sadness and fear better than those with poor interoceptive accuracy (Georgiou, Mai, Fernandez, & Pollatos, 2018). Further, judgements of the intensity of others’ naturalistic disgusted and painful facial expressions have been positively associated with cardiac interoceptive accuracy in a female sample, although accuracy distinguishing these expressions was unrelated to interoceptive accuracy (Dirupo, Corradi-Dell’Acqua, Kashef, Debbané, & Badoud, 2020). Increased empathy in those with higher interoceptive awareness may explain why higher interoceptive accuracy has also been found to correlate positively with altruism (Piech et al., 2017). Findings have been mixed, however, with Ainley, Maister and Tsakiris (2015) finding a non-significant association between HCT performance and self-reported empathy, perspective taking, or performance on the Reading the Mind in the Eyes Task, although this task included recognition of non-emotional mental states alongside emotional states.

Subjective ratings of the intensity and valence of one’s own emotional responses to others’ emotional expressions have also been associated negatively with systolic blood pressure (Pury, McCubbin, Helfer, Golloway, & McMullen, 2004), while accuracy of facial expression recognition has been found to correlate negatively with systolic and diastolic blood pressure, as well as total peripheral resistance (McCubbin et al., 2011). Similarly, heartbeat evoked potentials (an electroencephalography component thought to reflect the cortical processing of heartbeat sensations (Schandry, Sparrer, & Weitkunat, 1986; ; for a meta-analysis see Coll, Hobson, Bird & Murphy, 2020) vary in response to different observed facial expressions (Marshall, Gentsch, Schröder, & Schütz-Bobach, 2018). Where appraisal of others’ emotions is concerned, others’ facial disgust and fear expressions are rated as more intense when presented at the systole (ventricle contraction) phase of the cardiac cycle than at the diastole phase (when the ventricles relax and fill with blood) (Garfinkel et al., 2014; Gray et al., 2012). Cardiac timing of accuracy feedback has also been found to influence the speed with which good heartbeat perceivers learn the names of fearful faces, with
feedback at systole being more effective for learning than feedback at diastole (Pfeifer et al., 2017). For a more in depth discussion of this literature, see Critchley & Garfinkel (2015) and Garfinkel & Critchley (2016). Overall, this evidence suggests that both internal signals themselves, and one’s interpretation of these signals, relate to processing of others’ emotions.

Beyond behavioural responses, evidence suggests strong links between others’ emotions and one’s own physiological signals and neural responses in interoceptive cortex. Heart rate, for example, appears to change differentially in response to different emotions expressed by others, and these heart rate changes are associated with activation of brain regions involved in interoception, such as the amygdala, brainstem, OFC, AI, and ACC (Critchley et al., 2005; Gray et al., 2012). Similarly, responses of the AI and ACC to oesophageal stimulation appear to be affected by emotional context, whereby neural responses to oesophageal stimulation are greatest when intensely fearful facial stimuli are presented simultaneously (relative to less fearful or neutral facial expressions; Phillips, Gregory, et al., 2003). Further, neural responses to fearful faces are larger in the amygdala, insula and ACC, when the face is presented at systole than diastole (Garfinkel et al., 2014), and responses of the left OFT to disgusted faces, and the periaqueductal grey area to a range of emotional faces, are more strongly coupled to reductions in heart rate when stimuli are presented at systole than diastole (Gray et al., 2012). Non-invasive carotid artery stimulation also appears to decrease neural responses in the amygdala, hippocampus, thalamus and temporal fusiform areas while rating the intensity of fearful and neutral expressions (though increased lateral occipital and decreased temporal pole activity were observed while viewing the stimuli; Makovac et al., 2015). Further, carotid artery stimulation in this study led to higher ratings of emotional expression intensity, and this increase for fearful faces was associated with a decrease in activation in a number of brain areas including the insula, amygdala, and periaqueductal grey area. Others’ emotions therefore give rise to or alter interoceptive signals in observers and vice versa. As interoception is involved in one’s own emotional experience, it is likely that one’s internal state when encountering another’s emotion will affect one’s emotional response to the other’s emotion.
Findings such as these are consistent with a broad class of models that posit that one’s own emotions, and by inference the awareness of one’s own emotions, play a causal role in the recognition, understanding, and response to the emotions of others. Such models can be distinguished on the basis of the role that one’s emotions is assumed to play in processing those of the other (developmental or ‘on-line’). Under the former group of theories, infants learn to associate their own interoceptive states with their associated facial and vocal expression in others (through caregivers producing the associated expressions, for example), leading to accurate recognition of those states and an appropriate empathic response (Bird & Viding, 2014; Gergely & Watson, 1996; Heyes & Bird, 2007; Quattrocki & Friston, 2014). If interoception is impaired (either absent or noisy), then learning signals will be degraded, making associations between internal and external signals problematic, and leading to delayed or otherwise atypical learning. Within the ‘on-line’ group are ‘shared network’ and ‘embodied cognition’ models, in which another’s emotional state is mirrored, shared, or embodied in the self, and one’s own affective system is used to understand and recognise the state of the other (e.g. Barsalou, Niedenthal, Barbey, & Ruppert, 2003; Gallese, 2001; Goldman & Sripada, 2005; Niedenthal, Barsalou, Winkielman, Krauth-gruber, & Ric, 2005; Preston & de Waal, 2001). Interoception is perhaps of even greater importance in these models; mirroring another’s state is futile if one cannot then perceive that state in the self, and use it to understand the state of the other.

Evidence consistent with the role of interoception in one’s own affective processing also comes from neuroimaging and neuropsychological studies of interoceptive cortex. Obviously these structures are not solely concerned with interoception, and therefore evidence of affective impairment following damage to these areas can only ever suggest rather than confirm a role for interoception, but it is a useful literature to briefly survey nonetheless. The available literature from neuroimaging studies is replete with examples of activation of insula and ACC (interoceptive cortices) in affective tasks, suggesting functional overlap of the two processes. For example, both the AI and ACC are involved in processing one’s own emotions (Craig, 2009; Phillips, Drevets,
Rauch, & Lane, 2003), and activated during processing of others’ emotions (Calder & Young, 2005; Deng, Ma, & Tang, 2013; Wicker et al., 2003; Zaki, Davis, & Ochsner, 2012). Terasawa and colleagues identified numerous brain regions, namely the anterior insula, medial frontal cortex, lingual gyrus, temporo-parietal junction, and some brain stem regions, that were activated both in a condition where participants were asked to evaluate their current emotional state and in a condition where they evaluated their bodily signals, without an emphasis on emotion (Terasawa, Shibata, Moriguchi, & Umeda, 2012). In line with this, direct electrical stimulation of the insula has been found to increase recognition of facial expressions of anger, although this did not affect recognition of other basic emotions (Motomura et al., 2019). Fukushima, Terasawa and Umeda (2011) also demonstrated that cortical responses to one’s own heartbeats also predicted self-reported empathic traits. Similarly, a meta-analysis of the literature on empathy for pain revealed that the insula and ACC are the most reliably active structures in response to another’s pain. Finally, a recent multi-level kernel density analysis of fMRI studies found substantial overlap between emotional and interoceptive processing in bilateral insula, subgenual ACC, medial anterior temporal lobe, ventral mPFC, left basal ganglia, fusiform gyrus, and occipital cortex (Adolfi et al., 2017).

Lesion studies, especially studies of focal insula lesions, are not plentiful, and there is a resultant paucity of evidence for the role of interoceptive cortex. There are some reports of a reduction in the intensity of emotional experience following insula lesions (Berntson et al., 2011; Borg et al., 2013), but impairments in recognition and experience seem more common for disgust than other emotions. There is also evidence that insula lesions can result in the loss of the affective quality of pain; the patient can still ‘feel’ pain, but does not recognise it as aversive (‘asymbolia’; Berthier, Starkstein, & Leiguarda, 1988; Greenspan, Lee, & Lenz, 1999). This may be analogous to the amusia reported after insula lesions, whereby patients remain able to hear music, but lose their appreciation of its affective qualities (Griffiths, Warren, Dean, & Howard, 2004; Habib et al., 1995). Anterior insula lesions have also been associated with impaired recognition of pain in others, and reduced empathic arousal caused by pain in another (Gu et al., 2012). In line with this, in a meta-
analysis of studies on patients with fronto-insulo-temporal lesions, patients displayed both interoceptive and emotional impairments relative to healthy control participants (Adolfi et al., 2017). There is also evidence of both emotional and interoceptive impairments in frontotemporal dementia, where insula damage is common (Hobson, Brewer, Catmur, & Bird, 2019). Interoceptive cortex lesions therefore appear to impair processing of one’s emotional, as well as non-emotional, internal states. It should be noted, however, that there is some evidence suggesting that there may be other routes to emotion processing; as with non-emotional interoceptive abilities, typical emotional experiences have been reported in an individual with bilateral insula damage (Damasio et al., 2013).

3.1.1. The Role of Interoception in Emotional processing: Evidence from alexithymia

Given the importance of interoception for aspects of emotional functioning, research has begun to examine the relationship between interoceptive ability and alexithymia, a subclinical condition traditionally characterised by difficulties identifying and describing feelings, difficulties distinguishing emotional states from other bodily states, and a tendency to allocate attention to external rather than internal stimuli (Apfel & Sifneos, 1979; Nemiah, Freyberger, & Sifneos, 1976). While the cognitive aspects of alexithymia (difficulties identifying emotions) have been most commonly studied, it is worth noting that multiple types of alexithymia have been described, characterising individuals experiencing impairment at either the affective level only, where physiological signals themselves are atypical (Type III), the cognitive level only (Type II), or in both affect and cognition (Type I; Bermond, 1997; Moormann et al., 2008). Alexithymia can be seen as a neurodevelopmental condition in the majority of cases, occurring in the absence of neurological trauma, but can also be acquired following traumatic brain injury (Henry, Phillips, Crawford, Theodorou, & Summers, 2006; Wood & Williams, 2007; Hogeveen, Bird, Chau, Krueger, & Grafman, 2016). Although originally developed within a psychodynamic framework, the last few decades have seen alexithymia investigated by experimental psychologists, psychiatrists, and cognitive neuroscientists. While originally defined in terms of emotional deficits,
alexithymia was first observed in individuals with psychosomatic disorders (Sifneos, 1973), and we have recently proposed that many individuals with alexithymia experience a generalised interoceptive impairment, rather than one specific to emotional processing (Brewer et al., 2016a; Brewer, Cook, & Bird, 2016b). Indeed, alexithymia may be driven by interoceptive impairments in some individuals, and language or executive functioning impairments in others (Hobson et al., 2018, 2019; Murphy et al., 2018).

The conceptual links between alexithymia and interoception are clear; in Section 2.1 we outlined how interoception could be defined at a number of levels, and impairment at each of the conscious levels would affect emotional (as well as non-emotional) internal states, constituting alexithymia. Impaired conscious perception of a change of interoceptive state is likely to both result in, and be a result of, reduced attention to internal stimuli, and increased attention to external stimuli is thought to be a feature of alexithymia. Further, impaired conscious recognition of interoceptive signals would result in difficulties distinguishing emotional states from other, non-emotional interoceptive states such as hunger, and also result in a difficulty using interoceptive signals to distinguish between emotional states. Self-report alexithymia scales, such as the Toronto Alexithymia Scale (TAS-20; Bagby, Parker, & Taylor, 1994) and the Bermond-Vorst Alexithymia Questionnaire (BVAQ; Vorst & Bermond, 2001), therefore indirectly assess interoception, albeit specifically in the emotional domain.

There is direct and indirect evidence for alexithymia being closely associated with interoceptive atypicalities beyond the emotional domain. Most direct are studies that have employed objective tests of interoceptive abilities. A number of studies have utilised the most widely-used measure of interoceptive accuracy, the Heartbeat Counting Task, in order to investigate interoceptive ability as a function of alexithymia. Some evidence suggests that increased alexithymia was associated with decreased interoceptive accuracy; specifically that those with higher levels of alexithymia were less accurate at detecting their heartbeat (Herbert et al., 2011; Murphy et al., 2018; Näring & van der Staak, 1995; Shah, Hall, Catmur, & Bird, 2016). In line with
this, reductions in alexithymia have been associated with increases in cardiac perception accuracy following contemplative mental training (Bornemann & Singer, 2017). While Scarpazza, Sellitto, & di Pellegrino (2017), found the opposite pattern of results (that higher alexithymia was associated with better cardiac accuracy), it should be noted that this study did not include the standard control task for the Heartbeat Counting Task, meaning results should be interpreted with caution. Additionally, some studies have not observed relationships between cardiac accuracy and alexithymia in either direction using either the heartbeat counting task (e.g. Christensen, Gaigg, & Calvo-Merino, 2018; Nicholson et al., 2018; Nicholson et al., 2019; Zamariola, Vlemincx, Corneille, & Luminet, 2018) or discrimination task (Mul, Stagg, Herbelin, & Aspell, 2018; Mulcahy, Davies, Quadt, Critchley, & Garfinkel, 2019). While this evidence argues against the association between interoception and alexithymia, some of these mixed findings may also reflect the limitations of cardiac-based measures described in Section 2.2, or potential problems with task administration (e.g. Murphy, Brewer, Coll, et al., 2019; Nicholson et al., 2018). More recently, studies have investigated the relationship between alexithymia and interoceptive tasks outside of the cardiac domain. In one study, for example, alexithymia was found to correlate negatively with perception of taste, and muscular effort and positively with increased reliance on external cues for gauging respiratory output (Murphy et al., 2017). Similarly, another study reported that alexithymia was associated with reduced perception of thermal sensation (Borhani, Làdavas, Fotopoulou, & Haggard, 2017). However, not all studies are consistent with a relationship between alexithymia and interoceptive accuracy; for example, a recent study did not find a relationship between alexithymia and respiratory output or cardiac perception in a sample of autistic and neurotypical adults (Nicholson et al., 2019). However, differences in task administration of the respiratory output task (in comparison to its original description; Murphy et al., 2018) limit the conclusions that can be drawn from this study. Nevertheless, some evidence suggests no relationship between alexithymia and interoception across other domains (e.g., pain, affective touch, gastric perception; (Borhani et al., 2017; Jones, Roth, & Crowell, 2005). Indeed, a recent meta-analysis noted no simple
relationship between alexithymia and objectively measured interoceptive accuracy (Trevisan et al., 2019). However, the results of this meta-analysis are difficult to interpret; most included studies utilised the HCT and the inclusion or absence of a control task was not accounted for. Overall, these data emphasise a need for further study of the relationship between alexithymia and interoception across a range of populations and methodologies.

Individuals with high alexithymia also seem less able to report their state of arousal accurately (Gaigg, Cornell, & Bird, 2016; Nandrino et al., 2012; Stone & Nielson, 2001). In Gaigg et al.’s study, participants were required to report their degree of emotional arousal in response to a series of photographs, while objective measures of arousal including heart rate and galvanic skin response were recorded. Alexithymia scores predicted interoceptive accuracy, whereby increasing alexithymia predicted an increasing degree of disconnect between the subjective report of arousal level, and that measured objectively. Similarly, decoupling between self-reported emotional experience and autonomic responses has been observed in those with alexithymia (Connelly & Denney, 2007; Eastabrook, Lanteigne, & Hollenstein, 2013; Martin & Pihl, 1986; Newton & Contrada, 1994; Papciak, Feuerstein, & Spiegel, 1985; Pollatos, Werner, et al., 2011; Stone & Nielson, 2001).

Beyond objective measures of interoception, individuals with alexithymia have also been found to self-report impaired non-affective interoception, although findings differ according to whether interoceptive accuracy or attention is measured. Where interoceptive attention alone is measured (using the Body Perception Questionnaire; Porges, 1993), results have been mixed, with findings suggesting no correlation (Murphy et al., 2019), or a positive correlation (Betka et al., 2018; Ernst et al., 2014) with alexithymia. Self-reported interoceptive accuracy, on the other hand, has been shown to correlate negatively with alexithymia (Brewer et al., 2016a; Murphy, Catmur, et al., 2019). Relationships between alexithymia and other measures of self-reported interoception, such as the Multidimensional Assessment of Interoceptive Awareness (Mehling et al., 2012), the Self-Awareness Questionnaire (Longarzo et al., 2015), the Interoceptive Awareness Questionnaire
(van den Bergh, Bogaerts, Walentynowicz & Van Diest, 2012) and the Body Awareness Questionnaire (Shields et al., 1989), have also been observed (Longarzo et al., 2015; Muir, Madill, & Brown, 2017; Mul et al., 2018; Zamariola et al., 2018), but as these measures include items assessing both attention and accuracy, it is difficult to interpret these relationships. In line with these mixed findings, a meta-analysis of 66 independent samples concluded that the measures used to assess interoception unsurprisingly affect the relationship with alexithymia (Trevisan et al., 2019).

Although Trevisan et al. found a negative relationship between alexithymia and interoceptive ability over a range of subjective measures, a complex pattern of results emerged. Negative correlations were observed between alexithymia and both self-reported interoceptive accuracy and attention assessed by MAIA, while a positive correlation was seen between alexithymia and self-reported attention assessed by BPQ. These different relationships may be attributable to the fact that the BPQ and MAIA may assess negative and positive (e.g. mindfulness) aspects of interoception, respectively (Mehling, 2016).

Less direct evidence in support of a relationship between alexithymia and interoception adds to these findings. For example, individuals with alexithymia are delayed in seeking medical treatment in response to acute myocardial infarction, for example (Carta et al., 2013; Kenyon, Ketterer, Gheorghiade, & Goldstein, 1991). Similarly, those who self-report high levels of alexithymia consume caffeine (Lyvers, Duric, & Thorberg, 2014), alcohol (Thorberg, Young, Sullivan, & Lyvers, 2009), and other substances (de Haan, van der Palen, Wijdeveld, Buitelaar, & De Jong, 2014; Taylor, Parker, & Michael Bagby, 1990) more erratically than those with low self-reported alexithymia, suggesting difficulties perceiving the effects of these substances on one’s internal state. Indeed, recent evidence suggests that the relationship between self-reported interoception and alcohol consumption is mediated by alexithymia (Betka et al., 2018). Finally, alexithymia is associated with impaired abilities known to rely upon interoception, such as emotion regulation (Stasiewicz et al., 2012), emotion recognition (Grynberg et al., 2012), empathy (e.g. Enzi, Amirie, & Brüne, 2016; Gleichgerrcht, Tomashitis, & Sinay, 2015; Moriguchi et al., 2007a;
Silani et al., 2008), reward processing (Bibby & Ferguson, 2011; Goerlich et al., 2016; Morie et al., 2016) and decision-making, for example in the Iowa Gambling task (Ferguson et al., 2009; Kano, Masatochi, & Fukudo, 2011) (see Section 3.2 for a more detailed description of the role of interoception in decision-making). Notably, certain items on the TAS-20 are likely to assess interoceptive difficulties (e.g., “I have physical sensations that even doctors don’t understand”, “I am often puzzled by sensations in my body”, “I don’t know what’s going on inside me”).

Further support for the association between alexithymia and interoception comes from atypicalities in the structure and function of interoceptive cortex (AI and ACC) in alexithymia, which seem relatively specific to these regions, rather than diffuse across the brain (e.g. Borsci et al., 2009; Deng et al., 2013; Frewen, Pain, Dozois, & Lanius, 2006; Goerlich-Dobre, Bruce, Martens, Aleman, & Hooker, 2014; Grabe et al., 2014; Heinzel et al., 2010; Ihme et al., 2013; Jongen et al., 2014; Kano & Fukudo, 2013; Kano, Hamaguchi, Itoh, Yanai, & Fukudo, 2007; Kano et al., 2003; Moriguchi et al., 2007b; Paradiso, Vaidaya, Jatin, McCormick, Jones, & Robinson, 2008; Reker et al., 2010; Silani et al., 2008; Zhang et al., 2011). Evidence for the role of the AI in particular for the ontogenesis of alexithymia was provided by a large-scale (N = 129) study of patients with traumatic brain injury (Hogeveen, Bird, Chau, Krueger, & Grafman, 2016). Voxel-based lesion-symptom mapping was used to identify areas in the brain in which the degree of damage correlated with alexithymia severity. These analyses identified the degree of AI damage as a significant predictor of levels of alexithymia, with damage to the ACC approaching significance. In line with this, individuals with alexithymia also appear to exhibit atypical neural responses to interoceptive processes in these regions; higher alexithymia has been associated with greater insula and ACC activation in response to colonic distension (Kano et al., 2007), and increased insula activity but decreased ACC activity during interoceptive attention (Wiebking & Northoff, 2015). While studies have identified both positive and negative relationships between alexithymia and interoceptive cortex structure and function (Craig, 2009), likely owing to methodological
differences, these regions remain the most commonly implicated in studies of the neural bases of alexithymia.

Taken together with the findings presented in Section 3.1 above, there is substantial evidence that interoception contributes to both typical and atypical emotional experience. Interoception appears to contribute to one’s own emotional experience, the ability to recognise and label one’s own emotions, and the ability to recognise and empathise with others’ emotions. Where interoception is atypical, this has been associated with impairments in these emotional processes, including alexithymia. It can therefore be concluded that interoception is a fundamental process for typical emotion processing. Section 5.1 details how interoceptive atypicalities may, unsurprisingly, be related to the emotional difficulties that are observed in a number of clinical conditions.

3.2. The Role of Interoception in Typical Functioning: Learning and Decision-Making

It is intuitive that learning - particularly operant conditioning, in which a behaviour is associated with punishment or reward - relies on interoception. Operant conditioning requires the individual to perceive signals pertaining to punishment and reward, which are intrinsically interoceptive. The link is especially clear when one considers that poor interoception may be characterised not just by a failure to perceive any internal signal, but also by a failure to recognise and distinguish between interoceptive signals. Miscategorisation of interoceptive information may result in 1) a noisy or inconsistent learning signal, 2) systematic misclassification of the internal source of the punishment signal (e.g. pain as hunger), or 3) classification of a specific interoceptive signal which typically elicits a negative evaluation (e.g. pain, starvation, or muscle damage), as a positive reward signal (see Bevins & Besheer, 2014 for a thorough review of such ideas).

The association of interoception with motivation and decision-making is also intuitive. Classic theories of decision-making hold that, given two or more choices, an organism calculates the value of each, and selects the option with the highest value (e.g. Edwards, 1954; Bentham 1784-1832; Mill, 1773-1836). For many physiologically relevant stimuli (such as food and water), the value of the outcome associated with the stimulus depends on one’s present interoceptive state. The
more dehydrated one is, for example, the higher the value of water. It is therefore necessary to accurately perceive one’s internal state in order to determine the value of these stimuli. Even more fundamentally, value itself may be perceived as an interoceptive signal. This may be in a specific sense (such as the bodily changes associated with an attractive mate), or in a general sense of ‘wanting’ (Besheer, Palmatier, Metschke, & Bevins, 2004; Bevins, 2009; Pittenger & Bevins, 2013). Furthermore, under certain theories, decision-making is guided by stored representations of the bodily consequences of stimuli and responses, such as in Damasio’s (1994) Somatic Marker Hypothesis. These stored representations (‘somatic markers’), provided they can be perceived, are a further source of information when calculating the value of options.

A role for interoception in learning, motivation and decision-making was recognised by Pavlov (1849-1936) in his seminal work on classical conditioning. Pavlov noted that interoceptive states can enter into associations themselves, as either conditioned or unconditioned stimuli, and can also affect the acquisition and expression of learning by acting as contextual cues or occasion setters. Such ideas have been developed by Paulus and colleagues (Paulus, 2007; Paulus & Stein, 2010; Paulus & Stewart, 2014; Paulus, Tapert, & Schulteis, 2009), including their relevance to various psychiatric conditions (see Section 3.5), but there has been little direct investigation of the role of interoception in learning in humans.

One study that directly assessed the relationship between interoceptive accuracy and learning and decision-making was that of Werner, Jung, Duschek, & Schandry (2009), who found that higher scores on the heartbeat counting task predicted better performance on the Iowa Gambling Task, which relies on the ability to learn which of four options are advantageous, and which are disadvantageous. Interestingly, while this result has been replicated in typical adults, the opposite pattern has been observed in those with panic disorder (Wölk, Sütterlin, Koch, Vögele, & Schulz, 2014). Findings of a positive association between interoceptive accuracy and gambling task performance provide an intriguing explanation for a group of typical individuals described by Bechara and Damasio (2002), who displayed increased autonomic arousal when selecting
disadvantageous options (indicating implicit knowledge of which options produced positive and negative outcomes), but continued to select these options. It may be the case that these individuals, owing to poor interoceptive accuracy, were less able to perceive these bodily cues, and therefore less able to use them to aid their decision-making. Data consistent with this hypothesis were obtained by Dunn and colleagues (Dunn, Galton, et al., 2010), who used a modified version of the Iowa Gambling Task to demonstrate that when arousal cues favoured adaptive choices, those with better interoception made better choices, yet when arousal cues favoured maladaptive choices, individuals with better interoception made worse choices than those with poor interoception. Similarly, cardiac interoceptive accuracy has been found to correlate positively with the size of framing effects in a gambling task (the extent to which one’s decision to gamble is affected by whether identical outcomes are described as losses or gains) (Shah, Catmur, & Bird, 2016).

Rejection rates in the ultimatum game (in which participants are required to decide whether to accept or reject proposals from a partner concerning division of monetary rewards) have also been found to be predicted by electrodermal responses in those with good interoceptive accuracy, but not in those with poor interoceptive accuracy (Dunn, Evans, Makarova, White, & Clark, 2012). One’s ability to perceive interoceptive signals therefore appears to affect the extent to which internal cues are utilised in decision-making processes. Attention to interoceptive signals may also affect behaviour in the ultimatum game; listening to the timing of one’s own heart beats has been found to increase feelings of unfairness in response to unfair offers, and increase the likelihood of making unfair offers to other players (Lenggenhager, Azevedo, Mancini, & Aglioti, 2013).

The relationship between interoception and specifically risky decision-making has also been investigated. Previous work demonstrated that individual differences in physiological arousal correlate with individual differences in loss aversion (the overweighting of losses with respect to equal gains) during risky decision-making (Sokol-Hessner et al., 2009). Sokol-Hessner and colleagues (Sokol-Hessner, Hartley, Hamilton, & Phelps, 2014) predicted that the degree to which these physiological signals of arousal can be perceived (interoceptive accuracy) is likely to
modulate the impact of arousal on loss aversion during risky decision-making. Results confirmed their prediction, demonstrating that individuals with better interoceptive accuracy were more loss-averse, providing direct evidence for an impact of interoception on risky decision-making. These findings are consistent with the suggestion that the increased levels of risky decision-making during adolescence (e.g. Hoorn, Crone, & Leijenhorst, 2016; Smith et al., 2011; Steinberg, 2007) may be attributable to reduced interoceptive abilities during this stage (Murphy, Brewer, Catmur, & Bird, 2017). A recent study found that individuals with higher self-reported interoceptive attention (measured by the BPQ) were more conservative in a task where they filled a virtual body with air to earn money, at the risk of the body being filled to explosion (leading to the loss of the monetary reward), although this effect was not observed when the virtual item being filled was a balloon (Salvato, De Maio, & Bottini, 2019). The severity of the risk conveyed by the task may modulate findings, as a simple probability discounting measure of risk-taking (in which the choice is between small certain or large uncertain gains) was not associated with BPQ scores in one study (Herman, Critchley, & Duka, 2018). Interoceptive accuracy in the cardiac domain also appears to be related to risky decision-making in naturalistic scenarios; financial traders have been found to have higher accuracy than matched control participants, and interoceptive accuracy within the trader sample was positively related to profitability and years of experience (Kandasamy et al., 2016).

Further evidence for the role of interoceptive awareness during learning and decision-making was provided by Katkin, Wiens and Ohman (2001) who showed that the acquisition of fear conditioning with masked stimuli was dependent upon interoception. When masked images were associated with the delivery of an electric shock, individuals with high interoceptive accuracy were able to use the physiological arousal response associated with the stimulus to predict the likelihood of shock, whereas those low in interoceptive accuracy were not.

Improved learning by those with high interoceptive accuracy would be expected to result in improved recall. Consistent with this suggestion, interoceptive accuracy has been positively associated with recognition memory for previously presented affective stimuli (Pollatos &
Schandry, 2008; Werner, Peres, Duschek, & Schandry, 2010). Additional evidence for a role for interoceptive accuracy in memory was provided by Garfinkel, Barrett, Minati, Dolan, Seth, and Critchley (2013), who showed that the phase of the cardiac cycle during which a stimulus was presented was a significant predictor of subsequent memory, and that this effect was modulated by interoceptive accuracy. Similarly, higher heart rate and higher interoceptive accuracy on the HCT have been associated with performance on a prospective memory task, in which participants were required to identify targets while performing a concurrent 2-back task (Umeda, Tochizawa, Shibata, & Terasawa, 2016).

Internal signal changes such as cardiac deceleration, and increased pupil dilation and skin conductance, also tend to follow (e.g. Bastin et al., 2017; Hajcak, Donald, & Simons, 2003; Łukowska, Sznajder, & Wierzchoń, 2018) and even precede (Bury, García Huesca, Bhattacharya, & Herrojo Ruiz, 2019) errors, and recent work suggests interoceptive abilities are associated with error monitoring. In Sueyoshi and colleagues’ study, interoceptive accuracy, as measured by the heartbeat counting task, was associated with the degree of post-error behavioural slowing (assessed by reaction times), thought to be driven by an affective signal associated with error (Sueyoshi, Sugimoto, Katayama, & Fukushima, 2014). These results indicate that, by influencing the degree to which the error-related affective signal can be perceived, interoceptive ability affects feedback-related behaviour. Furthermore, interoceptive accuracy correlated with the size of the Pe event-related potential, an electrophysiological signal reflecting the later portion of the error-related neural signal, allowing the impact of interoceptive ability to be located in time. Finally, both alexithymia and interoceptive accuracy have been associated with the degree to which one engages cognitive control to prevent errors in tasks inducing response conflict (de Galan, Sellaro, Colzato, & Hommel, 2014).

While there is not an abundance of work directly testing the impact of interoception on learning, two bodies of work provide indirect evidence for this association, by demonstrating that common neural areas represent interoceptive information, and are involved in learning. The first of
these combines computational modelling of learning with cognitive neuroscience methods, in order to map the neural correlates of learning parameters. In these studies, individuals perform a learning task, and computational models are fit to participants’ data to derive learning parameters such as reward prediction errors and risk estimates. These parameters are then regressed against neural data to identify activity that covaries with learning parameters. Studies using this technique have demonstrated that activity in areas subserving interoception, including the insula and ACC, show patterns of activity consistent with the representation of risk (formally the variance associated with the outcome of a particular stimulus) and risk and reward prediction errors (Asahi et al., 2006; Mizuhiki, Richmond, & Shidara, 2012; Paulus, Rogalsky, Simmons, Feinstein, & Stein, 2003; Preuschoff, Quartz, & Bossaerts, 2008; Xiang, Lohrenz, & Montague, 2013; Xue, Lu, Levin, & Bechara, 2010). Additionally, insula and ACC activation have been found to increase with higher degrees of uncertainty concerning the outcome of one’s decision (Critchley, Mathias, & Dolan, 2001), and those with higher levels of intolerance of uncertainty exhibit greater anterior insula activation when reward is uncertain (Gorka, Nelson, Phan, & Shankman, 2016). Anterior insula activity has been associated with risk-free choices, and during risk-aversion mistakes (when an individual chooses not to take a risky decision, when it would have been advantageous) (Kuhnen & Knutson, 2005), and with individuals’ risk-taking propensity during risky decision-making; those with higher levels of harm avoidance and neuroticism exhibited greater insula activation (Paulus, Rogalsky, et al., 2003). Decision urgency (the increasing desire to make a decision as time passes) has also been associated with insula activation (van Maanen, Fontanesi, Hawkins, & Forstmann, 2016). The degree of insula activation during decision-making may also be a function of interoceptive accuracy; one study found that right insula activation was associated with Iowa Gambling Task performance only in individuals who performed well on the heartbeat counting task, and not in those with poor cardiac perception (Werner et al., 2013). It is worth noting, of course, that co-activation of large, differentiated neural structures during interoception and representation of
learning parameters does not necessarily indicate that the same neurons, or even populations of neurons, are involved in both processes.

A similar caveat should be applied to the second body of work, which reports the impact of damage to interoceptive cortex on learning in both human and non-human animals. A full examination of this literature is beyond the scope of this paper, but several authors have reported effects of insula lesions on learning and decision-making. Clark and colleagues, for example, reported that patients with insula lesions were less able to adjust their behaviour in a risky decision-making task, in response to changing levels of risk (Clark et al., 2008). Such an observation was supported by Weller et al. (2009), who observed that patients with insular lesions adopted lower risk decision-making strategies, even in situations in which risky decisions were optimal. Disadvantageous choices on the Iowa Gambling task in lesion patients have also been reported by Bar-On, Tranel, Denburg, & Bechara (2003) although it should be noted that the patient sample was heterogeneous, with those with insula lesions being grouped with those with ventromedial and amygdala lesions. The hypothesised role of the insula in learning is also supported by non-human animal studies, in which an experimental reduction in serotonin in the anterior insula impairs conditioned taste avoidance, while reduction of serotonin in posterior insula impairs the acquisition of conditioned disgust reactions (Tuerke, Limebeer, Fletcher, & Parker, 2012). Lesions to the rat insula have also been associated with reduced encoding of the incentive value of instrumentally conditioned behaviour to obtain food; while control rats performed fewer food-obtaining actions following low food deprivation than following high deprivation, lesioned rats did not alter their behaviour across these conditions (Balleine & Dickinson, 2000). Non-human animal studies also imply a role of the insula in risky decision-making. Ishii, Ohara, Tobler, Tsutsui, & Iijima (2012), for example, demonstrated that temporary inactivation of rat anterior insula cortex resulted in a reduction of risky choices across two separate gambling tasks. Similarly, insula lesions have been found to lead to reductions in performance on an adaptation of the Iowa Gambling Task in rats who previously performed optimally (Daniel et al., 2017).
Overall, the evidence presented in this section suggests a role for interoception in learning and decision making, across a range of tasks. As learning and decision-making impairments have been observed across a range of clinical conditions, these may again be explained by interoceptive atypicalities. Section 5.2 discusses the relationship between interoceptive atypicality and impairments in these domains, and their relevance to psychopathology.

4. Increased Prevalence of Interoceptive Impairments in Clinical Conditions

Sections 3.1 and 3.2 briefly surveyed the literature linking interoception to two major domains of functioning (emotional processing, and learning and decision-making). Having provided this evidence for the importance of interoception in typical functioning, the following sections detail evidence of interoceptive impairment in several psychiatric and neurological conditions.

4.1. Evidence of interoceptive impairment in psychiatric and neurological conditions

Atypical interoception is ubiquitous across psychiatric and neurological conditions. While certain conditions have long been associated with atypical interoception, such as Feeding and Eating Disorders (EDs) including anorexia Nervosa, Bulimia Nervosa, and obesity (Eshkevari, Rieger, Musiat, & Treasure, 2014; Fassino, Pierò, Gramaglia, & Abbate-Daga, 2004; Kinnaird, Stewart, & Tchanturia, 2020; Klabunde, Acheson, Boutelle, Matthews, & Kaye, 2013; Lapidus et al., 2020; Lattimore et al., 2017; Lutz et al., 2019; Pollatos et al., 2008; although see e.g. Ambrosecchia et al., 2017; Kinnaird et al., 2020), the full range of conditions impacted by interoceptive atypicalities is only just being realised (Barrett & Simmons, 2015; Brewer et al., 2016a; Brewer, Happé, Cook, & Bird, 2015). There is evidence of increased prevalence of atypical interoception in anxiety and panic disorders (Ehlers & Breuer, 1992; Paulus & Stein, 2010; Adrián Yoris et al., 2015), alcohol and substance abuse (Jakubczyk et al., 2019; Naqvi & Bechara, 2010; Paulus & Stewart, 2014; Paulus et al., 2009; Verdejo-Garcia, Clark, & Dunn, 2012), depression (Aaronson et al., 2017; Dunn, Dalgleish, Ogilvie, & Lawrence, 2007; Furman, Waugh, Bhattacharjee, Thompson, & Gotlib, 2013; Harshaw, 2015; Paulus & Stein, 2010; Pollatos, Traut-Mattausch, & Schandry, 2009), somatoform
disorders (Schaefer, Egloff, & Witthöft, 2012), Autism Spectrum Disorder (ASD; Garfinkel et al., 2016a; Hatfield, Brown, Giummarra, & Lenggenhager, 2017; Mul et al., 2018; Nicholson et al., 2019), Attention-Deficit/Hyperactivity Disorder (ADHD; Kutscheidt et al., 2019), Obsessive Compulsive Disorder (OCD; Lazarov, Dar, Oded, & Liberman, 2010; Schultchen, Zaudig, Krauseneck, & Pollatos, 2019; see also Stern, 2014 for a discussion), schizophrenia (Ardizzi et al., 2016), depersonalisation/derealisation disorder (Schulz et al., 2015; Sedeño et al., 2014; but see Michal et al., 2014), personality disorders (Mussgay, Klinkenberg, & Rüddel, 1999; although see Hart, Mcgowan, Minati, & Critchley, 2013), and those with high levels of psychopathic traits (Nentjes, Meijer, Bernstein, Arntz, & Medendorp, 2013). Notably, evidence for atypical interoception across each of these clinical groups has been mixed, with reports of increased, decreased and typical interoceptive abilities sometimes reported across the same condition. Such discrepancies are likely owing to individual differences in interoception within each population as well as difficulties with the measurement of interoception (see Section 2.2). The majority of studies investigating interoception in clinical groups have utilised measures of cardiac, respiratory and gastric internal signals, but evidence also exists for impairments of the perception of interoceptive signals such as taste and itch in clinical conditions (e.g. Caccavale, Bove, & Bove, 2016; Kazour et al., 2017; Kinnaird, Stewart, & Tchanturia, 2018).

Atypical interoception and high alexithymia have also been observed in neurological conditions with similar symptom profiles, such as Frontotemporal dementia (behavioural variant; García-Cordero et al., 2016; Salvato, Mercurio, Sberna, Paulesu, & Bottini, 2018; semantic variant; Sturm & Levenson, 2011), Alzheimer’s Disease (García-Cordero et al., 2016), Parkinson’s Disease (Assogna et al., 2012; Ricciardi et al., 2016), and Multiple Sclerosis (Chahraoui, Duchene, Rollot, Bonin, & Moreau, 2014; Prochnow et al., 2011). A number of medical conditions have also been associated with interoceptive difficulties (see Section 5.3.7). It is worth noting that much of the work in clinical populations has relied on measures of interoceptive accuracy in the cardiac domain,
or self-reported interoceptive attention. Future work should therefore aim to characterise each population in terms of interoceptive accuracy, attention, and metacognitive awareness.

Notably, the relationship observed between interoceptive accuracy and anxiety disorders and OCD is often in the opposite direction to that observed in other conditions. While the majority of clinical conditions are associated with reduced interoceptive accuracy, interoceptive accuracy has been found to be increased in anxiety and panic disorder (for reviews, see Domschke, Stevens, Pfleiderer, & Gerlach, 2010; Paulus & Stein, 2010) and in OCD (Yoris et al., 2017). This finding has not been consistent, however, indicating heterogeneity among anxiety patients with regard to interoceptive abilities (De Pascalis, Alberti, & Pandolfo, 1984; Van Der Does, Antony, Ehlers, & Barsky, 2000) with some recent studies suggesting that higher anxiety is associated with lower interoceptive accuracy (Ewing et al., 2017; Garfinkel et al., 2016b; Garfinkel et al., 2016).

Relatedly, elevated interoceptive accuracy when observed in those with anxiety disorders may be specific to the cardiac interoceptive domain, as respiratory interoceptive accuracy appears to be reduced in these individuals (Bogaerts et al., 2005; van den Bergh et al., 2004). Potentially, cardiac signals are particularly salient for (or preferentially attended to by) individuals with anxiety or panic disorders, leading to improved performance on this task specifically, rather than generally superior interoceptive accuracy (Ehlers, 1993). Relatedly, there is debate concerning the extent to which cognitive processes in anxiety disorder contribute to improved interoception; it is possible that anxiety and panic are associated with increased preoccupation with, atypical prediction of, and/or catastrophising of, internal signals, or superior detection of prediction errors, rather than increased accuracy in detecting these signals per se (Anderson & Hope, 2009; Clark, 1986; Ludewig et al., 2005; Paulus & Stein, 2006, 2010). Consistent with this conjecture, there is evidence to suggest that anxiety is characterised by high levels of attention to internal signals (as measured by the BPQ) relative to one’s cardiac interoceptive accuracy (Garfinkel et al., 2016b; Palser, Fotopoulou, Pellicano, & Kilner, 2018). Indeed, interoceptive exposure therapy is often used clinically to reduce attention to interoceptive cues in those with anxiety and panic disorders (Boettcher, Brake, &
Barlow, 2016; Boswell et al., 2013). Alternatively, internal cues may be stronger in those with panic/anxiety (e.g., higher arousal; Hoehn-Saric & Mcleod, 2000; Paulus & Stein, 2010). It is currently unclear, however, whether this arousal follows the initial misinterpretation of the interoceptive cue, or proceeds it and therefore aids detection. The recent development of an interoceptive framework emphasising the distinction between interoceptive accuracy and attention, both where self-report and behavioural measures are concerned (Murphy et al., 2017; Murphy et al., 2019), is likely to elucidate the relationship between interoception and anxiety in future research.

Beyond behavioural studies, interoceptive cortex atypicalities are also common in clinical samples. Atypical structure and function of the insula and ACC have been observed in those with eating disorders (Frank, 2015; Uher et al., 2004), ASD (Simms, Kemper, Timbie, Bauman, & Blatt, 2009; Uddin & Menon, 2009), OCD (Menzies, Chamberlain, Laird et al., 2008), anxiety disorders (Paulus & Stein, 2006; Rosso, Makris, Britton et al. 2010; Holzscheider & Mulert, 2011), depression and bipolar disorder (Drevets, Price, & Furey, 2008; Vizueta et al., 2012; L. Wang, Hermens, Hickie, & Lagopoulos, 2012), schizophrenia (Bouras, Kövari, & Hof, 2001; Reid et al., 2010; Shepherd, Matheson, Laurens, Carr, & Green, 2012; Wylie & Tregellas, 2010), substance abuse/addiction disorders (Luijten et al., 2014; Ma et al., 2011; Naqvi & Bechara, 2009, 2010; Zhou et al., 2011) and PTSD (Garfinkel & Liberzon, 2009). Indeed, a large-scale meta-analysis of brain morphology across six distinct psychiatric disorders identified left and right insula and dorsal anterior cingulate as the only areas of grey matter loss common to all disorders (Goodkind et al., 2015). Again, while this evidence is only indirect, it is consistent with the idea that interoceptive deficits are a common feature of psychiatric conditions.

Notably, while a linear relationship between interoceptive impairment and disorder severity is often assumed, it may also be the case that a quadratic relationship exists, whereby either atypically low or atypically high interoceptive abilities are associated with psychopathology, or where atypical interoception may be more likely to be observed in individuals with moderate rather than severe psychopathology (Dunn et al., 2007). The nature of the relationship may also differ, of
course, across clinical conditions. Further, while much research suggests an association between interoceptive abilities and psychopathology, this conclusion is based upon correlational findings. Longitudinal studies, as well as intervention studies, are required to determine whether interoceptive impairment causally affects disorder development. Similarly, whether the causal relationship between interoception and psychopathology is identical across disorders requires further investigation. As discussed in Section 7.2, the differences across disorders may be partially attributable to interoceptive impairment at different developmental stages, or in different interoceptive domains or dimensions. Whether specific atypicalities in interoception lead to particular disorders, and whether disorder symptoms can cause or exacerbate interoceptive impairment remains to be determined. It is possible, for example, that some disorders lead one to suppress internal signals; those with eating disorders may attempt to attend away from signals of hunger, thirst, and satiety, while those with depression may attempt to reduce negative affect by suppressing internal signals associated with aversive emotions. On the other hand, reduced attention towards, or accuracy of, detection of internal signals could contribute to flattened affect in depression, or the aversiveness of food restriction in eating disorders.

4.1.1. Interoceptive impairment in psychiatric and neurological conditions: Evidence from alexithymia

As alexithymia is closely associated with interoceptive impairment (see Section 3.1.1), the high number of psychological conditions that co-occur with alexithymia is also suggestive of impaired interoception across clinical populations. While this evidence is indirect, it is worthy of discussion while more direct evidence is still to be accumulated across clinical populations. Increased prevalence of alexithymia has been observed in eating disorders (Bourke, Taylor, Parker, & Bagby, 1992; Cochrane, Brewerton, Wilson, & Hodges, 1993; Jimerson, Wolfe, Franko, Covino, & Sifneos, 1994; Rozenstein, Latzer, Stein, & Eviatar, 2011), ASD (Berthoz & Hill, 2005; Hill, Berthoz, & Frith, 2004; Kinnaird, Stewart, & Tchanturia, 2019), schizophrenia (Henry, Bailey, von Hippel, Rendell, & Lane, 2010; Heshmati, Jafari, Hoseinifar, & Ahmadi, 2010), depression (Honkalampi,
Hintikka, Laukkanen, Lehtonen, & Viinamäki, 2001), post-traumatic stress disorder (Frewen, Dozois, Neufeld, & Lanius, 2008; Yehuda et al., 1997), substance abuse (Dorard et al., 2008; Mann, Wise, Trinidad, & Kohanski, 1995), somatisation/Medically Unexplained Symptoms (De Gucht & Heiser, 2003), chronic fatigue syndrome (van de Putte, Engelbert, Kuis, Kimpen, & Uiterwaal, 2007) and panic and anxiety disorders (Cox, Bourdeau, & Company, 1995; De Berardis et al., 2008; Marchesi, Fontò, Balista, Cimmino, & Maggini, 2005). Indeed, a recent study linked alexithymia to a range of self-reported psychopathological symptoms in a community sample of children and adolescents (Weissman et al., 2020). High levels of alexithymia in non-clinical developmental samples have been associated with high levels of anxiety (Sayin, Derinöz, Bodur, Şenol, & Şener, 2007), depression and interpersonal problems (Joybari, 2014), aggressive behaviour (Manninen et al., 2011) and delinquency (Zimmermann, 2006). Notably, while the increased alexithymia levels in anxiety appear to be at odds with the increased interoceptive accuracy often observed in this population, the relationship between interoceptive abilities and anxiety appears to be complicated (see Section 4). Perhaps alexithymia is common following a range of interoceptive atypicalities, including atypically high and low interoceptive accuracy and attention. Alexithymia levels are also elevated in those with some neurological conditions, such as frontotemporal dementia (Sturm & Levenson, 2011), and medical conditions, such as HIV (McIntosh et al., 2014), diabetes (Mnif et al., 2014; Topsever et al., 2006), and Multiple Sclerosis (Bodini et al., 2008; Chahraoui et al., 2014; Prochnow et al., 2011).

Consistent with the suggestion that alexithymia is characterised by general interoceptive impairment, the clinical disorders that have been associated with poor interoception overlap largely with those that co-occur with alexithymia. It is also interesting to note that in none of the studies listed above was alexithymia observed in every member of the clinical samples studied, and not all members of the clinical samples showed atypical interoception. This suggests that interoceptive impairment is not a universal symptom of those clinical conditions with which it has been associated, and that individual differences in interoceptive ability are likely to exist in clinical
populations (as they do in the typical population), and these individual differences in interoception may co-vary with alexithymia. It is worth noting that, while interoceptive impairments are likely to give rise to alexithymia, alexithymia may also result from linguistic impairments (Hobson et al., 2018, 2019), so evidence for interoceptive atypicalities using alexithymia measures should be treated as indirect.

5. The Link Between Interoceptive Impairment and Clinical Symptoms Across Disorders

A claim consistent with the evidence above is that interoception constitutes the ‘P’ Factor of psychopathology; that atypical interoception underlies various clinical symptoms seen across a multitude of psychiatric (and neurological) conditions. Of course, one does not need to agree with the existence of the P Factor to acknowledge the possible role of interoception across diagnostic and symptoms categories, it may be that different patterns of atypical interoception (heightened, reduced) across the domains (cardiac, gustatory, respiratory etc) and dimensions (accuracy, attention) of interoceptive processing result in a myriad of distinct conditions. Both positions are consistent with the evidence reviewed above: In Sections 3.1 and 3.2, we argued that interoception plays a significant role in typical cognition, particularly within the affective domain, and in learning and decision-making. In Section 4, we argued that atypical interoception is observed in a multitude of psychiatric and neurological conditions. Evidence for the link between interoception and emotional and learning and decision-making processes makes it plausible that at least some of the symptoms common to these conditions have a unitary, interoceptive basis, though the exact pattern of interoceptive impairments may dissociate across disorders (see Section 7.2). In the following sections, we outline existing evidence for this claim within selected domains of function. This review of evidence is not exhaustive, nor is it intended to cover all symptom domains. Instead, it is intended to demonstrate the range of common psychological difficulties that may be (partly) caused, or exacerbated, by atypical interoception.
5.1. Interoceptive impairment and affective deficits

As described in Section 3.1, almost all aspects of emotion processing rely on interoception. One would therefore expect affective impairment in all conditions characterised by interoceptive impairment. Major Depressive Disorder is the clinical condition most closely associated with affective deficits, so it is unsurprising that depression presence and severity (Dunn et al., 2007; Harshaw, 2015; Pollatos et al., 2009) and symptoms of depression such as anhedonia (Dunn, Stefanovitch et al., 2010; Lackner & Fresco, 2016), are associated with interoceptive impairment (though the relationship between interoception and depression may be non-linear; Dunn et al., 2007; see Sections 4.1 and 7.3). Indeed, stimulation of the vagus nerve (one of the core interoceptive neural pathways) has been successfully used as a means of treating depression (Aaronson et al., 2017). Evidence also suggests that affective impairments exist in disorders beyond depression; emotional difficulties such as emotion recognition, emotion regulation, and empathy deficits have been observed in a wide range of conditions, including Autism Spectrum Disorder (Harms, Martin, & Wallace, 2010; Harmsen, 2019; Weiss, Thomson, & Chan, 2014), ADHD (Friedman et al., 2003), Feeding and Eating Disorders (Brewer, Cook, Cardi, Treasure, & Bird, 2015; Brewer et al., 2019; Harrison, Tchanturia, & Treasure, 2010; Kerr-Gaffney, Harrison, & Tchanturia, 2019), Schizophrenia (Bonfils, Lysaker, Minor, & Salyers, 2016, 2017; Edwards, Jackson, & Pattison, 2002; Kimhy et al., 2016; O’Driscoll, Laing, & Mason, 2014), post-traumatic stress disorder (PTSD; Palgi, Klein, & Shamay-Tsoory, 2017; Plana, Lavoie, Battaglia, & Achim, 2014; Villalta, Smith, Hickin, & Stringaris, 2018), OCD (Cain, Ansell, Simpson, & Pinto, 2015; Daros, Zakzanis, & Rector, 2014; Yap et al., 2018), depression (Bourke, Douglas, & Porter, 2010; Harshaw, 2015; Schreiter, Pijnenborg, & aan het Rot, 2013), and Huntington’s Disease (Trinkler et al., 2017).

Although individuals with these conditions are more likely to exhibit emotional difficulties than typical individuals, such difficulties are not universal. Under our hypothesis, affective difficulties are more likely to be explained by the presence and severity of co-occurring interoceptive impairment, than of the disorder itself.
This hypothesis has been examined in our lab using a series of studies in which affective ability is examined in individuals with a clinical diagnosis (of ASD or ED for example) with varying degrees of alexithymia (likely an indicator of interoceptive impairment; see Section 3.1.1), and a non-clinical control group matched for alexithymia. Using this design, we have observed that impairments thought to be core characteristics of ASD, such as reduced empathic response of the insula (Bird et al., 2010) and impaired recognition of emotional facial expressions (Cook, Brewer, Shah, & Bird, 2013; Oakley, Brewer, Bird, & Catmur, 2016), are in fact due to co-occurring alexithymia and, by inference, potentially interoceptive impairment. A similar pattern of results have been observed in a neurotypical sample, whereby alexithymia largely explained the relationship between autistic traits and facial emotion recognition (Bothe, Palermo, Rhodes, Burton, & Jeffery, 2019). The ability to produce typical emotional facial expressions (Trevisan, Bowering, & Birmingham, 2016), and to recognise emotion from vocal cues (Heaton et al., 2012), also appear to be explained by alexithymia, rather than by ASD presence or severity itself. Similar findings have been observed in other clinical populations; again using groups matched according to alexithymia, we found that the emotion recognition impairment often reported in individuals with EDs is due to co-occurring alexithymia, rather than ED presence or severity itself (Brewer, Cook, et al., 2015). Alexithymia has also been linked to emotion recognition deficits in a sample of adolescents with EDs (Zonnevijlle-Bender, van Goozen, Cohen-Kettenis, van Elburg, & van Engeland, 2002). Alexithymia also seems to predict emotion recognition in individuals with Huntington’s Disease (Trinkler et al., 2017) and both emotion recognition and empathy deficits in those with traumatic brain injury (McDonald et al., 2011; Neumann, Zupan, Malec, & Hammond, 2014). High levels of empathic personal distress have been associated with high alexithymia, rather than ED presence or severity (Brewer et al., 2019), and empathic impairments (Maurage et al., 2011) and emotion regulation difficulties (Stasiewicz et al., 2012) appear to be predicted by alexithymia in those with alcohol addictions. Alexithymia also appears to contribute to the relationship between self-reported empathy and both autistic and schizotypal traits in the non-clinical population (Aaron, Benson, &
Park, 2015). It has also been suggested that high prevalence of alexithymia may account for negative experience of emotions, such as feeling ‘flooded’ by emotions or difficulties accepting one’s own emotions (Edel et al., 2010), in those with ADHD. This may be due to alexithymia reducing one’s ability to label emotional states, which has been shown to reduce the aversive nature of the state (Foland-Ross et al., 2010; Lieberman et al., 2007; Mazefsky & White, 2014). Finally, Martin (1985) hypothesised that the observed relationship between alexithymia and stress-related illness emerges as alexithymia leads to difficulties recognising that a situation is stressful, leading to prolonged exposure and therefore and increased physiological stress response.

Where interoception has been assessed using cardiac measures, it has been found to predict a number of affective abilities that are often impaired in clinical conditions, as described in Section 3.1. Interoceptive accuracy is associated, for example, with social anxiety (in anticipation of public speaking; Stevens et al., 2011), emotional lability (Wiens, 2005; Wiens et al., 2000), emotion regulation (Füstös et al., 2012; Kever et al., 2015), emotional memory (Pollatos & Schandry, 2008; Werner et al., 2010), emotional stability (Schandry, 1981), pain perception (Werner et al, 2009), and the intensity of one’s emotional experiences (Füstös et al., 2013; Herbert et al., 2010; Pollatos et al., 2007a, 2007b; Wiens et al., 2000). It is therefore possible that, where these abilities are atypical in clinical populations, interoceptive deficits may contribute to these impairments. While very little research thus far has examined the relationship between interoceptive abilities and emotional deficits in clinical populations, those that have support this hypothesis.

In an autistic sample, self-reported interoceptive attention (assessed by the MAIA) was associated with both self-reported and behaviourally measured empathy, and this relationship was mediated by alexithymia (Mul et al., 2018). This is consistent with the idea that atypical interoception may give rise to difficulties recognising one’s own emotions, which in turn leads to difficulties empathising with others’ emotions, although the correlational design prevents conclusions regarding causality from being drawn. Better metacognitive interoceptive awareness has also been associated with more accurate recognition of others’ vocal emotion in autistic
individuals (Mulcahy et al., 2019). Across a sample of autistic and neurotypical adults, interoceptive trait prediction error (ITPE; the relative difference between z-transformed BPQ scores and accuracy scores, both on the heartbeat counting task and heartbeat discrimination task) was negatively related to self-reported empathy (Garfinkel et al., 2016b). While ITPE was interpreted as indicating over- or under-estimation of one’s interoceptive abilities, this is complicated by the fact that the BPQ may be better characterised as a measure of interoceptive attention than accuracy. In this study, individuals with autism had increased interoceptive attention (BPQ scores) and decreased interoceptive cardiac accuracy, and individuals with this pattern of interoceptive tendencies had reduced self-reported empathy. Reduced interoceptive accuracy (assessed with the heartbeat counting task) and impaired recognition of facial emotion relative to typical controls have also been observed in a sample of individuals with somatoform disorder (Pollatos, Herbert, et al., 2011). Although correlations between these two measures were not reported, emotion recognition was found to correlate with heart rate variability (high frequency normalised units). Finally, individuals with schizophrenia report atypical bodily locations when asked where their emotions are experienced on a body map (Torregrossa et al., 2019). While this study did not utilise an explicit interoceptive measure, this suggests atypical links between internal representations and emotional experience. Therefore, while the relationship between interoception and emotional cognition is under-researched, early evidence does implicate interoception in some of the emotional and social difficulties experienced within clinical populations.

5.2. Interoceptive impairment and learning, reward and decision-making deficits

Section 3.2 describes evidence suggesting a role for interoception in learning and decision-making. This work suggests that interoceptive impairment common to a number of clinical conditions may underlie reports of learning-related impairment in those populations. Learning and decision-making impairments have been observed, for example, in ASD (Friston, Lawson, & Frith, 2013; Johnson, Yechiam, Murphy, Queller, & Stout, 2006; Luke, Clare, Ring, Redley, & Watson, 2011; Pellicano & Burr, 2012), Major Depression (Porter, Gallagher, Thompson, & Young, 2003), schizophrenia
(Waltz, Frank, Robinson, & Gold, 2007), EDs (Brogan, Hevey, & Pignattti, 2010), OCD (Joel et al., 2005), PTSD (Jovanovic, Kazama, Bachevalier, & Davis, 2012), and anxiety (Castaneda, Tuulio-Henriksson, Marttunen, Suvisaari, & Lonnqvist, 2008). Empirical tests of the hypothesis that these impairments are associated with interoception in clinical groups are scarce, although some studies have addressed the issue directly. Furman et al. (2013), for example, found that among patients with depression, those reporting difficulties with decision-making had poorer interoceptive accuracy (measured by the Heartbeat Counting Task) than patients not reporting problems with decision-making, and a healthy control group. Interestingly, Shah and colleagues found that, although interoceptive accuracy was positively correlated with the size of emotional framing effects while making gambling decisions in typical individuals, this relationship was not present in autistic individuals (Shah, Catmur, et al., 2016). It is therefore essential that future work explicitly investigates the role of interoception in multiple learning and decision-making contexts across psychiatric and neurodevelopmental conditions.

Several authors have suggested that interoception is of central importance in drug seeking, maintenance of drug use, and withdrawal behaviours, which are thought to be a product of atypical learning and decision-making (Gray & Critchley, 2007; Naqvi & Bechara, 2009; Paulus et al., 2009; Paulus, 2007; Schmidt, Eulenbruch, Langer, & Banger, 2013; Verdejo-Garcia et al., 2012). While recent models focus on the role of interoception with respect to craving, earlier accounts argued for a role for interoceptive states as occasion-setters, conditioned stimuli, or reinforcers (see Verdejo-Garcia et al., 2012 for a review). According to recent models, interoceptive abilities may govern the extent to which withdrawal-related states of anxiety and panic are perceived, or the extent to which the effect of certain drugs is perceived, and impacts craving (Schmidt et al., 2013). These models are supported by evidence for a role for interoceptive signals in cravings; smokers’ cravings, for example, appear to be reduced by anaesthesia of respiratory airways (Rose, Tashkin, Ertle, Zinser, & Lafer, 1985; Rose, Zinser, Tashkin, Newcomb, & Ertle, 1984).
In their review, Verdejo-Garcia et al. (2012) highlight that several models of addiction already posit a role for interoception. Under classical conditioning models of addiction, whereby drugs and drug-related stimuli act as conditioning stimuli for the hedonic, withdrawal, and craving aspects of drug use and addiction, individuals with high interoceptive abilities may be more likely to form associations, or be less able to ignore cravings, than those with poor perception of interoceptive signals (see also Verdejo-Garcia and Bechara, 2009; Navqi and Bechara, 2010). It has also been argued that typical individuals who are better at predicting interoceptive sensations are better at regulating their cravings for unhealthy food (Kruschwitz et al., 2019; although results may be due to memory for interoceptive states in this study rather than interoceptive predictions).

Another prominent model highlights how the change in addictive behaviour over time may result in part from interoceptive changes (Paulus et al., 2009). Central to this idea is that the reward value of a stimulus (how pleasurable it is perceived to be), is a function of the effect it has on the body’s current internal state, relative to an ideal state (Cabanac, 1971). It is claimed that drug addiction results in alterations of the body’s current internal state, and the representation of the ideal body state, such that the hedonic effects of drugs become less intense, and the craving and withdrawal effects become more intense, resulting in drug use changing from being impulsive to being compulsive. Goldstein et al. (2009) extend this approach, highlighting how interoceptive dysfunction may impair detection and recognition of interoceptive signals, which in turn affects emotional awareness, and potentially explains the tendency for some addicts to deny their addiction. While all of these models assign important roles to interoception, as highlighted by Vardejo-Garcia et al., a number of outstanding questions remain. For example, the mechanism underlying the relationship between interoception and addiction is unclear, especially as both high and low interoceptive abilities have been observed. While it is possible that a quadratic relationship between interoceptive ability and addiction susceptibility exists (whereby low interoceptive accuracy leads to misidentification of substances as more rewarding than is typical, and high interoceptive accuracy and/or attention lead to stronger craving sensations, for example), it may also be the case that
interoceptive abilities change, or play different roles, across the course of addictions. For example, whilst poor interoceptive accuracy has been implicated in the initial development of addiction, it has also been suggested that high interoceptive accuracy is involved in the maintenance of addiction (Verdejo-Garcia et al., 2012). It is also likely that the relationship between interoception and addiction varies depending on whether attention or accuracy is measured. Therefore, it is important for future research to consider whether it is the attention towards interoceptive signals, the recognition of such signals, or their regulation, that is of most importance for addiction, as well as the role of interoception in the vulnerability for, development of, and maintenance of addiction.

Research into the impact of insula lesions provides evidence consistent with the hypothesis that interoception underlies craving in addiction and learning-related abnormalities in other disorders. It should be acknowledged that the fact that interoception relies on the insula, and that the insula is damaged in these studies, does not necessarily mean that interoception itself is impaired. However, the studies may be informative as to the impact of atypical interoception. The seminal finding in the human literature was that of Navqi and colleagues (Naqvi, Rudrauf, Damasio, & Bechara, 2007), who demonstrated that insula lesion patients found it easier to stop smoking, reporting that they no longer experienced cravings, or that cravings were greatly reduced following their lesion. This finding was replicated in prospective studies by Suñer-Soler et al. (2012) and Gaznick and colleagues (Gaznick, Tranel, McNutt, & Bechara, 2014), although not in a shorter prospective study (Bienkowski, Zatorski, Baranowska, Ryglewicz, & Sienkiewicz-Jarosz, 2010). Similarly, inactivation of the insula in rats reduced drug-seeking behaviour (Contreras, Ceric, & Torrealba, 2007). That the insula is the neural basis for reward and cravings (Paulus, 2007) is further supported by the results of several imaging studies, which demonstrate that insula activity is correlated with craving-associated stimuli (Brody et al., 2002; Dom, Sabbe, Hulstijn, & Van Den Brink, 2005; Kilts et al., 2001; Kilts, Gross, Ely, & Drexler, 2004; Naqvi & Bechara, 2009), and craving intensity (G.-J. Wang et al., 1999). Similarly, insula and anterior cingulate activation have been found to be atypical in individuals with addiction disorders during decision-making and risk-
taking (Gowin et al., 2014; Hoffman, Schwartz, Meiri, Stevens, & Mitchell, 2008; London et al., 2005; Nestor, Ghahremani, Monterosso, & London, 2011; Paulus, Hozack, Frank, Brown, & Schuckit, 2003). Animal studies also support the role of the insula in addiction. Contreras et al. (2007), for example, demonstrated that inactivation of the insula in rats interfered with a previously learned place preference associated with amphetamine delivery. Interestingly, insula lesions are also linked to a reduction in the cognitive distortions linked to pathological gambling. While typical individuals exhibit an increase in gambling motivation in response to near-misses, and overestimate the frequency of one of two equally probable outcomes after a run of the alternate outcome, these effects are increased in those with problematic gambling behaviour (Fortune & Goodie, 2012; Michalczuk, Bowden-Jones, Verdejo-Garcia, & Clark, 2011). Patients with insula lesions, however, do not show these gambling-related cognitive distortions (Clark, Studer, Bruss, Tranel, & Bechara, 2014). This effect is consistent with neuroimaging work identifying insula involvement in near-miss situations (Akitsuki, Sugiura, Watanabe, Yamashita, & Sassa, 2003; Clark, Lawrence, Astley-jones, & Gray, 2009; Shao, Read, Behrens, & Rogers, 2013; Xue et al., 2010), and the involvement of the insula in risky decision-making reported in Section 3.2.

As argued in Section 3.1.1, we suggest that alexithymia may be indicative of a general failure of interoception, and there is an increased prevalence of alexithymia in almost all psychiatric conditions (see Section 4.1.1). Alexithymia research may therefore add to our understanding of the relationship between interoception and reward and decision-making, where addictive behaviours are concerned. In individuals with alcohol addiction, for example, alexithymia is positively associated with duration of alcohol use and scores on the Michigan Alcoholism Screening Test (Uzun, 2003), and use severity (Bruce, Curren, & Williams, 2012; Thorberg et al., 2009), including severity of alcohol-related behaviours following treatment (Cleland, Magura, Foote, Rosenblum, & Kosanke, 2005). Similarly, alexithymia predicts faster relapse, and lower levels of engagement with treatment in individuals with addiction disorders (Cleland et al., 2005; Stasiewicz et al., 2012). Further, Speranza and colleagues observed that, in individuals with Eating Disorders or Addictive Disorders,
high degrees of alexithymia predicted high levels of depression, which in turn predicted addictive
corporations (Speranza, Corcos, Ste, Halfon, & Jeammet, 2004). Alexithymia is also positively
related to craving in alcoholics (Junghanns et al., 2005), smokers (Sutherland, Carroll, Salmeron,
Ross, & Stein, 2013) and methamphetamine addicts (Saladin et al., 2012). This may be due to
increased interoceptive attention in alexithymia (Betka et al., 2018; Christensen et al., 2018; Ernst et
al., 2014; Longarzo et al., 2015; Zamariola et al., 2018), although findings have been mixed.

A small but growing literature exists on the impact of alexithymia on more explicit learning
and decision-making. Individuals with alexithymia, for example, have been found to make riskier
choices on the Iowa gambling task than those with low alexithymia levels, especially when less
information was available concerning their previous cumulative performance (Ferguson et al.,
2009). Kano and colleagues conducted a PET study and found that those with alexithymia made less
advantageous choices on the same task, and exhibited reduced interoceptive cortex (insula and
ACC) activation during the task (Kano et al., 2011). Alexithymia also predicted performance on a
version of the Iowa Gambling Task in pathological gamblers (Aïte et al., 2014), although not in a
sample of patients with de novo Parkinson’s Disease (Poletti et al., 2011). In Shah et al’s (2016)
study reported above, higher alexithymia was associated with smaller emotional framing effects in a
gambling task, but this was only observed in the typical, and not the ASD, sample. Recent evidence
also suggests that individuals with alexithymia are more likely that those without alexithymia to
select immediate smaller rewards relative to delayed larger rewards, although this tendency was also
associated with higher interoceptive accuracy, making interpretation of the underlying mechanisms
difficult (Scarpazza, Sellitto, & Pellegrino, 2017). Finally, alexithymia appears to be associated with
weaker fear conditioning, indicated by lower skin conductance in response to conditioned stimuli,
and faster extinction of this response in those with high, relative to those with medium and low,
levels of alexithymia (Starita, Làdavas, & di Pellegrino, 2016). While the literature on learning and
decision-making in alexithymia is highly heterogeneous in terms of methodology, evidence thus far
implicates alexithymia in difficulties with these processes, potentially due to atypical interoception.
5.3. Other domains potentially affected by interoceptive impairment

Sections 3.5.1 and 3.5.2 have described the way in which atypical interoception is associated with common impairments in clinical populations, with a focus on the emotional domain, and learning and decision-making. Notably, however, a number of other symptoms frequently observed in clinical groups may also be explained by interoceptive impairment. These are addressed in the following sections.

5.3.1 Sensory processing

Atypical sensory processing may be explained by interoceptive impairment, and is particularly relevant to ASD. Atypical sensory responsivity to the environment is now included as a diagnostic criterion for ASD in the DSM-5, and individuals with ASD often experience these sensory atypicalities (Ben-Sasson et al., 2009; Bettelheim, 1967; Crane, Goddard, & Pring, 2009; Green, Chandler, Charman, Simonoff, & Baird, 2016; Kanner, 1943; Lane, Molloy, & Bishop, 2014; Lane, Young, Baker, & Angle, 2010; Leekam, Nieto, Libby, Wing, & Gould, 2007; Marco, Hinkley, Hill, & Nagarajan, 2011; Tomchek & Dunn, 2007; Zachor & Ben-Itzchak, 2014). Many of these sensory domains can be considered interoceptive; ASD is associated with hyper- or hypo-sensitivity to taste, smell, and tactile stimulation (of which affective touch is interoceptive, as it is carried by interoceptive C fibres (Vallbo, Olausson, & Wessberg, 1999; Wessberg, Olausson, Fernström, & Vallbo, 2003)); see DuBois, Ameis, Lai, Casanova, & Desarkar (2016) and Hatfield et al. (2017) for reviews. Atypical processing of affective touch may also be associated with hyper- or hypo-sensitivity to particular fabrics or textures. Findings concerning sensitivity to tactile stimuli have been relatively inconsistent across ASD samples (Cascio et al., 2008; Tanidir, Mukaddes, & Gu, 2007; Tavassoli et al., 2006), in line with the suggestion that they arise due to impaired interoception (present in some but not all individuals with ASD), rather than ASD itself. Those with ASD also sometimes exhibit hypersensitivity to exteroceptive signals, such as visual and auditory stimuli (Green et al., 2016; Lane et al., 2014, 2010; Leekam et al., 2007; Marco et al., 2011; Tomchek & Dunn, 2007). While these stimuli are not interoceptive, it is possible that poor
interoception in some individuals with ASD leads to relative increases in the salience of exteroceptive stimuli, and therefore hypersensitivity. Alternatively, it may be the case that increased arousal in response to salient exteroceptive stimuli, such as intense light stimuli, may be incorrectly classified as interoceptive signals, such as pain.

It is possible that sensory atypicalities contribute to some of the core symptoms of ASD (Quattrocki & Friston, 2014), as sensory atypicalities have been associated with restricted and repetitive behaviours (Chen, Rodgers, & McConachie, 2009; Foss-Feig, Heacock, & Cascio, 2012) and need for sameness (Wigham, Rodgers, & South, 2015), as well as social and communication deficits (Foss-Feig et al., 2012). Similarly, Ben-Sasson et al. (2009), found that sensory responsiveness is associated with overall ASD symptom severity. Indeed, a recent study specifically investigated the relationship between interoceptive abilities and core features of ASD. Palser, Fotopoulou, Pellicano and Kilner (2019) found that social affective difficulties in autistic children were related to overconfidence in the heartbeat discrimination task, while restricted and repetitive behaviours were associated with higher heartbeat discrimination accuracy. This latter result may be driven by task features, such as the requirement to attend to multiple, repeated stimuli in the heartbeat discrimination task, which may be linked to restricted interests and repetitive behaviours. Alternatively, these findings may indicate a complex relationship between restricted interests/repetitive behaviours and the balance between interoceptive and exteroceptive attention or accuracy, and the extent to which these cues are integrated. Relatedly, autism appears to affect HCT performance differently at different intervals, with autistic children numerically outperforming neurotypical children at long intervals (Schauder, Mash, Bryant, & Cascio, 2014). Autistic traits may, therefore, be linked to particularly high or low interoceptive abilities in different task contexts. Taken together, these studies suggest that interoception and sensory atypicalities may link specifically to core autistic symptoms. As sensory atypicalities are not universal in those with ASD, however, interoceptive deficits likely represent one of many potential risk factors for ASD traits.
Although sensory hyper- and hypo-responsiveness are more common in ASD than in other neurodevelopmental disorders (Baranek, David, Poe, Stone, & Watson, 2006; Ben-Sasson et al., 2009; Leekam et al., 2007), it is worth noting that sensory atypicalities have also been observed in individuals with intellectual disability and other neurodevelopmental disorders (Green et al., 2003, 2016; Lane, Reynolds, & Dumenci, 2012; Tomchek & Dunn, 2007; Watling, Deitz, & White, 2001), with a systematic review suggesting that atypical olfaction may be a biomarker for a range of psychiatric disorders (Schecklmann et al., 2013). Consistent with the hypothesis that poor interoception is associated with a range of clinical symptoms, it is possible that interoceptive impairment contributes to atypical sensory processing across multiple clinical groups.

5.3.2. Pain sensitivity

Beyond sensory responsiveness in developmental disorders, hyper- and hypo-sensitivity to pain have been reported in a number of clinical populations, such as depression (Hermesdorf et al., 2016), schizophrenia (Stubbs et al., 2015), eating disorders (Papezová, Yamamotová, & Uher, 2005; Yamamotova, Bulant, Bocek, & Papezova, 2017), autism (Riquelme, Hatem, & Montoya, 2016), and OCD (Hezel, Riemann, & McNally, 2012). As pain is, by most contemporary definitions (Craig, 2002; Khalsa et al., 2018), an interoceptive signal, these atypicalities are unsurprising within the context of our proposal. It is likely that widespread interoceptive impairments across the clinical population contribute to atypical perception and recognition of pain in multiple psychiatric conditions. While pain sensitivity in individuals with chronic pain conditions has been associated with alexithymia (Glaros & Lumley, 2005; Lumley, Smith, & Longo, 2002), little research has investigated the relationship between pain hyper- and hypo-sensitivity and other interoceptive difficulties in these clinical groups. In individuals with medically unexplained symptoms (painful sensations in the absence of an organic underlying cause), cardiac interoceptive accuracy has been negatively associated with symptom severity (Schaefer et al., 2012), and symptoms appear to improve following cardiac perception training (Schaefer, Egloff, Gerlach, & Witthöft, 2014). This
population has also been hypothesised to exhibit increased interoceptive attention (Barsky, 1992), although this hypothesis has not been empirically tested.

5.3.3. Repetitive behaviours and compulsions

Repetitive behaviours, which are a core diagnostic feature of both ASD and OCD, are also common in Eating Disorders (e.g. repeated bingeing and/or purging behaviours) and addiction disorders (repeated addictive behaviours), and may be explained by atypical interoception. It is possible that difficulties perceiving or recognising the aversive consequences of maladaptive behaviours, and associating these consequences with the action, contribute to their maintenance. As behaviours are frequently learnt through conditioning processes, atypical processing of reward and punishment signals that are associated with a given behaviour means that those with poor interoception will be slower or less able to learn to repeat rewarded behaviours, and avoid punished behaviours. In the case of compulsions, it is possible that individuals with poor interoception misattribute the source of negative internal states, and experience an urge to perform compulsions in an attempt to reduce these signals. If, for example, one is preoccupied with cleanliness and contamination, one may classify any internal state of anxiety as arising due to cleanliness being compromised, and compulsively clean in order to reduce the aversive internal state. Forming inaccurate associations between internal states and their external triggers may, therefore, maintain compulsive behaviours. Equally, it is possible that these behaviours are explained by reduced recognition of the reward signals typically associated with novelty (e.g. Bardo, Donohew, & Harrington, 1996; Pierce, Crawford, Nonneman, Mattingly, & Bardo, 1990), causing the individual to engage in more predictable, and less novel, familiar, behaviours. Compulsions and repetitive behaviours may also arise owing to difficulties suppressing urges, which may be perceived as interoceptive signals themselves. Indeed, suppression of natural urges such as breathing and voiding involves interoceptive cortex activation (Banzett et al., 2000; Kuhtz-Buschbeck et al., 2005; Seseke et al., 2006). Multiple disorders are associated with difficulties suppressing urges, such as OCD, tics and
Tourette syndrome, hyperactivity and impulse control disorders, and addictions, and these conditions have been associated with atypical interoception, as described above. Interestingly, the premonitory urge to tic in individuals with Tourette Syndrome has been found to be positively associated with cardiac interoceptive accuracy and self-reported interoceptive attention (Ganos & Garrido, 2015; Rae, Larsson, Garfinkel, & Critchley, 2019). While the specific nature of the relationship may differ across populations, interoceptive atypicalities may account for repetitive behaviours, compulsions, and difficulties suppressing urges across a range of clinical conditions.

5.3.4. Self-other distinction

Recent theoretical work has suggested that interoception (in combination with exteroception; see Park & Blanke, 2019; Park & Tallon-Baudry, 2014) is crucial for developing an accurate representation of the self (Seth, 2013; Quattrocki and Friston, 2015). The implication that one’s own interoceptive experiences are being experienced by another body, for example, has been found to produce illusory distortion of one’s self-representation (e.g. Aspell et al., 2013; Botvinick & Cohen, 1998). Deficits in self-other distinction (the ability to represent the self as a distinct individual, separate from others) are observed in some clinical groups, and may again be explained by interoceptive deficits. The Rubber Hand Illusion of body ownership, for example, in which individuals experience a rubber hand as their own following stroking of their own hand while viewing simultaneous stroking of the rubber hand, is weaker in those with high levels of interoception, arguably due to more accurate representations of the self (Suzuki, Garfinkel, Critchley, & Seth, 2013; Tsakiris, Tajadura-Jiménez, & Costantini, 2011; although see Crucianelli, Krahé, Jenkinson, & Fotopoulou, 2018). Similarly, representations of one’s own face appear to be more malleable (assessed by the enfacement illusion; Sforza, Bufalari, Haggard, & Aglioti, 2010) in those with poor interoceptive accuracy (Tajadura-Jiménez & Tsakiris, 2014). Furthermore, lesions to the right posterior insula (an area associated with attention to interoceptive sensations; Schulz, 2016) have been associated with somatoparaphrenia (in which one’s limbs or other body parts are
not perceived as belonging to oneself). Taken together, these findings suggest that individual differences in interoception are associated with those in self-other distinction.

Where clinical groups are concerned, a reduced sense of self and distorted boundaries between representations of the self and others have long been implicated in, and are core features of, disorders such as schizophrenia (van der Weiden, Prikken, & van Haren, 2015) and depersonalisation/derealisation disorder (American Psychiatric Association, 2013). A number of other psychological conditions have also been associated with atypical representations of self and other, however, and it is possible that interoceptive differences account for these atypicalities and a range of related symptoms/behaviours across various clinical groups. It is possible that interoception-related atypicalities in self-other distinction, for example, explain the delusions of control and auditory hallucinations experienced by some individuals with schizophrenia, bipolar disorder, and depersonalisation/derealisation disorder (American Psychiatric Association, 2013). Delusions and hallucinations such as these are likely to arise owing to poor self-other distinction, such that one’s own behavioural commands are not represented accurately, leading to reduced precision of predictions concerning the consequences of these motor commands, and the perception that these behaviours were generated by an external agent (Frith, 2012; Seth, Suzuki, & Critchley, 2012). Evidence suggests, for example, that individuals with auditory hallucinations and passivity experiences perceive self-generated tactile sensations to be as intense as other-generated sensations, while self-generated stimuli are perceived as less intense than other-generated sensations by control participants (Blakemore, Smith, Steel, Johnstone, & Frith, 2000), likely due to reduced attenuation of sensory information, caused by reduced precision of predictions based on motor efferents. Similarly, those with schizophrenia experience the Rubber Hand Illusion more strongly than those without (Peled, Ritsner, Hirschmann, Geva, & Modai, 2000), and exhibit atypical somatosensory evoked potentials during this illusion (Peled, Pressman, Geva, & Modai, 2003), although see Cornielle & Lush's (2021) recent discussion of the potential role of demand characteristics in this
task. Recognition of images of the self also appears to be impaired in schizophrenia (Ferroni et al., 2019; Sandsten et al., 2020).

Further, contemporary models proposing that accurate interoception is crucial for representing the self as distinct from others (Quattrocki & Friston, 2014; Seth, 2013), suggest that poor interoception would lead to unintended mirroring of others, for example echopraxia and echolalia (Quattrocki & Friston, 2014). While evidence has been mixed, many have argued for atypical imitation in individuals with ASD (see Hamilton, 2013 for a review). Recent evidence suggests that autistic individuals have a tendency to over-imitate (Sowden, Koehne, Catmur, & Dziobek, 2015), in line with increased prevalence of echolalia and echopraxia in this population (Baltaxe & Simmons III, 1975; Cunningham & Dixon, 1961; Grossi, Marcone, Cinquegrana, & Gallucci, 2013; Jordan, 1993; Kanner, 1946; Koegel, Rincover, & Egel, 1982). Consistent with the high rates of alexithymia and interoceptive impairment in ASD, a relationship has been observed between both interoceptive accuracy (Ainley et al., 2014) and alexithymia (Sowden, Brewer, Catmur, & Bird, 2016), and the ability to inhibit automatic imitation in typical individuals. Poor interoception, and in turn poor self-other distinction may, therefore, explain atypical imitation in those with ASD. Notably, echolalia and echopraxia have also been observed in a range of clinical groups, including those with aphasia (Grossi et al., 1991; Trojano, Fragassi, Postiglione, & Grossi, 1988), Tourette syndrome (Eisenberg, Ascher, & Kanner, 1959; Michael, 1957) and intellectual disability (Critchley, 1964), and children who later develop schizophrenia (Bearden et al., 2000).

**Beyond echolalia and echopraxia**, reduced distinction between the self and other may also be responsible for the increased levels of emotion contagion and personal distress often observed in clinical groups (over-imitation in the affective domain). Emotion contagion (sharing another’s emotion) often follows observation of an emotional state in another, and can lead to personal distress if this ‘mirrored’ internal state is negative. Intact ability to distinguish between the self and other then allows the observer to attribute the emotional state to the observed individual, thereby reducing personal distress, and giving rise to empathic concern (Bird & Viding, 2014; Coll et al.,
Interestingly, personal distress (a marker of emotion contagion) is often increased in clinical conditions, while other facets of empathy (e.g., empathic concern) are often decreased in these groups. Self-reported personal distress has been found to be atypically high, for example, in those with eating disorders (Beadle, Paradiso, Salerno, & McCormick, 2013; Duchesne et al., 2012; Gutman & Laporte, 2002), schizophrenia (Bonfils et al., 2017), Tourette syndrome (Eddy, Macerollo, Martino, & Cavanna, 2015), OCD (Kang, Namkoong, Yoo, Jhung, & Kim, 2012), and depression (Banzhaf et al., 2018). As interoception is likely to play a role in self-other distinction, increased personal distress (due to decreased self-other distinction) across multiple clinical populations may arise owing to interoceptive difficulties in these groups.

5.3.5. Further social impairments

Social impairments (beyond emotion recognition and empathy difficulties) are common in many clinical populations, with ASD being the most notable; difficulties with social interaction and communication are a core diagnostic feature of ASD (American Psychiatric Association, 2013). One social impairment that has been consistently documented in those with ASD is reduced attention to social stimuli, such as faces (Dawson, Webb, & McPartland, 2005; Klin, Jones, Schultz, Volkmar, & Cohen, 2002; Nakano et al., 2010; Riby & Hancock, 2008). Quattrocki and Friston (2014) attributed this to interoceptive impairment, suggesting that if one struggles to perceive the interoceptive cues associated with a face, such as warmth and satiety during infancy, one will not learn the affective value of faces, and therefore experience reduced motivation to attend to them. Further, direct eye gaze has been described as inducing a range of negative internal states, such as pain, dizziness, headaches, nausea, and overheating (Trevisan, Roberts, Lin, & Birmingham, 2017), which may be due to misclassification of these associated internal cues. Decreased motivation to interact with others is also likely to exacerbate the communication difficulties experienced by those with ASD, via reduced opportunity for receiving social feedback on one’s communication.
A further social difficulty that is often observed in ASD is Theory of Mind (ToM) or mentalising, referring to the ability to infer another’s mental states (e.g. thoughts, feelings, and beliefs) (Brunsdon & Happé, 2014; Frith, Morton, & Leslie, 1991). Ondobaka and colleagues proposed that one’s understanding of one’s own internal states is likely to assist in inferring others’ internal states, which in turn may assist with determining their mental state (e.g. the observed individual is cold, therefore they want to go indoors) (Ondobaka, Kilner, & Friston, 2017). Empirical evidence relating ToM abilities to interoceptive abilities is scarce, but Shah, Catmur, and Bird (2017) tested this hypothesis in typical individuals using the Movie for Assessment of Social Cognition, a commonly utilised ToM task. While they found that cardiac interoceptive accuracy was positively associated with performance on items relating to others’ emotional states, it was unrelated to performance on ToM items that did not require emotional inference. Notably, this task does not include items that require inference of non-emotional internal states (mental states to be inferred were either emotional or unrelated to interoceptive signals). Further evidence relating to the role that interoception may play in ToM is clearly required, but it is possible that interoceptive abilities are involved in inference of others’ states where interoceptive or emotional signals are involved, but not when mental states do not relate to internal (including emotional) signals. While this theory has not been tested in clinical groups (beyond emotion recognition), interoceptive difficulties may explain some of the difficulties with inference of others’ internal states in a range of populations.

Beyond ASD, social difficulties have been observed in a number of psychological disorders, such as eating disorders (Cardi, Tchanturia, & Treasure, 2018), schizophrenia (Green, Horan, & Lee, 2015), bipolar disorders (Wolf, Brüne, & Assion, 2010), and depression (Romera et al., 2010). These may stem, at least in part, from the emotional impairments that are common in these populations. In individuals with Eating Disorders, for example, alexithymia predicts decreased self-reported social skills (Beales & Dolton, 2000). More research is required, however, into the nature of the relationship between interoception, emotional abilities, and social abilities outside of the emotional domain, especially relating to the causal nature of relationships across disorder.
populations. If interoceptive impairment follows development of a given disorder (as may be the case with FTD and AIDS, for example), it remains to be determined whether the disorder causes both interoceptive difficulties and social difficulties, or whether interoceptive impairment leads in turn to further emotional and social deficits. Similarly, it is unclear whether interoceptive impairment contributes directly to the development of both psychopathology and emotional deficits, or whether interoception leads to further emotional and social difficulties, and in combination these impairments lead to disorder development. It is likely that there are multidirectional relationships between interoception, psychopathology and socio-emotional abilities, but these may vary according to clinical diagnosis, making investigation of these associations across populations and across time necessary.

5.3.6. Eating behaviours

One area of psychopathology in which the role of interoception is clear is atypical eating behaviours. Clearly, difficulties detecting signals of hunger and satiety, and distinguishing them from other internal cues, would lead to difficulties determining appropriate quantities of food (Bruch, 1973). This is consistent with Herbert, Blechert, Hautzinger, Matthias, & Herbert's (2013) finding that those with higher interoceptive accuracy show more intuitive eating, an adaptive form of eating in response to internal cues of hunger and satiety. While the relationship between reduced or increased eating and perception or recognition of hunger and satiety cues is intuitive, atypical eating also refers to behaviours such as avoidance of particular tastes, flavours, and textures, and abnormal eating rituals and routines. While eating atypicalities are most strongly associated with Feeding and Eating Disorders (American Psychiatric Association, 2013), they are also common in individuals with ASD, (Martins, Young, & Robson, 2008; Schreck, Williams, & Smith, 2004), and depression, where changes in weight and appetite are a diagnostic criterion (American Psychiatric Society, 2013). Notably, while ASD and depression frequently co-occur with EDs (Marucci et al., 2018), many individuals with ASD or depression without co-occurring EDs still exhibit atypical eating behaviours that do not necessarily stem from weight and shape concerns or difficulties
processing hunger or satiety, such as dislike of particular food textures or flavours, and eating rituals in ASD (Sharp et al., 2013; Spek, van Rijnsoever, van Laarhoven, & Kiep, 2019), or appetite changes in depression, potentially linked to reduced reward (Coccurello, 2019). Reward-processing models of EDs have also been proposed; whereby food may be less rewarding for individuals who restrict calorific intake, but strongly associated with reward in those who binge eat (see Monteleone et al., 2018 for a review). Relatedly, atypical eating behaviour has also been related to emotional difficulties; emotional overeating is characterised by eating in response to negative emotions, potentially due to poor distinction between internal cues of hunger and of other states including emotions (Koch & Pollatos, 2014). Indeed, Koch and Pollatos (2014) found that emotional overeating is associated with interoceptive accuracy in overweight children, although in this longitudinal study emotional overeating predicted cardiac accuracy a year later, rather than vice versa, suggesting a complex and potentially bidirectional relationship. Importantly, taste, smell, and reward cues, can be seen as interoceptive signals, so the relationship between interoception and eating may be driven by processing of these cues, in addition to hunger, thirst and satiety signals.
5.3.7. Physical health

As described above, much research suggests a link between interoception and mental health. While physical health has received less attention in this field, interoception may play a role in the development of physical health conditions. Poor interoceptive accuracy has been observed, for example, in individuals with asthma (Khosravani, Alvani, Sharifi Bastan, Jamaati Ardakani, & Akbari, 2016), diabetes (Pauli, Hartl, Marquardt, Stalmann, & Srian, 1991), and congenital heart disease (Rietveld, Karsdorp, & Mulder, 2004). Heartbeat evoked potentials following resuscitation after cardiac arrest have also been found to be larger in survivors than non-survivors at 6 month follow-up (Schulz et al., 2018), and correlate with cardiac responses in patients with cardiac dysfunction (Gray et al., 2007). While interoception has not been specifically examined in a wide range of physical illnesses, the fact that high levels of alexithymia have also been observed in conditions such as human immunodeficiency virus (HIV) (McIntosh et al., 2014), diabetes (Mnif et al., 2014; Topsever et al., 2006), multiple sclerosis (Bodini et al., 2008; Chahraoui et al., 2014; Prochnow et al., 2011), and cancer (although evidence is limited; De Vries, Forni, Voellinger, & Stiefel, 2012) raises the possibility that interoceptive deficits may be common across a range of conditions. Poor interoceptive accuracy has also been linked to being overweight (Herbert & Pollatos, 2014). Interoception also appears to be associated with health-related decision-making, such as intuitive eating (Herbert et al., 2013) and regulation of physical load during activity (Herbert et al., 2007).

Research is lacking into whether interoception explains physical health problems that are common in those with psychological conditions, but several conditions warrant investigation. In particular, gastrointestinal difficulties are often observed in ASD (Klukowski, Wasilewska, & Lebensztejn, 2015; Wakefield et al., 2000; White, 2003; Williams et al., 2011) and irritable bowel syndrome commonly co-occurs with anxiety and depression (Fond et al., 2014; Tosic-Golubovic, Miljkovic, Nagorni, Lazarevic, & Nikolic, 2010). It is possible that a variety of physical health symptoms arise due to interoceptive difficulties in clinical populations.
5.3.8. Sleep

Disrupted sleep is commonly reported across a variety of psychiatric conditions, such as ASD (Elrod & Hood, 2015; Klukowski et al., 2015), depression (Alvaro, Roberts, Clinical, & Harris, 2013; Shanahan, Copeland, Angold, Bondy, & Costello, 2014), anxiety (Alvaro et al., 2013; Shanahan et al., 2014), OCD (Díaz-Román, Perestelo-Pérez, & Buela-Casal, 2015), PTSD (Kobayashi, Boarts, & Delahanty, 2007), and schizophrenia (Chan, Chung, Yung, & Yeung, 2017).

It is likely that interoception plays a role in sleep, as the vagus nerve (a core pathway for the transmission of interoceptive information; Craig, 2002, 2009) plays a key role in sleep and wakefulness (Leichnetz, 1972; Peñaloza Rojas, 1964), and sleep is accompanied by changes in bodily physiology, such as heart rate, blood pressure, and temperature (see Orem & Barnes, 2012). Whilst a new proposal, recent evidence supports a link between poor interoception and poor self-reported sleep quality across disorder groups, but not in the non-clinical control group (Ewing et al., 2017). Interestingly, in the non-clinical group, a positive association was observed between reported sleep difficulties and interoceptive accuracy, in line with findings that insomnia is associated with greater heartbeat evoked potential amplitudes (Wei, Ramautar, Colombo, Stoffers, & Gómez-Herrero, 2016). Furthermore, poor sleep quality is observed across multiple psychiatric conditions (Freeman et al., 2017) including depression, anxiety and alexithymia (e.g. Bauermann, Parker, & Taylor, 2008; Bazydlo, Lumley, & Roehrs, 2001; Engin, Keski, Dulgerler, & Bilge, 2010; Papadimitriou & Linkowski, 2005; Tsuno, Besset, & Ritchie, 2005). More recent evidence suggests that both depression and alexithymia uniquely contribute towards poor self-reported sleep quality in population samples even after controlling for self-reported anxiety symptoms (Murphy, Wulff, Catmur, & Bird, 2018). However, the causal relationship between interoception and sleep quality remains unclear; whilst it is possible that poor interoception results in risk of disrupted sleep, it is equally possible that poor sleep quality impacts negatively upon interoception (Ewing et al., 2017; Harshaw, 2015).
6. What gives rise to poor interoception? The Development of Typical Interoception and How it May Go Awry

The evidence that interoception plays a fundamental role in a range of abilities has prompted interest in the stability of interoception across development, and how early interoceptive deficits may contribute to later difficulties. Converging evidence from measures of interoception, as well as alexithymia, indicate variability across the lifespan that coincides with established risk periods for the development of psychopathology and changes in higher order cognition. In particular, fluctuations in interoception may be especially common in adolescence and in late adulthood (see Murphy et al., 2017 for a review).

While interoceptive abilities appear to develop early in life (Maister, Tang, & Tsakiris, 2017), it is thus far unclear how this developmental process occurs. Further work is required, for example, into understanding how infants learn to distinguish multiple (and constantly changing) internal signals from each other, and how they learn to associate particular signals or combinations of signals with appropriate verbal labels. It is likely that caregivers assist infants in this process by labelling internal states using contextual cues. Children’s behaviour, including infant cries, after a period without food or sleep, for example, may be labelled by caregivers as indicative of feeling hungry and tired, respectively. Learnt associations between internal states and these verbal labels may then allow the child to form representations of hunger and tiredness as distinct internal states. Difficulties perceiving internal cues would clearly hamper this process, as little interoceptive information would be available to be associated with verbal labels. Early difficulties perceiving interoceptive changes are therefore likely to exacerbate interoceptive difficulties later in life, as the individual would have less clearly defined categorical representations of internal states. Notably, if it is the case that linguistic input plays a role in the development of internal state representations, as has been proposed for emotional states (Barrett, 2006), interoceptive difficulties may also emerge in childhood following typical perception of internal signals during infancy. If, for example, caregivers are less able to correctly identify the child’s internal states, or provide less verbal information about
that state, the child would have less opportunity to refine their representation of that state, leading to intergenerational transfer of interoceptive impairment and alexithymia. Similarly, if an individual has intact interoceptive processing but limited verbal ability, they may not benefit from others providing verbal labels for their internal states (see Hobson et al. (2019) and (Way, Yelsma, van Meter, & Black-Pond, 2007) for reviews of the role of language abilities in the development of interoceptive difficulties and alexithymia). While the relationship between parent and child interoception has not been explicitly investigated, parental anterior insula activation while watching video recordings of parent-infant interactions (their child and others’ children) has been found to predict their reports of their child’s somatic symptoms six years later, suggesting that parent interoception may be associated with child interoception (Abraham, Hendler, Zagoory-sharon, & Feldman, 2019).

As in adulthood, interoception during childhood has predominantly been examined in the cardiac domain. Echoing findings from adulthood, evidence supports the association between interoceptive accuracy in childhood and physical and mental health (Eley, Stirling, Ehlers, Gregory, & Clark, 2004; Georgiou et al., 2015) and emotional abilities (Schaan et al., 2019). These findings are supported by a similar relationship between alexithymia and mental health outcomes in children (Housiaux, Luminet, Broeck, & Dorchy, 2010; Rieffe, Oosterveld, & Terwogt, 2006). Interoceptive ability in 6-11 year olds, however, appears equivalent to that in adults (Koch & Pollatos, 2014), indicating that if interoception develops over time, it may do so during early childhood years.

Whilst the factors that contribute towards individual differences in the development and stability of interoceptive ability across childhood remain unclear, twin studies of cardiac interoception in childhood indicate that a substantial portion of the variance is accounted for by nonshared environmental factors (Eley, Gregory, Clark, & Ehlers, 2007), a similar heritability pattern to that observed for alexithymia in adulthood (Jørgensen, Zachariae, Skytthe, & Kyvik, 2007). Indeed, more recent evidence suggests that a substantial portion of stability in interoceptive accuracy between the ages of 8 and 10 years is also accounted for by non-shared environmental factors.
(factors that make individuals within a family different). Whilst low stability was observed in this study, changes in interoception between the ages of 8 to 10 years were generally driven by improvements in interoceptive accuracy with age (Murphy, Cheesman, et al., 2019). Whilst the non-shared environmental factors that contribute to individual differences, stability and change in interoceptive accuracy have not been systematically identified, evidence that childhood adversity is associated with poor mental health (Kaplow & Widom, 2007; Macmillan et al., 2001; Repetti, Taylor, & Seeman, 2002), atypical interoceptive abilities (Schaan et al., 2019), and structural changes in interoceptive regions (Teicher, Anderson, Ohashi, & Polcari, 2014) suggests that early negative experiences may contribute towards atypical interoception (see Murphy et al., 2017).

Indeed, increased rates of alexithymia in adulthood have been associated with a number of negative experiences in childhood, such as physical or sexual abuse, emotional neglect, and poor family expressiveness (Aust, Härtwig, Heuser, & Bajbouj, 2013; Berenbaum, 1996; Kench & Irwin, 2000).

While adolescence (the period of transition from childhood to adulthood) is associated with a number of positive changes, such as the emergence of higher-order cognitive abilities (e.g., Blakemore & Choudhury, 2006), this developmental period is also associated with increased risk taking, accidents, and susceptibility to psychopathology (Crone, van Duijvenvoorde, & Peper, 2016; Giedd, Keshavan, & Paus, 2008; Kessler, Berglund, et al., 2005; Steinberg, 2007). Consistent with the proposal that both atypical learning and decision-making and the emergence of psychopathology are underpinned by interoceptive impairment, adolescence is associated with elevated rates of alexithymia (e.g. Gatta et al., 2014; Säkkinen, Kaltiala-Heino, Ranta, Haataja, & Joukamaa, 2007, although see Honkalampi et al., 2009; Joukamaa et al., 2007), particularly at the earlier stages of this period (Gatta et al., 2014; Karukivi, Pölönen, Vahlberg, Saikkonen, & Saarijärvi, 2014; Säkkinen et al., 2007). This spike in alexithymia prevalence in adolescence is consistent with the idea that interoceptive impairment represents a risk factor for, or is causally associated with, the development of psychopathology during this period of development (Murphy, Brewer, Catmur & Bird, 2017; Murphy, Viding, & Bird, 2019).
Whilst few studies have explicitly investigated interoception during this period, interoception appears to be related to a range of factors such as emotional abilities (Georgiou et al., 2018), body image (Todd, Aspell, Barron, & Swami, 2019), and health in adolescence, as in other stages of development. Atypical insula activation to interoceptive signals, for example, has been reported in adolescents with obesity (Delgado-Rico, Soriano-Mas, Verdejo-Román, Río-Valle, & Verdejo-García, 2013; Mata, Verdejo-Roman, Soriano-Mas, & Verdejo-Garcia, 2015) and those with substance use disorders (Berk et al., 2015; Migliorini, Stewart, May, Tapert, & Paulus, 2013). Little research, however, has investigated developmental change in neural activity evoked by interoceptive signals in the typical population, across childhood, adolescence and adulthood. Whilst May and colleagues (2014) observed greater bilateral posterior insula activation in adolescents than adults in response to affective touch, this was not specific to the interoceptive signal, as similar increases were observed for non-affective touch. 6-17 year olds have been found to activate similar neural substrates to adults while performing the heartbeat counting task, such as the insula, medial prefrontal cortex and inferior parietal lobule (Klabunde et al., 2019). In this study, age was positively associated with activation in the anterior cingulate cortex, the orbitofrontal cortex and the left prefrontal cortex, although it should be noted that these findings are based on a sample of 11 participants across a broad age range. In a larger sample, activation patterns appeared to change with age differently across neural regions, with age being positively and linearly related to dorsal anterior activation, but nonlinearly related to ventral anterior insular activation, in response to interoceptive signals (Li et al., 2017). Neither of these studies directly compared neural activation to that in adult participants, however, meaning further work is needed to determine the full developmental trajectory. In a sample of neurotypical and autistic individuals spanning a larger age range (8-54 years), quadratic and linear relationships were observed between age and neural activation while performing the heartbeat counting task in different subdivisions of the insula, and linear relationships between age and activation in a broader interoceptive network, although age was unrelated to behavioural task performance (Failla et al., 2020). Interoceptive accuracy does appear
to be associated with heartbeat evoked potentials in adolescence, as in adulthood (Mai, Ki, Georgiou, & Pollatos, 2018), but again work is needed that explicitly follows the developmental trajectory throughout childhood and into adulthood. It is also worth noting that these studies have used cross-sectional designs to investigate effects of age, so longitudinal designs following development within the same individuals are clearly required in order to determine trajectories accurately.

Another stage where interoceptive ability may fluctuate is late adulthood, the period from 65 years of age until the end of life (Levinson, 1978). Like adolescence, this transitional period is associated with a number of cognitive and physical changes that may be underpinned by changes in interoception (see Whitbourne, 2012). For example, increasing age is associated with physical difficulties, such as increased risk of dehydration (e.g. Silver, 1990), as well as social and emotional difficulties, including poor emotion recognition (Ruffman, Henry, Livingstone, & Phillips, 2008) and risky decision-making (Sparrow & Spaniol, 2016), which have been linked to, or are supported by, interoception (Dunn, Stefanovitch, et al., 2010; Füstös et al., 2012; Rainer Schandry, 1981; Sokol-Hessner et al., 2014; Terasawa et al., 2014; Werner, Jung, et al., 2009; Wiens et al., 2000). Consistent with the proposal that interoceptive changes during this period may contribute to these difficulties (see Murphy et al., 2017), aging is associated with both a decline in interoceptive accuracy on cardiac tests (Khalsa, Rudrauf, & Tranel, 2009; Murphy, Geary, Millgate, Catmur, & Bird, 2017 although see Ainley, Tajadura-Jiménez, Fotopoulou, & Tsakiris, 2012; Mikkelsen, O’Toole, Lyby, Wallot, & Mehlsen, 2019) as well as increased rates of alexithymia (Joukamaa, Saarijärvi, Muuriaisiemi, & Salokangas, 1996; Mattila, Salminen, Nummi, & Joukamaa, 2006; Paradiso et al., 2008 although see Gunzelmann, Kupfer, & Brähler, 2002). Impairments have also been observed in late adulthood in perceiving other interoceptive signals, such as pain, temperature, hydration and taste (Clark & Mehl, 1971; Gagliese, 2009; Silver, 1990; Stevens et al., 1995). Further, at the neural level, Paradiso et al. (2008) observed that increased age was associated with decreased volume of the anterior cingulate cortex. Whilst few studies have investigated the
consequences of poor interoception in the later stages of life, a body of evidence indicates that elevated alexithymia in older adults is associated with poor physical and mental health (Bamonti et al., 2010; Hintistan, Cilingir, & Birinci, 2013; Waldstein, Kauhanen, Neumann, & Katzel, 2002). However, more recent research found that whilst emotion recognition abilities showed a small association with interoceptive accuracy in individuals age 20-90 years, interoceptive accuracy did not uniquely predict emotion recognition abilities after multiple cognitive and affective factors were controlled for (Murphy, Millgate, Geary, Catmur, & Bird, 2019).

7. Outstanding Questions

Taken together, existing evidence suggests that interoceptive abilities are closely related to typical functioning, for example in the domains of emotional processing, learning and decision-making, with atypical interoception being associated with impairments in these domains. Importantly, atypical interoception has been observed across a range of clinical conditions, and related to a number of transdiagnostic symptoms, in line with the suggestion that atypical interoception may be a general risk factor for the development or maintenance of psychopathology. A number of outstanding questions remain, however, that require additional testing in order to fully comprehend the role of interoception in psychopathology. The following sections outline these unresolved issues, and highlight priorities for future research.

7.1. Is interoception a unitary ability?

Whether interoceptive abilities are unitary, i.e., whether an individual with good cardiac perception will necessarily have good respiratory, gustatory, or temperature perception, or whether interoceptive abilities can dissociate in a signal-dependent manner, is a crucial question that remains to be answered. While we have thus far treated interoception as a unitary ability, it is likely that distinct interoceptive signals are processed independently, and multiple signals across domains are integrated to create a representation of the body’s internal state as a whole (Quigley et al., 2021). Research into the factor structure of interoceptive ability has been limited thus far by the high degree of reliance on cardiac measures of interoceptive abilities, but as discussed in Section
2.2, recent efforts have been made to produce tasks that span a number of interoceptive signals. While findings have been mixed, these initial investigations do not suggest that interoceptive ability is a unitary construct. It is important to note, however, that the tasks utilised in these studies are not always directly comparable; they often vary substantially in their trial and response formats, and place differential demands on working memory and attention. Indeed, the low correlations that have been observed between tasks that measure interoception within the same domain (e.g. between the heartbeat counting task and the heartbeat discrimination task; (Hickman, Seyedsalehi, Cook, Bird, & Murphy, 2020) suggest that variation in task requirements can strongly influence relationships between tasks. Additionally, while a number of interoceptive domains have now been tested, there are numerous domains in which valid tests have not yet been developed. Future work is clearly required in this area, including further development of tasks assessing interoceptive abilities across a wide range of domains, which are ideally matched in terms of task demands. Once appropriate measures are available, it will be possible to determine the relationships between interoceptive abilities (both in typical and clinical populations), and whether training abilities in one domain leads to improvements in others.

Even if it is the case that perception of all interoceptive signals cannot be explained by a single ability, a small number of interoceptive clusters may exist, whereby processing of a given interoceptive signal is associated with processing of some, but not all, other interoceptive signals. If this is the case, the nature of clusters and how these are formed remains to be seen. Potentially, clusters could be based on the fibres that carry different signals (such as A delta fibres carrying signals of cold, touch, and first pain, C fibres carrying signals of second pain, warmth, itch, and muscle ache, and the vagus nerve carrying cardiac, respiratory information and gastric information (Craig, 2002)), or on the neurophysiological pathways that signals follow before being represented in the AI (e.g. A delta and C fibres projecting to lamina 1 of the spinal horn, and parasympathethic afferents such as the vagus and glossopharangeal nerve projecting directly to the nucleus of the solitary tract (Craig, 2002). Alternatively, clusters may represent psychological relationships, such
as signals relaying information concerning threat (e.g. cardiac and respiratory cues; Diest, Bradley, Guerra, Bergh, & Lang, 2009) forming a distinct cluster.

Whether interoception is a unitary ability will have a number of implications for both research and clinical practice. First, it will determine whether interventions for those with interoceptive impairment can be standardised across individuals, or must be tailored to each individual’s needs. While a number of clinical groups have previously been characterised by ‘interoceptive impairment’, evidence for dissociations among interoceptive abilities necessitates further investigation into multiple interoceptive domains in each clinical population. It may be the case that specific patterns of interoceptive impairment are common to a given disorder. Individuals with EDs, for example, may experience particular impairment in the gastric domain, while individuals with anxiety disorders may only struggle to perceive signals relating to threat, such as cardiac and respiratory cues. This seems a possibility, as many of the disorder symptoms described throughout this report are linked to particular aspects of interoception. This proposal is complicated, however, by the observation of cardiac difficulties (most commonly assessed by the HCT) in a wide range of clinical conditions. Further research is clearly required to determine whether interoceptive profiles differ across and within clinical groups. Interventions aiming to improve interoception-related behaviours may be required to begin with a full interoceptive profile for each individual. If some interoceptive abilities do cluster, however, then interventions targeting one ability may be sufficient to improve others within that cluster, and reduce all associated disorder symptoms.

Secondly, determining the conceptual structure of interoception will identify whether previous findings in a particular interoceptive domain (in particular, from studies relying on cardiac perception) are applicable to other interoceptive abilities. While cardiac perception has been associated with multiple abilities, such as emotion processing (Ferguson & Katkin, 1996; Pollatos, Gramann, & Schandry, 2007; Wiens, Mezzacappa, & Katkin, 2000; Critchley et al., 2004), and learning and decision-making (Sokol-Hessner et al., 2014; Werner, Jung, et al., 2009), as the relationship between interoceptive domains is unclear, it cannot be assumed that recognition of
other interoceptive signals also relates to these abilities. From a theoretical perspective, understanding the factor structure of interoceptive ability will aid in resolving the current controversy within the literature concerning the definition of interoception (Khalsa & Lapidus, 2016). While some define interoception in terms of neurophysiological pathways, and others in terms of the location of the origin of a signal, an understanding of the interoceptive signals that cluster together may lead to a more nuanced definition, encompassing multiple distinct domains. This relates to the debate over whether proprioception should be considered interoceptive. While early theories did discuss proprioceptive signals as within the definition of interoception (Cameron, 2001; Damasio, 2003; Vaitl, 1996), more recent work has excluded these signals (Craig, 2002; Craig, 2003; Critchley & Harrison, 2013). Proprioceptive information (similarly to exteroceptive information) is relayed by larger-diameter afferents than signals currently considered to be interoceptive, which project to a distinct area of the dorsal horn. An argument could be made for the inclusion of proprioception in the concept of interoception, however, based on the fact that these signals arise internally, and are processed by the insula (although more posteriorly to other interoceptive signals) (Bottini et al., 2001; Fasold et al., 2008; Ferrè et al., 2012; Mazzola et al., 2012; Petit & Beauchamp, 2003; Zu Eulenburg et al., 2013). Further, perception of peripersonal space appears to be associated with interoceptive accuracy (Ardizzi & Ferri, 2018). Determining which interoceptive cues are associated with each other would allow the umbrella term of ‘interoception’ to be expanded, with distinct clusters within this domain being clearly defined. Indeed, other signals beyond proprioception that are not currently considered interoceptive, but share a number of characteristics with interoception (e.g., olfaction) may also be included under a broader definition, with specified sub-domains. It is worth noting that if further sensory domains are included in the definition of interoception, interoceptive impairment may help to explain an even broader range of characteristics or symptoms across clinical conditions. If, for example, proprioception is treated as interoceptive, interoceptive difficulties could account for atypical biomotion and kinematics in those with ASD (Cook, Blakemore, & Press, 2013; Cook, 2016), and
may link interoceptive difficulties to developmental coordination disorder, characterised by difficulties acquiring and executing coordinated motor skills (American Psychiatric Association, 2013).

As well as the need to consider separable interoceptive domains (i.e., cardiac, respiratory, gastric, etc.), it is also necessary to determine separable interoceptive dimensions, as these may also relate differently to other abilities and psychopathology. As discussed above, we have recently proposed that one’s objective accuracy in perceiving internal states, when explicitly aiming to do so, is distinct from one’s attention towards interoceptive signals (Murphy et al., 2018; Murphy, Catmur, et al., 2019; Murphy, Brewer, et al., 2019). The latter may also be separable from one’s propensity to utilise or rely on these states to guide behaviour in everyday life, but this is yet to be empirically tested. The notion of metacognitive awareness of one’s own interoceptive abilities as described above (Garfinkel & Critchley, 2013) can be applied to each of these dimensions separately, in that an individual could have good or poor metacognitive understanding of their own accuracy, as well as of the degree to which they attend to internal signals (Murphy, Catmur, et al., 2019). Again, whether these dimensions relate to each other will inform understanding of whether interoception is a unitary ability.

If these dimensions (accuracy and attention, and their respective metacognitive abilities) are not strongly related to each other, it is possible that they relate differentially to psychological disorders. Whilst individuals with depression, for example, may have typical attention towards interoceptive information, they may struggle to perceive signals accurately. Alternatively, individuals with anxiety disorder may attend more to internal signals, while autistic individuals may attend to them less, leading to differences in the tendency to use these cues even if those with ASD and anxiety do not differ in their ability to objectively perceive internal signals. Of course, these dimensions may also interact or be causally related; reduced attention to internal signals, for example, may lead to reduced accuracy, owing to fewer opportunities to learn about one’s internal
signals. Profiles of abilities and difficulties across these interoceptive dimensions, as well as across interoceptive domains, therefore, should be examined across different clinical populations.

Research into the relationship between interoceptive accuracy and attention, and metacognitive awareness of each, is in its early stages. Much of this work is also limited by reliance upon comparing self-report measures of interoceptive attention with objective measures of interoceptive accuracy, making it difficult to determine relationships between them (Murphy et al., 2018). However, distinctions between these dimensions do appear to exist. Subjective reports of interoceptive attention, for example using the Body Perception Questionnaire and Body Awareness Questionnaire, do not appear to correlate well with objective interoceptive accuracy measures e.g. the Heartbeat Counting Task (e.g. Critchley, Wiens, Rotshtein, Öhman, & Dolan, 2004; Ferentzi, Drew, et al., 2018; Garfinkel et al., 2015), while self-reported interoceptive accuracy measures (questionnaires and confidence ratings) appear to correlate with each other as well as with objective measures of interoceptive accuracy, although the cardiac perception task used appears to affect results (Forkmann et al., 2016; Garfinkel et al., 2015; Murphy, Brewer, et al., 2019). Metacognitive awareness of one’s own accuracy (also termed interoceptive insight) also appears to be weakly associated with objective accuracy in the cardiac domain, although again discrepancies across the objective cardiac tasks have been observed (Forkmann et al., 2016; Garfinkel et al., 2015; Murphy, Brewer, et al., 2019). Where clinical groups are concerned, research is lacking concerning the stability of the relationships between these different dimensions across different disorders, and the relationship between these dimensions and disorder symptoms. Early evidence does suggest, however, that these dimensions may be separately related to clinical symptoms. In autistic individuals, for example, a pattern of increased self-reported interoceptive attention but decreased objective cardiac accuracy has been observed, with the discrepancy predicting anxiety severity (Garfinkel et al., 2016b). The existence and impact of such discrepancies across different clinical conditions therefore warrants further scrutiny.
Whether interoceptive ability is a unitary construct also relates to evidence for multiple forms of alexithymia. Although many see alexithymia as primarily reflecting cognitive impairment in the representation or labelling of emotions (e.g. Luminet, Vermeulen, Demaret, Taylor, & Bagby, 2006; Suslow & Junghanns, 2002), others have argued for the central role of affective difficulties, characterised by decreased ability to experience emotions (e.g. Bermond, 1997; Vorst & Bermond, 2001). The term ‘Type I’ alexithymia has been used to describe individuals who experience impairment in both the affective and cognitive domains, whereas ‘Type II’ alexithymia defines those with impairment only in the cognitive domain (Bermond, 1997). More recently, Moorman and colleagues also identified ‘Type III’ alexithymics, with affective but not cognitive impairment (Moormann et al., 2008). Impairment in the affective and cognitive domains has been associated, unsurprisingly, with different underlying neural atypicalities (Goerlich-Dobre, Votinov, Habel, Pripfl, & Lamm, 2015). While the majority of research focuses on the cognitive dimension, research into the affective dimension of alexithymia is becoming more common, and these two dimensions are likely to relate differently to interoceptive abilities.

In particular, it is possible that a similar distinction relates to interoception; that we can separate interoception at the perceptual level (comparable to the affective domain of alexithymia), from interoception at the cognitive level, involving representation, recognition and verbal labelling of interoceptive states. Some evidence does exist for a distinction between interoceptive perception and interoceptive recognition. Lesions to the left and right insula, for example, are associated with different consequences for taste perception and recognition; while perception is only impaired in the ipsilateral domain following either lesion, recognition is impaired ipsilaterally following right insula lesions, but bilaterally following left insula lesions (Cereda, Ghika, Maeder, & Bogousslavsky, 2002; Pritchard, Macaluso, & Eslinger, 1999). Similarly, dissociations have been found in patients with insula lesions concerning olfactory stimuli; patients have been observed who struggle to label smells (as sweet or non-sweet), despite remaining able to distinguish perceptually between smells (Stevenson, Miller and Thayer, 2008). Similarly, patients with insula lesions have been observed
who are able to perceive signals of pain, but no longer categorise these as painful (Berthier et al., 1988; Greenspan et al., 1999). These types of distinctions are similar to evidence suggesting that individuals with cognitive alexithymia are unimpaired at perceiving differences between emotional facial expressions, but impaired at classifying them, or labelling them as a particular emotion (Cook et al, 2013). There is clearly a need to distinguish between perception and recognition of internal states, despite the fact that these terms are often used interchangeably, or confounded, in the existing literature. Future work is required to determine whether these separable levels of interoceptive abilities relate differentially to psychopathology, or to social and emotional outcomes in the typical population.

A final component of the debate concerning whether interoception is unitary relates to the developmental origins of interoceptive impairment. While research into alexithymia indicates that developmental and acquired routes exist (Henry, Phillips, Crawford, Theodorou, & Summers, 2006; Hogeveen et al., 2016; Lemche, Klann-Delius, Koch, & Joraschky, 2004; Wood & Williams, 2007), little research has focused on this explicitly in the domain of non-emotional interoception. Interoceptive deficits following lesions, especially to the AI and ACC, are well documented (Ibañez et al., 2010), but the nature of the development of interoceptive difficulties in the absence of brain trauma requires further investigation, as described in Section 6. In particular, it is important to understand whether interoceptive impairments differ in their nature depending on whether they are neurodevelopmental, develop in response to environmental triggers, or develop in response to brain trauma. Similarly, whether interoceptive impairment is a stable trait, or a dynamic state requires investigation. Debate exists concerning whether alexithymia is a state or trait phenomenon; while many have observed stability over time (e.g. Mikolajczak & Luminet, 2006; Picardi, Toni, & Caroppo, 2005) even in clinical populations (Berthoz & Hill, 2005; de Timary, Luts, Hers, & Luminet, 2008; Martínez-Sánchez, Ato-García, & Ortiz-Soria, 2003; Saarijärvi, Salminen, & Toikka, 2006; Salminen, Saarijärvi, Toikka, Kauhanen, & Äärelä, 2006; Stingl et al., 2008), others argue that severity may vary, especially with disorder symptoms in clinical populations (e.g.
Honkalampi et al., 2004, 2001; Marchesi et al., 2005). Similarly, stability appears to differ depending on severity of alexithymia (de Haan et al., 2012), as well as across clinical populations (de Timary et al., 2008; Luminet, Bagby, & Taylor, 2001; Rufer et al., 2004; Saarijärvi et al., 2006; Stingl et al., 2008). Overall, it appears that while alexithymia scores can change (low absolute stability), relative stability is fairly high (see Lumley, Neely, & Burger, 2007 for a review). Where explicit measures of interoception have been used, while some degree of stability in interoceptive accuracy has been observed (e.g. Ferentzi, Drew, et al., 2018; Herbert et al., 2011; Koch & Pollatos, 2014; Pollatos, Traut-mattausch, et al., 2007; Van Dyck et al., 2016), situational factors can influence interoceptive accuracy, at least in the respiratory and cardiac domains (Bogaerts et al., 2005; van den Bergh et al., 2004; Wittkamp et al., 2018). Self-reported interoceptive attention (Ferentzi, Drew, et al., 2018; Murphy, Brewer, Plans, et al., 2019; Shields et al., 1989) and accuracy (Murphy, Brewer, et al., 2019) have also shown good test-retest reliability. Thus far, however, interoceptive abilities have not been monitored over a substantial time period, at least in adulthood. In childhood, modest correlations in HCT performance over one or two years have been observed (Koch & Pollatos, 2014; Murphy, Cheesman, et al., 2019). Assessing and comparing the development and stability of interoceptive abilities across multiple interoceptive domains and dimensions will help to determine whether interoception is a unitary or fractionated construct.

7.2. If interoception accounts for symptom commonalities across disorders, what are differences between clinical groups attributable to?

Thus far, we have argued that atypical interoception may lead to multiple manifestations of psychopathology, and that similar difficulties across disorder groups may be explained by a common underlying interoceptive impairment. Despite the substantial overlap in symptoms and transdiagnostic markers across clinical populations, however, there are clearly differences in the behavioural and neural profiles associated with distinct disorders. If it is the case that poor interoception is a contributing factor to psychopathology, the causes of varying disorder outcomes require further investigation. Notably, clinician bias may affect the attribution of formal diagnostic
labels (e.g. Schwartz, Docherty, Najolia, & Cohen, 2019), and there is an emphasis on the importance of individual case formulation, rather than relying on diagnostic categories, in clinical psychology (e.g. Bakker, 2019; Persons, 2008). This section is therefore focused on understanding the differences between individuals, and groups of individuals with similar manifestations of psychopathology, regardless of formal diagnosis.

As discussed above, clinical groups may be distinguished by impairment in separate domains or dimension of interoception, if interoceptive ability is not unitary. Relatedly, differences in the processing stage at which interoceptive impairment occurs may distinguish clinical populations, for example whether atypicalities exist in the physiological signal itself, perception (consciously or subconsciously) of the signal, or identification (labelling/recognition) of the signal (see e.g. Khalsa et al., 2018). Alternatively, it may be the case that the disorder that an individual develops is determined (at least in part) by the developmental stage at which interoceptive impairment is experienced. It is also possible that compensatory strategies for interoceptive impairment may be utilised differentially by clinical populations, leading to different patterns of difficulties. Notably, ‘state’ and ‘trait’ forms of interoceptive impairment may exist, whereby some individuals may be faced with these difficulties throughout their life, whereas others may experience impairments during specific periods only, potentially associated with changes in disorder symptomatology.

Finally, demographic group and environmental influences, for example sex, ethnicity, education, social economic status, parenting styles, exposure to trauma, and cultural context, may also influence the specific manifestation of psychopathology (e.g. Breslau, Kendler, Su, Gaxiola-Aguilar, & Kessler, 2005; Murphy, Viding, & Bird, 2019; Sayed, Iacoviello, & Charney, 2015; South & Krueger, 2011; Wadsworth & Achenbach, 2005). Further research should therefore investigate not only the interoceptive similarities across disorders, but also each of these potential explanations for varied behavioural and neurological manifestations of poor interoception, and its consequences, across clinical groups. Similarly, variation in interoceptive abilities within a
particular disorder population may be responsible for the heterogeneity observed within that diagnostic category.

While we propose interoceptive impairment as a general risk factor, and even a potential candidate for the P factor of psychopathological susceptibility should it exist, we acknowledge that interoceptive abilities vary within a given clinical population, and that interoceptive impairment is not universal in psychological disorders. While cross-disorder symptom similarity may result from a common interoceptive impairment in a large portion of individuals, specific symptoms that are central to a particular disorder are presumably responsible for the majority of differences between diagnostic categories. It is likely that, while interoceptive difficulties are associated with a tendency to develop a psychological disorder, additional routes to disorder development (both genetic and environmental in nature) also exist, and contribute to the differences between, as well as within, clinical groups. The relative contribution of interoceptive and additional difficulties across clinical populations is therefore also an important area for subsequent investigation.

7.3. What is the precise nature of the relationship between interoception and psychopathology?

The main claim of the current paper is that atypical interoception causally increases the risk of psychopathology. The majority of the existing work on interoception and psychopathology has relied, however, on correlational designs and the relationship between interoception and clinical symptoms is likely to be complex, potentially varying across conditions. While it seems plausible that atypical interoception may lead to the development of a number of disorders such as depression, anxiety, eating disorders, OCD, PTSD, etc., the relationship may be bidirectional. For example, difficulties recognising internal signals of hunger and satiety may contribute to development of eating disorders, and presence of an eating disorder may lead to deliberate suppression of these signals, for example to reduce distress caused by hunger when calorific intake is being reduced, leading to reduced interoception attention, and in turn, accuracy. It may be the case that in some individuals, atypical interoception causes disorder development, while in others
psychopathology leads to changes in interoception. It is likely that in many individuals, both of these processes are at play simultaneously, whereby the presence and severity of symptoms exacerbates interoceptive difficulties and vice versa. It is also important to consider neurodevelopmental disorders, which are lifelong conditions present from, or before, birth. In autism, for example, it is unlikely that interoception causes development of this condition. Autism may lead to interoceptive atypicalities, but it is also possible that both autism and interoceptive difficulties arise from widespread neural atypicality, making their co-occurrence highly likely. A similar explanation has been proposed for the co-occurrence between alexithymia and autism (Bird & Cook, 2013). While atypical interoception seems likely to cause psychopathology, testing the direction of the relationship through experimental manipulation (and longitudinal studies to some extent) across a range of clinical populations is clearly required.

Another aspect of the relationship between interoception and psychopathology that requires further investigation is its linearity. While it is often assumed that less accurate interoception, or lower attention to internal signals, leads to poorer mental health outcomes, this relationship has also been observed in the opposite direction, for example in the case of anxiety. This raises the possibility of a quadratic relationship between interoceptive abilities and psychopathology in certain conditions, which remains to be explicitly tested. Of course, this relationship could also vary across interoceptive dimensions and domains, so future work should aim to investigate this possibility using multiple measures of interoception.

7.4. Is intact interoception required online, or only during development?

Much of the evidence described thus far suggests that interoception is necessary for typical functioning in a number of domains (in particular, those of emotion processing and learning and decision-making), and that interoceptive impairment can lead to psychopathological outcomes. What remains to be seen, however, is whether interoception is only essential during development, for example while learning the affective value of particular cues, or whether it is crucial to continue to utilise interoception online. While interoceptive cues are clearly used continually, it may be that
learnt associations subsequently allow cognitive processing to occur without interoceptive information. If, for example, one has associated a facial expression of happiness with internal cues to happiness, it is unclear whether the facial expression could be recognised in the absence of interoceptive signals, as exteroceptive cues (physical properties of the face) have previously been associated with the label ‘happiness’. Simulationist models that emphasise ‘embodiment’ as important for emotion recognition (e.g. Barsalou et al., 2003; Gallese, 2001, 2003; Goldman & Sripada, 2005; Niedenthal et al., 2005; Preston & de Waal, 2001) would suggest that online interoceptive cues are necessary for this process. One way in which to determine the online contribution of internal cues would be to train or impede interoception in typical adults, and assess changes in other domains such as emotion recognition, or disorder symptom severity. Similarly, inducing internal states, for example during an emotion recognition task, and determining the extent to which internal signals bias perception and cognition, would add to our understanding of the online use of internal signals.

One indication that online interoception may not be required is the observation of intact emotional and interoceptive abilities in patients with insula and ACC lesions during adulthood (Damasio et al., 2013; Feinstein et al., 2016). Instead of suggesting that interoceptive cortex is not required for interoception, these findings may indicate that, assuming typical development, interoceptive information is not required online later in life. Potentially, if one has learnt to associate particular external stimuli with internal states, one can infer these internal states using these associations even in the absence of recognisable interoceptive information. If interoception is not required online, the possibility is raised that a sensitive period may exist during development, in which interoceptive learning is crucial. If an individual does not learn to associate interoceptive cues with exteroceptive cues, such as facial expressions and verbal labels, during this period, they may struggle to form these associations later in life. Similarly, if one does not attend to or perceive internal signals during a sensitive period, learning to do so subsequently may be particularly challenging. Whether interoceptive abilities develop during a sensitive period will have
consequences for the efficacy of interoceptive training later in life; if a sensitive period exists, interoceptive training in adults identified as experiencing interoceptive impairment may be less effective than training during childhood. Sensitive periods may also occur later in life, for example during puberty, pregnancy and menopause, when physical changes may disrupt interoceptive signals, and perception of these signals must be adapted or relearnt (Murphy, Viding, et al., 2019). Whether these periods exist requires investigation, and will impact on our understanding of the importance of interoceptive interventions at specific developmental stages. The role of interoception online, and potential critical periods, may also vary across skill domains, for example emotion recognition, reward processing, and learning and decision-making, making it crucial to investigate this question across a range of skills.

7.5. How can interoceptive abilities be improved?
If it is the case that poor interoception contributes to multiple clinical conditions and impairments in numerous aspects of cognition, an obvious aim should be to improve these abilities where they are deficient. While a number of attempts have been made to improve interoception, it remains to be determined whether improvements are long-lasting, generalisable across domains, and indeed helpful for reducing psychopathology (see Weng et al. (2021) for a review). Once questions concerning the nature of the relationship between interoception, psychopathology, cognition, and the structure of interoception, have been answered, determining the most efficient and effective interventions will be possible.

Thus far, attempts have been made to train interoception using heartbeat perception training, whereby feedback leads to improved perception of one’s heartbeats. Schandry and Weitkunat (1990), for example, demonstrated that auditory feedback presented 130ms after the EKG-R wave (such that it is perceived in synchrony with the heartbeat) increased both cardiac accuracy, and heartbeat evoked potentials on a subsequent heartbeat perception task, indicating increased interoceptive accuracy in the cardiac domain. More recently, Schaefer, Egloff, Gerlach, & Withhöft (2014) have used a similar procedure, involving presenting participants with visual cues 200ms
following the R wave, and a subsequent training phase where participants were required to press a button in synchrony with a particular heart beat (e.g. in three beats’ time), and were provided with immediate feedback on their accuracy. Schaefer et al. found that this training procedure improved heartbeat perception in patients with medically unexplained symptoms. What remains to be seen is whether training in the cardiac domain can improve interoceptive abilities in other domains of interoception (e.g., perception and recognition of signals relating to respiration, hunger, temperature, pain, etc.). This type of training has also not been compared to control training conditions, meaning neither its specificity nor its generalisability has been investigated. Clearly, this is likely to depend on whether or not interoception is a unitary ability, so cross-domain training studies will help to elucidate the structure of interoception as well as whether single domain interventions are likely to be useful clinically. Further, whether this type of training alters attention to, or propensity to use, internal signals, or simply interoceptive accuracy when explicitly instructed to attend to these signals, requires investigation. A greater understanding of the relationship between psychopathology and potentially distinct interoceptive domains and dimensions will inform the development of effective interventions.

Beyond explicit interoceptive training, it is often assumed that meditation may improve interoception, although evidence has been mixed. Meditation experience appears to be associated with increased volume of the insula (Hernández, Suero, Barros, González-Mora, & Rubia, 2016) and ACC (Boccia, Boccia, Piccardi, & Guariglia, 2015), and functional activation of these interoceptive regions is observed during meditation tasks (Boccia et al., 2015; Farb, Segal, & Anderson, 2013). Behavioural evidence has been mixed; while some studies have found that meditation experience is unrelated to cardiac accuracy (Khalsa et al., 2008; Melloni et al., 2013; Parkin et al., 2014), a recent study suggested that cardiac accuracy was improved following contemplative practice over a period of three months (Bornemann & Singer, 2017). The same contemplative practice has been found to reduce self-reported alexithymia scores in those with low levels of alexithymia (Bornemann & Singer, 2017). Experienced meditators also exhibit increased
levels of coherence between subjective and objective measures of emotional arousal (Sze, Gyurak, Yuan, & Levenson, 2010). Notably, improved respiratory perception has been observed in meditators, possibly due to the fact that meditation focuses more on respiration than on cardiac signals (Daubenmier, Sze, Kerr, Kemeny, & Mehling, 2013). A recent meta-analysis, however, did not find compelling evidence for an association between meditation and interoceptive accuracy, although only studies investigating the cardiac domain were included (Khalsa, Rudrauf, Hassanpour, Davidson, & Tranel, 2020). Again, determining whether a fractionated structure of interoception exists will shed light upon the most appropriate form of interoceptive training for individuals with specific interoceptive difficulties.

The distinction between interoceptive accuracy and attention is again relevant to the relationship between interoception and meditation or other forms of interoceptive training. The body-focused nature of meditation makes it likely that this practice leads to increased attention to, and use of, internal cues rather than increased interoceptive accuracy (Garland et al., 2012; Grossman, 2015), although increased attention may of course lead, in turn, to increased accuracy. Indeed, meditation has been found to increase self-reported interoceptive attention, as measured by the MAIA (Bornemann, Herbert, Mehling, & Singer, 2015), and confidence and perceived ease in completing the heartbeat discrimination task (Khalsa et al., 2008). It is important to utilise measures that distinguish between these dimensions of interoception, to enable interventions to be tailored more effectively to individuals’ interoceptive impairments. Importantly, whether interventions alter individuals’ attention to, or accuracy perceiving, internal cues may affect their efficacy differentially in different clinical populations.

As studies explicitly training interoceptive abilities are scarce, it is also worth briefly examining the literature on changes in alexithymia. Attempts have been made to reduce alexithymia through therapeutic intervention, with group therapy sessions seeming to reduce self-reported alexithymia in a range of clinical groups, such as those with heart disease (Beresnevaite, 2000) and bulimia nervosa (de Groot, Rodin, & Olmsted, 1995), and panic disorder (Michael Rufer et al.,
Indeed, a review of studies reporting the effect of psychological interventions on alexithymia found that reductions in alexithymia are common, in particular when interventions specifically targeted alexithymia (Cameron, Ogrodniczuk, & Hadjipavlou, 2014). Whether changes in interoceptive accuracy or attention underlie this reduction remains to be investigated. As the alexithymia construct is explicitly related to emotional states, it may be the case that improving general interoceptive abilities can reduce alexithymia, but that interventions specifically aimed at reducing alexithymia are not sufficient to improve interoceptive skills more generally. If this is the case, training in the interoceptive domain may be a more effective strategy for improving recognition of both emotional and non-emotional internal states.

While future research should aim to determine how interoceptive abilities across multiple domains and dimensions can be improved, it is worth noting that not all individuals or disorder groups may benefit from interoceptive training. As discussed above, atypically high interoceptive accuracy or attention may be detrimental to one’s mental health, in a similar way to atypically low interoceptive abilities. Determining the precise nature of the relationship between interoception and psychopathology, both in terms of directionality and linearity, will therefore be essential in order to develop the most beneficial interventions.

8. Conclusion

The current paper has reviewed evidence suggesting that interoception contributes to typical functioning, in terms of emotional abilities, reward-processing, learning and decision making, as well as findings indicating that atypical interoception is associated with impairments in these abilities. As many of these difficulties are shared by clinical conditions, we presented evidence for impaired interoception in a wide variety of these populations, and described possible mechanisms linking atypical interoception to disorder symptomatology. While a number of outstanding questions remain, it seems that interoceptive deficits represent a general risk factor for the development of psychopathology, perhaps even representing a plausible candidate for the P factor of psychopathology (Brewer et al., 2016b, 2016a; Murphy, Brewer, et al., 2017), whereby
difficulties developing interoceptive abilities early in life, or changes in interoceptive abilities later in life, may predispose one to developing a number of psychological and neurological disorders. Future empirical work should investigate the factors that determine which manifestations of psychopathology occur following atypical interoception, and whether distinct interoceptive domains and dimensions are associated with distinct clinical outcomes.
Acknowledgements

RB is supported by a New Investigator Grant from the Medical Research Council (MR/S003509/1).
References


Ainley, V., Tsakiris, M., Pollatos, O., Schulz, A., & Herbert, B. M. (2020). Comment on “Zamariola et al. (2018), Interoceptive Accuracy Scores are Problematic: Evidence from Simple Bivariate Correlations”—The empirical data base, the conceptual reasoning and the analysis behind this statement are misconceived and do not support. *Biological Psychology, 152*, 107870.


Ambrosecchia, M., Ardizzi, M., Russo, E., Ditaranto, F., Speciale, M., Vinai, P., … Gallese, V.


in chronic fatigue syndrome. Psychological Medicine, 38(7), 941–951.


Clark, L., Studer, B., Bruss, J., Tranel, D., & Bechara, A. (2014). Damage to insula abolishes cognitive distortions during simulated gambling. Proceedings of the National Academy of


Delgado-Rico, E., Soriano-Mas, C., Verdejo-Román, Río-Valle, J. S., & Verdejo-García, A. (2013). Decreased insular and increased midbrain activations during decision-making under risk in
adolescents with excess weight. *Obesity, 21*(8), 1662–1668.


---

125


Martin, J. B., & Pihl, R. O. (1986). Influence of alexithymic characteristics on physiological and


Psychophysiology, 30(2), 76–86.


Psychology, 101(1), 69–76.


Wigham, S., Rodgers, J., & South, M. (2015). The interplay between sensory processing...


**Figure Legends**

**Figure 1.** 2 x 2 dimensional structure of interoception, distinguishing between different measurement types (beliefs or performance) and interoceptive abilities (accuracy or attention). Metacognitive ability (referred to as interoceptive awareness; Garfinkel and Critchley, 2013, or interoceptive insight; Khalsa et al., 2018) is shown by the black arrows, reflecting the extent to which beliefs and performance are related, either within the accuracy or attention ability. Figure adapted from Murphy, Catmur, & Bird (2019).