Commentary on “Autism, oxytocin and interoception.”: Alexithymia, not Autism Spectrum Disorders, is the consequence of interoceptive failure

Rebecca Brewer, Francesca Happé, Richard Cook and Geoffrey Bird

*MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King’s College, London

*Department of Psychology, City University, London

Institute of Cognitive Neuroscience, University College London, London

*Corresponding author: Rebecca.1.brewer@kcl.ac.uk

MRC Social, Genetic & Developmental Psychiatry Centre

Institute of Psychiatry,
King’s College London,
16 De Crespigny Park,
London, SE5 8AF
Abstract
In “Autism, oxytocin and interoception” (Neuroscience and Biobehavioral Reviews 47, 410–430) Quattrocki and Friston present their theory of the role of oxytocin in interoception from multiple perspectives. The arguments contained therein are compelling, and highlight the fact that interoception, and the role of oxytocin in interoception, should receive more research attention. However, in addition to outlining the role of oxytocin in interoception the authors also suggest that Autism Spectrum Disorder (ASD) is a result of a failure of this system. It is this latter claim that we disagree with, instead suggesting that alexithymia, rather than autism, is most accurately characterised as a general failure of interoception. We review positive evidence that alexithymia produces several of the deficits identified as indicating a failure of interoception, and negative evidence that ASD (in the absence of comorbid alexithymia) is associated with these deficits. We highlight implications for the model, for oxytocin research, and for the clinical management of psychiatric conditions more generally.

Keywords: Interoception; Alexithymia; Autism; Oxytocin; Predictive Coding; Imitation; Theory of Mind; Empathy; Emotion Recognition.
In their “Autism, oxytocin and interoception” paper, Quattrocki and Friston present a theory of the role of oxytocin in interoception from anatomical, genetic, computational, psychological and behavioural perspectives. The theory they outline is developmentally specific, addressing how an initial insult to the oxytocin system can result in higher-level socio-cognitive deficits over a protracted period of development. There is much to praise about the theory, indeed the practice of specifying a theory simultaneously at multiple levels of explanation should become the gold standard for theoretical papers in the future. However, in addition to outlining the role of oxytocin in interoception the authors also suggest that Autism Spectrum Disorder (henceforth “autism”) is the result of a failure of this system. It is this latter claim that we disagree with, instead believing that alexithymia, rather than autism, is most accurately characterised as a general failure of interoception.

1. The Model as Outlined by Quattrocki and Friston

1.1. Oxytocin, interoception and Predictive Coding

Quattrocki and Friston present interoception, the process of sensing the internal state of the body, within a predictive coding framework in which the brain aims to construct an internal probabilistic model of the state of the body. The brain tests the accuracy of its model of the body’s state by comparing predictions derived from the model against incoming sensory data. Any discrepancy between model predictions and sensory data (prediction error) is then used to update the model of the state of the body as necessary, or engage autonomic reflexes to change sensory (interoceptive) data. The balance between updating and reflexive responding is set by the precision of, or confidence in, predictions relative to prediction error.

It is claimed that oxytocin plays a crucial role in interoception within this predictive coding model – particularly with respect to attenuation of prediction errors with resultant impact upon context-dependent associative learning of links between interoceptive states and exteroceptive stimuli. For example, if there is a top-down prediction from an established model that blood levels of glucose should be $X$, where $X$ is the optimum level for healthy functioning, and sensory signals suggest that glucose levels are in excess of $X$, then interoceptive prediction errors relating to glucose levels will be generated. These prediction errors can be resolved in one or two ways, either inference about glucose levels can be revised to induce a (sub-personal) percep of hyperglycaemia or insulin can be released to lower glucose levels. In both cases, prediction errors are eliminated but only the autonomic reflexes releasing insulin will maintain homeostasis. Which of the two processes is evoked by the presence of prediction errors is determined by the relative precision of prediction and prediction error.\(^1\) If the precision of the prediction is higher than that of the prediction error then insulin is released, whereas if the sensory precision is higher than prediction precision then the model of glucose levels is updated to produce beliefs in line with sensory

\(^1\) Strictly speaking, in hierarchical predictive coding schemes the precision pertains to the prediction errors at each level of the hierarchy. In what follows, we will refer to the precision of sensory input as sensory precision and the precision of higher level prediction errors as the precision of predictions.
data (i.e., I am hyperglycaemic). This makes intuitive sense; when we have high confidence in our sensory data and there are sensory prediction errors, it suggests we have an incorrect view of the world and should therefore update our model. In contrast, when we are very confident in our predictions, we do not trust our sensory data and take steps to ensure that the incoming sensory data matches our expectations.

The first claim of Quattrocki and Friston’s paper is that within the interoceptive domain, oxytocin plays a key role both in the formation of context-dependent models of the internal state of the body and in homeostasis itself. To continue with the earlier glucose example, oxytocin plays a role in the learning process that produces a model of glucose levels (by selectively augmenting and attenuating prediction errors relating to glucose concentration and therefore influencing the likelihood that generative models will be updated, and by facilitating the synaptic plasticity required for the development of generative models), which in turn provides a prediction corresponding to a healthy level of glucose\(^2\). In order to ensure that any prediction error derived from a mature model is met with a release of insulin to lower glucose levels rather than an updating of the model, the sensory precision is attenuated – a process that is also reliant on oxytocin. The role of oxytocin in balancing the precision of predictions and sensory prediction error can be likened to interoceptive attention; the assignment of precision or confidence to various sensory channels can be considered the formal homologue of attentional selection, in the sense that precise sensory information is selected over imprecise information (prediction errors). Similarly, the attenuation of sensory precision may correspond to sensory attenuation; namely, the attenuation of the intensity of the sensory consequences of self-produced sensations.

1.2. Development and Autism Spectrum Disorders

Associative learning of social saliency and attachment is mediated by oxytocin under this model. Specifically, oxytocin mediates the context-dependent associative learning between interoceptive states and their exteroceptive cues in the environment, while also modulating the precision of interoceptive cues leading to the learning of saliency for biologically-relevant exteroceptive cues. The example highlighted in “Autism, oxytocin and interoception” is that infants may learn attachment behaviours, and that their caregiver’s face is a salient stimulus, via oxytocin-mediated associative learning in which their caregiver’s face is associated with oxytocin-modulated interoceptive signals of warmth and satiety.

Thus, the second claim of Quattrocki and Friston’s paper is that should this system function atypically, an individual would exhibit the following impairments:

\[^2\] It is worth noting that the tonic effects of mature generative models on precision of prediction errors and the phasic effects produced by context-dependent associative links may not necessarily be mediated by the same mechanisms. This is a topic worthy of future elaboration.
• Reduced ability to learn about socially relevant stimuli, resulting in reduced social attention and imitation and a resultant impairment in understanding the mental states of others: If impairment in the oxytocin system prevents one from associating social cues, such as others’ faces, with positive interoceptive states, one is unlikely to develop a preference for social stimuli. Reduced attention to others would necessarily lead to decreased opportunity to learn about their behaviour (leading to reduced likelihood of imitating this behaviour). Similarly, decreased social attention would impair one’s ability to associate particular mental states with specific contexts, or behavioural cues.

• Poor representation of emotional states: As oxytocin is argued to modulate the association between interoceptive and exteroceptive cues, impairment in this system would prevent the interoceptive state associated with each emotion from becoming associated with particular contexts and behaviours. A lack of markers for different emotional categories is likely to lead to decreased ability to distinguish between these categories, making emotional representation challenging (Bird and Viding, 2014).

• Impairments in the ability to distinguish between Self and Other, coupled with a cognitive egocentrism: Reduced attention to other individuals, via reduced oxytocin-mediated associative learning linking others with positive interoceptive states, would result in an egocentric cognitive bias. Furthermore, limited ability to attenuate interoceptive cues from one’s own body may result in self-related interoceptive inferences conflicting with those related to other individuals - the salience of Other-related cues and the attenuation of Self-related cues are learned as part of the generative model in typical development – with impairments leading to ‘blurred’ boundaries between Self and Other.

The third claim of the Quattrocki and Friston paper is that this pattern of impairments is exhibited by individuals with autism, and that therefore autism can be characterised as a disorder of interoception caused by oxytocin abnormality. While the evidence for the role of oxytocin in interoception is compelling, it is this latter claim with which we do not agree. Instead, we suggest that the literature demonstrates that individuals with ASD do not exhibit the majority of these characteristics, and that where observed in the autistic population, impairments are actually due to co-occurring alexithymia. Indeed, the pattern of impairments across autism and alexithymia lead us to question the second of Quattrocki and Friston’s claims of a developmental cascade of impairments due to interoceptive impairment. Instead, we suggest that the symptoms they identify as resulting from interoceptive impairment should be considered as two independent clusters of symptoms that differentially characterise autism and alexithymia.

2. Alexithymia, but not autism

Quattrocki and Friston suggest that alexithymia, characterised by an inability to recognize and describe one’s own emotions, is perhaps the primary deficit of the impaired
interoception that characterises ASD. Furthermore, they suggest that an inability to modulate interoceptive prediction errors results in a failure to learn associations between an internal affective state and expressions of those states in others, impairing attention to the eye region of faces and the recognition of and attention to the emotional expressions of others.

This is our main point of departure from the Quattrocki and Friston model; despite their frequent co-occurrence, **alexithymia and ASD are independent constructs**. Alexithymia is neither necessary nor sufficient for an autism diagnosis, nor is it universal among autistic individuals; current estimates suggest that approximately 50% of individuals with autism also suffer alexithymia (e.g. Hill, Berthoz, & Frith, 2004). Conversely, many individuals show severe degrees of alexithymia without demonstrating autistic symptoms. In our view, the supposed affective impairments in autism are in fact due to comorbid alexithymia (Bird and Cook, 2013).

This debate would be of little practical relevance were it not for the fact that several studies suggest that the impact of autism and alexithymia are independent, precisely in those domains identified as reliant upon interoception. For example, in an early study of **emotional awareness** in individuals with autism, Silani et al. (2008) demonstrated that the level of anterior insula activation was predictive of emotional awareness, in accordance with the Quattrocki and Friston model, but that the degree of anterior insula activation during emotional interoception was predicted by alexithymia and not by autism. Similarly, Bird et al. (2010) demonstrated that the degree of ‘empathic’ anterior insula activity evoked by the pain of another was associated with alexithymia and not autism. These results are supported by analysis of brain structure – while alexithymia was associated with coherence of anterior insula-based networks, autism symptom severity was associated with coherence of networks including medial prefrontal and temporoparietal cortices (Bernhardt et al., 2013). **Emotion recognition** is dependent on appropriate functioning of the oxytocin system throughout development under the Quattrocki and Friston model. In common with the above studies, it has been demonstrated that alexithymia and not autism predicts the ability to recognise emotion from both facial and vocal stimuli (Cook, Brewer, Shah, & Bird, 2013; Heaton et al., 2012; Jones et al., 2011). These results strongly support the role of associative learning in linking awareness of one’s own emotional states to recognition of those states in another, as highlighted by Quattrocki and Friston (see also Allen & Heaton, 2010; Bird & Viding, 2014; Gergely & Watson, 1996; Heyes & Bird, 2007). All of these studies adopted a ‘four group’ design, crossing the presence/absence of autism and alexithymia, and so could specifically separate out the effects of autism and alexithymia.

A further area of functioning of relevance to the Quattrocki and Friston model is **orienting towards social and emotional stimuli**. Under their model, atypical functioning of the interoceptive system and oxytocin-mediated associative learning result in reduced salience of social and emotional signals, with social orienting towards faces and eyes highlighted as two domains of particular relevance. While the social attention hypothesis of autism has attracted a lot of interest (Klin et al., 2002; Pelphrey et al., 2002; Riby and Hancock, 2009, 2008) recent data have cast doubt on some aspects of the theory (Johnson, 2014). Complicating the picture is the fact that there may be three dissociable aspects of orienting that impact upon social functioning; an early-developing sub-cortical attentional orienting
system towards face-like stimuli (Johnson, 2014), a later-developing attentional bias towards faces (Frank et al., 2014), and a separate system governing attention towards the eyes which impacts upon the ability to recognise emotion (Aviezer et al., 2008; Calder et al., 2000; Smith et al., 2005; Wong et al., 2005). With respect to the sub-cortical orienting system, recent research suggests that this system is intact in both adults (Shah et al., 2013) and children (Fischer et al., 2014) with autism. The endogenous allocation of attention towards faces, and eyes within faces, seem to dissociate however; when directly compared it was found that autism, but not alexithymia, predicts attention towards faces, while alexithymia, but not autism, predicts attention towards eyes (Bird et al., 2011). Similarly, despite reduced attention towards social stimuli, individuals with autism prioritize salient information within social stimuli in a typical manner (Chevallier et al., 2013). This result suggests that the common conflation of the social and emotional ‘symptoms of autism’ may be unhelpful.

Finally, Quattrocki and Friston suggest that an inability to modulate interoceptive precision leads to a less well-defined sense of self – evident in a reduced ability to differentiate between the self and another, and a reduced ability to represent mental states, both one’s own and those of other individuals (‘theory of mind’; ToM). While the precise mechanism by which this results from an inability to modulate the precision of interoception is understandably less well specified than other parts of the model, it is speculated that it may result, in part, from reduced social orienting and propensity to imitate in autism. While we have already questioned the existence of an absolute deficit of social orienting in autism, there is some evidence for reduced attention to faces in autism (Klin et al., 2002; Pelphrey et al., 2002; Riby and Hancock, 2009, 2008). The picture relating to imitation in autism is clearer; recent evidence suggests that individuals with autism are as capable of imitating as non-autistic individuals (Bird et al., 2007; Dinstein et al., 2010; Gowen et al., 2008; Hamilton et al., 2007; Leighton et al., 2008; Press et al., 2010), but instead have a problem with modulating (Cook & Bird, 2012) or inhibiting imitation, shown both by laboratory studies (Spengler et al., 2010) and by the increased prevalence of echopraxia and echolalia in this population (Baltaxe and Simmons III, 1975; Cunningham and Dixon, 1961; Jordan, 1993; Kanner, 1946; Koegel et al., 1982). This is consistent with the proposed failure to contextualize – or selectively attend to – sensory cues during interactions with others.

Interestingly however, the problems experienced by autistic individuals in inhibiting imitation may provide support for another claim made by Quattrocki and Friston, that those with autism have a problem with individuation – the ability to separate representations of the self and others. Brass, Ruby and Spengler (2009) argued that individuation is a common requirement of a number of socio-cognitive tasks including ToM (where one must keep separate the contents of one’s own mental states and those of another) and the ability to inhibit imitation (in which one must keep one’s own motor plan separate from that of another). This claim was based upon the observation that imitation inhibition (but not inhibition of other over-learned responses) activated the same network of brain areas as involved in ToM (Brass et al., 2005). In support of this conjecture a recent study demonstrated that training to inhibit imitation increased the ability to take another’s visual perspective, thought to rely upon the same process of individuation as ToM and imitation inhibition (Santiesteban et al., 2012). Importantly, the degree to which individuals with autism are capable of inhibiting imitation has been shown to correlate with both
behavioural performance on a verbal ToM task and the degree of activation in ToM-related brain networks during performance of a different, non-verbal ToM task (Spengler et al., 2010). These results, along with findings of increased personal distress in response to another’s pain in autism (Rogers et al., 2007; see Smith, 2009 for an overview), support the existence of problems with individuation in autism. They are therefore in line with the autistic symptoms outlined by Quattrocki and Friston, but are not associated with imitation deficits per se, rather with deficits in learning the distinction between self and other – a distinction that is necessary for inhibiting imitation.

3. Two (or more) Systems?

Thus far we have argued that the affective deficits suggested to result from an inability to modulate the interoceptive sensory precision under the Quattrocki and Friston model characterise individuals with alexithymia not autism. We have presented positive evidence for this claim, with a series of studies demonstrating that deficits in emotional awareness, emotion recognition, empathy, and eye-contact are a product of alexithymia rather than autism (and hence seen in non-autism groups with alexithymia). We have presented evidence of intact imitation and sub-cortical orienting in autism, and acknowledged the deficits in theory of mind and self-individuation (characterised by impaired ToM, increased empathic personal distress, and a reduced ability to inhibit imitation) associated with the condition. This pattern of intact and impaired ability in autism, in domains that are all suggested to rely on interoceptive systems, suggests that there may not be a unitary interoceptive system underlying all of social cognition but that emotional awareness, emotion recognition, empathy and eye-contact may form one cluster of interoception-dependent abilities, and imitation inhibition and ToM another. It is also possible that (at least sub-cortical) social orienting and imitation may form a third cluster. At present it is not known how the motor atypicalities in autism, which are associated with the degree to which those with autism are impaired socially (Cook, Blakemore, & Press, 2013), are related to any of these abilities, but the atypical kinematics identified in the movements of autistic individuals by Cook et al., (2013) were not associated with the severity of their alexithymia. It is possible therefore, that proprioception may also contribute uniquely to some of the processes underlying social competence and mask, or exacerbate, interoceptive impairment. Such a fractionated view of interoception-related impairments would be consistent with evidence that, even within autism, social and non-social symptoms have different aetiologies and show little phenotypic correlation (Dworzynski et al., 2009; Robinson et al., 2012; Taylor et al., 2014).

4. Clinical implications of the alexithymia hypothesis

As demonstrated above, alexithymia is not a symptom of autism, but rather an independent construct that frequently co-occurs with autism. Interestingly, high rates of co-occurring alexithymia are also seen in a number of other clinical conditions including schizophrenia, eating disorders, substance abuse, and post-traumatic stress disorder (Bird and Cook, 2013; Grynberg et al., 2012). Although little research has investigated the reasons underlying this
co-occurrence, it is likely that genetic vulnerability to atypical neural development is associated with numerous conditions, including alexithymia, and that the locus or timing of atypical development determines the nature of the cognitive impairment. ‘Pure’ forms of alexithymia, or clinical disorders without co-occurring conditions, are therefore possible following very circumscribed patterns of atypical neural development, but co-occurrence between conditions as a result of diffuse patterns of abnormalities would be expected to be more frequent (Bird and Cook, 2013). A further possibility is that a particular deficit, for example in the oxytocin system, may lead to differing phenotypes depending on the point at which it occurs in the cortical hierarchy and the number of levels it spans. Impairment at one level, for example, may be associated with ASD or alexithymia alone, while impairment at another may be associated with both conditions. Thus far, evidence concerning these possibilities is lacking, making investigation of the mechanisms underlying the co-occurrence crucial.

Recent data from our lab show that the impaired emotion recognition often seen in individuals with eating disorders is a product of co-occurring alexithymia rather than eating disorders per se (Brewer et al., 2015). This raises the question as to whether the inconsistent reports of emotion recognition and empathy impairments across a range of disorders, such as schizophrenia and substance abuse (Abramowitz et al., 2014; Edwards et al., 2002; Kim et al., 2011; Kornreich et al., 2003) can be explained by alexithymia. Such a pattern of impairment across a wide range of conditions in which alexithymia levels are increased is more consistent with Quattrocki and Friston’s model of impairment of a core neural mechanism impacting upon multiple systems than is a deficit linked to autism only.

More speculatively, if the model proposed by Quattrocki and Friston is correct and an interoceptive failure can produce a number of deficits, then one can imagine that alexithymia is potentially linked to decreased awareness of intoxication levels and altered reward processing in substance abuse, to reduced awareness of (i.e., ability to selectively attend to) hunger and satiety cues in eating disorders and obesity respectively, and with the movement atypicalities and sensory abnormalities seen in a number of conditions, for example in those with autism. Furthermore, the elevated levels of alexithymia in those with PTSD and social anxiety support the developmental models proposed by Quattrocki and Friston whereby interoceptive deficits, mediated by an atypical oxytocin system, may be related to the inability to suppress aversive memories (Boccia and Baratti, 2000) and with anxiogenic properties when in socially stressful situations (Alvares et al., 2012, 2010; de Oliveira et al., 2012a, 2012b; Heinrichs et al., 2001).

5. Summary and outstanding questions.

We wish to highlight how highly we regard the original “Autism, oxytocin and interoception” paper by Quattrocki and Friston. The paper provided a comprehensive multi-level account of the role of oxytocin in interoception and an account of how atypical functioning of the system may impact behaviour over development. It made testable predictions, and even outlined how those predictions might be tested. While persuaded by arguments as to the

---

3 We are grateful to an anonymous Reviewer for this suggestion.
importance of oxytocin to interoception, and by the claims of its role within a predictive coding account, we have suggested that the model outlined by the paper should be revised in two ways. First, we discussed evidence that alexithymia, rather than autism, is best characterised by an interoceptive failure. We reviewed positive evidence that those with autism but without alexithymia demonstrate typical emotional awareness, emotion recognition ability, eye-contact and empathy. In contrast, individuals with alexithymia, whether also autistic or not, are impaired in each of these domains. We also reviewed negative evidence demonstrating that neither imitation nor sub-cortical orienting to faces is impaired in autism.

Second, we questioned whether the full range of impairments specified by Quattrocki and Friston would arise from a core failure of interoception. We noted those with alexithymia, who have problems with interoception within the affective domain at least, are neither impaired at imitating others, nor at representing their mental states, and are as likely as non-alexithymic individuals to attend to faces. In contrast, individuals with autism are unimpaired at emotion recognition, imitation, and eye-contact, but are impaired at ToM, endogenous orienting to faces, and imitation inhibition. Accordingly, we speculate that there may be two or more interoceptive systems and that they may be selectively impaired.

Lastly, we highlighted the clinical relevance of Quattrocki and Friston’s model coupled with the alexithymia hypothesis, demonstrating its impact upon affective processing across a number of conditions and speculating that it may underlie a range of symptoms including anxiety, maladaptive eating and addictive behaviours, and atypical social reward and fearful memory processing. An interesting outstanding question related to these speculations is the extent to which alexithymia represents a specific deficit of interoception in the affective domain or whether multiple interoceptive domains are affected. We have been interested in this question for some time and preliminary data from our lab, as well as findings of decreased heart rate awareness (Herbert et al., 2011), longer delays to seek treatment after heart problems (Carta et al., 2013; Kenyon et al., 1991), erratic caffeine use (Lyvers et al., 2014), and an association with diabetes (Abramson et al., 1991), lead us to suggest that alexithymia may best be considered a general disorder of interoception.

Finally, if the Quattrocki and Friston model is correct then we may hope to relieve a multitude of symptoms across a number of disorders. Although they were admirably and appropriately cautious in advocating treatment options suggested by their model, if treatments (whether pharmacological, behavioural, or combined) can be designed to address the deficits they identify, the alexithymia hypothesis predicts that these will be efficacious for eating disorders, substance abuse problems, empathy and emotion recognition deficits, PTSD, anxiety and more. Whether such an achievement is feasible remains to be seen, but the theoretical implications of Quattrocki and Friston’s model applied to alexithymia are clear in terms of treatment possibilities. At the very least, alexithymia may serve to explain the poor reliability of oxytocin effects. While several studies have found positive effects of oxytocin, several other studies have failed to replicate these effects (Lischke et al., 2012; Scheele et al., 2012; Weigand et al., 2013), even when experimental tasks have been shown to rely upon interoception and to activate the anterior insula (Singer et al., 2008). As demonstrated by Luminet and colleagues (2011), the effects of oxytocin may be limited to those high in alexithymia, suggesting that measuring
alexithymia may represent an easy, a priori, method for identifying those who may be helped by oxytocin administration.
Acknowledgements

We wish to gratefully acknowledge the time and effort spent by Karl Friston in helping us to understand the intricacies of predictive coding and for his comments on an earlier draft of this paper.
References


antecedents subjects compared with normal controls. Psychiatry Res. 119, 251–260.
doi:10.1016/S0165-1781(03)00130-6


doi:10.1016/j.neuropsychologia.2010.07.012


doi:10.1016/j.cognition.2011.11.004


