Decisions and inferences on internal, social and probabilistic information: insights from pharmacological challenge with citalopram and atomoxetine

JAMES J. A. LIVERMORE

Thesis submitted for the degree of Doctor of Philosophy

UNIVERSITY OF SUSSEX

October 2019

SUMMARY

Monoamines are essential neurotransmitters in the functioning of the central nervous system, and monoaminergic agents are widely used in clinical psychopharmacology. They have been linked by previous research to processes including decision making under uncertainty and in social environments, and affective inferences directed at the self and others, all of which show characteristic alterations in psychiatric conditions. However, important gaps remain in understanding the underlying neurocognitive mechanisms that elicit these effects, which if bridged may offer greater insight into the roles of monoamines in both the healthy human brain and in its dysfunction.

This body of work investigates serotonergic links within three overlapping themes of inference on internal, probabilistic and social information, using a combination of pharmacological challenge, behavioural testing, magnetic resonance imaging and measures of cardiac activity. Citalopram, a selective serotonin reuptake inhibitor, was used to manipulate serotonin levels, while the noradrenergic reuptake inhibitor atomoxetine was used to alter levels of noradrenaline and prefrontal dopamine. Using healthy volunteers in double-blind placebo-controlled designs, we carried out a series of behavioural experiments with citalopram and atomoxetine, and neuroimaging experiments with citalopram.

We determined that citalopram but not atomoxetine affected cardiac interoceptive awareness and decisions to sample information prior to making a choice. Pharmacological challenge on both drugs differentially affected paired social interactions depending on whether participants were in the same pharmacologically-induced state or not. We also showed changes in activation of brain regions implicated in interoceptive and emotional processing while carrying out related tasks, with some pharmacological effects, and which showed common areas between the two. This thesis therefore extends the understanding of monoaminergic contributions to essential inferential processes, as well as providing further evidence for shared neural substrates of emotion and interoception.

STATEMENT

This thesis is written in a 'papers-style' format, and each experimental chapter is largely self-contained, with a theoretical overview and general discussion that ties them together. All empirical work contained in this thesis is original research that I carried out during the period of registration on the doctorate programme. I am the lead author of all chapters, and held primary responsibility for experimental design, data collection, analysis and writing. Dr Daniel Campbell-Meiklejohn was the sole supervisor of this project. In this role he is senior author of all experimental chapters, contributing to the design of studies and editing of manuscripts. Prof Hugo Critchley and Prof Sarah Garfinkel provided feedback on chapter 4. Data collection was assisted by PhD students Jo Cutler and Kristian Adamatzky, Masters students Freia Culla-Perarnau. Marusa Levstek and Clare Holmes, and undergraduate students Helen Findlay and Andriana Boudouraki. Medical cover was provided by Dr Gyorgy Moga, Dr James Brittain, Dr Andrew Barritt, Dr Sofia Toniolo and Dr Marco Bozzali. Research was funded by School of Psychology, a grant from Clinical Imaging Sciences Centre for scanning costs, and a grant from Lundbeck Foundation held by Dr Campbell-Meiklejohn.

Some of the work within has been or will be submitted for publication in peer-reviewed journals. Chapter 2 was originally submitted to *Neuropsychopharmacology* and was rejected following review. It has undergone significant revisions since that submission including the incorporation of reviewer feedback and the addition of further analysis.

I hereby declare that this thesis has not been, and will not be, submitted in whole or in part to another University for the award of any other degree.

Signature:.....

ACKNOWLEDGEMENTS

Cliché or no, it remains a fact: this thesis would not have been possible without the help and support of a great many people over the years.

First, I would like to thank my supervisor **Dr Daniel Campbell-Meiklejohn**, whose invaluable advice and guidance helped craft the project. I think, despite the various hurdles along the way, we've managed to do some research to be proud of. Thanks also to all the students in our lab that helped with the testing, without whom I'd be testing for another 4 years.

To my **parents**, who have taken my occasional career 180s in their stride and given love and support regardless. To **Uncle Roy**, for all you've done for me and my education over the years.

To all the great friends I made along the way, to the times we shared and those yet to come:

The **Avocados** (**Tasmin, Callum, Abi, Heather, Meike, Mateo, Becks, Vangelis, Mikey**), the most work-hard-play-hard peers (and honorary peers) that any PhD student with a penchant for distractions could hope to have. Good friends one and all – it's been a blast! Let all of Europe now be our playground.

Lina, and our many great conversations over bottles of wine - I'll convince you about Quantum-Integrated-MicroHomunculus-Theory[1] yet! Thanks for always being there when things were rough.

Diana – five years since the start of the masters, and now we've made it! The coming years will be glorious, you'll see. **Michelle**, another compatriot of that era – your turn now :P

Kiki, a great friend and a fountain of wisdom for all things psychopharmacology and neuroimaging. Thanks for helping no matter how inane the inquiry.

[1] This very serious new theory of consciousness, now appearing in academic print for the first time, will be expanded on in due course

My labmates and fellow travellers – **Jo, Jolyon** and **Lina**. May we still be ranting together five years from now.

My housemate **Felix** – because a PhD student in the midst of writeup is neither a friendly nor a tidy individual.

Claire for listening when I needed it.

Finally, to a certain **Hub** of **Sci**ence, when our library journal subscriptions didn't cut it. Science should be free at the point of consumption.

James

CONTENTS

ABBREVIATIONS USED IN THIS THESIS:

5-HT – serotonin (5-hydroxytryptamine)

5-HTTLPR – serotonin-transporter-linked polymorphic region

AIC – Akaike Information Criteria

ATD – acute tryptophan depletion

ACC – anterior cingulate cortex

BA – Brodmann area

BOLD – blood-oxygen-level-dependent

CMS – cortical midline structures

CSF – cerebrospinal fluid

DMN – default-mode network

DRN – dorsal raphe nucleus

FLAME – FMRIB's Local Analysis of Mixed Effects

FWHM – full-width at half maximum

fMRI – functional magnetic resonance imaging

GABA – gamma-aminobutyric acid

GLM – generalised linear model

IFG – inferior frontal gyrus

IPL – inferior parietal lobule

MDD – major depressive disorder

MPFC – medial prefrontal cortex

NET – norepinephrine (noradrenaline) transporter

PD – Prisoner's Dilemma

PET – positron emission tomography

PCC – posterior cingulate cortex

PFC – prefrontal cortex

RCT – randomised controlled trial

ROI – region of interest

RDBPC – randomised double-blind placebo-controlled

SERT – serotonin transporter

SNRI – selective noradrenaline reuptake inhibitor

SSRI – selective serotonin reuptake inhibitor

SFG – superior frontal gyrus

STG – superior temporal gyrus

TOM – theory of mind

UG – Ultimatum Game

VMPFC – ventromedial prefrontal cortex

1 THEORETICAL OVERVIEW

1.1 INTRODUCTION

The human brain is constantly engaged in assessing information from various modalities, which provide critical input allowing us to maintain homeostasis and navigate an uncertain environment. The central nervous system has evolved an array of neurotransmitters to relay this information and trigger action or further deliberation. Dysfunction in neurotransmitter activity has been repeatedly linked with various psychopathological conditions, with evidence coming from genetics, analysis of endogenous pharmacological substrates and external pharmacological challenge.

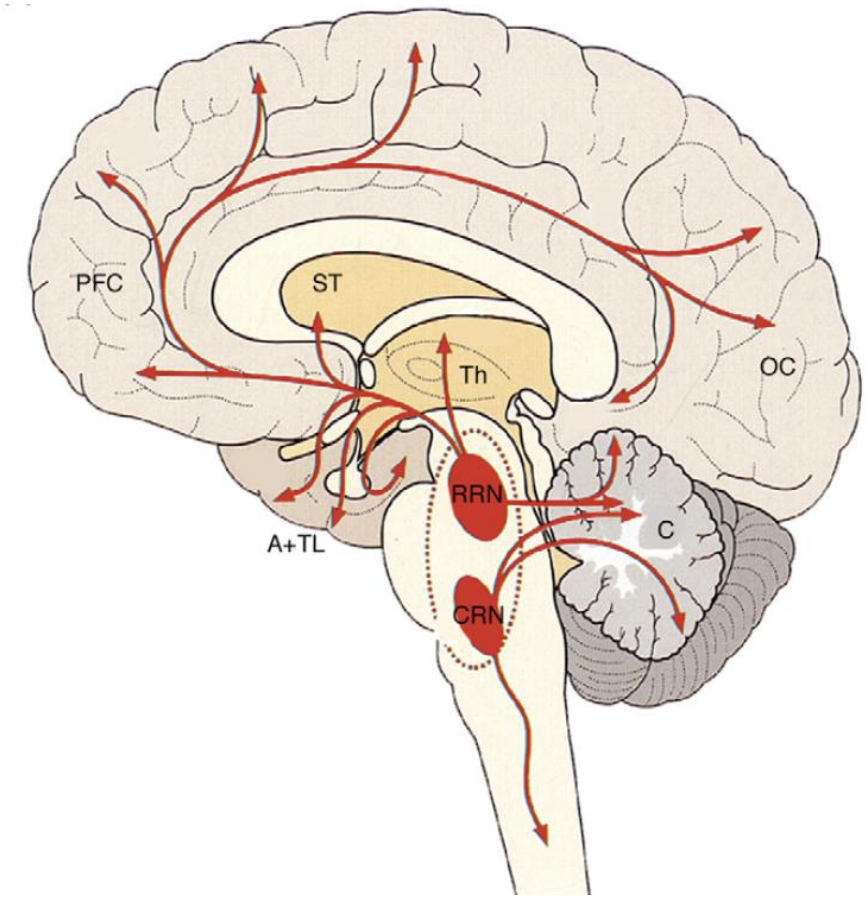
This work focusses on the serotonin system, which has been linked to an array of key functions including maintenance of homeostasis (Ray et al., 2011), emotion (Cools et al., 2008; Catherine J. Harmer, 2008), decision making and learning from reward and punishment (Cools et al., 2011; Daw et al., 2002), and social cognition (Canli & Lesch, 2007; Crockett & Fehr, 2014). While distinct, these functions show considerable overlap aside from common neurochemistry, with somatic and affective influences in decision making (Bechara & Damasio, 2005), common neural substrates of social and monetary reward (Levy & Glimcher, 2012), and close links between emotions and basic physiological signals (Craig, 2009; Seth, 2013).

Our aims in this research were to probe serotonergic roles in neural processes based on internal, social and probabilistic information, while also drawing comparisons with contributions of other monoamine systems. To do this, we conducted a series of double-blind, placebo-controlled studies using two drugs – the selective serotonin reuptake inhibitor (SSRI) citalopram and the selective noradrenaline reuptake inhibitor (SNRI) atomoxetine. This approach allowed us to make direct causal inferences, and where both

drugs were used, gave the ability to examine the specificity of serotonergic effects by contrasting with those of noradrenaline and prefrontal dopamine which are atomoxetine's main targets. Our studies were in healthy populations, but in using drugs with well-researched clinical effects as well as types of task that have shown altered behaviour in clinical samples, we hoped to gain insight into conditions associated with serotonergic dysfunction.

1.2 OVERVIEW OF CENTRAL SEROTONIN SYSTEM

Figure 1: diagram of serotonergic system in the human brain (from Cools, Roberts, & Robbins, 2008). Abbreviations: CRN, caudal raphé nuclei; RRN, rostral raphé nuclei; C, cerebellum; Th, thalamus; A, amygdala; TL, temporal lobe; ST, striatum; PFC, prefrontal cortex; OC, occipital cortex



Serotonin (5-hydroxytryptamine, 5-HT) is a neurotransmitter with a long evolutionary history, and is present in nervous systems across the animal kingdom

(Turlejski, 1996). It is synthesised by the body from the amino acid tryptophan, both centrally and peripherally, with the main source of serotonin release in the brain taking place in the raphe nuclei (Kandel, 2013, p.1040).

In humans, serotonin receptors are situated in high densities across cortical, subcortical and brainstem areas (Pazos et al., 1987a, 1987b). There are at least 17 distinct types of serotonin receptor (Cools et al., 2008), with both inhibitory and excitatory modulatory effects mediated by other neurotransmitters such as dopamine (Alex & Pehek, 2007; Di Matteo et al., 2008), glutamate and GABA (Ciranna, 2006). Figure 1 shows the major serotonin pathways in the human brain. Serotonergic neurons in dorsal and median raphe nuclei in the midbrain project to areas throughout the cortex and basal ganglia. High receptor densities are found in areas related to emotion and valenced learning, including the amygdala, nucleus accumbens and putamen. There are also descending pathways to raphe nuclei from various cortical and subcortical regions (Peyron et al., 1997) including the medial prefrontal cortex (Celada et al., 2001) which exert both excitatory and inhibitory control over the serotonergic system. This may indicate a connection between abnormal prefrontal activity and serotonin dysregulation in depression (Celada et al., 2002): this link has been causally tested with optogenetic activation in rat to induce depressive-like behaviours (Warden et al., 2012).

1.3 REVIEW OF SEROTONERGIC ROLES IN COGNITION, BEHAVIOUR AND EMOTION

1.3.1 Serotonin, emotion and mood disorders

Serotonin's role in emotion and its dysfunction has been researched for over 50 years (Coppen, 1967). Key findings (reviewed in Owens & Nemeroff, 1994) of reduced serotonin metabolites in depressed patients, and the widespread use of serotonin-enhancing SSRIs as first-line treatments for depression and anxiety, formed the basis of the serotonin hypothesis of depression. More evidence came from rapid relapse of remitted depressives in response to acute tryptophan depletion (ATD; for further meta-analysis of ATD in current and remitted depressives see Booij et al., 2002). ATD is a technique to temporarily lower serotonin levels through depletion of the amino acid precursor tryptophan (e.g. Young, Smith, Pihl, & Ervin, 1985) which lowers central serotonergic activity (Crockett et al., 2012). It has however been suggested that lowered tonic activity may increase the dynamic range of phasic activity and give rise to more phasic effects (Cools et al., 2008), with the two modes potentially signalling different valences (Daw et al., 2002).

In mood disorders, serotonergic modulation of patterns of activity across the human brain have been shown. PET studies have shown reduced 5-HT1A receptor binding compared with healthy controls in the raphe nuclei (Drevets et al., 1999) and insula of depressives (Ling Wang et al., 2016), and in raphe nuclei, insula, amygdala and anterior cingulate cortex of social anxiety patients (Lanzenberger et al., 2007). Dutta et al. (2019) showed alterations to resting-state network connectivity in major depressive disorder (MDD) patients compared with controls, which were restored to patterns seen in controls by a single dose of citalopram. Cheng et al. (2017) showed similar changes in MDD patients following 4 weeks of citalopram treatment.

Genetic research on genes affecting serotonergic action, most notably the SERT transporter gene SLC6A4, has provided insight into the link. The S-allele of the 5-HTTLPR polymorphic region, associated with reduced expression of the transporter and lowered serotonergic functioning (Reist et al., 2001; Williams et al., 2001), has been linked in candidate gene studies to depressive symptoms in response to life stress (Caspi et al., 2003) and trait anxiety (Lesch et al., 1996; though possibly only certain subtypes – meta-analysis in Sen, Burmeister, & Ghosh, 2004). A caveat to this is the failure to replicate effects of 5-HTTLPR genotype on depression in two large-scale studies (Border et al., 2019; Culverhouse et al., 2018), indicating that the genetic link with SERT polymorphisms remains controversial and potentially unreliable[2]. Zhang, Liu, Li, Song, & Liu (2015) showed that compared with L-allele carriers, S-allele homozygote males (but not females) showed weaker amygdala-insula resting-state functional connectivity, greater anxiety, and a mediation of 5-HTTLPR effect on anxiety through amygdala-insula connectivity. Social anxiety disorder patients carrying the S-allele showed less activity in amygdala and anterior insula in response to seeing aversive images compared to those homozygous in L-allele, and those with a greater insula response also showed greater intolerance of uncertainty. Recent work in marmoset monkeys showed that individuals homozygous in the low expressing AC/C/G haplotype of the SERT gene showed reduced 5-HT2A receptor binding and lower associated RNA expression in right posterior insula, and higher anxious traits (Santangelo et al., 2019).

While genetic studies and research on individuals with mood disorders offer valuable insight into the importance of serotonergic processes, these should be understood in the context of long-term alterations with concomitant changes in receptor and

[2] These studies are in line with critiques of many candidate gene and candidate gene-environment studies (i.e. studies looking at the effects of specific genetic variants rather than the genome as a whole) as being underpowered and showing publication bias, thus having unfeasibly large effect sizes and inflated false discovery rates (Border et al., 2019; Dick et al., 2015). While the replication failures specifically apply to depression, caution should be taken in the interpretation of results from studies using this approach, which includes all genetic research presented in this chapter.

transporter expression in both serotonin and other neurotransmitter systems, as well as psychological changes induced by long-term changes in neurotransmitter activity. Other evidence illustrates a complex relationship between serotonin, mood and mood disorders that cannot be captured by a purely serotonergic account. Some widely used drugs in the treatment of depression predominantly affect catecholamines with only weak or negligible effects on serotonin, including bupropion and desipramine (Richelson, 1994). Additionally, studies on acute serotonergic manipulation in healthy individuals show somewhat different characteristic patterns to clinical populations[3]. Studies using ATD have shown varying results: Young, Smith, Pihl, & Ervin (1985) showed a general lowering of mood in a male sample, but Benkelfat, Ellenbogen, Dean, Palmour, & Young (1994) in a male sample showed differential effects, with individuals with a multigenerational family history of mood disorders experiencing a negative mood response but not those without. Ellenbogen, Young, Dean, Palmour, & Benkelfat (1996) showed a negative mood effect on a female sample without family history of mood disorder. A review (Van der Does, 2001) concluded that while current and remitted depressives broadly show increase negative mood under ATD, similar findings for healthy individuals with vulnerability were more robust than for those without. In various single dose SSRI studies, no effect on mood was seen despite changes in neural activity (Arce et al., 2008; Del-Ben et al., 2005; Matthews et al., 2010; S. E. Murphy, Norbury, et al., 2009). However, acute studies have also shown various effects on cognition and processing of stimuli, often with direct relevance to characteristic patterns seen in mood disorders. These are reviewed in the following sections.

1.3.2 Serotonin, decision making and learning

With emotion an integral driver of decision making (Lerner et al., 2015), a clear serotonergic link to decision making can be made. In probabilistic decision-making, Murphy

[3] All human psychopharmacology studies reviewed in this chapter used randomized double-blind placebo control (RDBPC) crossover designs with healthy participants of both genders, unless stated otherwise in the text.

et al. (2009) showed that 14 days of tryptophan supplementation reduced loss aversion, which they saw as in keeping with previous research showing serotonergic enhancement reducing sensitivity to negatively-valenced information (Catherine J. Harmer et al., 2004; S. E. Murphy et al., 2006). Crockett, Clark, Smillie, & Robbins (2012) examined the effects of ATD on decisions to sample information, finding that participants increased costly (but not free) sampling as a result, which they posited was due to attenuation of serotonergic avoidance of sampling costs.

The links between serotonin and probabilistic learning have been observed in multiple studies. Chamberlain et al. (2006) and Skandali et al. (2018) showed probabilistic reversal learning deficits with acute citalopram and escitalopram administration respectively, in both cases with an increased tendency to respond to misleading feedback by altering behaviour. Both papers suggest that transiently lowered serotonergic activity may be responsible. Den Ouden et al. (2013) in a large-scale genetic study looking at (amongst others) allelic variations of 5-HTTLPR on the SERT gene showed that individuals homozygous in the L-allele displayed increased lose-shift behaviour in response to punishment. They posited that their findings were compatible with a serotonergic role in attenuating response to punishment, due to L-homozygotes showing increased SERT binding (Willeit & Praschak-Rieder, 2010) and therefore potentially decreased extracellular serotonin levels. This interpretation however conflicts with findings of higher CSF levels of serotonin metabolites in L-carriers compared with S-homozygotes (Williams et al., 2001), where serotonin metabolite levels provide an indirect measure of serotonin function. Evers et al. (2005) showed that ATD was associated with significantly higher activation in dorsomedial prefrontal cortex in male participants during reversal switch errors relative to correct baseline responses.

Research showing low serotonergic activity in impulsivity disorders (reviewed in Dalley & Roiser, 2012), linked with deficits in behavioural inhibition and aversive processing, motivated theories of serotonergic contributions to both (Cools et al., 2011;

Dayan & Huys, 2008). However, perhaps paradoxically, depression is linked to increased aversive processing and reduced behavioural vigour. Dayan & Huys (2008) address this by casting serotonergic action as a mechanism of arresting thoughts and actions that are predicted to lead to aversive outcomes, and that the failure of this system leads to unsignalled (and therefore unpredictable and uncontrollable) negative outcomes from poor decisions leading to a learned helplessness state. Other authors have proposed an opponent process between serotonin and dopamine, with tonic serotonin signalling long-run average reward and phasic serotonin prediction error for future punishment (Daw et al., 2002; Nakamura, 2013).

A complicating factor in understanding decision-making is the widely-shown biases that individuals exhibit when faced with uncertain rewards or punishments. Instead of maximising average reward, most people show a degree of risk aversion (Arrow, 1971; Dohmen et al., 2011; Holt & Laury, 2002) – a sure reward is valued more highly than a probabilistic reward with equal expected value (so receiving a definite £1 will generally be favoured over a 50% chance of £2 vs nothing). Much of the previous literature son information sampling problems has not modelled this, either implicitly assuming risk neutrality or observing probabilities of decisions without modelling the value of their outcome. Importantly, assessing the effectiveness of a decision should be seen in the context of the decider's preferred outcome to gain a normative understanding of whether decision making is impaired under pharmacological challenge – for example, the finding of Crockett et al. (2012) that sampling is increased under ATD cannot alone determine whether this increase would be better or worse for the decider. Increased sampling could be variously consistent with an attenuated serotonergic drive to avoid the costs of sampling (as they posit), or a greater drive for certainty due to either a fear of losing or a deficit in the decider's ability to accurately process the probabilistic information available to them. Disambiguating these possible explanations would necessitate a further measure of the individual's

preferred level of risk, and calculation of the consequent value of the outcomes they receive on that basis.

1.3.3 Serotonin, social perception and social decision-making

Processing of social information is altered in major depression, with negative biases in perception of others' emotions (Bourke et al., 2010; Weightman et al., 2014) and deficits in theory of mind (Weightman et al., 2014), along with altered patterns of neural activity (Cusi et al., 2012). Depression is also widely characterised by impairments in social functioning (Hirschfeld et al., 2000) which may persist even after remission (Kennedy et al., 2007).

Considerable research has looked at serotonergic contributions to social behaviour (for review see Kiser, Steemers, Branchi, & Homberg, 2012). Several studies found serotonergic effects on social cue processing. Lower serotonin levels induced by ATD in a study was shown to reduce appraisals of both attractiveness of positively-valenced faces and intensity of threatening faces compared with placebo (Beacher et al., 2011), suggesting lower sensitivity to social cues, while ATD decreased fear recognition in female but not male participants in a between-subjects study (Harmer, Rogers, Tunbridge, Cowen, & Goodwin, 2003). Acute administration of citalopram increased sensitivity and lowered response latency to expressions of fear and happiness (Harmer, Bhagwagar, et al., 2003) and fear but not happiness (M. Browning et al., 2007) in two between-subjects studies. Simonsen et al. (2014) found in females that acute citalopram made participants rate faces as less trustworthy and adjust their ratings more to the negative judgements of others, suggesting serotonergic effects on both social cues themselves and on conformity to others' opinions.

Selvaraj et al. (2018) looked at the neural correlates of processing fearful expressions under a single-blind acute citalopram challenge in males, using fMRI and PET. They found increased bilateral amygdala activation in response to fearful vs neutral faces under citalopram vs placebo, and a significant correlation between left amygdala BOLD

response to both fearful and happy faces (vs neutral) and 5-HT1A receptor availability in the dorsal raphe nucleus (DRN). They interpret the latter finding in the context of presynaptic 5-HT1A autoreceptors in the DRN inhibiting 5-HT release and neural firing (Barnes & Sharp, 1999; Bosker et al., 1996; Sprouse & Aghajanian, 1987). By contrast, Murphy, Norbury, O'Sullivan, Cowen, & Harmer (2009) showed reduced right amygdala activity to fearful faces in a between-groups study. Cools et al. (2005) showed a comparable effect of fearful faces on bilateral amygdala activation during ATD on male participants as a function of threat sensitivity as measured with the Behavioural Inhibition Scale, and Hariri et al. (2002) showed that fearful faces generated higher right amygdala activity in 5-HTTLPR S-carriers compared to L-homozygotes. Klumpp et al. (2014) found that social anxiety disorder patients homozygous in 5-HTTLPR S-allele displayed greater insula activation to fearful faces than those homozygous in the L-allele.

In social decision-making behaviour, increased punishment of social conspecifics in response to unfair or provocative choice behaviour was shown from ATD in two studies (Crockett, Clark, Tabibnia, Lieberman, & Robbins, 2008; Marsh, Dougherty, Moeller, Swann, & Spiga, 2002), with the latter study also showing the opposite effect from acute tryptophan augmentation. Two studies using tryptophan augmentation administered over 15 and 12 day periods showed decreased quarrelsome behaviour, the first in individuals selected for high trait quarrelsomeness, although in the latter study only in participants administered the placebo condition first (aan het Rot et al., 2006; Moskowitz et al., 2001). Citalopram was seen to decrease the acceptability of hypothetical outcomes where the participant would cause harm to others, as well as decreasing rejection of unfair offers in an ultimatum game (UG) (Crockett, Clark, Hauser, & Robbins, 2010). The latter finding evidences an opposite effect from Crockett et al's earlier ATD study (Crockett et al., 2008), consistent with serotonergic enhancement. Similarly, Tse & Bond (2002) showed increased cooperative behaviour and communication following two weeks of citalopram administration. Wood, Rilling, Sanfey, Bhagwagar, & Rogers (2006) showed that ATD reduced decisions to

cooperate, although only for the treatment order where participants received ATD treatment in the first session.

While offering useful insights, previous serotonergic work on social decision-making largely used confederate designs with a set response pattern of the other party. While this approach is widely used to avoid confounds of stimuli differing between conditions, it has limited generalisability to most real social decision-making problems where interaction is a dynamic process characterised by reciprocal effects of interactors on one another's behaviour. This is particularly pertinent due to specific psychosocial dysfunction in depression that occurs independently of cognitive and mood effects (Knight & Baune, 2019).

1.4 INTEROCEPTIVE PROCESSES AND POSSIBLE SEROTONERGIC LINKS

1.4.1 Interoceptive processes, emotion and dysfunction

Interoception is the perception of internal bodily states. A large and growing body of literature has assessed its importance in a variety of cognitive and affective processes, and mapped out characteristic biases that are shown in psychopathology. Considerable theoretical work has also been done to place a central role for interoception in affective disorders.

Key drive states to maintain homeostasis, and therefore continued survival of the organism, are conveyed by interoceptive information. The primary path of these signals is through afferent cranial nerves, most notably the vagus nerve. These convey motivationally-relevant information including hunger, thirst, temperature and dyspnoea (Burki & Lee, 2010; Critchley & Harrison, 2013; Székely, 2000; Zimmerman et al., 2019) from organs such

as the heart and gastrointestinal tract to the brain, with outputs in the nucleus of the solitary tract, parabrachial nucleus, periaqueductal gray matter and ventromedial posterior thalamus. The latter region relays information to amygdala and insula, and viscerotopic organisation of both thalamus and insula has been shown (Aleksandrov & Aleksandrova, 2015; Cechetto & Saper, 1987). Within the insula, primary interoceptive representations are localised in posterior regions, and are then re-represented and integrated across modalities in the anterior insula to give rise to subjective interoceptive sensations (Craig, 2009).

Converging evidence places interoceptive processes at the heart of emotion (Craig, 2009; Garfinkel & Critchley, 2013; Quadt et al., 2018; Seth, 2013). The theoretical framework of predictive processing posits hierarchical organisation of cognitive functions where the brain generates top-down inferential models (*priors*) of the causes of current states which are then tested against sensory evidence, with their corresponding precisions (inverse variance) determining how priors and evidence are weighted in the resulting percept. Integrating interoception within this framework, some theories place interoceptive inference as the basis of selfhood (Seth, 2013; Seth Anil K. & Friston Karl J., 2016) with emotions as arising from predictive models of the causes of interoceptive signals.

In keeping with this link, disorders of emotion would be expected to show altered interoception. This has been shown in a variety of mood disorders including depression and anxiety (reviewed in Paulus & Stein, 2010), panic disorder (Ehlers, 1993) and eating disorders (B. Herbert & Pollatos, 2019; Klabunde et al., 2013). Depressive symptoms have been shown to blunten interoceptive sensitivity (Furman et al., 2013; Pollatos et al., 2009) and anxiety symptoms to heighten it (Pollatos et al., 2009), with a possible interaction (Dunn, Stefanovitch, et al., 2010). In panic disorder patients compared to controls, pharmacological challenges including yohimbine (α2-adrenoreceptor antagonist; Gurguis, Vitton, & Uhde, 1997), isoproterenol (β-adrenoreceptor agonist; Pohl et al., 1988), caffeine (adenosine receptor antagonist; Charney, Heninger, & Jatlow, 1985), and carbon dioxide

(Rassovsky & Kushner, 2003) have been shown to increase the intensity of interoceptive sensations and cause panic responses. Drawing on this evidence, Paulus and colleagues (Paulus et al., 2019; Paulus & Stein, 2010) cast disorders such as anxiety and depression as a disorder of interoception. In this theory, anxiety is characterised by excessive weights placed on interoceptive over exteroceptive signals (such as tachycardia in response to an ambiguous stimulus). Coupled with hyper-precise priors excessively weighting a negative model of the world (e.g. expecting a threat rather than an innocuous outcome) and context rigidity which generalises negative models to new contexts, the result is exaggerated threat responses in a variety of situations which do not adapt sufficiently when the ambiguity is resolved in favour of an unthreatening outcome. With persistence of these states, learned helplessness and depression result. Importantly, this framework distinguishes between the effects of transient physiological states or changes in interoceptive processing, and long-term effects, with the former leading to dysfunction only with repeated exposure or if biased interpretation has been established through other means (e.g. genetic predisposition).

In the human brain, interoceptive processes have been closely linked to activation in the insula. An fMRI meta-analysis by Schulz (2016) showed overall significance in bilateral clusters in the insula for interoceptive focussed tasks, and 8 of the 9 component studies also identified this area. The insula has also been implicated in affective states from emotional stimuli (Phillips et al., 2003) and emotional recall (Phan et al., 2002; Reiman et al., 1997), while depressed individuals show both altered insula activity (Drevets et al., 1999; Hamilton et al., 2012) and decreased grey matter volume (H. Zhang et al., 2016). Interaction between interoceptive and emotional experience in the insula has been shown (Terasawa, Shibata, et al., 2013; Terasawa et al., 2014).

Serotonin receptors are reduced in brain regions including the insula in depressives (meta-analysis in Wang et al., 2016) and social anxiety patients (Lanzenberger et al., 2007), while individuals carrying genes associated with lower serotonergic function also show weaker insula connectivity (L. Zhang et al., 2015). Moreover, previous research located

specific effects of SSRI challenge on the insula during emotional processing related to the self and others (Arce et al., 2008; Matthews et al., 2010). There are therefore good reasons to suspect that interoceptive processes in the brain may be modulated by serotonergic activity, though this has not yet been tested. With the links shown between emotional and interoceptive processes, an understanding of serotonergic influence may help to develop a fuller picture of how these are mediated by neural systems.

1.4.2 Interoceptive processes and probabilistic decisions

Research has linked interoceptive processes to the domain of risk-related (i.e. probabilistic) decisions. Bechara & Damasio (2005) review evidence that decision making is guided by 'somatic markers', or changes in bodily (somatic) states that are associated with emotion. They posit a mechanism by which primary inducers (innate or learned stimuli with valenced association) trigger somatic states via the amygdala, which cause signals in areas such as insula that correspond to the subjective feeling of the state. Subsequently, secondary inducers are formed from these feelings, which are memories or hypothetical mental imagery that then may also trigger the corresponding somatic state when recalled or imaged. Hence according to the theory, interoceptive processes that detect and utilise the somatic markers are crucially involved in this appraisal process and in consequent decision making.

In line with this, Werner, Jung, Duschek, & Schandry (2009) found that higher interoceptive ability predicted better success in a gambling task. Dunn, Galton, et al. (2010) found that interoceptive ability could be associated with positive or negative effects on decision making depending on whether bodily responses were in line with better or worse decisions, with the latter potentially varying as a function of pre-existing biases. Kandasamy et al. (2016) found that for investment traders (whose occupations involved continual probabilistic decision making, potentially allowing negation of unhelpful biases through experience), interoceptive ability was positively related to profitability and career success.

Clark, Studer, Bruss, Tranel, & Bechara (2014) found that damage to insula, an area with strong links to interoception (Schulz, 2016), attenuated cognitive distortions in gambling tasks showed by healthy controls. Thus some evidence, viewed through a scope of interoception, suggests differential contributions of bodily signals based on existing biases. The interoceptive reappraisal would then lend weight to (potentially unconscious) signals predicting a valenced outcome, but the results of this process may be dependent on established contingencies, accurate or otherwise. Hence, as with the Paulus et al. (2019) theoretical framework discussed in the previous section, transient changes in interoceptive processes may not manifest in changes to decision making. Any changes may instead be contingent on a longer-term change in learned biases, potentially as a function of experience.

1.4.3 Interoceptive processes and social cognition

Understanding others' behaviours has been shown to have interoceptive correlates. Direct links were shown between interoceptive sensitivity and recognition of emotion on faces (Terasawa et al., 2014), and cognitive and affective empathy for the pain of others (Grynberg & Pollatos, 2015). Shah, Catmur, & Bird (2017) found that cardiac interoceptive ability was positively correlated with empathetic (but not cognitive) theory of mind, while Bernhardt & Singer (2012) review evidence that these processes recruit a set of common brain areas including anterior insula and dorsal anterior to mid-cingulate cortex.

In social decision-making, Piech et al. (2017) showed that higher interoceptive sensitivity was associated with more altruistic behaviour, while Lenggenhager, Azevedo, Mancini, & Aglioti (2013) showed that exposing participants to sounds from their own heart (rather than from another's heart or footsteps) biased offers in an Ultimatum Game towards greater self-centredness.

Autistic spectrum individuals experience characteristic difficulties in social cognition. Garfinkel et al. (2016) showed that individuals with high functioning Autistic

Spectrum Disorder exhibited a dissociation between objective and subjective measures of interoception, with poorer interoceptive accuracy but high subjective ratings of interoceptive sensation, though later work implicated specific alexithymic traits rather than the general autism construct (Mul et al., 2018; Shah et al., 2016).

1.5 PHARMACOLOGICAL MANIPULATION OF NEUROTRANSMITTER SYSTEMS

In this section, I outline the characteristics of the two psychoactive drugs used in this research – citalopram and atomoxetine. Specific details of the studies conducted, the hypotheses and rationale behind using each of these drugs are given in the next section.

1.5.1 Citalopram – selective serotonin reuptake inhibitor

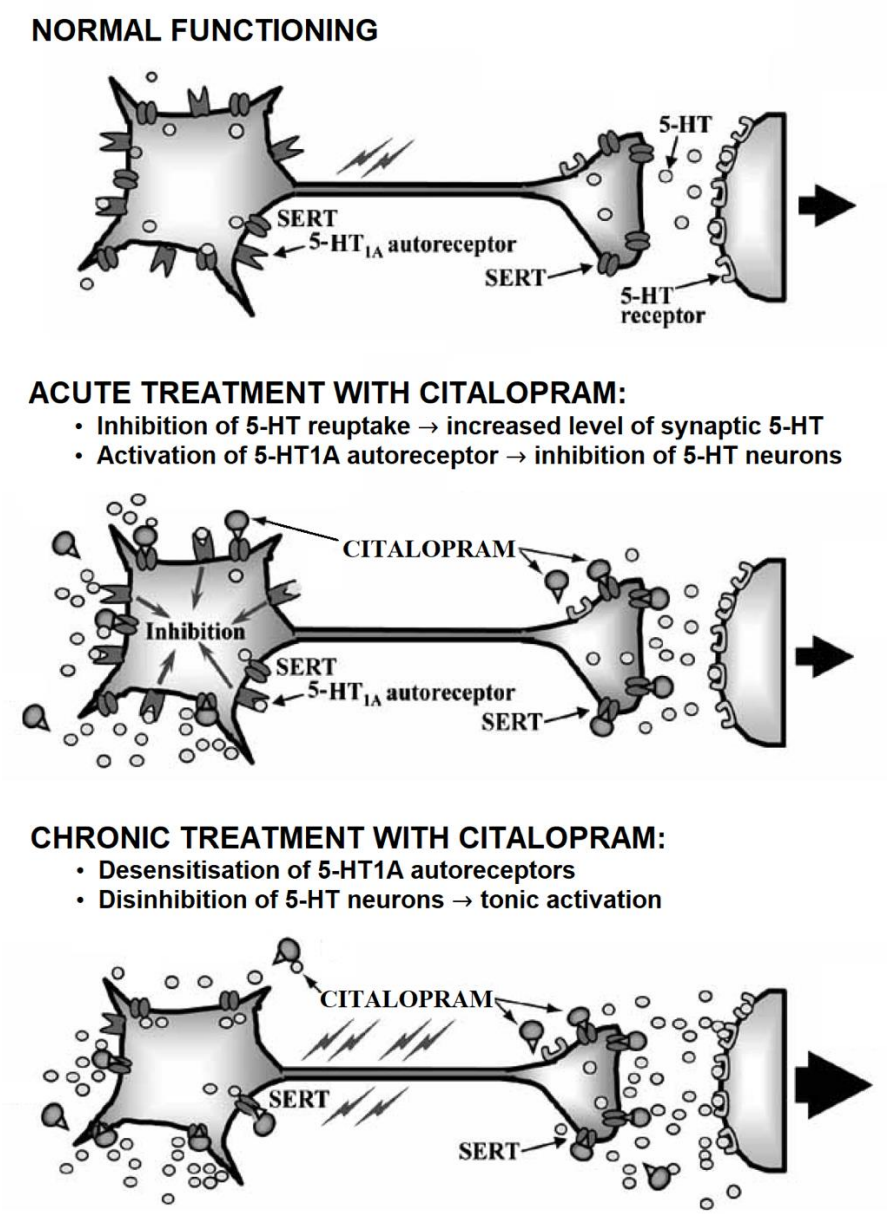
1.5.1.1 Pharmacology and mechanism of action

Citalopram is the racemic mixture of 50% (R)-(−)-citalopram and 50% (S)-(+)-citalopram. The S-enantiomer is considered to be the active component (Hyttel et al., 1992), with the R-enantiomer inhibiting the S-enantiomer's effect (Sánchez et al., 2004). It is a highly selective serotonin reuptake inhibitor, which binds to the serotonin transporter (SERT) with 3,870 times the affinity than to the norepinephrine transporter and 10,300 times the affinity than to the dopamine transporter (Michael J Owens et al., 2001). It is used clinically to treat a range of mood disorders including depression and anxiety (Kupfer et al., 2012; Nutt, 2005).

Plasma serum concentrations of citalopram reach peak levels at between 2 and 4 hours after oral administration (Lundberg et al., 2007; Milne & Goa, 1991). Figure 2 shows the hypothesised mechanism of action (Hjorth et al., 2000; Ohno, 2010). In summary, under normal functioning, synaptic release of serotonin after presynaptic neuronal firing allows serotonin to bind to postsynaptic receptors and trigger firing of the postsynaptic neuron, but presynaptic transporters then remove (*reuptake*) the neurotransmitter to limit further activity. When citalopram is introduced, it occupies SERT in the presynaptic neuron axon terminals and reduces reuptake of 5-HT, leaving more in the synapse to continue activation

of the postsynaptic neuron. In parallel, SERT occupancy in the somatic dendrites of the presynaptic neuron allows more binding of 5-HT to 5-HT1A autoreceptors which inhibits 5-HT synaptic release, reducing the effect of the drug. In chronic administration, autoreceptors desensitise, removing the inhibitory effect and leading to a net increase in postsynaptic neural activation.

Figure 2: theoretical mechanism of action of citalopram (adapted from Ohno, 2010). Note that relative levels of post-synaptic neuronal activation in the acute phase are debated, and that desensitization of the 5-HT1A receptor can be caused by decreased 5-HT1A receptor density (as shown) or by functional desensitization (reduced signal transduction, not shown). Autoreceptors on the presynaptic axon may also play a more limited role (Hjorth et al., 2000)



Using PET with the [^{11}C]MADAM radioligand in males, acutely administered citalopram has been shown to rapidly bind to the 5-HT transporter (Lundberg et al., 2007), occupying 66-78% of sites after 6 hours depending on brain region, comparable with the ~80% after 4 weeks of administration (Meyer et al., 2004). Nonetheless, as illustrated in Figure 2, citalopram also acts on autoreceptors of the presynaptic neuron to inhibit further serotonin release. Studies in rats showed decreased firing of serotonergic dorsal raphe neurons after short-term administration (Chaput et al., 1986; El Mansari et al., 2005), although in doses much higher than those clinically used in humans. There is evidence that 5-HT1A autoreceptor blockade through co-administration of SSRIs with 5-HT1A antagonist UH-301 (Hjorth, 1993) and 5HT1A/β-adrenoreceptor antagonist pindolol (Romero et al., 1996) increases extracellular 5-HT concentrations in rats, with pindolol also showing acceleration of SSRI clinical effectiveness in humans (Ballesteros & Callado, 2004). However, evidence for delayed clinical effectiveness of SSRIs alone is mixed, with large-scale meta-analyses showing improvement after one week of sustained dosage (Taylor et al., 2006), although less than at 6 weeks. The literature on single-dose administration also shows evidence consistent with either reduced or increased serotonergic function (see section 1.3).

1.5.1.2 Dosages and side effects

Citalopram in clinical treatment for depression is often started at a daily dose of 20-40mg (Milne & Goa, 1991), though modern clinical guidelines recommend 20mg (NICE, 2019b). Research in healthy volunteers using orally-delivered acute citalopram has frequently used 20mg (M. Browning et al., 2007; Grillon et al., 2007; S. E. Murphy, Norbury, et al., 2009), although dosages of 30mg have also been used (Chamberlain et al., 2006; Crockett et al., 2010; Nandam et al., 2011). Due to a need to balance experimentally-relevant effects with adverse side effects, as well as ethical requirements, we used 20mg for all citalopram studies.

Citalopram is generally considered to be well-tolerated, but with some common side effects including nausea, headache and dizziness (Ekselius et al., 1997). The former is in keeping with peripheral changes in serotonin activity (Anderson, 2004).

1.5.2 Atomoxetine – selective noradrenaline reuptake inhibitor

1.5.2.1 Pharmacology and mechanism of action

Atomoxetine is an SNRI. It binds to the norepinephrine transporter (NET) with 15.4 times greater affinity than the serotonin transporter and 290 times greater than the dopamine transporter, and in rat models it did not show an effect on extracellular serotonin levels (Bymaster et al., 2002; Koda et al., 2010). It is used as a treatment for attention deficit hyperactivity disorder (ADHD).

After oral administration, plasma levels of atomoxetine reach their peak in 1-2 hours (Sauer et al., 2005). As a reuptake inhibitor, atomoxetine blocks NET to prevent its reuptake action. However, unlike SERT, NET also transports dopamine (Horn, 1973; Raiteri et al., 1977). NET shows high concentrations in the prefrontal cortex, and is the main mechanism for dopamine clearance there (Morón et al., 2002). Atomoxetine has been shown to raise prefrontal dopamine as well as noradrenaline levels in mouse (Bymaster et al., 2002) and rat (Koda et al., 2010; Swanson et al., 2006), and greater dopamine and noradrenaline-related neural firing in the region in an inverted U-shape dose-response curve in monkey (Gamo et al., 2010). As with citalopram, autoreceptor effects attenuate the effect of the drug, but the net change in neurotransmitter levels of prefrontal regions remains positive (Swanson et al., 2006). Unlike dopaminergic agents however, it does not increase dopamine levels in limbic areas such as striatum and nucleus accumbens (Bymaster et al., 2002) due to sparse NET in those regions. The prefrontal effects of atomoxetine are thought to be behind its clinical mechanism, redressing characteristic deficits in behavioural inhibition and attentional regulation that are present in ADHD (Arnsten, 2009).

1.5.2.2 Dosages and side effects

For the treatment of ADHD, current guidelines for dosage of atomoxetine is 40mg daily (NICE, 2019a). Single-dose research on healthy volunteers has frequently used either 40mg (Chamberlain et al., 2009; Warren et al., 2017) or 60mg (Chamberlain et al., 2006; Marquand et al., 2011; Nandam et al., 2011). Similarly to citalopram, the side effect profile of atomoxetine includes nausea and dizziness (Caballero & Nahata, 2003; Michelson et al., 2003). There is evidence associating atomoxetine with increases in heart rate and blood pressure (Wernicke et al., 2003). Due to a need to balance associated risks, we used 40mg in studies involving this drug.

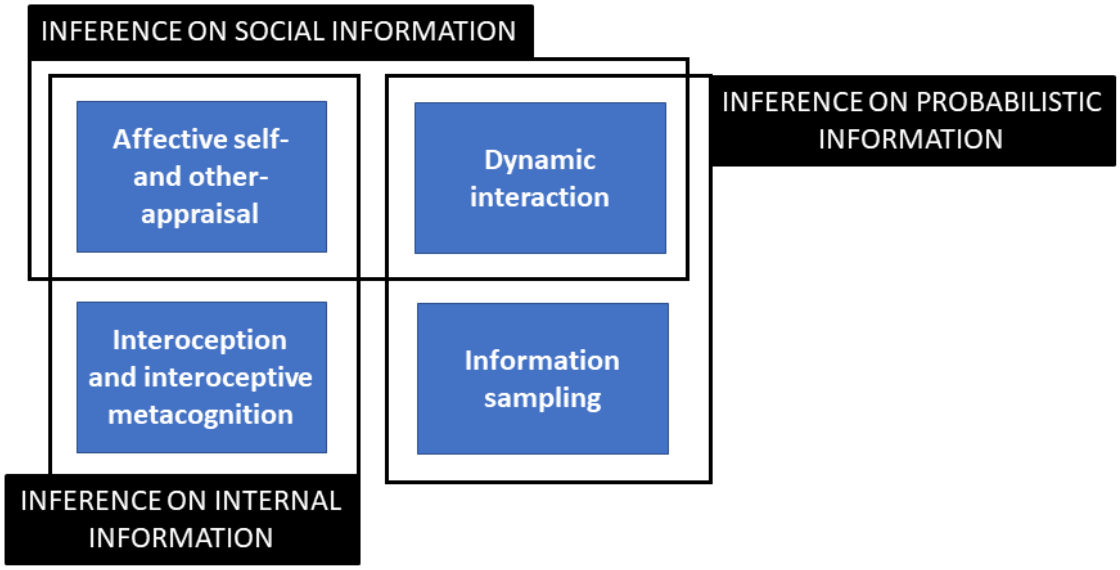
1.5.2.3 Utility as a high-level control

As another monoaminergic reuptake inhibitor, atomoxetine has useful properties to facilitate comparisons with citalopram for specific serotonergic effects against catecholaminergic effects and non-specific drug effects. The citalopram-atomoxetine comparison has been used to infer differential effects in several previous studies (Chamberlain et al., 2006; Nandam et al., 2011, 2014; Ye et al., 2016).

1.6 OVERVIEW OF THIS RESEARCH

1.6.1 Aims of this research

Figure 3: overview of the research themes



In reviewing the current literature on serotonergic contributions to decisions and inferential processes, we identified several areas in which important gaps remain in understanding the underlying neurocognitive mechanisms that elicit these effects, which if bridged may offer greater insight into the roles of serotonin in both the healthy human brain and in its dysfunction. Specifically, we found that:

1. Previous research into decisions to sample probabilistic information under serotonergic challenge had not been able to quantify normative effects through modelling characteristic decision patterns under risk. This gave limited explanatory power to understand whether alterations in patterns of information sampling could also change the effectiveness of decisions in terms of value to the decision maker.

2. Despite indications of important links between serotonergic systems and social decision-making, prior research did not incorporate dynamic modelling of interaction – where interacting individuals' choices affect one another's behaviour,

and thus repeated interaction leads to changes in incentives to cooperate and compete. This gave them limited generalisability to actual decisions in a social environment.

3. The neurochemical mechanisms behind interoception had not been identified despite its pivotal importance to affective processes and clinical disorders. There are strong reasons to suspect serotonergic involvement, but to date no studies have tested this hypothesis. Additionally, while research showed that interoceptive decisions (like decisions in other perceptual modalities) demonstrate dissociations between first-order and metacognitive levels, neurotransmitter contributions at different levels had not been examined, nor had their interactions with metacognitive processes in other domains.

4. Despite the relevance of self- and other-related affective processing to mood disorders that present serotonergic dysfunction and are linked to interoceptive processes, their neural correlates and overlap with interoceptive processes in the brain had not been mapped out under serotonergic challenge.

This body of work aimed to address each of these points with a series of experiments using psychopharmacological challenge in healthy volunteers. The central aim was to examine the effect of serotonergic manipulation with citalopram on these processes, but research using atomoxetine is also presented here. It was used with the same experimental tasks as citalopram, as a high-level control or to provide a contrast between serotonergic action and that of other monoamines. Where this was drug was used (in chapters 2 and 3), the individual chapters give details of its hypothesised effects in each context.

1.6.2 Overview of empirical work

Testing of participants was conducted in three periods. The first was a behavioural study using citalopram, tested in the period February and March 2017. The second, a

behavioural study using atomoxetine with an identical set of tasks to the first study, was tested in September 2017. The third consisted of behavioural and neuroimaging components, and was tested in the period June to November 2018. Chapters 2 and 3 are based on data from the first two testing periods. Chapter 4 is based from data from the second and third testing periods. Chapters 5 and 6 are based on data from the third testing period.

This thesis has five experimental chapters. Throughout, we used a double-blind placebo crossover design, where participants were tested in two sessions. Assignment to treatment order was randomised and counterbalanced for gender by a researcher with no contact with the participants, while all researchers, medical personnel and participants involved in testing had no knowledge of treatment order. However, participants were aware of which drug was being tested during that testing period.

1.6.3 Overview of studies

This section provides an overview of the experimental chapters and their main findings. For further discussion see the individual chapters and the general discussion in chapter 7.

Chapter 2 is titled '**Acute citalopram administration results in impaired information sampling'**. It focusses on the links between serotonergic activity and decisions to sample information, contrasting the citalopram manipulation with atomoxetine. We adapted an existing task, the Information Sampling Task (Clark, Robbins, Ersche, & Sahakian, 2006), with a new analysis which furthered Bayesian approaches by taking account of their prior updating as the task progressed. We also tested participants on a separate task to measure risk preferences, using the results of this to parameterise individual utility functions. We modelled changes in their behaviour as assessed by both an existing measure - their probability of giving the correct response given the information available – and the computed utility of their decisions. We found that citalopram reduced

both the utility of choices and the probability of being correct, indicating a failure to reduce or increase information seeking in response to collected evidence. It did not change the overall information gathered. Atomoxetine did not influence any IST measure.

Chapter 3 is '**Monoaminergic challenge in strategic decision making**'. This chapter looks at the effects of citalopram and atomoxetine in a strategic decision-making task. Participants were tested in pairs in an actual repeated interaction (without confederate or pre-determined responses) through a task in which their rewards depended on both of their actions. The task was designed to have conflicting incentives to compete and cooperate, with a bonus trial structure to manipulate the relative levels of these incentives. We found that neither drug significantly affected behaviour when examined on an individual basis, but a post-hoc analysis based on whether the pair members were in the same drug condition or not (i.e. the drug of the testing period, or placebo) found that scores were higher (indicative of more cooperative outcomes) when they were in the same than different conditions. This was the case for both drugs, and was a better predictor of scores than individual drug condition or whether the participants were of the same gender (tested to eliminate possibilities of same vs different gender changing levels of cooperation). Additionally, we developed a learning model that generalised reinforcement learning to also reinforce the other pair member's reward. We found that this model fit better than either basic reinforcement learning, the Experience-Weighted Attraction model from previous game theory literature, or a null model based on static probabilities. Analysis on parameters of this model showed specific effects on the weighting of the other's reward, which was more positive when pairs were in the same condition, while other parameters did not significantly differ.

Chapter 4, '**Serotonergic effects on interoception and interoceptive metacognition**', tests a serotonergic theory of interoception. Testing participants on two widely-validated interoception tasks, the heartbeat discrimination and tracking tasks, we looked at effects of citalopram on measures of three levels of interoception – accuracy,

sensibility (mean confidence) and awareness (trial-by-trial correspondence of the two). We found that citalopram enhanced interoceptive awareness as measured by the discrimination task, and interoceptive accuracy measured by the tracking task. We thus found evidence for selective serotonergic effects on interoception.

Chapter 5 is titled '**Neural correlates of serotonergic effects on interoception**'. In this study, we carried out functional neuroimaging with a citalopram challenge, using a task that contrasted interoceptive focus in two modalities (heart- and stomach-focus) with exteroceptive visual focus on a target. We replicated findings from the literature during interoceptive focus in several brain regions: precentral, postcentral, superior temporal, middle temporal and lingual gyri, precuneus and cuneus. We also found an effect of citalopram increasing activity during heart and stomach focus in lateral orbitofrontal cortex, and during stomach focus alone in superior frontal gyrus, inferior parietal lobule and posterior cingulate. These findings partially overlapped with the default mode network (DMN), which has a variety of self-related processing roles, and showed a parallel with previous research showing hypoactivation of lateral orbitofrontal cortex during interoception in depressives.

Chapter 6, '**Emotional self- and other-directed appraisal: neural substrates and effects of citalopram challenge'**, is based on a second neuroimaging task. Participants were displayed pictures of neutral and fearful faces, and directed to focus on either their own feelings or on the feelings of the person displayed in the picture. We found recruitment of several areas differentially increased during self- compared to other-focus, including DMN structures in anterior cingulate/medial prefrontal cortex and temporoparietal junction, and the insula. Connectivity analysis revealed citalopram effects on connectivity between superior frontal gyrus and parahippocampal gyrus that were higher in other- than self-focus. We also identified several clusters in the conjunction between activations in self-focus and interoceptive-focus data from the previous chapter, again mostly in DMN areas,

that provided further evidence for common substrates of emotional appraisal and interoceptive focus.

2 ACUTE CITALOPRAM ADMINISTRATION RESULTS IN IMPAIRED INFORMATION SAMPLING

2.1 ABSTRACT

Background

Gathering and evaluating information leads to better decisions, but often at cost. The balance between information seeking and exploitation (i.e. making a choice) features in in neurodevelopmental, mood, psychotic and substance-related disorders. Serotonin's role has been shown through experimental reduction of its precursor, tryptophan. We tested the boundaries and applicability of these effects by asking whether they would be observed with modern analysis, after controlling for specific decision factors, and with clinically-relevant serotonergic and catecholaminergic treatments.

Methods

Behaviour on an established Information Sampling Task (IST) showed that that normal sampling was consistent with a degree of risk aversion, suggesting that risk preferences should be incorporated into IST analysis. Next, we tested two groups of healthy volunteers on the IST using a modified task and a within-subject, double-blind, placebo-controlled design. One group was tested on/off 20mg of citalopram and the other on/off 40mg of atomoxetine, both orally administered. They were also tested for risk preference in a separate task. We introduce a new analysis method that also incorporates possible learning of likely states of sampling spaces across trials on the IST, and modelled their responses both with an existing measure of probability of correctness and with a expected utility approach to look at the normative effectiveness of information gathering choices accounting for their risk preferences.

Results

Citalopram reduced the utility of choices and the probability of being correct, indicating a failure to reduce or increase information seeking in response to collected evidence. It did not change the overall information gathered. Atomoxetine did not influence any IST measure.

Conclusions

Clinically-relevant shifts of serotonin impair the use of acquired information for choices to sample again.

2.2 INTRODUCTION

How does one decide they have enough information to form a belief and make a choice? Efficient sampling and evaluation of environments will enhance the accuracy, efficiency and net benefit of choices, though this can come at the cost of energy, forgone opportunity and, though more apparent in other species, exposure to harm and predation.

Problems from sampling too little or too much information occur in a variety of psychopathologies. Examples include 'jumping to conclusions' with insufficient information in psychosis (Dudley et al., 2016), and sampling less information in depression (Tavares et al., 2007), Parkinson's patients (Djamshidian et al., 2013), substance use disorders (Clark et al., 2006) and binge drinking (Banca et al., 2016). Increased information gathering is observed in patients with obsessive-compulsive disorder and people high on the compulsivity spectrum (Hauser, Moutoussis, Dayan, et al., 2017; Hauser, Moutoussis, Iannaccone, et al., 2017). Formal experimental models of such biases provide specific insight into decision processes associated with these disorders and new targets for their treatment.

Information sampling relates to the psychological construct of 'reflection impulsivity', the tendency to make decisions without gathering and effectively evaluating

information. Reflection impulsivity is dissociable from motor and temporal impulsivity forms (Caswell et al., 2015). In recent decades it has been studied with the Information Sampling Task (IST; Clark et al., 2006) and urn tasks (FitzGerald et al., 2015). In each, participants choose a quantity of information to sample, at a defined cost, before making a decision that will lead to a reward or penalty. These relate to the explore-exploit learning theory framework, as an individual explores the information available, before committing to a choice by exploiting this information (Averbeck, 2015). More samples detract from the potential reward (in the 'Decreasing Win' condition of the IST), but increase the likelihood of a correct decision. The two competing accounts must strike an optimal balance.

Serotonin has received attention for its links to various closely related concepts: impulsivity in general (Dalley & Roiser, 2012), avoidance of aversive outcomes (Cools et al., 2011), and risk preferences (S. E. Murphy, Longhitano, et al., 2009). Murphy et al. showed tryptophan supplementation reduced loss aversion in probabilistic decision making, while acute tryptophan depletion (ATD) manipulations have been shown to have effects on risk preferences in rats (Koot et al., 2012) and macaques (Long et al., 2009). Crockett, Clark, Smillie, et al. (2012) examined the effects of ATD on IST behaviour. Participants increased costly sampling when tryptophan (serotonin's precursor) was depleted, while sampling without cost was unaffected. It was posited that serotonin may reduce avoidance of local costs, relative to the prospect of a future global loss. Similarly, in other contexts, ATD has been shown to increase the deferral of complex decisions to make purchases (Lichters et al., 2016).

Comparable studies using clinical doses of serotonergic medication have not yet been reported. Citalopram is a highly selective SSRI (Michael J Owens et al., 2001) widely used in the treatment of depression and various anxiety disorders. Chamberlain et al. (2006) explored the effects of acute citalopram, a commonly prescribed selective serotonin reuptake inhibitor, on probabilistic reversal learning. The research showed that a single 30mg dose of citalopram could impair learning about a changing reward environment. This

led to our hypothesis that serotonin, by way of a clinical dose of citalopram, may also change learning in information sampling contexts. To contrast serotonergic effects with those of other monoamines and to employ an active control with similar side effect profile, we also tested a group with atomoxetine. Atomoxetine is a specific noradrenaline reuptake inhibitor that increases prefrontal noradrenaline and dopamine levels (Bymaster et al., 2002; Koda et al., 2010; Swanson et al., 2006) and is associated with improved inhibitory control (Chamberlain et al., 2009; Rae et al., 2016). It has been linked to lowered impulsivity, but crucially in other subtypes of motor and temporal impulsivity (Bizot et al., 2011; Chamberlain et al., 2006) rather than the reflection sub-type. The Chamberlain et al. study that showed the effect of citalopram on reversal learning also showed no effect of atomoxetine. Furthermore, tests of acute atomoxetine (40mg oral dose) on a task that modelled both random and directed exploration showed effects on random but not directed exploration (Warren et al., 2017). As sampling is a directed strategy, i.e. a strategy seeking information that can be used to obtain future reward (Wilson et al., 2014), we predicted a null effect on this task.

Recent advances in the field have reformulated behavioural models of information sampling and resulting decisions. From the point of view of the decider, the probability of a decision being correct on the IST, based on information available, can be formalised as a Bayesian inference problem (Bennett et al., 2016; FitzGerald et al., 2015). Through this, a normative set of choices (an 'ideal observer') can be formulated as an upper bound on individuals' ability to effectively predict outcomes from gathered information. As real-life decision making is often suboptimal, this allows quantification of how an individual or group may differ from the ideal – a measurement of how effectively they can use information. While Bayesian approaches have been used to analyse decisions on this task (Axelsen et al., 2018; Bennett et al., 2016), existing models have not incorporated updating of the prior with repeated trials.

We developed a variant of the IST designed to assess the effects of pharmacological challenge on information sampling decision. Our approach sought to disentangle possible confounds that we identified in earlier work. The first of these lies in individual risk preferences. A simple approach could assume that decision makers should seek to maximise the average likely reward (*expected value*), accounting for the costs of sampling, in their choice to sample information. However, considerable evidence shows that most decision makers exhibit a degree of risk aversion (Arrow, 1971; Dohmen et al., 2011; Holt & Laury, 2002) – we tend to prefer a higher degree of certainty even if that reduces our average payoff. *Expected utility,* unlike expected value, incorporates this (Von Neumann & Morgenstern, 1944). Thus to model biases in models of reflection impulsivity, a normative conceptual framework should target expected utilities of rewards rather than expected value (absolute payoffs), to distinguish between manipulations that affect ability to effectively use available information (i.e. departures from the ideal observer due to effects on the ability to quickly learn from the information set) from those that may affect risk or loss aversion.

On the IST, risk aversion would lead to more samples than needed to maximise expected value. To determine whether this is shown in previous research, we performed Monte Carlo simulations determining the optimal numbers of samples for different risk preferences, to compare these with actual choices. For this study, we separately tested participants on their risk preference in a different decision task, and then our variant of the IST. In the analysis we used their elicited risk preference to parameterise individual utility functions, and determine whether their decisions would differ in utility under the two pharmacological challenges compared with placebo. Our version of the IST used purely positive or zero payoffs to remove the possible confound of loss aversion, taking the trial as a whole as opposed to the cost of each sample as the unit of analysis (a reasonable approach considering the short duration of the trial and the fact that sampling is integral to an informed decision). We also employed a fixed interval between the sampling and decision

periods to negate temporal discounting effects (i.e. the tendency to sample less through impatience for receiving reward).

We hypothesised our simulation would show that sampling behaviour in previous studies would evidence risk aversion. Acute citalopram was predicted to reduce the efficacy of using the information presented, as in the Chamberlain et al. (2006) finding. We predicted that atomoxetine would not affect outcome measures, in line with its lack of effect on directed exploration (Warren et al., 2017).

2.3 METHODS AND MATERIALS

The IST is designed to provide an index of information sampling with negligible demands on visual processing and working memory and an intuitive interface (Clark et al., 2006). Subjects are presented with a grid of 25 grey boxes, which conceal underlying squares in one of two colours, and told to open as many as they wish before deciding which of the two was in the majority. In the Decreasing Win condition which is in the scope of this chapter, each sample taken has a fixed cost, with a positive payoff for a correct decision and no payoff for an incorrect one. Ten trials are presented, with no instruction to the subject as to the underlying distribution of the numbers or locations of squares.

2.3.1 Simulation of original IST

We first generated a Monte Carlo simulation of the number of points that participants would be expected to receive if they chose a given number of samples on average on this task, using a simple choice algorithm. This was used to compare with previous research and understand whether participants showed behaviour consistent with maximising total points received or a degree of risk aversion. Payoffs in points were then mapped to utilities using a utility function parameterised with *r* determining the curvature of the function, as shown in Figure 4. *r* is a risk aversion parameter, with positive values implying risk aversion (higher being more risk averse), zero implying risk neutrality, and

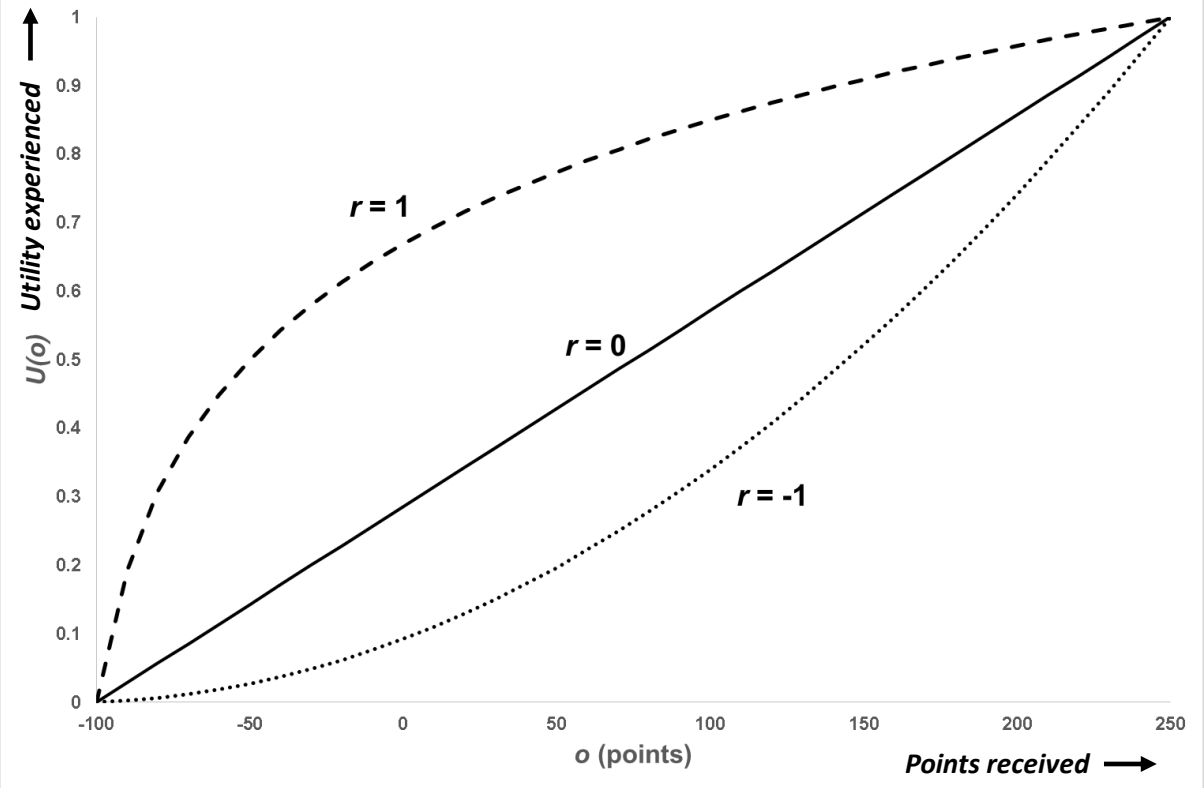
negative values implying risk seeking. Outcomes were transformed with the addition of a constant to make them strictly positive, then scaled from zero to one such that a wrong decision had utility of zero and the highest possible value had utility of one.

$$U' = \begin{cases} \frac{(\text{outcome} + \text{c})^{1-r}}{1-r} \text{ if } r \neq 1 \\ \log(\text{outcome} + \text{c}) \text{ if } r = 1 \end{cases}$$

$$E(U') = p(correct) \times U'(\text{outcome})$$

Full details are shown in Supplemental Methods.

Figure 4: example utility functions at r=-1, 0 and 1, mapping points scored in each trial to utility experienced by the participant



For the simulation, we assumed that losses are treated equivalently to gains. While theories such as Prospect Theory (Kahneman & Tversky, 1979) incorporating loss aversion suggest that the two are treated differently, there is evidence that loss aversion is absent or

reversed in experiments with small absolute outcomes (Harinck et al., 2007; Mukherjee et al., 2017). Additionally, this approach has the advantage of a single free parameter, *r*.

2.3.2 Behavioural study

2.3.2.1 Participants

Ethical permission was granted by University of Sussex Sciences & Technology C-REC (ER/JL332/6, ER/JL332/7). Potential subjects were screened with a health questionnaire (see Supplemental) and the Mini-International Neuropsychiatric Interview (Sheehan et al., 1998). Exclusion criteria included: age under 18 or over 35 years; history of psychiatric disorder (including anxiety disorder, depression, eating disorder, psychosis and substance abuse disorder); presence of significant ongoing medical condition (including migraine, diabetes, epilepsy, glaucoma and hypertension); pregnancy or breastfeeding; currently taking any medication (excluding contraceptive pill); first-degree family history of bipolar disorder; Mini International Neuropsychiatric Interview (MINI) indication of: major depressive episode, manic episode, panic disorder, social phobia, OCD, PTSD, alcohol dependence, substance dependence, mood disorder with psychotic features, psychotic disorder, anorexia nervosa, bulimia nervosa, generalised anxiety disorder, or antisocial personality disorder. They were also instructed to abstain from alcohol or caffeine in the preceding 12 hours before the start of test sessions.

53 healthy subjects aged 18-35 were recruited for this study, 28 for the citalopram group and 25 for the atomoxetine group. Of those, 1 from the citalopram and 2 from the atomoxetine group did not complete the study due to adverse side effects, and the data of 1 participant from the atomoxetine group were excluded as they took no samples in all but one of the trials. This left 49, 27 for the citalopram group (11 males, age M=23.4 SD=4.70) and 22 for the atomoxetine group (10 males, age M=23.1 SD=2.93). The groups (citalopram and atomoxetine) were tested consecutively, and participants were aware in advance which group they were being recruited into. The groups were matched for age and gender.

Subjects were tested on two sessions at least 7 days apart (days between sessions *M*=9.60, *SD*=4.26). Assignment to treatment order was double-blind and counterbalanced, with the drug treatment administered in one session and the placebo in another.

2.3.2.2 Procedure

Doses in the drug treatment conditions consisted of 20mg citalopram and 40mg atomoxetine. These doses have been shown to elicit cognitive changes in previous studies (M. Browning et al., 2007; Chamberlain et al., 2009; Grillon et al., 2007; Warren et al., 2017), and were chosen to balance active drug effects of interest against unwanted side effects.

Drug and placebo doses were delivered in gelatine capsules, indistinguishable from one another, with the capsule filled with microcrystalline cellulose (in addition to the active drug in the drug conditions). Drug and placebo doses were all manufactured according to good manufacturing practice (GMP) guidelines.

During the first session, following drug administration they were given the risk preference elicitation task immediately after the first dose, but before the treatment was absorbed. They were also given visual analogue scales at three timepoints – immediately following the dose, preceding the start of tasks, and following the end of tasks. Scales (from 0-100) were given to assess three somatic effects (nausea, headache and dizziness) and five emotion/arousal related effects (pairs of antonyms: alert–drowsy, stimulated–sedated, restless–peaceful, irritable–good-humoured, anxious–calm) to measure whether the drug was affecting these measures. To allow for drug levels to reach peak absorption (Milne & Goa, 1991; Sauer et al., 2005), the citalopram group commenced behavioural testing after 3 hours from the drug/placebo dose, and the atomoxetine group after 1.5 hours. They then carried out a set of tasks including the Modified Information Sampling Task. Following the end of behavioural testing and the final scales, participants in the atomoxetine group were monitored for a further 1.5 hours, resulting in the same length of testing session for each group.

Risk preference elicitation (RPE) task

This task was used to elicit the risk preferences of participants and calculate values of *r* (degree of risk aversion) for each participant according to Equation 2, which were then used in the computation of decision utility for each participant during IST performance. The task was adapted from Eckel & Grossman (2002, 2008), using a simple choice of gambles. Participants were told to select their preferred gamble and informed that at the end of the session, a fair coin would be flipped to determine whether they received the low or high payoff. The amount won was added to their participation fee. The gambles are shown in Table 1, and were structured to put bounds on a range of parameter *r* for the utility functions of each participant. Further details of these calculations are in Supplemental Methods.

Table 1: gamble choices in Risk Preference Elicitation task

gamble	low payoff	high payoff	*r* range	*r* value interpolated/extrapolated
1	£1.15	£1.15	$r > 4.28$	5.78
2	£1.00	£1.50	$2 < r < 4.28$	2.97
3	£0.90	£1.80	$.861 < r < 2$	1.32
4	£0.80	£2.00	$.382 < r < .861$	0.61
5	£0.30	£2.90	$-.317 < r < .382$	0.06
6	£0.05	£3.00	$r < -.317$	-0.74

Modified Information Sampling Task

Figure 5: time course of task

We designed a modified version of the Information Sampling Task (Clark et al., 2006). To make the decision dependent nature of task winnings more salient, potential winnings in monetary form (starting from £3.00 and decrementing by £0.10 for each sample taken) were displayed to subjects during their sampling decisions. These operated solely in the gain domain (i.e. an incorrect decision resulted in no change, and a correct decision resulted in a positive payoff) to understand whether serotonergic effects were present where loss aversion was not a factor. The sampling time was also fixed at 15 seconds with a decision required at that time regardless of how many samples were taken, to minimise

possible temporal discounting for reward delay. The true generative probability was a discrete uniform distribution of the majority colour occupying between 13 and 16 squares. Participants were informed in advance that the winnings from one of the trials selected at random would be paid to them at the end of the study.

2.3.3 Measures

We computed a set of measures to index behaviour in the task, based on and extending Bayesian models of the probability of making a correct choice based on information available – a measure termed *p(correct)* in the literature. Previous models for behaviour on the IST were formulated by Axelsen, Jepsen, & Bak (2018) and Bennett et al. (2016). As there were relatively few trials and a distribution was not specified to subjects in advance, there is little experience to build up an accurate prior, and the discrete structure has few possible outcomes, so it is feasible for participants to keep approximate track of the outcomes whether deliberately or otherwise. We sought a principled approach that did not require a complex inference and made minimal assumptions. This was the categorical distribution, where observed frequencies of outcomes added to the probability mass for the prior of the next decision. To avoid the problem of the unknown personal prior (i.e. the participant's personal a priori interpretation of the probability structure of the task) for the first trial in which no experience had been gained of the board, data from the first trial were excluded from analysis, so no information or distributional assumptions beyond the feedback presented and the current trial's information set was required. This is referred to as the *learned prior* model. Full details of the three models are given in supplemental methods.

The expected values of decisions were calculated by multiplying probabilities from each model with the potential winnings available after the cost of the samples taken had been subtracted. Next, expected utilities were calculated with Equation 2. We tested both

fixed utility functions across the entire dataset, and individual utility functions computed from the results of the RE task.

To test whether somatic effects of the drugs were potentially a source of changing performance, we compared the visual analogue scale measures. Details of this comparison are given in the Supplemental section.

2.4 RESULTS

2.4.1 Simulation of original IST

Figure 6: utilities on simulated data from the original IST at two values of r. If participants were risk neutral on average, utilities would be maximised by taking a mean of 4 samples. Using the actual mean of 8.8 samples, the risk aversion parameter that maximises utilities is 0.57.

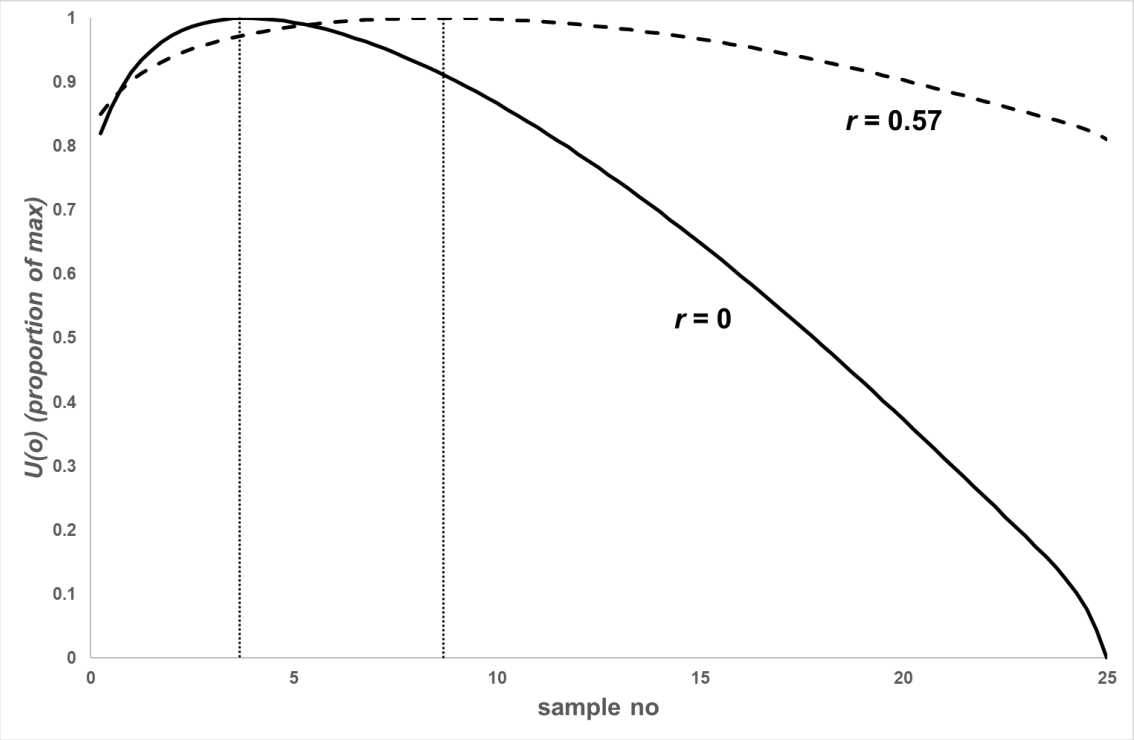


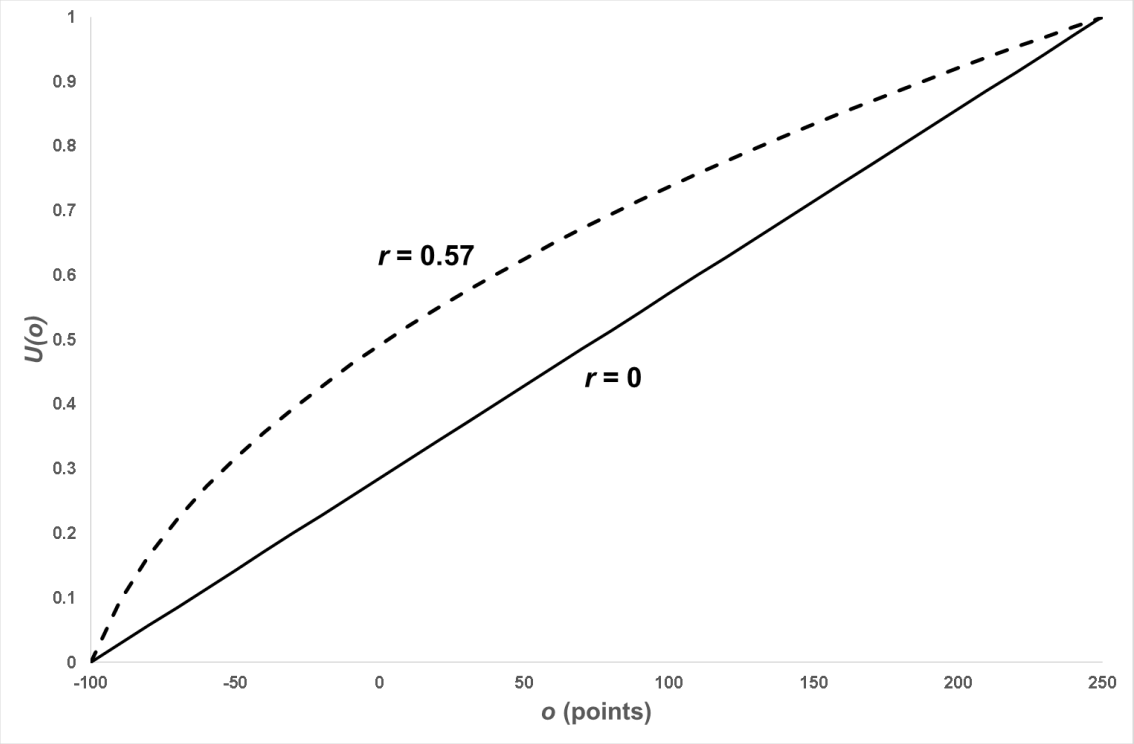
Table 2: results of previous studies using the original IST task

Reference	Group size	Sample number mean (SD)
(Clark et al., 2006)	26	7.5 (2.8)
(Tavares et al., 2007)	25	9.5 (3.7)
(Chamberlain et al., 2007)	20	7.5 (3.0)
(Clark, Roiser, Robbins, & Sahakian, 2009)	19	8.9 (2.5)
(Delazer et al., 2011)	58	9.6 (4.3)

Monte Carlo simulations of the expected value of decisions made by binomial choice algorithms showed a maximum expected value with a parameter P=0.16, corresponding to a mean of 4 samples taken.

We searched the literature for data using the IST on healthy participants, excluding those using either an adolescent or older-aged subject population, or where the standard deviation of sampling decisions was unavailable. Five studies were located including the original paper that introduced the task, which are shown in Table 2. The weighted mean sample number was 8.84. Solving the inverse problem of what utility function would be required on average for this number to maximise expected utility using Equation 2, we found that expected utility was maximised with r=0.57. Figure 7 shows utilities of decisions with r=0.57 against expected values (no risk aversion), demonstrating the degree of risk aversion that is shown by prior participants performing the original IST.

Figure 7: utilities from possible payoffs from the modelled utility function r=0.57 indicate a degree of risk aversion compared to risk-neutral utility



2.4.2 Experimental results

Risk preference elicitation task

Table 3: numbers of participants choosing each gamble in the RPE

gamble	number of choices
1	5
2	15
3	5
4	9
5	5
6	9

The results of the RPE show that 34 of 48 participants show a degree of risk aversion (choosing gambles 1-4), while 5 are approximately risk neutral (gamble 5) and 9 are risk seeking (gamble 6).

Information Sampling task

The mean baseline (placebo) sample number across the dataset was 8.37 (3.89). Testing against the mean behaviour of the previous literature, there was no significant difference between the two ($t(48)= 0.837$, $p=.40$) – behaviour on our task variant was not significantly different from the original.

We transformed each participant's decision utilities using Equation 2, with the individual implied utility function parameters computed from the RPE task used as the participant's r parameter. These were computed separately using the three probability models: *binomial prior, flat prior, and learned prior*. Neither treatment order nor gender showed significance when tested as a between-subjects factor for any group comparisons.

Table 4: results of within-subject comparisons of drug and placebo conditions for each drug group, and between-subject comparisons of drug-placebo differences (where negative numbers indicate atomoxetine has higher values). P < .05 bold, p < .1 italic.

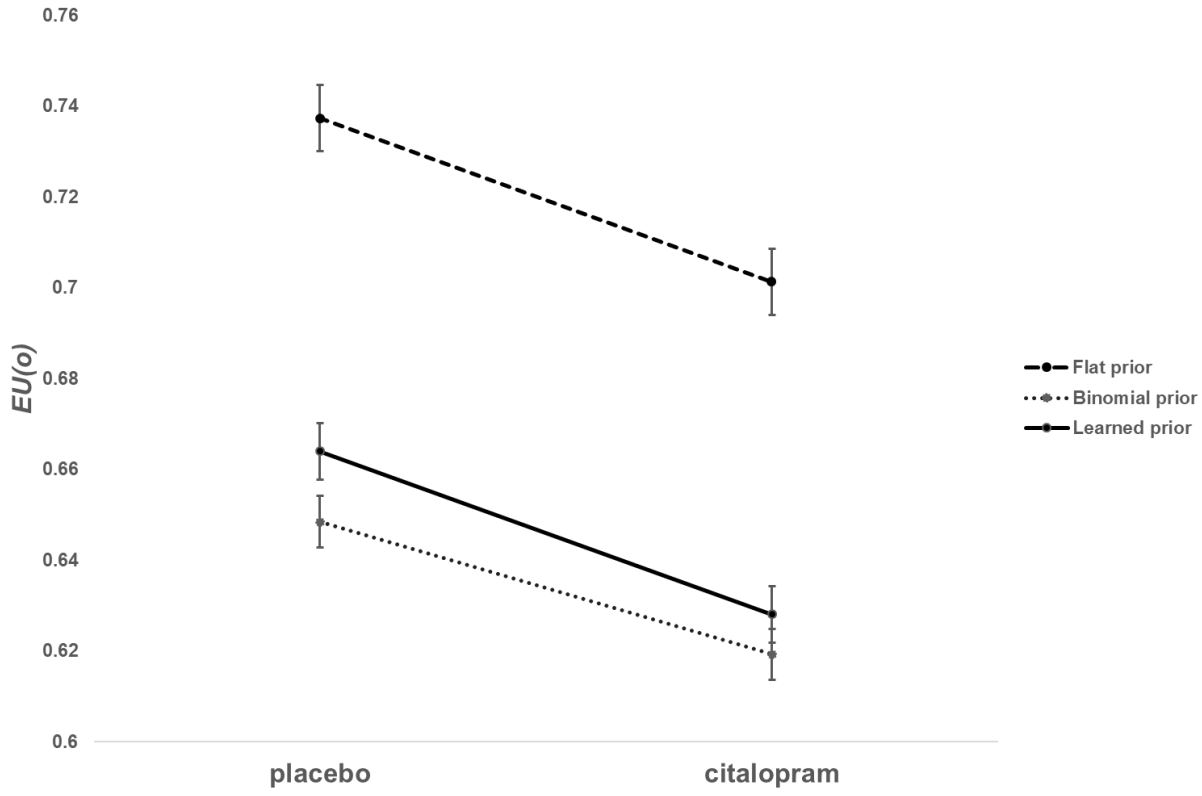
CITALOPRAM GROUP (n=27)

	model	*EU(o)* individual *r*		*EU(o)* group *r*		*p(correct)*	
		Placebo	**Drug**	**Placebo**	**Drug**	**Placebo**	**Drug**
learned prior	**mean (SD)**	0.672 (0.133)	0.633 (0.149)	0.639 (0.047)	0.608 (0.043)	0.758 (0.069)	0.723 (0.074)
	***t*-stat**	2.88		2.71		2.37	
	p	**.008**		**.012**		**.025**	
	Cohen's d	0.55		0.52		0.46	
flat prior	**mean (SD)**	0.737 (0.142)	0.701 (0.157)	0.702 (0.068)	0.675 (0.061)	0.829 (0.052)	0.8 (0.067)
	***t*-stat**	2.49		1.96		2.33	
	p	**.019**		*.061*		**.028**	
	Cohen's d	0.48		0.38		0.45	
binomial prior	**mean (SD)**	0.648 (0.13)	0.619 (0.14)	0.615 (0.038)	0.595 (0.032)	0.731 (0.061)	0.709 (0.074)
	***t*-stat**	2.59		2.34		1.87	
	p	**.015**		**.027**		*.073*	
	Cohen's d	0.50		0.45		0.36	

ATOMOXETINE GROUP (n=21)

		EU(o) individual *r*		*EU(o)* group *r*		*p(correct)*	
		Placebo	**Drug**	**Placebo**	**Drug**	**Placebo**	**Drug**
learned prior	**mean (SD)**	0.606 (0.118)	0.614 (0.166)	0.615 (0.032)	0.617 (0.053)	0.718 (0.051)	0.735 (0.076)
	***t*-stat**	-0.23		-0.04		-1.07	
	p	0.82		0.97		0.3	
	Cohen's d	-0.05		-0.01		-0.23	
flat prior	**mean (SD)**	0.69 (0.127)	0.683 (0.167)	0.702 (0.057)	0.69 (0.065)	0.815 (0.035)	0.817 (0.063)
	***t*-stat**	0.4		0.91		-0.28	
	p	.70		.37		.78	
	Cohen's d	0.09		0.20		-0.06	
binomial prior	**mean (SD)**	0.597 (0.112)	0.6 (0.149)	0.607 (0.03)	0.605 (0.042)	0.708 (0.041)	0.721 (0.072)
	***t*-stat**	-0.26		0.04		-1.2	
	p	.80		.97		.25	
	Cohen's d	-0.06		0.01		-0.26	

Figure 8: expected utilities under each probability model

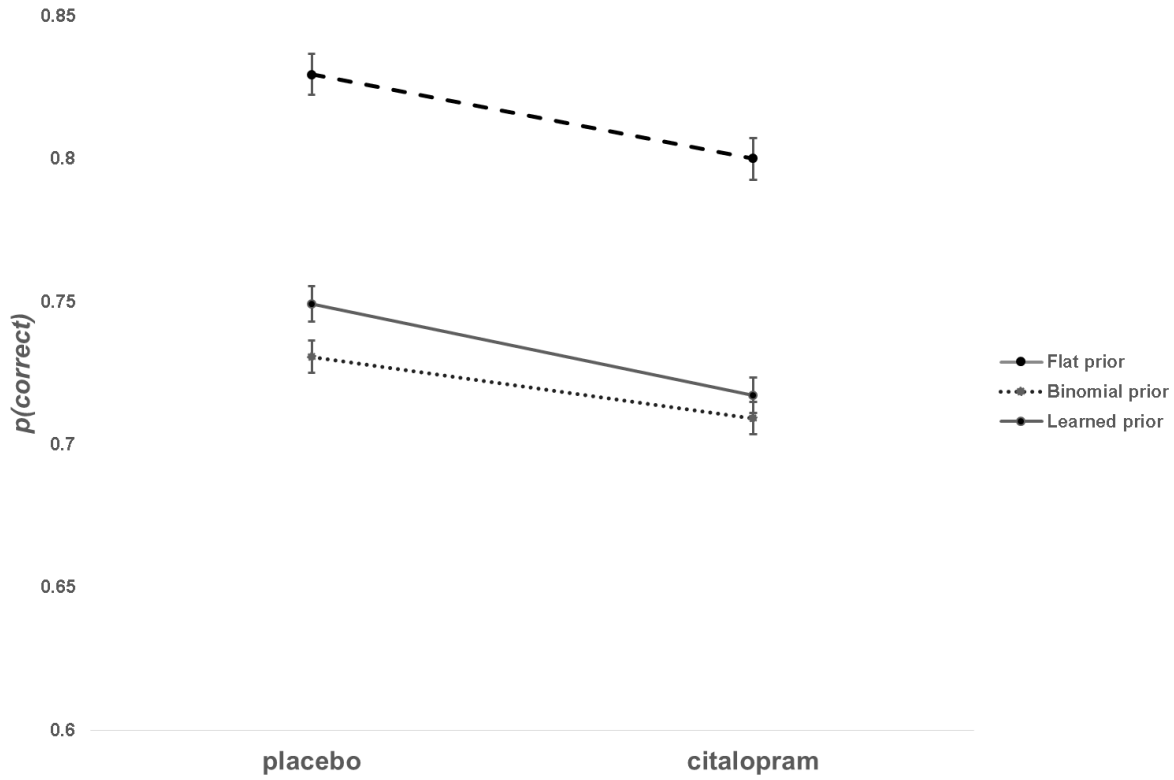


The three expected utility derivations were highly correlated ($corr(EU_{learned}, EU_{flat}) = 0.950$, $corr(EU_{learned}, EU_{binomial}) = 0.980$, $corr(EU_{flat}, EU_{binomial}) = 0.970$, all p<.001). Comparing within-subjects means of each score by drug treatment, pairwise t-tests showed significant differences between placebo and citalopram for all three measures (Table 4). In each case, the expected utilities of decisions made under citalopram were lower than those made under placebo. By contrast, those between placebo and atomoxetine were non-significant.

Next, to relax the assumption that the risk preferences determined by the RPE task would be the same as those utilised in this task, due to domain specificity of risk preferences and the possibility that pharmacological challenge may alter risk preferences, we assumed fixed risk preferences across the participant set. The mean baseline (placebo) sample number across the dataset was 8.37 (SD=3.89). Using the same method as the previous

section, we determined that utilities were maximised at $r = 0.68$, compared with $r = 0.58$ on the simulation of the original task. Transforming these utilities with this parameter, pairwise t-tests showed significant differences between placebo and citalopram for learned and binomial priors, while the model with the flat prior approached significance at the .05 level. No significant differences were shown under atomoxetine.

Figure 9: probabilities of correct choices at decision time under each probability model



To compare findings with previous research and understand the causes of utility shifts demonstrated, we tested mean probabilities at decision time between drug and placebo for each model (*p(correct);* Clark et al., 2006*s*). Pairwise t-tests showed significant differences between placebo and citalopram for learned and flat prior models – probabilities of correct decisions were lower under citalopram than placebo. The binomial prior model showed the same trend, although not reaching significance at the .05 level. It should be noted however that the assumptions of the binomial prior model are not fully met in this variant of the information sampling task, as the true underlying distribution of

probabilities for each trial follow a truncated uniform distribution as shown in Methods and Materials. Once again, for atomoxetine there were no significant differences, while the trend differences were in the opposite direction from citalopram.

Finally, we looked at the numbers of samples chosen and erroneous decisions, the latter defined as decisions choosing either the minority colour or where the numbers of both colours were equal (as in this situation taking an extra sample would increase expected utility). There were no significant effects from either drug, shown in Table 5.

Table 5: results of sample number and erroneous decisions

	CITALOPRAM GROUP					ATOMOXETINE GROUP				
	n	**placebo**	**drug**	***t*-stat**	***p***	**n**	**placebo**	**drug**	***t*-stat**	***p***
sample no	27	8.71 (3.52)	8.73 (3.45)	0.036	.97	23	7.91 (2.98)	8.42 (3.82)	1.03	.32
erroneous decisions	27	0.062 (0.077)	0.074 (0.093)	0.65	.52	23	0.053 (0.067)	0.074 (0.081)	1.07	.29

2.5 DISCUSSION

Acute serotonin reuptake inhibition was shown to cause reductions in the utility of choices made and probability of correct decisions in an information sampling context. Models showed similar effect sizes regardless of choice of prior, and the effect on expected utility and probabilities were also of similar magnitude. This effect was not observed following similar inhibition of the noradrenaline transporter, despite a similar side effect profile of the treatments.

We first used simulation to show that behaviour on both the original and our modified version of the IST is consistent with a moderate degree of risk aversion. Despite our task version operating solely in the gain domain and with a stronger control of temporal discounting, the numbers of samples taken and the risk aversion parameter implied by control data was similar to the 5 sampled studies in the literature, suggesting that, in

general, that loss aversion is not a prerequisite for observed task performance and validating the single-parameter expected utility approach in modelling task behaviour.

With our revised paradigm and a clinical dose of citalopram we extend the findings of Crockett et al. (2012) and significantly develop our understanding of serotonin's role in reflection impulsivity in the context of information sampling. As effects on decision probability and expected utility occurred without a shift in sample number, and as utility falls under various utility functions, the change does not appear to be related to shifts in risk preference, instead indicating that citalopram reduces the effective use of sampled information: in other words, using all the information available from the environment to guide decisions, improve the likelihood of an optimal outcome and determine if more or less information sampling should take place.

The effect may be presynaptic or postsynaptic influences on serotonin at the synapse. Acute citalopram blockade of the serotonin transporter increases postsynaptic serotonin levels (David et al., 2003; Moret & Briley, 1996), but also increases serotonin availability at presynaptic autoreceptors which inhibit further release (Chaput et al., 1986; El Mansari et al., 2005; Nord et al., 2013). A clear consensus of net effect has not been established. In our results, the effect of citalopram on decision probability were consistent with and in the opposite direction to that ATD, as demonstrated by Crockett et al. By this, the effect may be expected to be postsynaptic enhancement of serotonin's influence – increasing participant tolerance to the short term (i.e. local) costs of information sampling.

There are also alternative explanations, supporting a presynaptic account. Research by Chamberlain et al. (2006) and Skandali et al. (2018) showed probabilistic learning deficits from acute citalopram and escitalopram, where misleading feedback (i.e. feedback not in line with current contingencies, such as a loss when contingencies give rewards with 80% probability) was more likely to cause a shift of action than would be optimal. They posit that presynaptic serotonin autoreceptor activity may be responsible. Complementing

this interpretation, Lottem et al. (2018) showed that activation of serotonergic neurons during foraging in mice promote exploitation of a rewarding patch rather than exploration of an alternative action. From a perspective of task switching and information gathering, our results are consistent with this presynaptic interpretation. While the prepotent behaviour is to exploit in these studies and explore in the present, one (speculative) explanation may lie in covariation of serotonin activity and the amount of information required to switch behaviour (to explore and exploit, respectively). ATD studies of loss chasing behaviour are also consistent with this, where ATD lowered the threshold to switch to a loss-making decision rather than gamble further and risk further losses (Campbell-Meiklejohn et al., 2011). To probe this further and understand ATD effects on decision utility, a follow-up study using ATD with our methodology would be necessary.

The Chamberlain et al. (2006) study with a comparable citalopram effect on probabilistic learning also showed a null effect on that task with the SNRI atomoxetine, consistent with our findings. They showed that atomoxetine reduces motor impulsivity (failure to inhibit unwanted or premature motor actions), in line with noradrenergic and prefrontal dopamine influences on brain areas responsible for inhibitory signalling. Other research showed that atomoxetine similarly reduces temporal impulsivity or delay discounting (choice of a smaller immediate reward over a larger delayed reward) (Bizot et al., 2011). Both are distinct impulsivity subtypes dissociable from reflection impulsivity (Caswell et al., 2015). By contrast, directed exploratory decisions in this task to seek out information for future gain are more deliberative, and the temporal aspect of reward delay was removed by design. While care must be taken in interpreting the null results for atomoxetine, they are in line with this dissociation.

While this study used a healthy population, these findings raise important implications for SSRI pharmacotherapy. Our findings of impairment in decision making based on optimal understanding of the environment adds to understanding of early stage

treatment effects that clinicians should take into account when balancing treatment needs with adverse cognitive effects.

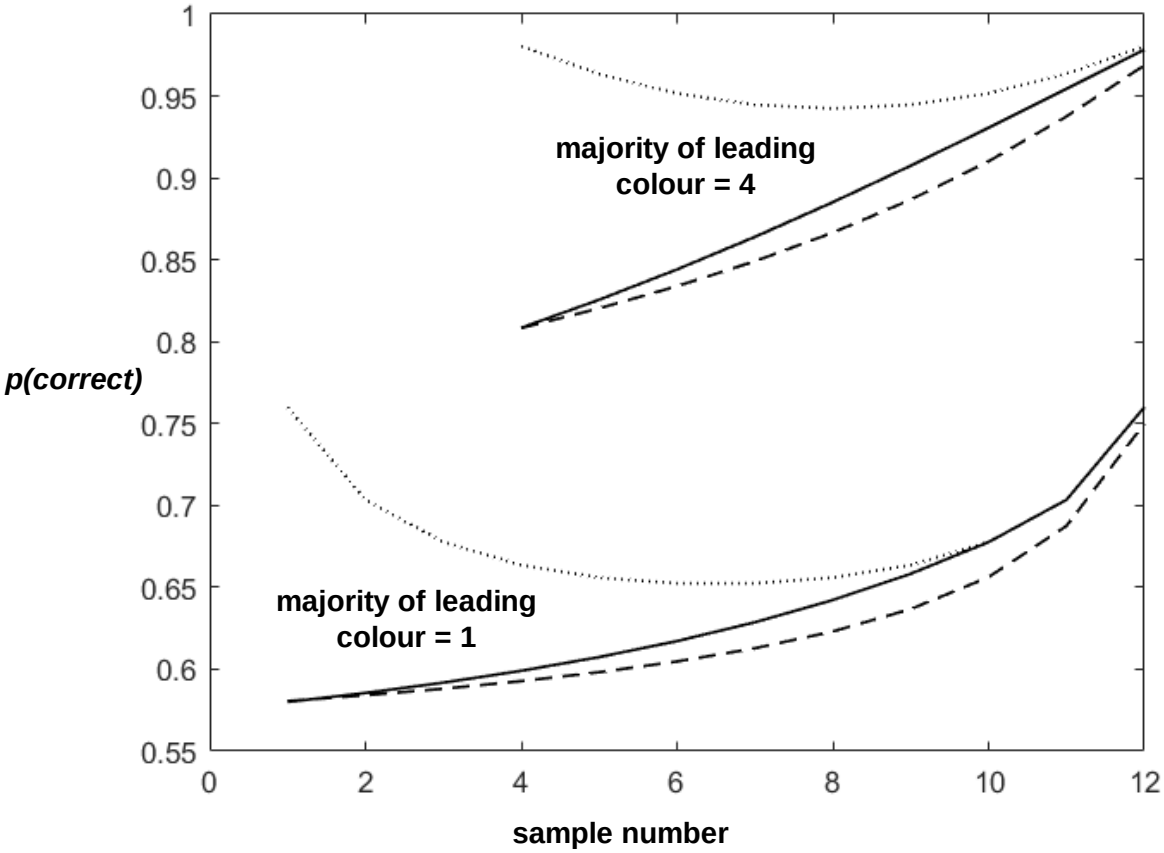
These findings suggest that serotonin may play an important role in the shift from seeking information to making a choice. While this has clear implications for our knowledge of decision-making in associated disorders, and SSRI effects, it also hints at the role of serotonin for determining thresholds of 'sufficient knowledge' for the generation of beliefs. In a world of ever-growing sources of information and persuasion, research on this mechanism may play and increasingly important role for predicting the choices and navigation of an individual's life.

2.6 SUPPLEMENTAL METHODS

2.6.1 Bayesian inference on available information

Formalised as an inference problem, decision making requires a combination of displayed information (numbers of each colour) and, optimally, an inference on the likelihood of various proportions occurring. This is because the true range of proportions is not equiprobable across the potential proportions, but is constrained by an upper bound (proportions higher than a given level do not occur). The choice of prior results in quite different probabilities inferred from the available information, as shown in Axelsen et al. (2018). This underscores the importance of being prior-agnostic in modelling, as evidence from our data indicates that participants vary in their choice of priors.

Figure 10: probability of correct decision for each sample number, where solid lines show the underlying generative probabilities in our IST variant, dashed lines are probabilities inferred from a binomial prior, and dotted lines are inferred from a flat prior



To link the model of inference on information to optimality of decision making, the cost of sampling should be factored in, with the simplest model looking at the expected value of making decisions with the chosen information set – the probability of a correct answer multiplied by the winnings available based on the number of samples, minus possible losses. This method is used in the simulation approach of Averbeck (2015).

$$EV_{decision} = p\,|\,info \times winnings - (1 - p|info) \times loss \quad (1)$$

2.6.2 Monte Carlo simulation approach

We conducted Monte Carlo simulations of three types of choice algorithms choosing samples for 10 trials, iterated 100,000 times. With sufficient samples, mean winnings w in this simulation converged with true expected values (Brooks et al., 2011). The proportion of squares of the majority and minority colours was generated from a binomial distribution truncated so that proportions above 20:5 were not included, as specified in Axelsen et al. (2018), with samples from this distribution chosen using a Knuth shuffle algorithm (Knuth, 1969) to assign equal probability to each permutation. The algorithm drew samples based on a binomial distribution with fixed parameters P, corresponding to the probability of choosing each of the 25 squares:

$$n_i \sim Binomial(25, P)$$

The mean number of samples was therefore 25P, and values of P in intervals of 0.01 from 0 to 1 were simulated. The algorithm chose whichever colour was in the sample majority, or a random colour if there were equal numbers of both.

2.6.3 Transforming points into utility

Payoffs *o* were transformed into utilities using a utility function with constant relative risk aversion (CRRA; e.g. Chiappori & Paiella, 2011). As the equation required strictly positive inputs, a constant *c* was also added prior to transformation, based on the minimum payoff o_{min} in the game design (i.e. the outcome for an incorrect decision) and the cost of a sample *s*, which is taken as the smallest unit of outcome. The resulting utilities were then normalised so that the utilities of the maximum and minimum outcomes (250 and -100 respectively for the original IST, 3 and 0 for the modified version) were set to 1 and 0, allowing them to be compared with one another.

$$U' = \begin{cases} \frac{(outcome+c)^{1-r}}{1-r} \; if \; r \neq 1 \\ log(outcome + c) \; if \; r = 1 \end{cases} \quad (2)$$

$$E(U') = p(correct) \times U'(\text{outcome})$$

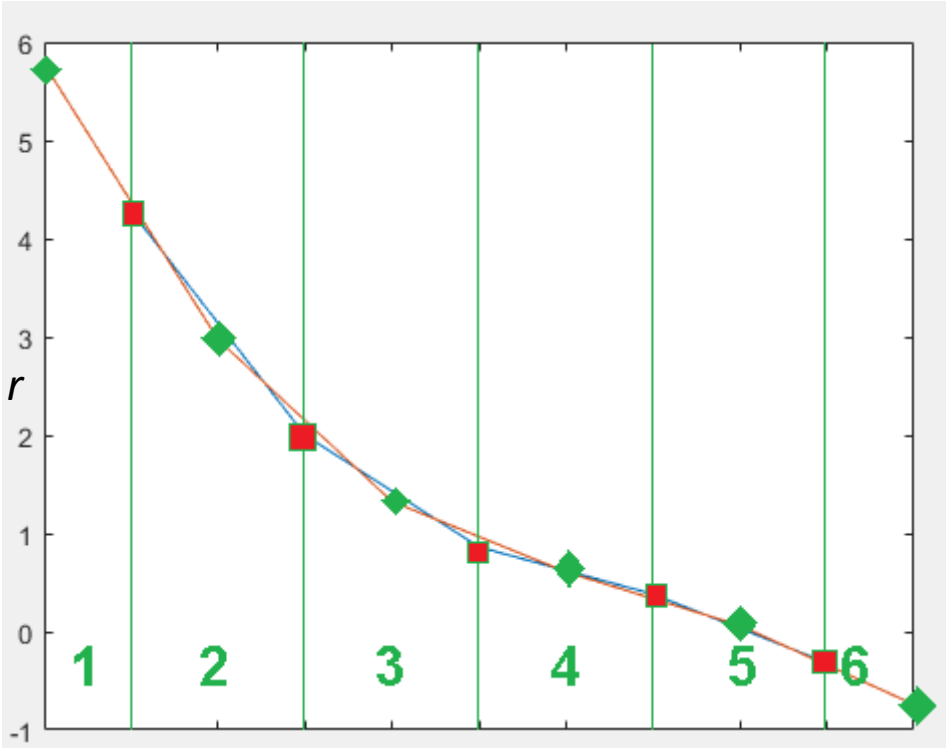
2.6.4 Risk Preference Elicitation task

The rationale behind this task is based on the principle that if the participant is assumed to have a CRRA utility function as in Equation 2, a value of *r* can be calculated at the indifference point between two gambles *a* and *b* (with probability 0.5 of a high outcome given by subscript *H* and a low outcome by subscript *L*) as follows:

$$0.5\frac{{a_H}^{1-r}}{1-r} + 0.5\frac{{a_L}^{1-r}}{1-r} = 0.5\frac{{b_H}^{1-r}}{1-r} + 0.5\frac{{b_L}^{1-r}}{1-r}$$

If the gamble chosen is 2-5, the gamble can be considered as chosen over the gambles on either side, so the indifference points mark the upper and lower bounds of *r*. With 1 and 6, only a lower and upper bound respectively are specified. To produce a point estimate of *r* for calculation of utilities, as the ranges implied were non-linear, we used Matlab's (R2017b, The MathWorks, Inc., Natick, Massachusetts, United States) PCHIP algorithm (Piecewise Cubic Hermite Interpolating Polynomial) to interpolate and extrapolate values within the ranges specified, as shown in Figure 11.

Figure 11: point r values for each gamble number (green diamonds), generated by interpolating (values 2-5) and extrapolating (values 1 & 6) from the bounds implied by the gamble chosen (red squares)



2.6.5 Bayesian computations of decision probabilities

$$posterior = \frac{likelihood \times prior}{model\ evidence}$$

Bayes' theorem provides the method of using available information to compute the probability of an outcome. All three models use the hypergeometric distribution as likelihood (inference on current information), being the discrete probability distribution that describes the probability of a given number of outcomes when a fixed number of draws are made without replacement from a finite population. The models differ in their choice of priors or information that comes from before the current trial, such as previous experience from other trials.

Flat prior model (Bennett et al., 2016): this assumes that each combination of possible proportions of the majority to minority colour from 25:0 to 13:12 are equally likely, and that this assumption is maintained throughout the trials.

$$P(\theta|n_1, n_2) = \frac{\binom{\theta}{n_1}\binom{25-\theta}{n_2}}{\sum_{j=n_1}^{25-n_2}\binom{j}{n_1}\binom{25-j}{n_2}}$$

Binomial prior model (reformulated by Axelsen et al. (2018), equal to the original *P(correct)* measure in Clark et al., 2006): this assumes a personal prior on the underlying generative process of p=0.5, i.e. that on average each colour is equally likely, resulting in a binomial distribution on the two colours. Notably, this means that extreme values of proportions are considered much less likely than values where the majority and minority colours have similar numbers.

$$P(\theta|n_1, n_2) = \frac{\binom{\theta}{n_1}\binom{25-\theta}{n_2}\binom{25}{\theta}}{\sum_{j=n_1}^{25-n_2}\binom{j}{n_1}\binom{25-j}{n_2}\binom{25}{j}}$$

Learned prior model, developed in this paper: this assumes that on trial T, information about true proportions given in the feedback for trials 1 to T-1 are incorporated in the form of a categorical distribution, where observed numbers of proportions are assigned a probability according to the number of times they were observed, and unobserved proportions are assigned a zero probability. Only trials $T \geq 2$ are considered for analysis.

$$C_M = \sum_{t=1}^{T-1} I_t \ \ where\ I_t \begin{cases} 1 \text{ if } \theta = M \\ 0 \text{ otherwise} \end{cases}$$

$$P(\theta|n_1, n_2) = \frac{\binom{\theta}{n_1}\binom{25-\theta}{n_2}\frac{C_M}{2(T-1)}}{\sum_{j=n_1}^{25-n_2}\binom{j}{n_1}\binom{25-j}{n_2}\sum_{\Theta=0}^{25}P(\Theta|n_1, n_2)} \text{ where } T \geq 2$$

All models then compute the probability of a correct decision by summing probabilities that the proportion of the chosen colour is 13 or higher, i.e. that the chosen colour is in the majority.

$$P(correct) = P(\theta \geq 13|n_1, n_2) = \sum_{M=13}^{25} P(\theta = M|n_1, n_2)$$

2.7 SUPPLEMENTAL RESULTS

2.7.1 Comparing probability measures to actual outcomes

To empirically confirm the validity and unbiasedness of the probability measures, which as Bayesian approaches should converge to true probabilities of winning the given amount, we first calculated mean winnings across all decisions in the dataset and compared them with the expected values from each method. Mean winnings were £1.56 (1.00), while the expected value from the learned prior model was £1.55 (0.354) and the binomial prior model was £1.53 (0.350) – the lower standard deviations being from the comparison of actual outcomes with expected values. By contrast, the flat prior expected value was £1.76 (0.482), as the model systematically overestimated probability of decision success.

2.7.2 Testing VAS differences

There was a small but significant difference in the nausea scale between drug and placebo conditions for atomoxetine and a marginally significant difference for citalopram,

corresponding to a change of 3.9 (CIT) / 6.6 (ATX) on a 100-point scale. There were no significant differences between the two drugs in drug-placebo differences.

*Table 6: VAS score comparisons at test time, with scores on a 100-point scale. On antonym pairs higher numbers are closer to the second term. P - placebo condition, D – drug condition, * $p < .05$*

	CITALOPRAM GROUP				ATOMOXETINE GROUP				BETWEEN GROUPS	
scale	**P**	**D**	***t*-stat (df=26)**	***p***	**P**	**D**	***t*-stat (df=21)**	***p***	***t*-stat (df=47)**	**p**
nausea	4.32 (6.63)	8.24 (11.5)	-2.06	.050	3.36 (3.94)	9.91 (14.0)	-2.31	**.031***	-0.79	.43
headache	10.8 (15.9)	11.26 (16.8)	-0.18	.86	4.39 (4.40)	7.66 (12.6)	-1.49	.15	-0.85	.40
dizziness	7.28 (12.7)	8.89 (8.7)	-1.01	.32	5.16 (5.66)	9.55 (11.8)	-2.00	.058	-1.05	.30
alert – drowsy	36.8 (17.0)	41.2 (18.8)	-1.15	.26	43.9 (14.3)	45.6 (18.5)	-0.377	.71	0.44	.66
stimulated – sedated	41.7 (13.3)	44.2 (13.7)	-0.91	.37	46.4 (11.4)	47.7 (14.4)	-0.367	.72	0.27	.79
restless – peaceful	65.3 (13.9)	62.7 (18.7)	0.80	.43	63.8 (12.0)	60.4 (17.3)	1.25	.23	0.19	.85
irritable – good-humoured	64.7 (15.4)	66.6 (14.5)	-0.76	.45	70.9 (13.9)	65.6 (13.7)	1.84	.08	1.88	.066
anxious – calm	70.8 (13.3)	68.7 (14.6)	0.63	.53	70.6 (13.8)	68.1 (15.0)	0.95	.35	0.10	.92

3 MONOAMINERGIC CHALLENGE IN STRATEGIC DECISION MAKING

3.1 ABSTRACT

Achieving cooperation in repeated interaction is a necessary challenge in humans and other social species. Previous research found roles of monoaminergic systems in related cognitive processes such as learning and social cognition, but is limited in its modelling of the dynamical nature of true interaction. This study bridged the gap by testing pairs of participants in an interactive task with separable competitive and cooperative incentives, using the selective serotonin reuptake inhibitor citalopram and selective noradrenaline reuptake inhibitor atomoxetine as pharmacological challenge in a double-blind placebo crossover design. We found for both drugs that better outcomes were achieved when both participants were in the same condition, whether drug or placebo, than when they were in different conditions. Learning models on behaviour indicated that when in the same condition, participant actions were reinforced more when they benefited the other party as well. The results suggest that coordination is more easily achieved when pair members are in the same pharmacologically-induced state, with important implications for how real-life interactions achieve coordinated outcomes.

3.2 INTRODUCTION

Humans and other social species must make decisions where interacting conspecifics have both individual and shared incentives, often requiring coordination to achieve both ends. Keeping track of reciprocal interactions and each other's intentions is a necessary step for coordination to succeed. Research to date has shown many instances of reciprocity in game theoretic tasks – tasks that classical decision theoretic models suggest should elicit purely self-interested behaviour (Fehr et al., 2002). A variety of decision tasks,

including ultimatum, trust and Prisoner's Dilemma games, have been used to understand the extent to which separable motives of self-interest and sociality interact (Fehr & Krajbich, 2014), with the additional consideration that in repeated interactions, reputation for cooperation may lead to improved outcomes for oneself (C. F. Camerer & Hare, 2014).

Psychopharmacology studies have revealed the contributions of various neurotransmitters to social behaviour and learning. Of particular interest are the effects of serotonin, which has been linked to social cognition by both genetic at psychopharmacological work (Canli & Lesch, 2007; Siegel & Crockett, 2013), and prefrontal dopamine, linked with adaptation to complex contingency changes (Frank et al., 2007; Puig & Miller, 2012) but also with behaviour towards social conspecifics (Sáez et al., 2015).

Studies manipulating serotonin have generally found a positive association between serotonergic activity and cooperativeness. Three studies used variants of the Prisoner's Dilemma (PD) game. The PD game involves individuals simultaneously choosing to cooperate or defect. Individual 'defection' (choosing an option that does not achieve the best group outcome) always leads to better outcomes for the individual regardless of the other person's action, but mutual defection has lower outcomes for both parties than mutual cooperation. Tse & Bond (2002) used the selective serotonin reuptake inhibitor (SSRI) citalopram in 2 week repeated administration, with a modified repeated PD game played with a confederate. Citalopram increased cooperative behaviour on the task. Lowering serotonin levels through acute tryptophan depletion (ATD) decreased cooperation with a confederate in repeated PD (Wood et al., 2006). Two studies using the Ultimatum Game (UG), a task in which participants are offered a varying share of a pot of money by another individual in which acceptance means both are rewarded and rejection results in no rewards to either party, showed serotoninergic effects on responses to unequal treatment independently of mood (Crockett et al., 2008, 2010). ATD increased rejection of unfair offers, while acute citalopram decreased rejection. In learning, recent research in mice showed that optogenetic activation of serotonergic neurons in the dorsal raphe nucleus

increased learning rates (Iigaya et al., 2018). Serotonin transporter polymorphisms show differential effects, with individuals homozygous in the L-allele of the 5-HTTLPR region choosing to shift strategies more following a loss than S/L and S/S carriers (den Ouden et al., 2013). Acute citalopram administration caused an increased rate of errors in a probabilistic reversal learning task compared with placebo (Chamberlain et al., 2006).

Studies using manipulation of prefrontal dopamine have taken two approaches: direct manipulation of prefrontal dopamine using catechol-O-methyltransferase (COMT) inhibition (where COMT degrades catecholamines such as dopamine, and its inhibition increases prefrontal dopamine levels; Tammimaki, Aonurm-Helm, Kaenmaki, & Mannisto, 2016; Tunbridge, Bannerman, Sharp, & Harrison, 2004), and inhibition of noradrenaline reuptake. Selective noradrenaline reuptake inhibitors (SNRIs) atomoxetine and reboxetine target the norepinephrine transporter (NET), which also transports dopamine (Horn, 1973; Raiteri et al., 1977). NET shows high concentrations in the prefrontal cortex and acts as the main mechanism for dopamine clearance there (Morón et al., 2002). Animal models have shown atomoxetine to raise prefrontal dopamine as well as noradrenaline levels in mouse (Bymaster et al., 2002) and rat (Koda et al., 2010; Swanson et al., 2006), and reboxetine to increase prefrontal dopamine (Linnér et al., 2001) and noradrenaline (Sacchetti et al., 1999) levels in rat. Neither drug are specific, and thus inferences on them may also reflect changes in noradrenergic activity, but both appear to show robust effects on prefrontal dopamine levels. Crockett et al.'s (2010) study, which showed serotonergic mediation of rejection rates in UG, showed no effect of atomoxetine on the same social measures, though an increase in measures of executive function. Tse & Bond (2003) showed that reboxetine left behaviour on a PD game unchanged, but increased cooperativeness as reported by participants' flatmates. Sáez and colleagues (2015) used the COMT inhibitor tolcapone to increase prefrontal dopamine levels and tested participants on a Dictator Game task, a variant of the UG in which no rejection can take place (so providing a purer measure of concern for another individual's rewards in the absence of strategic considerations). They

found that tolcapone increased inequity aversion compared to placebo. In models of learning, Chamberlain et al's (2006) study showed no effect of atomoxetine in contrast with the detrimental effect of acute citalopram. Within the area of strategic interaction, Set et al. (2014) showed that the COMT gene modulated degree of anticipatory learning of the actions of others across participants.

Learning models have distinct advantages over traditional statistical analysis, providing simple, plausible and testable algorithms to model evolving decision behaviour across the course of an experiment. They have demonstrated strong correspondence with measures of brain activity (reviewed in Daw & Tobler, 2014). However, little psychopharmacology research has used this technique to understand problems of learning during pairwise interaction. Another important gap in previous research is the lack of truly interactive studies in this area. Schilbach et al. (2013) note the distinction between studying detached individuals in a socially-posed problem and an interactive approach, the latter being formalised as a dynamic, self-organised and emergent process (Dale et al., 2013). Previous studies with confederates do not incorporate mutual adaptation to one another's behaviour and therefore cannot capture the effect of pharmacological challenge on dynamic social processes.

We sought to bridge this gap by testing healthy participants on a strategic interaction task designed with both competitive and cooperative incentives. We used a varying pattern of available rewards with both unpredictable and predictable bonuses, in order to separately elicit competitive and cooperative approaches respectively. This used a larger strategy space than provided by the PD game to provide greater ability for learning models to characterise the type of learning approach used (e.g. based on simple reinforcement, or beliefs on other's actions; Salmon, 2001). We used the drugs citalopram and atomoxetine to understand the contributions of monoamine systems on this behaviour. With the two drugs showing similar side effect profiles, they also act as high-level controls for one another to control for generalised drug effects (such as somatic symptoms) that are

not specific to each drug's mechanism of action. Previous research showed that COMT gene polymorphisms associated with higher prefrontal dopamine was linked to greater learning from the actions of others (Set et al., 2014), while heightened prefrontal dopamine through tolcapone administration was associated with more egalitarian behaviour (Sáez et al., 2015). However, previous SNRI studies showed neither effects on social behaviour nor social learning. Coordination (requiring an understanding of the other's behaviour) and cooperation (requiring concern for the other's outcomes) was designed in this task to lead to higher outcomes. We thus hypothesised that atomoxetine would either show a small increase in cooperative behaviour and higher outcomes, or no effect.

3.3 METHODS AND MATERIALS

3.3.1 Participants

Ethical permission was granted by University of Sussex Sciences & Technology C-REC (ER/JL332/6, ER/JL332/7). Potential participants were screened with a health questionnaire (see supplemental appendix) and the Mini-International Neuropsychiatric Interview (Sheehan et al., 1998). Exclusion criteria included: age under 18 or over 35 years; history of psychiatric disorder (including anxiety disorder, depression, eating disorder, psychosis and substance abuse disorder); presence of significant ongoing medical condition (including migraine, diabetes, epilepsy, glaucoma and hypertension); pregnancy or breastfeeding; currently taking any medication (excluding contraceptive pill); first-degree family history of bipolar disorder; Mini International Neuropsychiatric Interview (MINI) indication of: major depressive episode, manic episode, panic disorder, social phobia, OCD, PTSD, alcohol dependence, substance dependence, mood disorder with psychotic features, psychotic disorder, anorexia nervosa, bulimia nervosa, generalised anxiety disorder, or

antisocial personality disorder. They were also instructed to abstain from alcohol or caffeine in the preceding 12 hours before the start of test sessions.

53 healthy participants aged 18-35 were recruited for this study, 28 for the citalopram group and 25 for the atomoxetine group. Of those, 1 from the citalopram and 2 from the atomoxetine group did not complete the study due to adverse side effects. This left 50, 27 for the citalopram group (11 males, age *M*=23.4 *SD*=4.70) and 23 for the atomoxetine group (11 males, age *M*=23.1 *SD*=2.87). Data from two sessions in the citalopram group and three sessions in the atomoxetine group could not be used as participants were paired with individuals who were not used in the study, so only the unaffected session was analysed from each of those participants, and they were excluded from pairwise analysis. The groups (citalopram and atomoxetine) were tested at separate times, so pairs of participants were always in the same group but varied separately as to whether each participant was in drug or placebo condition, and participants were aware in advance which group they were being recruited into. Participants were tested on two sessions at least 7 days apart (days between sessions *M*=9.31, *SD*=4.41). Assignment to treatment order was double-blind and counterbalanced, with the drug treatment administered in one session and the placebo in another.

3.3.2 Procedure

Doses in the drug treatment conditions consisted of 20mg citalopram and 40mg atomoxetine. These doses have been shown to elicit cognitive changes in previous studies (M. Browning et al., 2007; Chamberlain et al., 2009; Grillon et al., 2007; Warren et al., 2017), and were chosen to balance active drug effects of interest against unwanted side effects.

Drug and placebo doses were delivered in gelatine capsules, indistinguishable from one another, with the capsule filled with microcrystalline cellulose (in addition to the active drug in the drug conditions). Drug and placebo doses were all manufactured according to good manufacturing practice (GMP) guidelines. Participants were instructed not to discuss

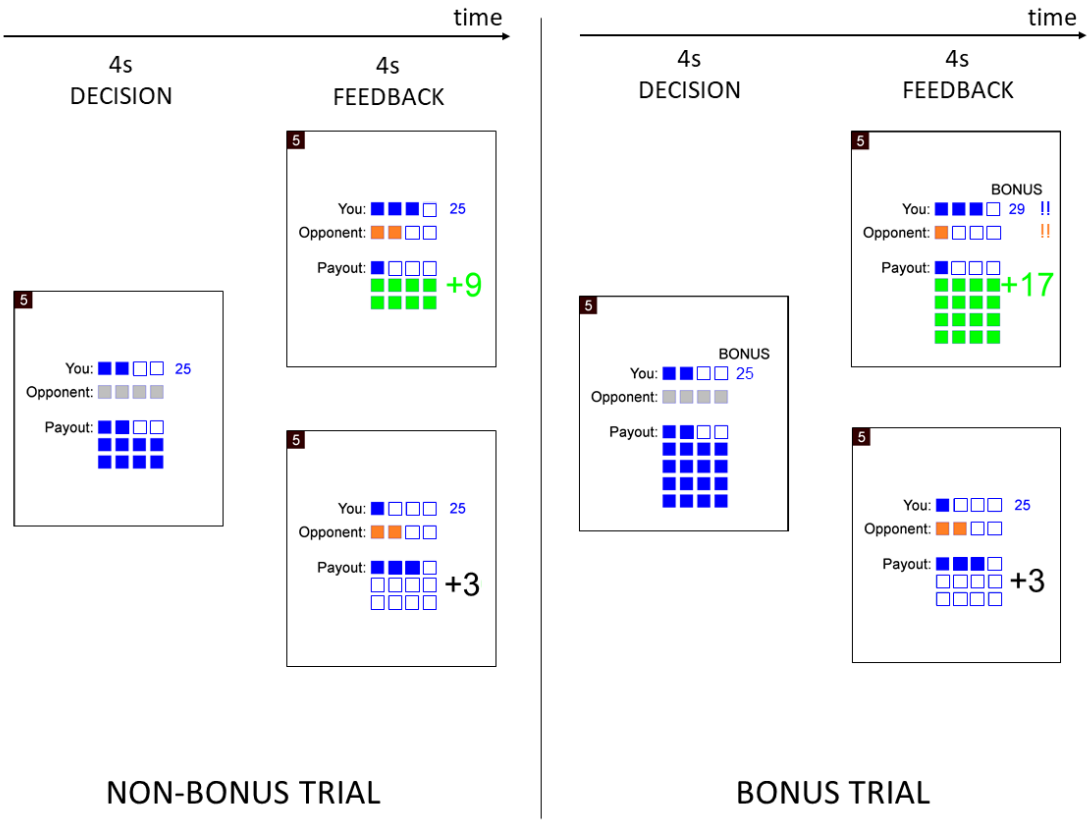
any aspect of the study including their subjective effects or the tasks, and were monitored at all times to ensure that this did not take place.

Participants were also given eight visual analogue scales at three timepoints – immediately following the dose, preceding the start of tasks, and following the end of tasks. Visual analogue scales (from 0-100) were given to assess three somatic effects (nausea, headache and dizziness) and five emotion/arousal related effects (pairs of antonyms: alert–drowsy, stimulated–sedated, restless–peaceful, irritable–good-humoured, anxious–calm) to measure whether the drug was affecting these measures. To allow for drug levels to reach peak absorption (Milne & Goa, 1991; Sauer et al., 2005), the citalopram group commenced behavioural testing after 3 hours from the drug/placebo dose, and the atomoxetine group after 1.5 hours. They then carried out a set of tasks including the task for this study. For this task, participants were tested in pairs which were randomly assigned, and the pair assignment differed between sessions such that no participant was paired with the same other participant on both sessions. During the course of the task, pairs were seated alongside one another. The task was displayed on a single screen, with a divider placed between pair members so that they could not see the other member nor their display. Participants wore ear defenders rated for 37dB standard noise reduction to prevent any aural cues from the other pair member.

Following the end of behavioural testing and the final scales, participants in the atomoxetine group were monitored for a further 1.5 hours, resulting in the same length of testing session for each group.

3.3.3 Task design

Figure 12: order of stimulus presentation on the two types of trial



This task was programmed in Matlab (version 2017a, Mathworks, Natick, MA, USA) using Psychtoolbox-3 (Kleiner et al., 2007) and two standard keyboards to collect responses. We developed a strategic interaction game broadly based on the Patent Race game (Rapoport & Amaldoss, 2000; Lusha Zhu et al., 2012) but with key differences in the payoff structure. Each trial, participants were shown a pot of tokens that they could win, which varied depending on the type of trial and between participants. They then simultaneously chose to play a number of tokens between zero and four, without any information about what the other participant's choice was at that stage. if one participant played more than the other, they would receive their pot of tokens plus any tokens that they had not played in that trial. If both played the same amount then neither won the pot. At the end of the trial, their outcome was displayed along with the other participant's choice.

We varied the incentive structure over the course of the experiment to elicit different degrees of competitive and cooperative behaviour, by using a bonus structure. Bonuses applied to one participant, both participants or neither participant in a trial. Participants were informed at the start of each trial if it was a bonus trial for them, and at the end of the trial if it had been a bonus trial for the other participant. The pot available for a participant was set at 8 tokens for a non-bonus trial and 16 tokens for a bonus trial. Table 7 shows the game's payoff matrix. These rules were explained in advance to participants, who also were shown a demonstration of the task before the experimental trials, and were questioned to ensure comprehension. No deception was employed.

Table 7: payoff matrix for participant given other participant's choice, with bonus trial payoffs given in square brackets where different from non-bonus trials. Units are tokens allocated/won

		Other participant's choice				
		0	**1**	**2**	**3**	**4**
Participant's choice	**0**	4	4	4	4	4
	1	11 [19]	3	3	3	3
	2	10 [18]	10 [18]	2	2	2
	3	9 [17]	9 [17]	9 [17]	1	1
	4	8 [16]	8 [16]	8 [16]	8 [16]	0

To manipulate incentives to compete and cooperate across the course of the experiment, the presentation of bonus trials was determined by block, with different blocks for coordinated and predictable bonus trials, and random and unpredictable ones. There were three blocks always presented in order: *no bonus, unstructured bonus* and *structured bonus*. The *control (C)* block had no bonus trials for either participant. In the *unstructured bonus (U)* block, bonus trials were randomised and therefore could occur either simultaneously or separately for each participant. In the *structured bonus (S)* block, bonus trials cycled in a repeating pattern of **P1 bonus → P2 bonus → no bonus**, such that bonus trials for the two participants were always on different trials. Participants were not informed about the block structure in advance, nor were informed during play when a new block started. They played 60 trials of each of 3 block types, for a total of 180 trials lasting

approximately 25 minutes. Every 20 trials, the game paused for 12 seconds to allow participants to rest.

Participants were informed that they would be paid £1 for each 1000 tokens they won, plus a split of £5 based on their share of the total number of tokens won by both participants. This was also designed to provide incentives for both competitive and cooperative behaviour.

The design allows for various approaches to action choice in the repeated game. Participants could choose actions randomly or solely based on the experienced reward for each action. Alternatively, they could take account of the other's action, either to maximise the joint reward or try and score more than the other. The addition of bonus trials allowed for coordination, where participants could play lower choices for non-bonus than bonus trials in the *U* and *S* blocks, allowing the other to win these trials and producing a higher combined score. Alternatively, a competitive approach would mean playing higher choices on these trials in order to prevent the other from winning their own bonuses, although this allowed for reciprocal punishment by the other participant. The predictable structure of the *S* block made strategic options particularly salient, although coordination on bonus trials were an option in either. Our analysis was designed to disentangle these strategies.

3.3.4 Analyses

3.3.4.1 Game theoretic approach

Formally, this game constitutes a game of imperfect and incomplete information (Harsanyi, 1967) as neither the other participant's action nor their potential rewards (due to the bonus structure) are known in advance. A simple one-stage mixed strategy Nash equilibrium model provides a null model without learning or coordination to which more complex models can be compared.

As the *C* block has no bonus trials for either participant, which is salient during play although not informed in advance, the *C* block trials can be approximated by a mixed strategy Nash equilibrium of complete information and symmetric payoffs (Dutta, 1999 p.104). *U* and *S* blocks have incomplete information (Dutta, 1999 p.316), requiring an additional consideration of the unknown probability of each trial being a bonus trial for the other participant. The frequency of these (which is 1 in 3 trials) can be approximated from the feedback of a few trials. With the assumption that this frequency is learned quickly, probabilities approximate the true frequency of payoffs. If the trial structure in *S* block is learned, the game is once again one of complete information, as the bonus trial status of the other participant can be deduced with certainty. Probabilities of choices are summarised in Table 8. For learning models, these choices were used as null models for comparison, referred to as **mixed strategy NL** for the model in which the bonus structure was not learned, and **mixed strategy L** where it was learned. Importantly, with both participants using the same strategy, *mean scores are 4 regardless of the block or bonus trial status, and regardless of whether S block probabilities are learned*. This provides an important benchmark by which performance can be measured.

Table 8: probabilities of making each choice according to one-stage mixed strategy Nash equilibria. † where bonus structure of S block is learned

	probability of choice				
block	**0**	**1**	**2**	**3**	**4**
***C, S* other non-bonus †**	1/8	1/8	1/8	1/8	1/2
U	3/32	3/32	3/32	3/32	5/8
***S*, no learning**	3/32	3/32	3/32	3/32	5/8
***S*, other bonus †**	1/16	1/16	1/16	1/16	3/4

3.3.4.2 ANOVA models

We conducted paired ANOVAs for each drug group on *score*, with *drug condition* as a repeated-measures factor and *order* (order of drug and placebo conditions for each participant) as a between-subjects factor.

3.3.4.3 Mixed models

All models were fitted in the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2019) and computing statistics for parameter estimates using lmerTest (Kuznetsova et al., 2017). For the specification of random effects we followed a backward-selection heuristic throughout (Matuschek et al., 2017), using the criteria of model convergence and improvement of model fit, while constraining models for each outcome variable to the same random effects structure to allow for comparison between models.

We estimated a series of mixed models on two measures – participant choices and scores on each trial. First, to test the validity of the bonus trial structure against the null of choices being independent of whether bonus were available, as well as the hypothesis that choices would be lower for non-bonus trials and higher for bonus trials in blocks *U* and *S*, we modelled choices on factor variables for *block type* and *bonus*. These were also admitted as random effects slopes with *participant id* as the random effect grouping variable. As models on mean choice have limited ability to capture the interactive nature of the task due to coordination being possible outside of the bonus structure (for example, alternating high and low choices), further analysis of choice was carried out using the learning model approach detailed in the following section.

To analyse the determinants of task performance, we next estimated models on trial score, removing trials where both participants had bonus trials simultaneously (~3.9% of trials) to facilitate comparison between blocks, as these trials could only occur in the *U* block. The base models used the same predictors as for *choice* (*block type* and *bonus*). Next, models were estimated with other variables of interest added individually to the base models as fixed factors, estimating main effects and interactions with *block type* and *bonus.* These were (all dummy coded): *drug* indicating whether the participant was in drug or placebo condition, *order* to indicate the session order of drug and placebo conditions, *pair condition* indicating whether both participants were in the same condition (drug or placebo)

as each other or not, and *same gender* indicating whether the participants were of the same gender or not. We used model comparison with Akaike Information Criteria (AIC) weights to determine which models provided the most accurate account of task performance – see section 3.3.4.5 for details.

3.3.4.4 Learning models

We estimated a series of learning models on the choices of token allocation, using models from the existing literature, as well as formulating a new model which we term the *Own-Other Reward model.* Models were estimated separately for each participant session by minimising negative log-likelihood, then compared by computing AIC. We fitted two models from the existing literature – a basic Rescorla-Wagner reinforcement learning model (e.g. Daw & Tobler, 2014) and Experience Weighted Attraction (C. Camerer & Ho, 1999). Full details are available in Supplemental Methods.

We also designed a new model which we term *Own-Other Reward Learning*. This is a generalisation of reinforcement learning allowing rewards of both the participant themselves and of the other pair member to reinforce actions. Equation 3 shows the updating equation, specifying the value of an action *V(A)* at each time step, where *R* refers to the participant's own reward and *O* is the reward of the other pair member. This has two free parameters: $\boldsymbol{\alpha}$ for the learning rate, bounded between 0 and 1; and $\boldsymbol{\gamma}$ as the weight on the other's reward, bounded between -1 and 1, with positive and negative values respectively increasing and decreasing reinforcement of an action with a higher reward to the other. Reinforcement learning from own reward alone is thus a special case of this model with $\boldsymbol{\gamma} = 0$.

$$V_{t+1}(A) = \begin{cases} V_t(A) + \alpha[R_t + \gamma O_t - V_t(A)] & \text{if action A is chosen} \\ V_t(A) & \text{otherwise} \end{cases} \quad (3)$$

Probabilities of making actions were computed from action values for all learning models through a softmax function, which added an additional free parameter (see

Supplemental Methods). Null models were also estimated, corresponding to the empirical distribution of responses across the participant set and a uniform probability of playing each action. Model comparisons were carried out using AIC weights (shown in the next subsection). Following model comparison, we carried out non-parametric comparisons of parameters from the winning model.

3.3.4.5 Model comparison

AIC provides a useful metric for model comparison, penalising extra parameters to prevent overfitting (Burnham & Anderson, 2002). To quantify the degree of evidence in favour of winning models (AIC_{best}, i.e. the model with lowest AIC), AIC weights were computed across the candidate model set through application of Equations 4 and 5 (Burnham & Anderson, 2002), which are interpretable as the probability of the model being the best given the dataset and the other candidate models (Wagenmakers & Farrell, 2004).

$$\Delta AIC_m = AIC_m - AIC_{best} \quad (4)$$

$$P(m) = wAIC_m = \frac{e^{-0.5\Delta AIC_m}}{\sum_{n \in M} e^{-0.5\Delta AIC_n}}, m \in M \quad (5)$$

3.4 RESULTS

3.4.1 Demographics

Table 9: demographic statistics for each group

	CITALOPRAM GROUP	ATOMOXETINE GROUP
N	27	23
age	23.4 (4.7)	23.2 (2.9)
gender	16 females, 11 males	12 females, 11 males

Table 9 shows the demographic statistics of each group. Groups did not significantly differ on age ($t(48) = 0.246, p = .81$) or gender ($\chi^2(1,\ N\ =\ 50) = 0.25, p = .62$). Tests on VAS scores showed small but significant differences between drug and placebo conditions

on nausea, but no difference between the two drug groups in drug-placebo differences (see Supplemental for details).

3.4.2 ANOVAs

Table 10: mean scores by drug condition

	CITALOPRAM GROUP			ATOMOXETINE GROUP		
	n	**placebo**	**drug**	**n**	**placebo**	**drug**
score	27	4.40 (1.69)	4.74 (1.74)	23	4.84 (2.22)	4.38 (1.64)

ANOVAs on *score* showed no significant effects of *drug condition* for either citalopram group ($F(1,23) = 0.336, p = .57$) or atomoxetine group ($F(1,18) = 0.631, p = .44$). There were no *order* interactions for either drug group.

3.4.3 Mixed models

Following the model selection procedure in 3.3.4.3, the mixed models all included *bonus* as a random effect and *participant id* as a random effect grouping variable.

3.4.3.1 Choice model

*Table 11: parameter estimates for choice model, where units are in tokens allocated. Note that as bonus trials are only in U and S blocks, block type * bonus interactions are only estimated for U block and coefficients represent contrasts with S block scores. *** p < .001*

group	parameter	fixed effect [95% CI]	*t*-stat	*p*
ALL	**(intercept)**	2.61 [2.44,2.78]	29.463	**<.001*** **
	block type	***U:*** -0.184 [-0.241,-0.127]	-6.292	**<.001*** **
		S: -0.173 [-0.23,-0.115]	-5.905	**<.001*** **
	bonus	0.986 [0.809,1.163]	10.92	**<.001*** **
	block type * bonus	***U:*** 0.022 [-0.087,0.131]	0.393	.69
CITALOPRAM	**(intercept)**	2.65 [2.46,2.85]	26.486	**<.001*** **
	block type	***U:*** -0.145 [-0.222,-0.069]	-3.723	**<.001*** **
		S: -0.229 [-0.305,-0.152]	-5.867	**<.001*** **
	bonus	1.04 [0.796,1.28]	8.439	**<.001*** **
	block type * bonus	***U:*** -0.065 [-0.21,0.08]	-0.884	.38
ATOMOXETINE	**(intercept)**	2.56 [2.26,2.86]	16.826	**<.001*** **
	block type	***U:*** -0.231 [-0.318,-0.145]	-5.232	**<.001*** **
		S: -0.105 [-0.192,-0.018]	-2.377	**<.001*** **
	bonus	0.924 [0.663,1.185]	6.938	**<.001*** **
	block type * bonus	***U:*** 0.127 [-0.037,0.292]	1.519	.13

The model for *choice* against *block type* and *bonus* showed that participants made lower choices for non-bonus trials in blocks including bonuses compared with the control block, while they made higher choices in bonus trials (Table 11). The interaction was not significant in any model. *Order* was tested as a main effect and in interactions with other variables, but did not significantly enter any model.

3.4.3.2 Score models

Table 12: AIC values for each model and probability of model being better fitting than all other models for the group specified (where best fitting model $AIC_m = AIC_{min}$). Bold indicates best fitting model

		ALL		CITALOPRAM GROUP		ATOMOXETINE GROUP	
model	**df**	ΔAIC_m	**P(m)**	ΔAIC_m	**P(M)**	ΔAIC_m	**P(m)**
(base)	9	261	<.001	120	<.001	147	<.001
order	13	257	<.001	119	<.001	144	<.001
drug	13	261	<.001	81	<.001	111	<.001
pair condition	13	0	**>.999**	0	**>.999**	0	**>.999**
same gender	13	266	<.001	64	<.001	32	<.001

Table 12 shows the model fits for models on *score*. Although the *drug* model was not the best fitting, to test hypotheses of drug effects, we examined the parameter estimates of this model. The main effect of *drug* was not significant in any model (ALL: $t(16370) = -0.855$, $p = 0.39$; CIT: $t(8969) = 0.485$, $p = .63$; ATX: $t(7400) = -1.80$, $p = .072$).

Pair condition models for across both groups and for each drug group provided the best fit. Table 13 shows the summary statistics split by *pair condition.* Parameters from these models were analysed further.

Table 13: summary statistics of score in each block split by pair condition and bonus trial, excluding trials where both participants had bonus trials simultaneously (see section 3.3.4.3)

		score				
		C block	***U block***		***S block***	
group	**pair conditions**	**(no bonus)**	**(no bonus)**	**(bonus)**	**(no bonus)**	**(bonus)**
ALL	***all***	4.15	3.73	8.48	3.45	7.90
	same	4.28	4.08	9.13	3.80	8.63
	different	3.92	3.14	7.40	2.87	6.70
CIT	***all***	4.03	3.59	8.22	3.43	8.49
	same	4.28	4.08	9.15	3.80	9.28
	different	3.72	2.97	7.04	2.96	7.49
ATX	***all***	4.28	3.89	8.79	3.47	7.19
	same	4.29	4.09	9.11	3.79	8.02
	different	4.27	3.42	8.05	2.73	5.30

*Table 14: parameter estimates for score models, where units are in tokens won. Note that as bonus trials are only in U and S blocks, block type * bonus interactions are only estimated for U block and coefficients represent contrasts with S block scores. * p < .05, ** p < .01, *** p < .001*

group	parameter	block	fixed effect [95% CI]	df	*t*-stat	p
ALL	***(intercept)***		3.833 [3.51,4.16]	116	23.2	**<.001*** **
	block type	***U***	-0.770 [-1.06,-0.482]	16334	-5.24	**<.001*** **
		S	-1.06 [-1.35,-0.769]	16334	-7.2	**<.001*** **
	bonus		3.04 [2.32,3.76]	81	8.28	**<.001*** **
	pair condition		0.539 [0.250,0.808]	4767	3.72	**<.001*** **
	block type * bonus	***U***	0.291 [-0.084,0.666]	16337	1.52	.13
	block type * pair condition	***U***	0.562 [0.203,0.921]	16334	3.07	**.002** **
	bonus * pair condition	***S***	0.575 [0.216,0.934]	16334	3.14	.002
			2.34 [1.82,2.85]	2376	8.96	**<.001*** **
CIT	***(intercept)***		3.75 [3.38,4.12]	70	19.9	**<.001*** **
	block type	***U***	-0.761 [-1.12,-0.398]	8943	-4.11	**<.001*** **
		S	-0.750 [-1.11,-0.386]	8943	-4.05	**<.001*** **
	bonus		3.97 [3.17,4.78]	44	9.65	**<.001*** **
	pair condition		0.550 [0.188,0.912]	2116	2.98	**.003** **
	block type * bonus	***U***	-0.400 [-0.910,0.109]	8945	-1.54	.12
	block type * pair condition	***U***	0.575 [0.098,1.05]	8943	2.36	**.018***
	bonus * pair condition	***S***	0.263 [-0.214,0.739]	8943	1.08	.28
			2.11 [1.45,2.76]	884	6.32	**<.001*** **
ATX	***(intercept)***		3.95 [3.39,4.52]	52	13.8	**<.001*** **
	block type	***U***	-0.730 [-1.20,-0.257]	7390	-3.02	**<.001*** **
		S	-1.66 [-2.13,-1.18]	7390	-6.86	**.003** **
	bonus		1.92 [0.735,3.11]	41	3.17	**.003*** **
	pair condition		0.487 [0.049,0.924]	2521	2.18	**.030***
	block type * bonus	***U***	1.13 [0.576,1.68]	7391	4.00	**<.001** **
	block type * pair condition	***U***	0.477 [-0.082,1.04]	7390	1.67	.094
	bonus * pair condition	***S***	1.20 [0.644,1.76]	7390	4.22	**<.001** **
			2.56 [1.76,3.36]	1215	6.26	**<.001** **

Table 14 shows parameter estimates for the *pair condition* models. All models showed similar positive and significant coefficients on the main effect of *pair condition*, corresponding to an increased score on non-bonus trials for participants paired with another participant in the same condition, as well as *bonus*. Coefficients on the *bonus* main

effect *bonus * pair condition* interaction were also positive and significant, corresponding to an increase in bonus trials for participants paired in the same condition, which was larger for the atomoxetine than citalopram group. The main effect of *block type* was significant and negative for all groups, while there was a positive interaction of *block U * pair condition* for all except the atomoxetine group, and *block S * pair condition* for all except the citalopram group. The atomoxetine group also showed a significant *block U * bonus* interaction.

3.4.4 Learning models

Figure 13: model fits of each learning model across the relevant data sets

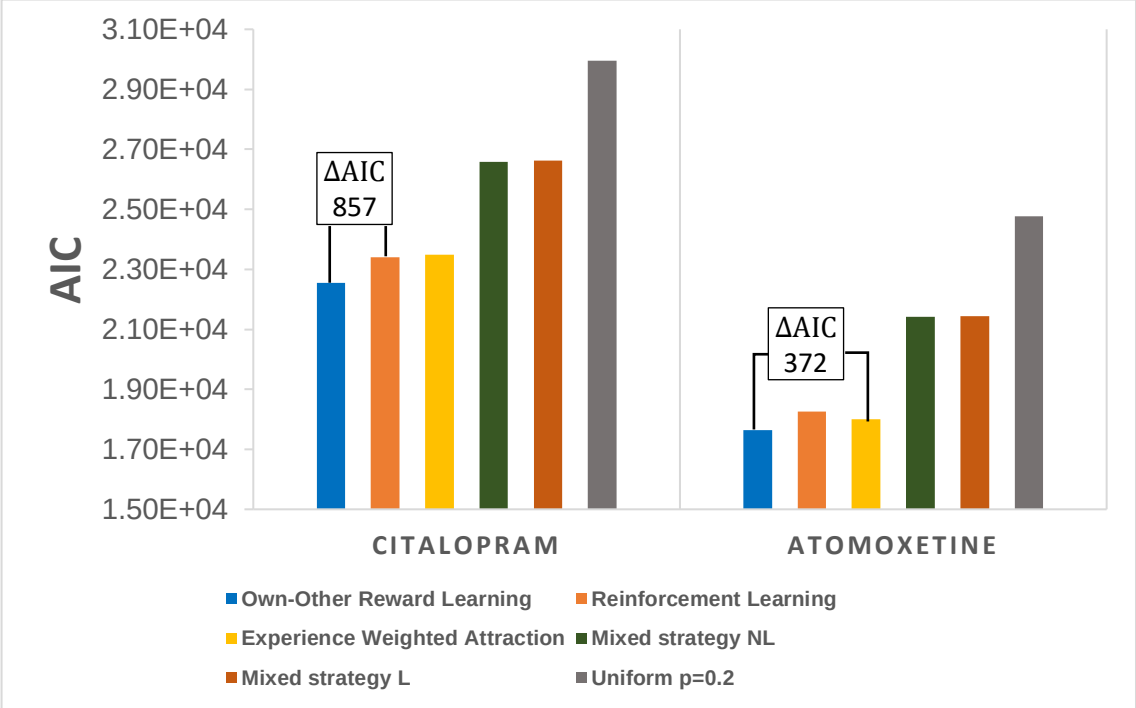
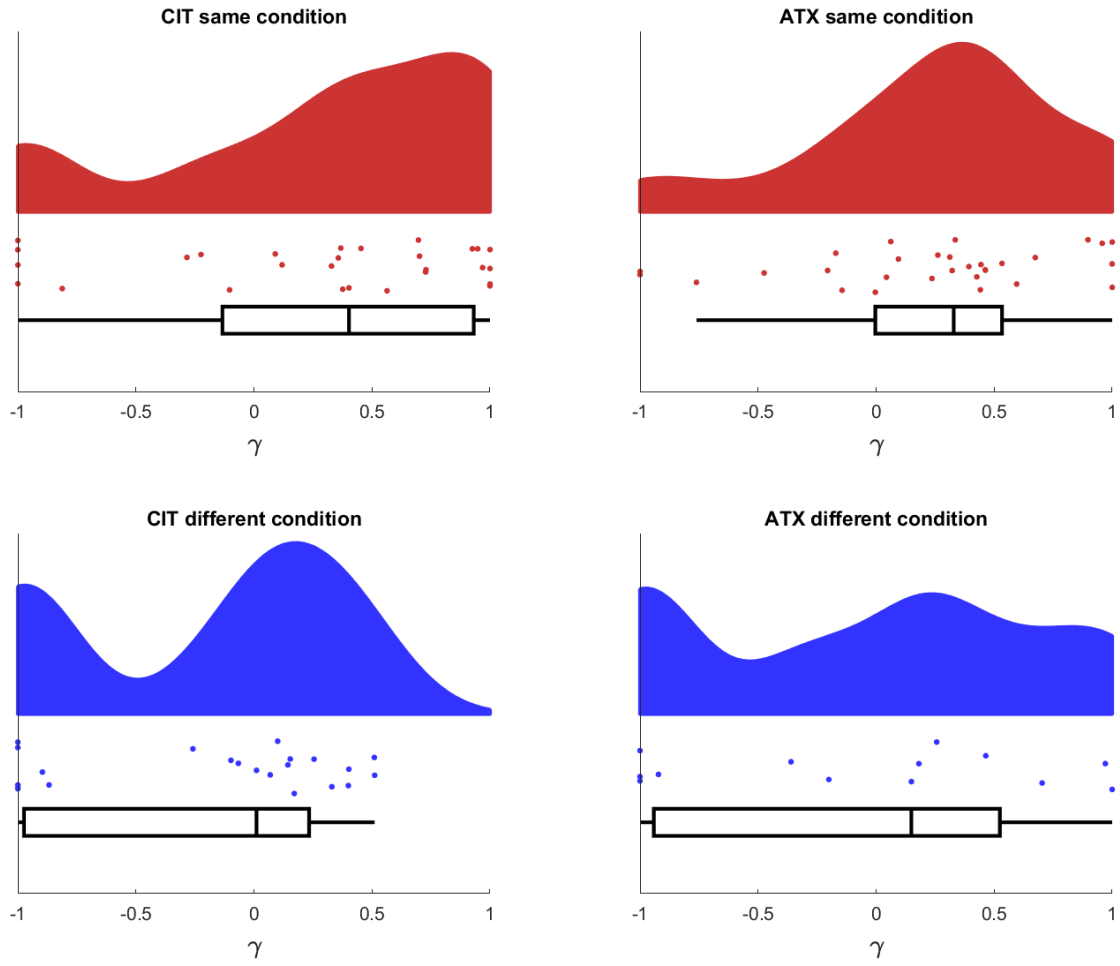


Figure 13 shows the fits of each learning model. The Own-Other Reward Learning model provided better fits than all others for both citalopram and atomoxetine, with probability > 0.999 of being the best model (see Supplemental Results for details).

*Table 15: comparisons of parameter estimates for Own-Other Reward Learning model by same/different condition, Mann-Whitney U-tests. ** p < .01*

group	param	same condition mean (SD)	diff condition mean (SD)	N same	N diff	*W*-stat	p
ALL	α	0.094 (0.183)	0.192 (0.249)	59	36	1258	.135
	β	2.48 (2.63)	3.18 (3.33)			1167	.43
	γ	0.263 (0.616)	-0.164 (0.645)			661	.002**
CIT	α	0.121 (0.691)	-0.223 (0.588)	29	23	391	.29
	β	2.71 (2.99)	3.97 (3.80)			407	.18
	γ	0.287 (0.691)	-0.223 (0.588)			178	.004**
ATX	α	0.068 (0.120)	0.144 (0.199)	30	13	223	.47
	β	2.27 (2.25)	1.79 (1.63)			169	.51
	γ	0.240 (0.544)	-0.058 (0.750)			151	.24

*Figure 14: plots of **γ** parameter estimates, with points showing individual session estimates, area showing kernel density estimates and box plots showing medians and interquartile ranges*



We performed parameter analysis on the split of sessions for pair condition, *as this was shown in the mixed models on* score *to provide the best predictor of outcome. As inspection of the distributions of parameter estimates showed multimodality in some conditions, we analysed parameters using Mann-Whitney U tests, summarised in Table 15. The* γ *parameter was significantly different for* pair condition *sessions, with higher* γ *when participants were in the same condition than in different conditions.*

Figure 14 illustrates these parameter estimates. Additionally, as with the *score* models, we tested parameter estimates split by *drug*. None were significantly different between drug and placebo (see supplementary Table18).

3.5 DISCUSSION

The *choice* model showed evidence of participant choices being interdependent overall, as choices were lower for non-bonus trials in blocks where bonus trials were present compared to the control block when they were absent, which is consistent with turn-taking based on bonuses – playing lower on non-bonus trials and higher on bonus trials. As this game is repeated, a strategy to try and win all trials would be open to reciprocal punishment, so restricting higher choices to trials with a higher potential outcome is potentially a social strategy that also solves the coordination problem. The lack of significant interaction between *block type* and *bonus* shows however that mean choices were not largely affected by the bonus structure. This suggests that the structured bonuses of *S* block were not generally learned in the limited numbers of trials available, so participants had limited ability to react to the other person's bonus trials (as opposed to coordination on the basis of their own trial incentives), and hence did not show adaptation of behaviour to the bonus structure either to enhance coordination or cooperation compared with *U* block. However, as noted, this is not a prerequisite for coordination to occur based on the knowledge of own bonus availability.

Models on scores show that being in the same condition as the other pair member was a better predictor of scores than any other factors considered. The task setup was designed to eliminate any cues from the other pair member aside from their behaviour on

the task, so other perceived similarities such as gender should not play a strong role. The *pair condition* factor showed a significantly positive main effect in all drug group models (all participants, citalopram and atomoxetine groups), as shown in Table 13. This demonstrates an effect independent of the bonus trial structure, with higher scores when participants were in the same condition as one another.

The effect of bonus trials was positive and strongly significant in all three. Notably this is not an artefact of the bonus trial structure alone, as shown in section 3.3.4.1 – playing at the one-stage mixed strategy Nash equilibrium would result in mean scores of 4 regardless of the type of trial. Thus, as with choice models, behaviour evidences a turn-taking rather than competitive approach overall. The interaction between *bonus* and *pair condition* was also strongly positive and strongly significant in all models, suggesting that turn-taking was more apparent when participants were in the same condition. As a corollary, main effects of both *U* and *S* blocks were negative – lower scores resulted when it was not the participant's 'turn' to win, in keeping with the choice model's finding of lower choices overall in *U* and *S* against *C* block. Interactions show higher scores in *pair condition* in both *U* and *S* blocks overall, in *U* block for the citalopram group and *S* block for the atomoxetine group, although the non-significant *block type * pair condition* interactions also trend in the same direction. These can be interpreted in the light of coordination in the same condition taking time to evolve and so being more apparent in the later blocks, while the presence of the *bonus* * *pair condition* interaction may also model some of the variance of the block effect due to bonus trials only being present in *U* and *S* conditions.

Learning models on choices showed that the Own-Other Reward Learning was the best fitting. The only parameter of this model showing significant differences between *pair condition* levels was $\boldsymbol{\gamma}$ or the weight and valence on the other's reward. For participants in the same condition, this was significantly higher in the full set and the citalopram group, while trending in the same direction for the atomoxetine group. Notably, mean $\boldsymbol{\gamma}$ values were positive when participants were in the same condition and negative when in different

conditions. Thus, when participants were in the same condition, choices that lead to greater reward for the other pair member were on average more positively reinforcing, suggestive of more cooperative behaviour, with the opposite effect when they were in different conditions.

These findings suggest that coordination was more easily achieved when individuals were in the same state induced by pharmacological challenge than when in different states. As reciprocity is a factor in repeated interactions due to 'greedy' behaviour being susceptible to punishment, reinforcement of the other's reward may be entirely self-interested; nonetheless, successful coordination requires an accurate assessment of the other individual's pattern of behaviour which may be facilitated by being in a similar state. While it is possible that response to the bonus trial structure may be to some extent in line with a framing effect, where non-bonus trials in *U* and *S* blocks have comparatively lower reward to that available in bonus trials and so prompt lower choices, there is no clear reason why this would differentially change by pair condition. Additionally, the main effect of pair condition on score, and the fact that scores already tended to be higher in the control block prior to any bonus trials for same condition pairs, suggests that coordination provides a better picture of observed effects.

Research on pharmacological effects of learning in paired interactions has been limited, and previous studies (Tse & Bond, 2002, 2003; Wood et al., 2006) all used a confederate design with only the participant on the drug. This study (to our knowledge) provides the first comparison of interactions between participant pairs in same vs different treatment conditions, so the finding of markedly different outcomes and learning patterns between the two in the absence of significant effects of the drugs individually is an intriguing result.

Parallels lie in models of communication as mutual prediction and joint action (Friston & Frith, 2015; Garrod & Pickering, 2009). Friston & Frith (2015) build a theory of

communication based on convergence of generative models of behaviour between interactors to minimise mutual prediction error. Cast in this light, our finding suggests that both pair members may be better able to predict one another's behaviour when in the same pharmacological state, with more coordinated play as the result.

The learning model developed here did not explicitly model prediction. However, since outcomes are dependent on both parties' actions, and the learning model is based on the outcomes of previous actions rather than a forecast of unknown future behaviour, accurate inferences on the reward to the other from taking an action are dependent on their behaviour being stable and thus predictable. Otherwise, an apparently cooperative action (for example, playing zero to allow the other person to take the reward from a trial because it was 'their turn' to win) may lead to a suboptimal outcome for both sides (if the other person then plays zero themselves) and thus not be reinforced. Hence our finding on the **γ** parameter is consistent with enhanced mutual prediction under the same pharmacological state.

Notably and not in line with our hypotheses, drug condition did not show any significant differences in mixed models on score or in learning models on choices. The discrepancy from previous research may lie in the interactive method that we used. While this may have increased the noise of the drug comparisons and thus inflated the type II error risk of not finding a true effect (and for atomoxetine there was a trend effect on *drug* for score as well as the **α** parameter of the learning model), it may also speak to limitations of previous research that did not incorporate an interactive approach. As real-life interactions are dynamic in nature, with the actions of both interactors contributing to the outcome rather than one party showing static behaviour, we believe that this study gives a more realistic picture of interaction under pharmacological challenge.

3.5.1 Limitations of this study

The study was designed to look at the contrast between behaviours under pharmacological challenge against placebo, with the findings on pair condition being post hoc. Assignment to the factor was incidental, so in all groups (particularly the atomoxetine group) numbers assigned to same vs different condition pairs were different, and the more powerful repeated-measures approach was not used for this factor. As a result, the atomoxetine comparisons on pair condition were underpowered, potentially explaining the lack of significant effects in the contrast of learning model parameters. However, as shown in Supplemental Table 16, proportions of *order*, *gender* and *same gender* did not significantly differ between levels of *pair condition*. As models were also estimated for each of these factors but were all significantly worse fitting than the *pair condition* model, the inference that pair condition plays a stronger role than any other factor considered remains sound.

3.6 SUPPLEMENTAL METHODS

3.6.1 Learning models

Action values in all learning models were initialised to zero, then were fitted for the 179 trials of each participant session (excluding the first trial), thus estimating a set of parameters separately for each participant's drug and placebo condition. As the highest number of free parameters in any model was 4, the ratio of sample size to free parameters was always at least 44.75, so AIC is considered a reliable approach to quantifying fit corrected for number of free parameters (Burnham & Anderson, 2002).

3.6.1.1 Q-learning

We fitted a standard reinforcement learning model, in which only the participant's own reward *R* is reinforced on each trial. This has a single free parameter $\boldsymbol{\alpha} \in [0,1]$ for the learning rate.

$$V_{t+1}(A) = \begin{cases} V_t(A) + \alpha[R_t - V_t(A)] & \text{if action A is chosen} \\ V_t(A) & \text{otherwise} \end{cases}$$

3.6.1.2 Experience-Weighted Attraction

The Experience-Weighted Attraction (EWA) model is a learning model featuring both reinforcement learning and fictitious play, the latter being reinforcement of unchosen strategies based on rewards they would have yielded given the other player's action. It is governed by two updating equations – the first determining the strength of prior beliefs, and the second updating values. There are four free parameters: $\boldsymbol{\varphi} \in [0,1]$ decaying previously learned values, $\boldsymbol{\delta} \in [0,1]$ weighting hypothetical rewards of unchosen actions, $\boldsymbol{\delta} \in [0,1]$ decaying prior weighting, and $\boldsymbol{\rho} \in [0,1]$ the discount rate for the strength of past experience controlling the influence of prior beliefs.

$$N_t = \rho N_{t-1} + 1$$

$$V_{t+1}(A) = \begin{cases} \dfrac{\varphi N_t V_t(A) + \pi_i[A_i, A_{-i}]}{N_t} & \text{if action A is chosen} \\ \dfrac{\varphi N_t V_t(A) + \delta \pi_i[A_i, A_{-i}]}{N_t} & \text{otherwise} \end{cases}$$

3.6.1.3 Own-Other Reward Learning

The Own-Other Reward Learning model (discussed in detail in section 3.3.4.4) reinforces chosen actions based on both rewards gained by the participant, and the reward O_t received by the other participant in the same trial. This has two free parameters, $\boldsymbol{\alpha} \in [0,1]$ for the learning rate, and $\boldsymbol{\gamma} \in [-1,1]$ for the weight on the other's reward.

$$V_{t+1}(A) = \begin{cases} V_t(A) + \alpha[R_t + \gamma O_t - V_t(A)] & \text{if action A is chosen} \\ V_t(A) & \text{otherwise} \end{cases}$$

3.6.1.4 Computing probabilities

For all learning models, probabilities of making each action were computed from values of the five available choices *j* using a softmax function (e.g. Sutton & Barto, 1998). This required estimation of an additional free parameter of decision temperature $\boldsymbol{\tau}$.

$$P_t(A_x) = \frac{e^{\frac{V_t(A_x)}{\tau}}}{\sum_{j \in J} e^{\frac{V_t(A_j)}{\tau}}}, x \in j$$

These probabilities were used to compute log-likelihoods based on actual participant choices at each trial, summed across all trials *T*.

$$LL(P_i, S_s) = \sum_{t \in \{2,3\ldots\}} \log P_t(A_t)$$

3.6.1.5 Null models

To check that the learning models chosen provided a better fit than static probabilities, we also fitted two null models based on probabilities in Table 8.

3.6.1.6 Model fitting

Learning models were fitted using Matlab's built-in fmincon function with Sequential Quadratic Programming algorithm (Nocedal & Wright, 2006) to find maximum likelihood parameters.

3.7 SUPPLEMENTAL RESULTS

3.7.1 Frequencies of factors within groups

Table 16: frequencies of factors across pair condition levels

		no of participant sessions					
		order		*gender*		*same gender*	
group	**pair conditions**	**1**	**2**	**F**	**M**	**N**	**Y**
CIT	**same**	12	17	18	11	12	17
	different	13	10	13	10	10	13
ATX	**same**	15	15	14	15	20	10
	different	7	6	8	5	8	5

As the *same drug* factor considered in the analysis was a post hoc categorisation that had not been controlled by design, we tested proportions of participant sessions to ensure there were no imbalances that could bias inferences. We computed Fisher's exact test statistics on each of the 2×2 contingency tables, *pair condition* against each other factor in Table 16. There were no significant differences in proportions of any other factor (all $p>.4$).

3.7.2 Learning model comparison

Table 17: ΔAIC_m for each candidate model (where best fitting model $AIC_m = AIC_{min}$)

		CITALOPRAM GROUP		**ATOMOXETINE GROUP**	
	Free parameters	$\Delta \boldsymbol{AIC_m}$	**P(m)**	$\Delta \boldsymbol{AIC_m}$	**P(m)**
Own-Other Reward Learning	3	0	**>.999**	0	**>.999**
Reinforcement Learning	2	857	<.001	630	<.001
Experience Weighted Attraction	4	935	<.001	372	<.001
Mixed strategy NL	0	4026	<.001	3790	<.001
Mixed strategy L	0	4080	<.001	3817	<.001
Uniform *p*=0.2	0	7413	<.001	7147	<.001

Table 17 shows ΔAIC_m for each candidate model. Through the comparison procedure in section 3.3.4.5, no other model had a probability higher than 0.001 of fitting better than the full Own-Other Reward Learning model.

3.7.3 Learning model parameter tests by drug

Table18: comparisons of parameter estimates for Own-Other Reward Learning model by drug condition, paired Wilcoxon signed-rank tests

group	parameter	placebo condition mean (SD)	drug condition mean (SD)	N	*W*-stat	p
ALL	*α*	0.130 (0.206)	0.139 (0.235)	45	556	.67
	β	2.72 (2.93)	2.79 (3.06)		525	.94
	γ	0.114 (0.650)	0.072 (0.668)		512	.85
CIT	*α*	0.148 (0.245)	0.185 (0.272)	25	135	.47
	β	3.10 (3.53)	3.40 (3.46)		152	.79
	γ	0.062 (0.698)	0.031 (0.716)		152	.97
ATX	*α*	0.106 (0.144)	0.082 (0.169)	20	153	.076
	β	2.24 (1.92)	2.03 (2.34)		120	.60
	γ	0.177 (0.596)	0.124 (0.618)		115	.73

No significant differences were found between any parameter estimates in drug vs placebo conditions.

3.7.4 Testing VAS differences

There were small but significant differences in the nausea scale between drug and placebo conditions for atomoxetine and a marginally significant difference for citalopram, corresponding to a change of 3.9 (CIT) / 6.1 (ATX) on a 100-point scale. There were no significant differences between the two drugs in drug-placebo differences.

*Table19: VAS score comparisons at test time, with scores on a 100-point scale. On antonym pairs higher numbers are closer to the second term. P - placebo condition, D – drug condition, * $p < .05$*

	CITALOPRAM GROUP				ATOMOXETINE GROUP				BETWEEN GROUPS	
scale	**P**	**D**	***t*-stat (df=26)**	***p***	**P**	**D**	***t*-stat (df=22)**	***p***	**t-stat**	**p**
nausea	4.32 (6.63)	8.24 (11.5)	-2.06	.050	3.37 (3.85)	9.48 (13.8)	-2.23	.036*	-0.668	.51
headache	10.8 (15.9)	11.26 (16.8)	-0.18	.86	4.91 (4.99)	7.33 (12.4)	-1.07	.30	-0.59	.56
dizziness	7.28 (12.7)	8.89 (8.7)	-1.01	.32	5.07 (5.55)	9.13 (11.7)	-1.92	.068	-0.94	.35
alert – drowsy	36.8 (17.0)	41.2 (18.8)	-1.15	.26	43.8 (14.0)	46.2 (18.3)	-0.534	.60	0.337	.74
stimulated – sedated	41.7 (13.3)	44.2 (13.7)	-0.91	.37	46.0 (11.3)	47.9 (14.1)	-0.569	.58	0.125	.90
restless – peaceful	65.3 (13.9)	62.7 (18.7)	0.80	.43	64.5 (12.1)	60.9 (17.1)	1.33	.20	0.21	.84
irritable – good-humoured	64.7 (15.4)	66.6 (14.5)	-0.76	.45	71.2 (13.7)	65.9 (13.5)	1.88	.073	1.89	.064
anxious – calm	70.8 (13.3)	68.7 (14.6)	0.63	.53	70.9 (13.5)	68.4 (14.7)	0.969	.34	0.09	.93

4 SEROTONERGIC EFFECTS ON INTEROCEPTION AND INTEROCEPTIVE METACOGNITION

4.1 ABSTRACT

Interoception is perception and interpretation of internal physiological states. It steers decisions, emotions, learning and experience of 'self'. Computational theories describe top-down weighting of interoception in cognition, emotion and behaviour according to the experienced precision of the interoceptive sensation. This process depends on accurate assessment of the interoceptive experience, shown to be disrupted in certain psychiatric disorders. Many of these same disorders respond to serotonergic treatment. A within-participant placebo-controlled test of forty-eight healthy adults demonstrated that acute serotonin reuptake inhibition increases awareness of interoceptive precision, without parallel effects in an exteroceptive domain. This motivates a theory of serotoninergic effects on cognition though shifts of assessment of interoceptive sensation, tying serotonergic function, disorders and treatments to allostatic responses to perceived changes of homeostasis.

4.2 INTRODUCTION

Interoception is the perception and interpretation of internal physiological states. Allostasis is the cognitive and behavioural response to interoceptive sensations to restore homeostasis, a system of self-guided survival. Since the 19th century visceral states have been shown to be accessed by human cognition and linked to human decisions, attention, arousal, memory and emotion (Craig, 2009; Cameron, 2002). They are closely related to the corporeal and psychological sense of 'self', such as the anticipation of how one will

experience and respond to future events (H.-D. Park & Tallon-Baudry, 2014; Seth, 2013; Tsakiris, 2017). Problems arise when interoceptive signals are too strong, weak, or misinterpreted. Correspondingly, interoceptive changes occur in psychiatric disorders (DuBois, Ameis, Lai, Casanova, & Desarkar, 2016; Ehlers, 1993; B. Herbert & Pollatos, 2019; Klabunde, Acheson, Boutelle, Matthews, & Kaye, 2013; Paulus & Stein, 2010). Advances in computational psychiatry provide new routes for understanding of these deficits, positing interoception as our most intimate measure of danger, arousal and reward in ourselves to which other information is compared, unless otherwise proven unreliable (Allen & Tsakiris, 2018; Seth, 2013). If one cannot assess the reliability of interoceptive cues, they are relied on too much, or too little. For example, inaccurate inappropriate fear-related interoceptive responses can potentiate further anxiety (Khalsa & Feinstein, 2019; Paulus & Stein, 2010). Inappropriate interpretation of hunger cues may be instrumental in eating disorders (B. Herbert & Pollatos, 2019). A key advancement for understanding interoceptive deficits has been the discovery of independent variance of different interoceptive processes (Garfinkel et al., 2015). Three component processes are considered: interoceptive accuracy (detection of interoceptive signals), sensibility (subjective assessment of one's ability to detect them), and awareness (the correspondence of subjective assessment to reality; Garfinkel et al., 2015).

Many disorders currently licensed for effective treatment with serotonin reuptake inhibitors (SSRIs), which include depression, panic disorder, social anxiety, generalised anxiety disorder and bulimia nervosa, are associated with changes of interoceptive experience (Ehlers, 1993; B. Herbert & Pollatos, 2019; Paulus & Stein, 2010). With depression and social anxiety, reduced 5-HT_{1A} receptor activity is present in anterior insula cortex (Drevets et al., 1999; Lanzenberger et al., 2007), a region consistently identified as a neural mediator of interoception (Craig, 2009; Schulz, 2016). There is also evidence of serotonin changes altering the coupling between certain neural cardiac measures (Mueller et al., 2012). However, no causal link between serotonin and interoception has been

established and therefore a unifying theory of serotonin effects by moderation of interoceptive processes could not be founded. Serotonin is historically associated with mood, aggression, impulsivity, and adaptive decision-making (Cools et al., 2011), but findings are often inconsistent using current interpretations, demonstrating the potential for an underlying effect that may help to explain individual differences and, for instance, the delay of effect on mood (Harmer & Cowen, 2013). The recalibration between cognition and interoceptive cues may be that common thread.

Metacognitive sensitivity in decision making refers to the ability of an individual to have insight into their own performance of an 'object-level' task (Fleming & Dolan, 2012), where an object-level stimulus is one at the sensory level or a memory. Consistent findings in exteroceptive tasks have demonstrated a dissociation with object-level performance (Fleming et al., 2010; Rouault et al., 2018; Song et al., 2011), and a dissociation between domains of exteroceptive metacognition and metamemory (Fleming et al., 2014). The same framework has been laid out in the interoceptive domain, with a similar dissociation between object- and meta-level performance (Garfinkel et al., 2015; 2016), with the latter study showing correspondence between interoceptive metacognition of cardiac and respiratory signals but a dissociation from tactile sensory metacognition. However, it remains unknown whether visual metacognitive performance is associated with interoceptive metacognition or if separate neural systems underlie the two. Prior psychopharmacological research has demonstrated that noradrenergic blockade using the β-adrenoreceptor antagonist propranolol (but not dopaminergic blockade using amisulpride) can improve perceptual metacognitive efficiency without affecting object-level performance (Hauser, Allen, Purg, et al., 2017). No comparable experiments have been carried out with serotonergic agents.

In the present study, we contrasted serotonergic effects on interoception with placebo, as well as an alternative pharmacological manipulation. Participants were tested using citalopram, employing a double-blind placebo crossover design. Citalopram is a highly

selective SSRI, binding to the serotonin transporter with 3,800 times the affinity to the norepinephrine transporter and 10,000 times the affinity to the dopamine transporter (Michael J Owens et al., 2001). This gives a high degree of confidence around the serotonergic nature of the manipulation.

We tested participants on two well-validated procedures to measure interoception – the heartbeat discrimination and tracking tasks (Garfinkel et al., 2015; Katkin et al., 1982; Schandry, 1981). Each assessed three dissociable measures of performance: interoceptive accuracy, awareness and sensibility. There is some evidence that the two tasks measure different underlying processes (Garfinkel, Manassei Miranda F., et al., 2016), so using both allowed for wider comparison with previous literature.

Interoceptive ability was measured using the heartbeat discrimination task (Whitehead et al., 1977). It measures an individual's ability to identify whether auditory tones presented were in or out of sync with the participant's heartbeat. Accuracy, sensibility (i.e. confidence in accuracy judgements) and awareness measures were taken. By current standards, this provides most precise definition of interoceptive processes at different levels (Garfinkel et al., 2015). Participants also performed a heartbeat tracking task counting perceived beats over an interval of time (Garfinkel et al., 2015; Schandry, 1981). The latter task has known confounds related to heart rate changes (Zamariola et al., 2018), so analysis of this task used cardiac measures to statistically control for these effects. We hypothesised that citalopram would lead to enhanced interoception, in line with its antidepressant effect countering the bluntened interoceptive sensitivity seen in depression (Furman et al., 2013; Pollatos et al., 2009). We assessed visual metacognitive ability on a separate control task (from Fleming et al., 2014), on which we had no hypothesised effect of citalopram. We predicted that drug effects on interoception would be independent of changes to visual metacognition, in line with earlier research showing dissociations of metacognition across modalities (Garfinkel, Manassei Miranda F., et al., 2016).

4.3 METHODS AND MATERIALS

4.3.1 Experimental Design

This study used a double-blind placebo repeated-measures design, with 20mg citalopram delivered orally. This dosage has been shown to elicit cognitive effects in previous research (M. Browning et al., 2007; Grillon et al., 2007), and were chosen to balance active drug effects of interest against unwanted side effects. Participants underwent two test sessions under medical supervision, ingesting the active drug dose in one session and a placebo (an identical capsule containing cellulose) in the other session, with a randomised counterbalanced order of presentation. No-one who had contact with participants was aware of the treatment order, which was pseudo-randomised, balanced for gender, and coded by a researcher who was not present during testing.

4.3.2 Participants

Fifty-one participants were recruited. Citalopram group testing was conducted in two separate testing periods and locations, as one set was also included in a neuroimaging study with scanning carried out after the tasks in this study were completed; only behavioural results are presented in this paper. On a separate occasion prior to testing, prospective participants undertook a screening session with a health questionnaire, heart rate and blood pressure monitoring by a medical doctor, and a structured interview to determine any undiagnosed psychiatric conditions (Mini International Neuropsychiatric Interview; MINI, Lecrubier et al, 1997).

Exclusion criteria included: age under 18 or over 35 years; history of psychiatric disorder (including anxiety disorder, depression, eating disorder, psychosis and substance abuse disorder); presence of significant ongoing medical condition (including migraine, diabetes, epilepsy, glaucoma and hypertension); pregnancy or breastfeeding; currently

taking any medication (excluding contraceptive pill); first-degree family history of bipolar disorder; Mini International Neuropsychiatric Interview (MINI) indication of: major depressive episode, manic episode, panic disorder, social phobia, OCD, PTSD, alcohol dependence, substance dependence, mood disorder with psychotic features, psychotic disorder, anorexia nervosa, bulimia nervosa, generalised anxiety disorder, or antisocial personality disorder. They were also instructed to abstain from alcohol or caffeine in the preceding 12 hours before the start of test sessions.

This study received ethical approval from the University of Sussex Sciences & Technology Cross-Schools Research Ethics Committee (ER/JL332/3, ER/JL332/9).

4.3.3 Procedure

Each participant performed a battery of tests including the interoception tasks at the estimated peak absorbency of each drug (Milne & Goa, 1991; Sauer et al., 2005): between 3 and 5 hours after administration for citalopram. Citalopram can exhibit side effects (usually mild at the dose used here) including nausea, headache and dizziness (Ekselius et al., 1997). Visual analogue scales (VAS; from 0-100) were given to assess for the presence of these three somatic effects. Additionally, five emotion/arousal related effects were also assessed with pairs of antonyms: alert–drowsy, stimulated–sedated, restless–peaceful, irritable–good-humoured, anxious–calm. Sets of measures were taken three times each, immediately following dosing and at the start and end of behavioural testing. Mean scores for the two testing times were used in analyses, with paired t-tests to analyse whether significant differences occurred between drug and placebo conditions.

4.3.4 Tasks

For both tasks, participants were connected to a fingertip pulse oximeter (Xpod with 8000SM sensor, Nonin Medical Inc., Minnesota, USA) and undertook the two tasks run in Matlab (version 2018a, Mathworks) using a variant of the tasks originally developed in Hart, McGowan, Minati, & Critchley (2013), while their heart rate was monitored.

Heartbeat discrimination task

The discrimination task (Garfinkel et al., 2015; Katkin et al., 1982) is a two-alternative forced choice task. Participants were instructed beforehand that the computer would play a set of tones that would be either in or out of sync with their heartbeat. During each trial, their heartbeat was measured in real-time, while a computer played a set of ten tones at either the beginning of the rising edge of the pressure wave at the pulse oximeter (~250ms after R-wave; Payne, Symeonides, Webb, & Maxwell, 2006), or an interval of 300ms later. These correspond respectively to judgements of maximum and minimum simultaneity between stimulus presentation and heartbeat (Wiens & Palmer, 2001). Following each trial, the participant was directed to respond whether the tones were in or out of sync with their heart, and how confident they were in that answer using a Likert scale ranging from 'total guess' to 'complete confidence'. Synchronous and asynchronous trials were presented in pseudorandomised order. Twenty trials were carried out in each session.

Heartbeat tracking task

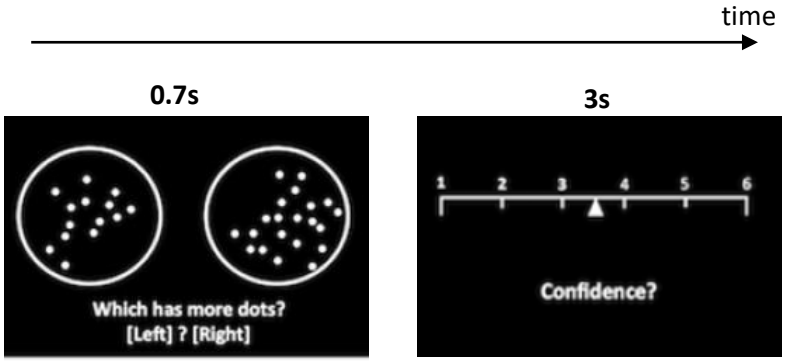
The tracking task involved trials of varying intervals (25-50 seconds), in which the participant was directed to count the number of heart beats that they felt within the interval and report this to the experimenter. As with the discrimination task, they then reported how confident they were in that answer using a Likert scale. Participants were not aware of the length of each interval in advance of each trial, and were directed when to begin and end counting by a voice recording. Six trials were carried out in each session.

Participants were not permitted to use any external cues such as monitoring their pulse using a finger. After the second session, the participants were asked whether they had been able to feel their pulse in the finger attached to the probe. Three participants' data were excluded due to the presence of these sensations.

Technical errors prevented a full set of trials being completed for two participants on the tracking task and one participant on the discrimination task. This left 47 participants considered in the discrimination task analysis (16 males, age M=23.3, SD=4.0) and 46 in the tracking task (15 males, age M=23.2, SD=3.9).

Visual metacognition task

Figure 15: order of stimulus presentation for visual metacognition task



The visual metacognition task was taken from Fleming et al. (2014). Participants were shown circles containing dots and instructed to indicate which contained more. Following each trial, they were asked to indicate their confidence in the previous response on a Likert scale. 200 trials were conducted in 8 blocks, with a self-timed rest every 25 trials. The difficulty was staircased over the course of the trial, with the difference in numbers of dots continually adjusted to so that the mean rate of correct answers was 70%, in order to prevent ceiling effects.

4.3.5 Analysis

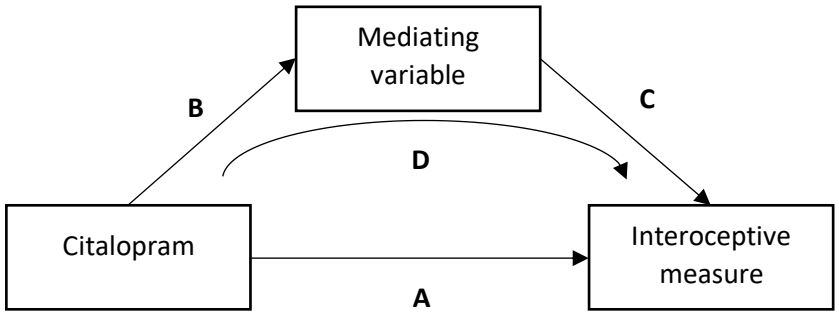
For the heartbeat discrimination task, discrimination accuracy scores were calculated by taking the mean number of correct responses for the session and dividing by the number of trials. Discrimination awareness was calculated through receiver operating characteristic (ROC) curve analysis (Green & Swets, 1966; Hajian-Tilaki, 2013), measuring the correspondence of trial-by-trial relative changes in confidence to accuracy independent of individual differences in overall level of confidence, fitted using Matlab. Heart rate (HR) in each interval was calculated, as well as heart rate variability (the standard deviation of HR across intervals). Tracking accuracy scores across the six trials in each session were computed as Knoll & Hodapp (1992):

$$accuracy = 1 - \frac{1}{6}\sum \frac{\left|beats_{real} - beats_{reported}\right|}{beats_{real}}$$

Sensibility scores were computed as the mean of the trial-wise confidence VAS measure in each task, coded 0-10. For the visual metacognition task, visual metacognitive efficiency (VME) measures were calculated as the ratio of meta-d' to d' (Fleming & Lau, 2014; Green & Swets, 1966) were calculated in Matlab.

We then used GLMs to model within-subject differences between each drug and its placebo, including treatment order and gender as effect-coded covariates to both model potential effects of these confounds and determine the significance of main effects decorrelated from confounds.

Figure 16: potential mediation pathways



To test if significant interoceptive changes were mediated by changes in heart rate, visual metacognitive efficiency or subjective report measures, we additionally modelled mediation using the Judd, Kenny, & McClelland (2001) approach. This was carried out for all potential mediators showing either significant differences in the mediating variable between drug and placebo ($\Delta M_{mediator}$, i.e. showing an effect through path B in Figure 16), or significant correlations between their drug-placebo differences and those of task variables (ΔM_{task}, an effect through path C). Additionally, to test the hypothesis of a drug effect mediated by general metacognitive processes, we carried out mediation analyses on this for all significant interoceptive effects.

Mediation analyses were carried out by first regressing $\Delta M_{mediator}$ with order and gender confounds to determine the drug effect on the mediator (path B). Next, a GLM was carried out on the task measure including both mediator and confounds, where the beta values on ΔM_{task} and $\Delta M_{mediator}$ determined effects on paths A and C respectively. The indirect effect was calculated as the product of the two $\Delta M_{mediator}$ betas (path D). This allowed us to statistically separate out the direct (unmediated) effect of the drug from the indirect (mediated) effect which the drug might influence through somatic or psychological changes.

4.4 RESULTS

4.4.1 Drug effect on task performance

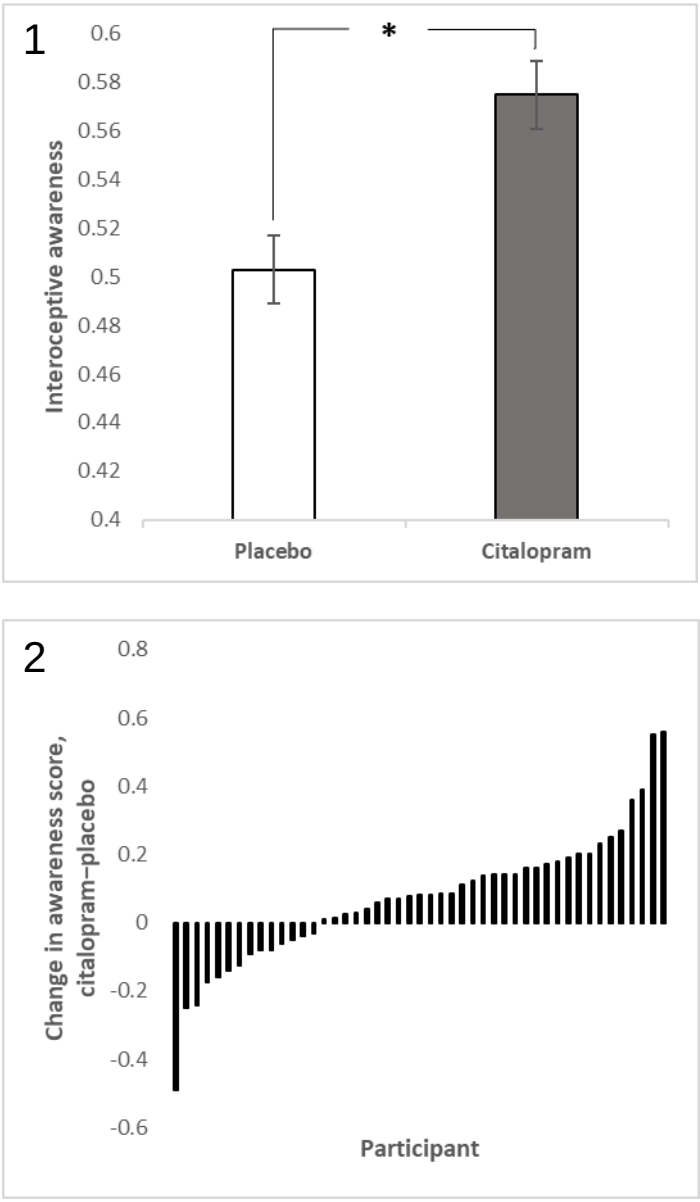
*Table 20: results of all task measures for model with order and gender as covariates. * p < .05, ** p < .01.*

	condition			
measure	**placebo**	**drug**	***F*-stat**	***p***
discrimination accuracy	0.562 (0.131)	0.583 (0.155)	0.09	.76
discrimination awareness	0.503 (0.181)	0.575 (0.165)	6.58	**.014***
discrimination sensibility	5.03 (1.84)	5.47 (1.48)	2.48	.12
visual metacognitive efficiency	0.789 (0.367)	0.766 (0.334)	0.004	.95

Table 20 shows the results for the interoception discrimination and visual metacognitive task. GLMs with drug condition as a within-subject factor and order and gender as covariates confirmed a significant difference between placebo and citalopram for measures of awareness on the discrimination task. Interoceptive awareness was higher on drug than placebo (Figure 17). Interactions with order and gender showed small and non-significant effects (order: p = .56, gender: p = .65). There was no effect of citalopram on interoceptive accuracy or confidence on the discrimination task, only their correspondence. Tests on a restricted set with very low changes in somatic effects still resulted in the observed effect on interoceptive awareness (see Supplemental).

As the tracking task effect showed a trend correlation with changes in heart rate ($r(47) = -.261$, $p = .080$), we restricted analysis on this task to mediation analysis.

Figure 17: Effect of citalopram on cardiac interceptive awareness. (1) Mean interoceptive awareness in each condition. Error bars are within subject 95% CI. (2) Change of interoceptive awareness for each participant.



While the tracking task offers only 6 trials to generate an interoceptive awareness measure, and so would be underpowered to draw a strong inference on its own, the effect of citalopram on this measure correlated between the two tasks, across subjects (after controlling for changes in heart rate, partial $r(42) = .29$, $p = .027$ one-tailed). This indicated generalisability of the SSRI effect on interoceptive awareness.

Significant interactions were seen with order for tracking accuracy ($p = .049$) and metacognitive efficiency ($p = .019$).

4.4.2 Drug effects on mediators and mediation analysis

*Table 21: results of mediation analyses on discrimination task interoceptive awareness. HR – heart rate, VME – visual metacognitive efficiency, † in units of awareness scores, ‡ in units of the mediator, * p < .05, ** p < .01*

	mediator		
	HR	**nausea**	**VME**
direct drug effect on interoceptive awareness (path A) †	0.0844	0.0746	0.0769
***t*-stat**	2.33	2.31	2.61
p	**.025***	**.026***	**.013***
effect of drug on mediator (path B) ‡	-4.05	3.91	0.0041
***t*-stat**	-3.76	2.30	0.06
p	**<.001****	**.026***	.95
effect of mediator on measure (path C) †	0.0012	0.00018	0.0973
***t*-stat**	0.271	0.06	1.46
p	.79	.95	.15
indirect effect through mediator (path D) †	-0.0051	0.00071	0.00040

No changes in mediation variables between drug and placebo showed significant or trend correlations with change in interoceptive awareness (see supplemental results) – notably, cardiac variables showed almost zero correlation (heart rate: $r(47) = -.005$, $p = .98$; heart rate variability: $r(47) = -.004$. $p = .98$). However, there was a significant reduction in heart rate (measured in beats per minute: $F(1,44) = 14.1$, $p < .001$, $\Delta M = -4.05$, 95% CI [-6.22,-1.88]) and increase in nausea under citalopram (on 100-point scale: $F(1,44) = 5.31$, $p = .0259$, $\Delta M = 3.91$, 95% CI [0.491, 7.33]) compared with placebo, so both these and visual metacognitive efficiency were tested with mediation analysis. This revealed no significant effects of mediators on interoceptive measure, while the direct effect remained significant throughout. Indirect effects were all negligible compared to the direct effect, with the largest indirect effect for HR at ~6% of the size of the direct effect and in the opposite direction.

*Table 22: results of mediation analyses on tracking task interoceptive accuracy. HR – heart rate, HRV – heart rate variability, VME – visual metacognitive efficiency, † in units of awareness scores, ‡ in units of the mediator, * p < .05, ** p < .01*

	mediator			
	HR	**HRV**	**nausea**	**VME**
direct drug effect on interoceptive accuracy (path A) †	0.0510	0.0581	0.0549	0.0671
***t*-stat**	2.08	2.71	2.45	3.30
p	**.044***	**.010***	**.019***	**.002****
effect of drug on mediator (path B) ‡	-4.05	-0.72	3.91	0.0041
***t*-stat**	-3.76	-1.55	2.30	0.06
p	**< .001****	.13	**.026***	.95
effect of mediator on measure (path C) †	-0.004	-0.009	0.003	0.0818
***t*-stat**	-1.21	-1.40	1.32	1.80
p	.23	.17	.19	.079
indirect effect through mediator (path D) †	0.015	0.0068	0.011	0.00034

For the tracking task, changes in both heart rate and heart rate variability showed trend correlations with changes in tracking accuracy. These and the mediators shown significantly different between drug and placebo were all tested in mediation analyses. Results are shown in Table 22. Once again, indirect effects were considerably smaller than the direct effect, and no mediators showed a significant effect on tracking accuracy. There was no significant effect of drug on visual metacognitive efficiency, and a near zero effect compared to mean metacognitive efficiency (M = 0.778, SD=0.264). The effect of changes in metacognitive efficiency on interoceptive measures was larger but non-significant, and the indirect effects were considerably smaller than the direct effect of drug on interoceptive measures.

4.5 DISCUSSION

The central finding of this study demonstrates the serotonergic modulation of afferent interoceptive signal processing within the central nervous system. Serotonin changes can cause a change of interoceptive awareness in the absence of changes to visual metacognition and independent of changes in objective cardiac measures or subjective effects. This is the first demonstration of a controlled pharmacological manipulation of the brain's access to and metacognition of cardiac interoceptive cues, consistent with modern computational theory and interoceptive deficits in serotonin-related disorders.

Serotonin has previously been linked to measures of cardiac-neural coupling, where tryptophan depletion reduced evocation of an EEG potential linked with cardiac changes in response to feedback (Mueller et al., 2012), as well as other aspects of homeostatic function (Ray et al., 2011). Depression and impulsivity, both linked with serotonergic dysfunction (Dalley & Roiser, 2012; M. J. Owens & Nemeroff, 1994) are associated with lowered interoceptive ability (Herman et al., 2019; Pollatos et al., 2009). In our study, we found that serotonergic manipulation resulted in greater levels of accuracy about and awareness of interoceptive function, suggesting a strengthening of neural-cardiac links. This therefore showed opposing effects to those characteristically seen in conditions associated with lowered serotonin levels, although in the same direction as those seen in anxiety-related disorders (Ehlers, 1993; Köteles & Doering, 2016; Pollatos et al., 2009). This is in line with findings of SSRI treatment resulting in relatively early improvement in depressive symptoms (Taylor, Freemantle, Geddes, & Bhagwagar, 2006) but anxiogenic effects (Nutt, 2005). Notably, while sensibility (or overall confidence in responses) trended in the same direction, it did not show significant increases on either task, showing limited ability to adapt to these changing signals.

Prominent computational theories of anxiety and depression characterise them as excessive weighting of bodily evidence leading to inability to adapt expectations when the

external environment changes. Coupled with overly precise expectations of negative outcomes through a negative model of the world, the result is persistent miscalculation of the likelihood of negative events occurring (Paulus et al., 2019). In the long run these can become entrained into depressive biases, with frequent comorbidity. This makes altered interoception as potentially key to the aetiology of these disorders. Our finding reveals that serotonergic pathways influence interpretation of interoceptive signals, and that serotonergic manipulation could therefore increase the weight of internal over external signals. Notably however, we did not find significant increases in anxiety through subject state report, consistent with earlier observations of acute citalopram influences on behaviour (M. Browning et al., 2007; S. E. Murphy, Norbury, et al., 2009), as our healthy population would not have the characteristic biases needed to bias interpretation of the upweighted signals as threatening. The lack of state anxiety change in our healthy participants may relate to the finding that consistent evidence connecting interoception with anxiety is limited to clinical cohorts (Domschke et al., 2010) and is predicted by mood disorder in a first degree relative (Harada et al., 2014). A direct link between serotonin and interoception may help to explain how anxiety symptoms accompany the initiation of selective serotonin reuptake inhibitor (SSRI) treatment, but wane with the later onset of therapeutic effects (M. Browning et al., 2007; Harada et al., 2014; Kent et al., 1998). The vulnerability could be psychologically (as well as biologically) determined: a negative-bias in the processing of somatic information is more prevalent in clinical presentations of mood disorders (Paulus & Stein, 2010). Specifically, vulnerable individuals may be more likely to interpret serotonin-mediated increased interoceptive precision of cardiac sensations as a signature of threat as in Paulus et al. (2019), while updating of internal models towards a low threat external environment requires repeated exposure and therefore a delay. This research therefore may inform how serotonergic treatments work, differ between individuals, are delayed in their clinical effectiveness (Frazer & Benmansour, 2002) due to relearning of interoceptive processing, and have cognitive, affective and at times

paradoxical anxiogenic effects in the short term (Harmer & Cowen, 2013) due to new access or interpretation of interoceptive events. The benefit of such a theory is that it is a closer to basic biological processes than current cognitive measures, making findings more accessible to general theories of serotonin function in the brain (Dayan & Huys, 2009).

While previous pharmacological research has provided direct evidence for the link between noradrenergic-modulated sympathetic arousal and anxiety in panic disorder (Gurguis et al., 1997; Pohl et al., 1988), and the β-adrenoreceptor antagonist propranolol is used as a treatment for anxiety (Steenen et al., 2016), our findings are (to our knowledge) the first identified pharmacological effect directly linked to interoceptive processing itself rather than modulation of bodily information. Accumulating evidence shows that interoceptive therapies improve patient outcomes in symptoms including anxiety and depression (Khoury et al., 2018). Thus understanding interoception's neural substrates provides a leap forward in our understanding of mood disorders including depressive, anxiety, eating, dissociative and somatoform disorders which affect at least 322 million people worldwide (WHO, 2017).

We also offer important insight into the differing roles of interoceptive processing at different cognitive levels. A large body of literature in perceptual decision making revealed dissociable processes of perception and metacognitive oversight (Fleming & Dolan, 2012). Focal metacognitive deficits have been shown from psychiatric (Hauser, Allen, Rees, et al., 2017; Rouault et al., 2018) and neurological (Fleming et al., 2014) conditions, and metacognitive processes are central to consciousness and selfhood in Higher-Order Thought theory (Lau & Rosenthal, 2011). By contrast, despite a similar dissociation being mapped out and experimentally verified in interoception (Garfinkel et al., 2015; Garfinkel, Manassei Miranda F., et al., 2016), we provide the first insight into pharmacological pathways of interoceptive metacognition. Notably, previous work has linked bodily self-awareness (Blanke et al., 2002; Guterstam et al., 2015; Ionta et al., 2011) and specific cardiac mediation of self-related thoughts (Babo-Rebelo, Richter, et al., 2016) to the brain's default

mode network (DMN). These areas have been repeatedly linked to self-related processing (Qin & Northoff, 2011) and show altered patterns of activity in depression that are restored with SSRI treatment (Cheng et al., 2017; A. Dutta et al., 2019). Our finding motivates future work to determine whether serotonergic effects on interoceptive metacognition may be mediated by common structures with other aspects of self-related processing (see next chapter). It also lays the groundwork for future clinical work targeting these higher-order processes for treatment of psychiatric conditions related to interoception.

The impact of serotonin on interoception was pronounced. This does not preclude the involvement of other neurotransmitters changing access to cortical representations of internal state. For example, the concentrations of the inhibitory neurotransmitter GABA within insular cortex (measured using magnetic resonance spectroscopy; MRS) correlates with interoceptive experience (Wiebking et al., 2014). Serotonin modulates specific aspects of GABA activity (Ciranna, 2006) and correspondingly a subset of the therapeutic effects of SSRIs is suggested to be GABA-mediated (Luscher et al., 2011). Further research would be needed to disentangle these effects.

The findings on interoceptive accuracy in the citalopram group were partially mediated by heart rate changes, so may be partially related to underlying vascular changes. Nonetheless, the connection between shifts in vascular activity and interoceptive processes does not negate their potential contribution to behaviour – if altered interoceptive signals can affect emotional states, this effect may be present regardless of whether neurally or physiologically mediated. Further research is needed to see how any of these effects persist with chronic changes in serotonin, and if so the extent to which they can affect emotion and other processes that are linked with interoception.

We demonstrated that cardiac interoceptive processing is causally affected by pharmacological manipulation of serotonin. This finding advances the neurochemical understanding of the central experience of the body and self, and provides important new

implications for understanding the mechanism of effect of serotonin on emotion, decision-making, interoception-related disorders and their treatment.

4.6 SUPPLEMENTAL RESULTS

4.6.1 Somatic and psychological effects

*Table 23: VAS scales, mean score at test time. † full sample, ‡ restricted sample, * p < .05, ** p < .01*

	scale	placebo	citalopram	*t*-stat (df=46† / 37‡)	*p*
FULL SAMPLE	nausea	4.47 (7.29)	9.18 (11.8)	-2.93	.005**
	headache	12.3 (17.3)	12.2 (17.6)	0.05	.96
	dizziness	9.44 (13.0)	11.02 (12.1)	-1.22	.23
	alert – drowsy	44.0 (18.2)	47.8 (19.5)	-1.27	.21
	stimulated – sedated	46.4 (15.6)	47.7 (14.9)	-0.54	.59
	restless – peaceful	64.3 (18.6)	61.2 (18.7)	1.02	.31
	irritable – good-humoured	63.9 (17.7)	64.8 (16.1)	-0.49	.62
	anxious – calm	71.9 (15.9)	67.6 (15.7)	1.75	.086
RESTRICTED SAMPLE	nausea	4.74 (7.71)	5.1 (6.45)	-0.46	.65
	headache	12.5 (18.3)	10.2 (16.9)	1.46	.15
	dizziness	9.35 (11.4)	10.1 (12.1)	-0.63	.53
	alert – drowsy	43.3 (17.8)	46.6 (19.2)	-1.11	.27
	stimulated – sedated	46.6 (15.2)	47.4 (15.0)	-0.31	.76
	restless – peaceful	62.7 (19.2)	64.0 (17.8)	-0.44	.66
	irritable – good-humoured	63.5 (17.8)	66.11 (16.1)	-1.4	.17
	anxious – calm	70.7 (16.9)	70.0 (15.6)	0.27	.79

Table 23 shows the difference between scores in drug and placebo conditions on subjective ratings at test times. There were small but significant differences in the sample on the subjective nausea rating, corresponding to a mean 4.7 difference on a 100 point scale. To determine whether this could affect inferences on the drug effect (for example, increased nausea causing participants to become more aware of their bodily sensations), we conducted a second analysis restricted to participants with less than a 10-point change in nausea (38 participants). The citalopram effect on interoceptive awareness was similarly significant in this sample ($F(1,35) = 5.62$, $p = .023$).

4.6.2 Correlations

Table 24: correlations between drug-placebo changes in cardiac/self-report variables and interoception measures

	discrimination awareness		tracking accuracy	
	r(47)	p	r(46)	p
heart rate	-.0036	.98	-.261	.080
heart rate variability	-.0037	.98	-.269	.071
nausea	-.1367	.36	.0992	.51
headache	-.0041	.98	-.103	.50
dizziness	.0343	.82	-.123	.41
alert–drowsy	.0175	.91	-.127	.40
stimulated – sedated	-.0823	.58	-.0062	.97
restless – peaceful	.0098	.95	-.0610	.69
irritable – good-humoured	-.0112	.94	-.0591	.70
anxious – calm	.0216	.89	.0312	.84

No changes in cardiac or self-report variables between drug and placebo showed significant correlation with discrimination awareness or tracking accuracy changes.

5 NEURAL CORRELATES OF SEROTONERGIC EFFECTS ON INTEROCEPTION

5.1 ABSTRACT

Interoception, or processes directed at internal physiological states, is altered in anxiety and depression which show serotonergic dysfunction. We found previously that acute challenge with the selective serotonin reuptake inhibitor citalopram could enhance measures of interoceptive ability. We sought to test whether the same manipulation would show changes in neural activity in areas previously linked to interoception and self-processing. We tested 22 healthy volunteers using 20mg citalopram in a double-blind placebo crossover design. We used an interoceptive task with conditions of both heart and stomach focus, along with a visual control task. We found increases in activity during interoceptive over exteroceptive focus, in nodes of the default mode network including posterior cingulate and inferior parietal lobule, as well as lateral orbitofrontal cortex and superior frontal gyrus. There was considerable overlap between interoceptive focus in heart and stomach modalities, with the latter showing more areas differentially recruited under citalopram. Our findings are suggestive of an early mechanism by which serotonergic treatment can affect interoception, with important implications for disorders presenting with interoceptive dysfunction.

5.2 INTRODUCTION

Interoceptive processes (processes directed at internal physiological states) are strongly linked to experience of bodily selfhood and emotion (Craig, 2009; Critchley & Garfinkel, 2017; Seth, 2013; Strigo & Craig, 2016). They are altered in various psychopathological conditions that are associated with serotoninergic dysfunction (Ehlers,

1993; Furman et al., 2013; Paulus & Stein, 2010; Pollatos et al., 2009). We have presented evidence in the preceding chapter that a single dose of citalopram enhanced interoceptive awareness. This motivated the current study to understand whether interoceptive processes showed altered neural signatures under serotonergic challenge, and compare these effects with patterns shown in disorders of serotonergic function.

The neural correlates of interoceptive processes have been studied in healthy populations (for an fMRI meta-analysis of 9 studies see Schulz, 2016). Brain regions correlated with interoceptive processing versus exteroceptive control tasks include the insula (8 of 9 studies analysed in Schulz 2016 plus overall significance in 3 clusters of the meta-analysis), in particular the posterior subdivision (all 3 significant insula clusters, 2 in right hemisphere) corresponding to Brodmann area (BA) 13. Other significant areas in the meta-analysis included precentral gyrus and medial frontal gyrus. Areas not significantly activated in the meta-analysis but found active during interoceptive focus in multiple previous studies include postcentral gyrus (Caseras et al., 2013; Critchley et al., 2004; Simmons et al., 2013; Wiebking & Northoff, 2015), superior temporal gyrus (STG; Caseras et al., 2013; Critchley et al., 2004; Wiebking & Northoff, 2015), inferior parietal lobule (IPL; Critchley et al., 2004; Pollatos, Schandry, Auer, & Kaufmann, 2007; Wiebking & Northoff, 2015) and cingulate gyrus (Critchley et al., 2004; Wiebking & Northoff, 2015; Zaki et al., 2012). Schulz noted that the meta-analysis combined tasks involving heartbeat counting (Caseras et al., 2013; Pollatos et al., 2007; Wiebking et al., 2010; Wiebking & Northoff, 2015; Zaki et al., 2012), tone synchrony judgements (Critchley et al., 2004) and non-specific attention to the heart (Avery et al., 2014; Kuehn et al., 2016; Simmons et al., 2013), with potentially different patterns of activity, which could explain why some areas did not show significance on the meta-analysis as a whole (which included tasks of both types) despite activation in multiple studies. We used significant areas in the meta-analysis for our hypothesised areas of serotonergic effect under acute citalopram.

Other evidence, from researched focussed on bodily awareness and selfhood, have linked these processes to default mode network (DMN) activity (Tsakiris & Critchley, 2016). The DMN is comprised of the interconnected regions of the brain that show activity increases during rest and decreases during externally-oriented attention. Studies using out-of-body illusions implicated the DMN nodes of IPL and posterior cingulate cortex (PCC) in locating the body in spatial surroundings (Blanke, Ortigue, Landis, & Seeck, 2002; Guterstam, Björnsdotter, Gentile, & Ehrsson, 2015; Ionta et al., 2011). Furthermore, specific DMN responses to cardiac activity have been shown to affect conscious visual perception (Hyeong-Dong Park et al., 2014) and encode selfhood in spontaneous thoughts (Babo-Rebelo, Richter, et al., 2016). This has led to proposals that both DMN and insula participate in connecting bodily awareness and selfhood through monitoring of visceral information (Babo-Rebelo, Wolpert, et al., 2016). With DMN showing altered connectivity in depression that is restored with SSRI treatment (Cheng et al., 2017; A. Dutta et al., 2019), this raises the possibility of serotonergic modulation of interoceptive processes through DMN node activity.

Studies in populations with disorders compared to healthy controls have found altered activation patterns, in keeping with the findings of behavioural studies. In a population with major depressive disorder, Avery et al. (2014) found decreased activity bilaterally in the dorsal mid insula during interoceptive focussing, as well as decreased activity in bilateral orbital frontal cortex, right caudate, right amygdala and left superior parietal lobule. Wiebking et al. (2015) found a similar decrease in right dorsal and ventral anterior insula and bilateral posterior insula. Kerr et al. (2016) contrasted anorexic females in remission with healthy controls, showing dissociations between insula regions and modalities of interoceptive focus (heart, stomach and bladder) with distinct differences between patients and controls – left dorsal mid-insula showing significant differences in stomach but not heart nor bladder, with higher activity in healthy controls, and right anterior insula showing less activity in heart but not stomach or bladder in healthy controls.

Interestingly, their findings on healthy controls showed lower BOLD signal in anterior insula in all interoceptive conditions compared with exteroceptive. While this appears to contradict the results of studies showing increased insula activity during heart focus (8 of 9 studies in meta-analysis by Schulz, 2016), there were important differences in the control trials. Previous studies used either counting (e.g. counting the number of appearances of a target within an interval; Avery et al., 2014; Pollatos, Schandry, Auer, & Kaufmann, 2007; Simmons et al., 2013; Wiebking et al., 2015; Wiebking & Northoff, 2015), oddball/target detection in auditory tones (Caseras et al., 2013; Critchley et al., 2004), or non-specific tone monitoring (Kuehn et al., 2016; Zaki et al., 2012). Conversely, Kerr et al's control trials involved monitoring of target intensity, which shifted from black to shades of grey at a rate approximating normal heart rate. This was chosen to provide a closer match to the interoceptive task, which was also intensity focussed. As the anterior insula is a component within the salience network mediating stimulus-driven attentional control (Uddin, 2015), the monitoring and reporting of varied visual stimuli intensity may have particularly recruited this area. In this study, we used a modified version of the Kerr et al. task.

While interoception in healthy and disordered populations has been studied in previous literature, little research has focussed on the neurochemical underpinnings of interoceptive processes. Wiebking et al. (2014) used magnetic resonance spectroscopy (MRS) to examine the relationship with GABA, showing that concentration of the inhibitory neurotransmitter GABA in left insula was correlated with neural response to interoceptive stimuli as contrasted with exteroceptive, proposing that insula inhibition during interoceptive awareness inhibits attention to external events in order to enhance internal focus. To our knowledge however, no previous research has used pharmacological challenge to causally link changes in brain activity with interoceptive processes.

Notably, both depression and anxiety are linked with serotonergic dysfunction (Akimova et al., 2009; M. J. Owens & Nemeroff, 1994), and serotonergic drugs such as selective serotonin reuptake inhibitors (SSRIs) are widely used in the treatment of both

(Kupfer et al., 2012; Nutt, 2005). This is particularly relevant as there is a lack of consensus around the neurocognitive changes seen in acute SSRI challenge, with both positive and negative biases in emotional processing shown (M. Browning et al., 2007; C. J. Harmer, Bhagwagar, et al., 2003), as well as anxiogenic effects in early treatment (Harada et al., 2014; Kent et al., 1998).

We adapted the task from Kerr et al. (2016), using two interoceptive focus conditions of heart- and stomach-focus, together with the control condition involving monitoring the intensity of a target changing between shades of grey, and used a double-blind placebo crossover design contrasting 20mg citalopram with placebo. Our specific hypotheses were that we would replicate findings on previously identified areas involved in interoceptive over exteroceptive focus, including medial frontal gyrus and precentral gyrus which were identified in the Schulz meta-analysis. As the Kerr et al. study showed reduced insula activity in the interoceptive-exteroceptive contrast, but other studies showed increased activity, we remained agnostic about the overall direction of effect. However, we predicted that the drug condition would show enhanced activity in insular and other interoceptive areas during interoception compared with placebo, in line with our previous behavioural findings of enhanced interoceptive processing under citalopram. We also predicted that changes might be seen in the activity of areas within the DMN, due to prior research connecting DMN activity with bodily selfhood and cardiac function. Participants tested on this task also formed part of the sample tested with interoceptive behavioural tasks discussed in the previous chapter. We hypothesised that changes in neural activity under citalopram would correlate with changes in interoceptive awareness on the heartbeat discrimination task, testing this hypothesis with a group-level covariate in a separate analysis.

5.3 METHODS AND MATERIALS

5.3.1 Participants

Ethical permission was granted by University of Sussex Sciences & Technology C-REC (ER/JL332/9). Potential participants were screened with a health questionnaire (see appendix) and the Mini-International Neuropsychiatric Interview (Sheehan et al., 1998). Exclusion criteria included: age under 18 or over 35 years; history of psychiatric disorder (including anxiety disorder, depression, eating disorder, psychosis and substance abuse disorder); presence of significant ongoing medical condition (including migraine, diabetes, epilepsy, glaucoma and hypertension); pregnancy or breastfeeding; currently taking any medication (excluding contraceptive pill); first-degree family history of bipolar disorder; Mini International Neuropsychiatric Interview (MINI) indication of: major depressive episode, manic episode, panic disorder, social phobia, OCD, PTSD, alcohol dependence, substance dependence, mood disorder with psychotic features, psychotic disorder, anorexia nervosa, bulimia nervosa, generalised anxiety disorder, or antisocial personality disorder; or scanner contraindications (e.g. metallic implants). They were also instructed to abstain from alcohol or caffeine in the preceding 12 hours before the start of test sessions.

24 healthy participants aged 18-35 were recruited for this study. Neuroimaging data from two participants were excluded due to excessive motion (>6% of volumes identified as motion outliers for either session scan; see section 5.3.4.3 for details), leaving 22 participants (7 males, age $M = 24.0, SD = 3.29$).

5.3.2 Study procedure

We used a repeated-measures design with separate drug and placebo sessions. held at least 7 days apart (days between sessions $M = 10.3$, $SD = 6.87$). Assignment to treatment order was double-blind and counterbalanced, with the drug treatment administered in one session and the placebo in another. Doses in the drug treatment

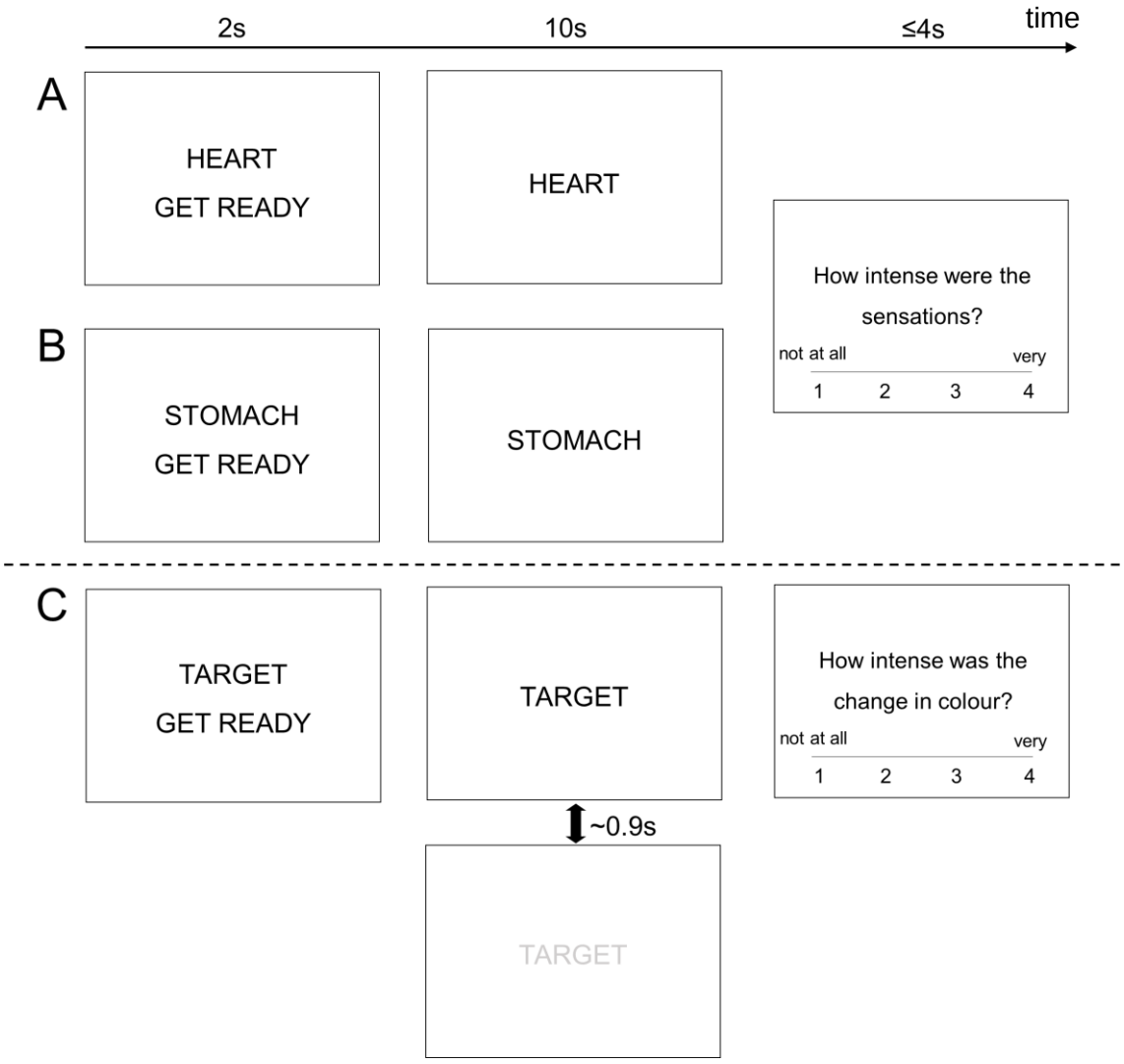
conditions consisted of 20mg citalopram. Drug and placebo doses were delivered in gelatine capsules, indistinguishable from one another, with the capsule filled with microcrystalline cellulose (in addition to the active drug in the drug conditions). Drug and placebo doses were all manufactured according to good manufacturing practice (GMP) guidelines.

Participants were also given eight visual analogue scales at four timepoints – immediately following the dose, preceding the start of the scan, following the end of the scan, and at the end of the test session. Scales (from 0-100) were given to assess three somatic effects (nausea, headache and dizziness) and five emotion/arousal related effects (pairs of antonyms: alert–drowsy, stimulated–sedated, restless–peaceful, irritable–good-humoured, anxious–calm) to measure whether the drug was affecting these measures.

Following the dose but prior to testing, participants were given instructions and practice on the task. To allow for drug levels to reach peak absorption (Milne & Goa, 1991; Sauer et al., 2005), participants commenced behavioural testing (not listed in this chapter) 3 hours after drug/placebo administration, with neuroimaging starting approximately 45 minutes later. The neuroimaging tasks included the interoceptive focus task described in the following section, and an emotional appraisal task, the full results of which are given in the next chapter.

5.3.3 Task design

Figure 18: Interoception scanner task, in (A & B) interoceptive conditions focussing on heart and stomach respectively and (C) exteroceptive baseline condition. Each condition was presented 15 times, and the ratings phase was presented 8 times for each of conditions A & B and 7 times for condition C.



The scanner task was based on the heart, stomach and target conditions of Kerr et al. (2016). Prior to the experimental trials, participants were given the following instructions: '*While the word 'HEART' or 'STOMACH' is shown on the screen, focus attention on the intensity of the sensations experienced from the area of your heart or stomach. When the word 'TARGET' is shown on the screen, it will sometimes change to from black to different shades of grey. Focus attention on the amount that the colour changes. When a CROSS is shown*

on the screen, you can rest. Please keep your eyes open and try not to think of anything in particular. You will sometimes be asked to rate how intense the sensations or the colour changes were, using the buttons shown. Please make sure that you breathe evenly throughout the task. You must NOT hold your breath at any time.'

Figure 18 shows the task design. The *Heart* and *Stomach* cue blocks corresponded to interoceptive conditions, and the *Target* blocks to an exteroceptive baseline condition. In the *Target* condition, after a short interval the target started changing from black to grey (0.15-0.85 between black and white) and continued cycling between shades for a pseudorandomly varied length of time between 0.7s and 1.1s, designed to mimic the approximate frequency of normal heart rate. On half of the trials, participants were asked to rate the intensity of the sensations from the area of focus (interoceptive trials) or the intensity of the change in colour (exteroceptive trials). This was done in order to maintain participant attention on the task. Rating trial presentation, condition order and shades of grey for the TARGET blocks were pseudorandomly varied. There were 15 trials in each condition, for a total duration of approximately 15 minutes.

5.3.4 Neuroimaging

5.3.4.1 Neuroimaging procedure

Testing took place at the Clinical Imaging Sciences Centre (CISC), Brighton and Sussex Medical School. MRI data were acquired using a Siemens Magnetom Prisma 3T scanner (Siemens Healthcare GmbH, Erlangen, Germany) and a 32 channel head coil. Stimuli were displayed using NNL's (NordicNeuroLab, Bergen, Norway) MRI-safe 40-inch monitor (1920×1080 resolution) at the head of the bore and a mirrored headset. Responses for participants on behavioural measures were recorded using two NNL MRI-safe 2 button response boxes affixed together, operated with their right hand.

5.3.4.2 Image acquisition

Functional images were acquired using T2*-weighted echo-planar imaging (multiband factor 4, echo time 37ms, repetition time 1500ms, voxel size 2.2 × 2.2 × 2mm, 104 × 104 voxels per slice, 72 slices, FOV 205 × 205mm^2, flip angle 52°), with a varying number of volumes according to the speed of task completion (*M* = 580, *SD* = 36.8). Prior to the task fMRI sequence, pairs of phase-encode reversed images were acquired for distortion correction (FOV 228 × 228mm^2 104 × 104 voxels per slice, echo spacing 0.54ms). On one session, a structural T1-weighted magnetization-prepared rapid acquisition gradient echo (MPRAGE) image was acquired (echo time 2.2ms, repetition time 2400ms, 0.8mm isotropic voxels, 300 × 320 voxels per slice, 208 slices, FOV 256 × 256mm^2, flip angle 8°).

To test for changes in resting perfusion which may accompany the pharmacological challenge and confound task-related changes in BOLD signal (Detre et al., 2012; D. J. J. Wang et al., 2011), pulsed arterial spin labelling (ASL) images were acquired using a FAIR-QII sequence (4 label-control image pairs, echo time 16ms, repetition time 4600ms, voxel size 1.5 × 1.5 × 3mm, 126 × 128 voxels per slice, 40 slices, interleaved slices, FOV 192 × 192mm^2, flip angle 180°, inversion time 1990ms, bolus duration 700ms).

5.3.4.3 Pre-processing

Functional MRI images were pre-processed with a standard FSL pipeline including motion correction using the middle volume of the timeseries as reference volume, identification of motion outlying volumes using FSL Motion Outliers (root mean squared intensity difference of adjacent volumes method), distortion correction using phase-encode reversed image pairs, high-pass filtering at 80 seconds, brain extraction and co-registration to subject structural images. Structural images were co-registered to standard space (Montreal Neurological Institute MNI-152 stereotactic template). Independent Component Analysis (ICA) decomposition was carried out using FSL-MELODIC (Beckmann & Smith, 2004) on unsmoothed data, and denoising was carried out manually, with noise

components identified and removed using criteria laid out in Griffanti et al. (2017). Following spatial smoothing with 5mm full-width at half maximum (FWHM) Gaussian kernel, a second denoising step was carried out to remove residual noise. ASL images were processed using Bayesian Inference for Arterial Spin Labelling (BASIL) to produce perfusion-weighted images.

5.3.4.4 Analyses

Functional images were all analysed using FSL FEAT v6.0.0 (Woolrich et al., 2001, 2004). We employed a three-level generalised linear model (GLM), with the first level as individual sessions using events corresponding to trial condition (*Heart*, *Stomach* and *Target*), along with regressors of no interest for parametrically modulated change in target intensity, fixation cross, ratings periods, button presses and outlying motion volumes. Second level models were within-subjects contrasts between drug and placebo sessions for participants, and third level was the group-wise analysis across participant session contrasts using FMRIB's Local Analysis of Mixed Effects (FLAME) 1 & 2. In a separate analysis, we also used data from the previous chapter of drug-placebo differences in interoceptive awareness as a group-level covariate.

First-level contrasts were employed on *Heart vs Target* and *Stomach vs Target*. Z (Gaussianised T/F) statistic images were thresholded non-parametrically using clusters determined by $Z > 3$, chosen to maintain Type-I error rates below the .05 level (Eklund et al., 2016), and a (family-wise error rate corrected) cluster significance threshold of $p < .05$. Group level analyses contrasted drug effects, employing covariates for order and gender. Following this, we performed a conjunction analysis (Nichols et al., 2005) on significant clusters from each contrast to determine the extent to which significant clusters overlapped between the interoceptive conditions. Additionally, 2 two-level GLMs were performed separately on the task contrasts in drug and placebo conditions, to compare the extent of significant activity.

Perfusion-weighted images were analysed at the group level in a paired t-test of drug and placebo sessions using Statistical Parametric Mapping (SPM12; https://www.fil.ion.ucl.ac.uk/spm/software/spm12), with cluster-wise family-wise error rate of $p < .05$, and with an additional exploratory threshold of $p < .001$ with extent > 50 voxels per cluster.

Behavioural analysis of drug-placebo differences on self-reported intensity and VAS scales was carried out using paired t-tests.

5.4 RESULTS

5.4.1 Behavioural results

Paired t-tests showed no significant differences between drug and placebo conditions on intensity measures (Table 25) nor on VAS self-reported measures (see Supplemental).

Table 25: results of self-reported intensity (normalised from 0 to 1)

condition	placebo	citalopram	*t*-stat (df=21)	*p*
heart	0.58 (0.13)	0.54 (0.13)	1.326	.20
stomach	0.46 (0.13)	0.44 (0.16)	0.765	.45
target	0.61 (0.11)	0.58 (0.08)	1.182	.25

5.4.2 ASL

The analysis of ASL image pairs showed no significant clusters at the familywise error rate, suggesting that citalopram compared to placebo did not affect cerebral blood flow in general, and that any effects on BOLD responses were not mediated by general effects on blood flow. For the uncorrected threshold, a single significant cluster was shown in the occipital cortex (peak voxel MNI coordinate: (18, -96, 14), $p_{uncorrected} = .001$, $p_{FWE} = .10$).

5.4.3 Task fMRI results: main effects and interaction

Table 26-Table 27 and Figure 19-Figure 21 show the main effects of *Heart vs Target* and *Stomach vs Target* contrasts and their interaction with drug condition: (*Heart/Stomach > Target) × (Citalopram > Placebo)*. There were no significant clusters in the inverse interaction.

Table 26: significant clusters in contrasts of Heart vs Target condition: mean across citalopram and placebo treatments, and drug interaction. STG – Superior Temporal Gyrus, SFG – Superior Frontal Gyrus, SPL – Superior Parietal Lobule, preSMA – pre-Supplementary Motor Area

Contrast	Hemis.	Region	BA	# voxels	Z score (max)	MNI coordinates (at max Z) x	y	z
Heart > Target	R	Precuneus	7	5347	6.69	20	-46	62
	L	STG	40	3644	6.27	-62	-24	14
	R	Cuneus	18	2909	7.18	10	-94	26
	R	Precentral Gyrus	44	2814	7.23	56	8	4
	L	SFG	9	1543	5.54	-8	64	30
	L	Middle Temporal Gyrus	39	617	5.59	-46	-74	28
	L	Parahippocampal Gyrus	36	352	4.61	-22	-18	-18
	L	Postcentral Gyrus	3	259	5.75	-48	-14	50
	L	STG	38	231	4.4	-48	14	-32
	R	Parahippocampal Gyrus	36	222	6.05	18	-14	-22
	R	Lingual Gyrus	18	196	5.06	10	-72	0
	R	STG	38	149	4.19	46	10	-28
	L	Cerebellum	-	135	4.07	-26	-60	-58
Target > Heart	L	Fusiform Gyrus	19	5983	7.12	-40	-76	-18
	R	Middle Frontal Gyrus	6	4853	6.13	34	6	60
	R	SPL	7	3849	9.33	40	-52	54
	R	Fusiform Gyrus	19	3570	6.8	46	-70	-16
	L	Intraparietal Sulcus	40	1263	7.11	-46	-42	42
	R	preSMA	8	845	5.37	6	26	52
	R	Caudate	48	419	4.69	12	2	16
	R	Cerebellum	-	202	4.38	2	-50	-34
	L	Middle Frontal Gyrus	10	179	4.86	-38	48	0
	L	Cerebellum	-	166	4.71	-18	-34	-42
	L	Cerebellum	-	165	4.23	-40	-44	-40
	R	Insula	13	161	5.16	36	20	-2
	R	Cerebellum	-	127	5.97	26	-38	-44
(Heart > Target) × (Citalopram > Placebo)	L	Lateral Orbitofrontal Cortex	10	102	3.94	-24	50	-2

Table 27: significant clusters in contrasts of Stomach vs Target condition: mean across citalopram and placebo treatments, and drug interaction. STG – Superior Temporal Gyrus, SFG – Superior Frontal Gyrus, SPL – Superior Parietal Lobule, IPL – Inferior Parietal Lobule

Contrast	Hemis.	Region	BA	# voxels	Z score (max)	MNI coordinates (at max Z) x	y	z
Stomach > Target	L	STG	42	3862	6.67	-60	-36	20
	R	Medial Frontal Gyrus	6	3643	7.72	4	-10	62
	L	Cuneus	18	3608	8.00	-8	-94	26
	R	IPL	40	2864	6.02	54	-28	32
	L	Medial Frontal Gyrus	10	2327	6.26	0	56	0
	L	Parahippocampal Gyrus	36	743	5.13	-14	-10	-22
	L	Middle Temporal Gyrus	39	716	6.16	-42	-74	36
	R	Parahippocampal Gyrus	36	535	6.15	18	-14	-20
	L	Inferior Frontal Gyrus	47	270	5.27	-38	32	-20
	R	STG	38	260	4.95	46	20	-34
	R	Lingual Gyrus	19	143	4.06	30	-68	6
Target > Stomach	L	Fusiform Gyrus	37	11738	8.32	-50	-68	-12
	R	SFG	9	9164	7.46	44	40	28
	R	Intraparietal Sulcus	40	4372	11.5	42	-50	44
	L	SPL	7	2142	5.88	-26	-74	54
	R	Caudate	-	1318	5.79	10	-2	14
	L	Middle Frontal Gyrus	9	545	5.88	-48	10	36
	L	Cerebellum	-	543	6.07	-24	-32	-42
	R	Substantia Nigra / VTA	-	476	5.03	4	-16	-20
	L	Caudate	-	440	4.59	-8	6	12
	L	Middle Frontal Gyrus	46	419	5.55	-40	44	0
	R	Cerebellum	-	172	5.09	22	-34	-40
(Stomach > Target) × (Citalopram > Placebo)	L	SFG	9	777	4.67	-4	48	48
	L	Lateral Orbitofrontal Cortex	10	432	4.48	-18	56	-6
	L	IPL	40	161	4.44	-50	-52	32
	L	Posterior Cingulate	30	130	4.42	0	-48	24

Figure 19: results of Heart vs Target contrast. Heart > Target – red-yellow, Target > Heart – dark blue-light blue

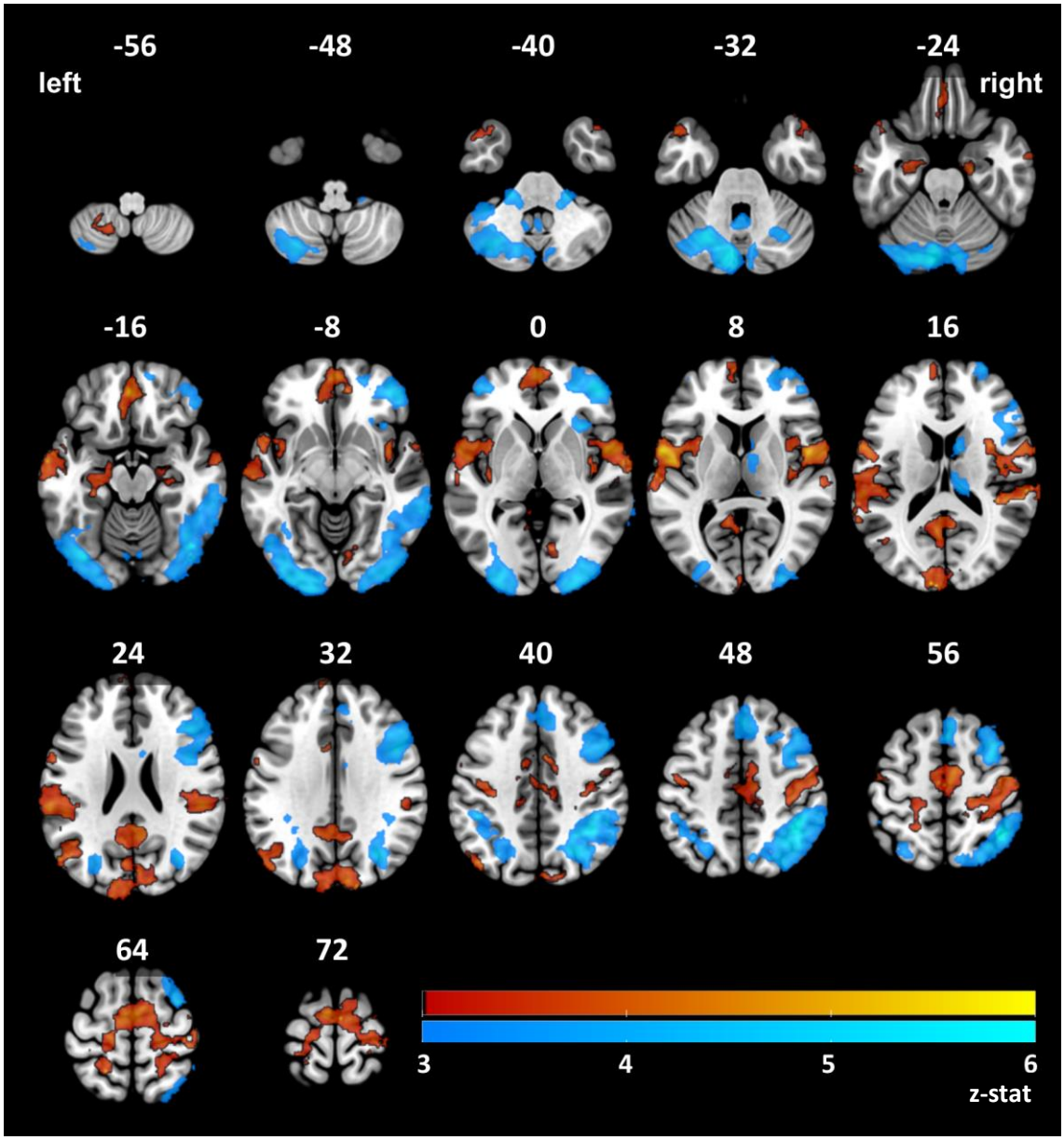


Figure 20: results of Stomach vs Target contrast. Stomach > Target – red-yellow, Target > Stomach – dark blue-light blue

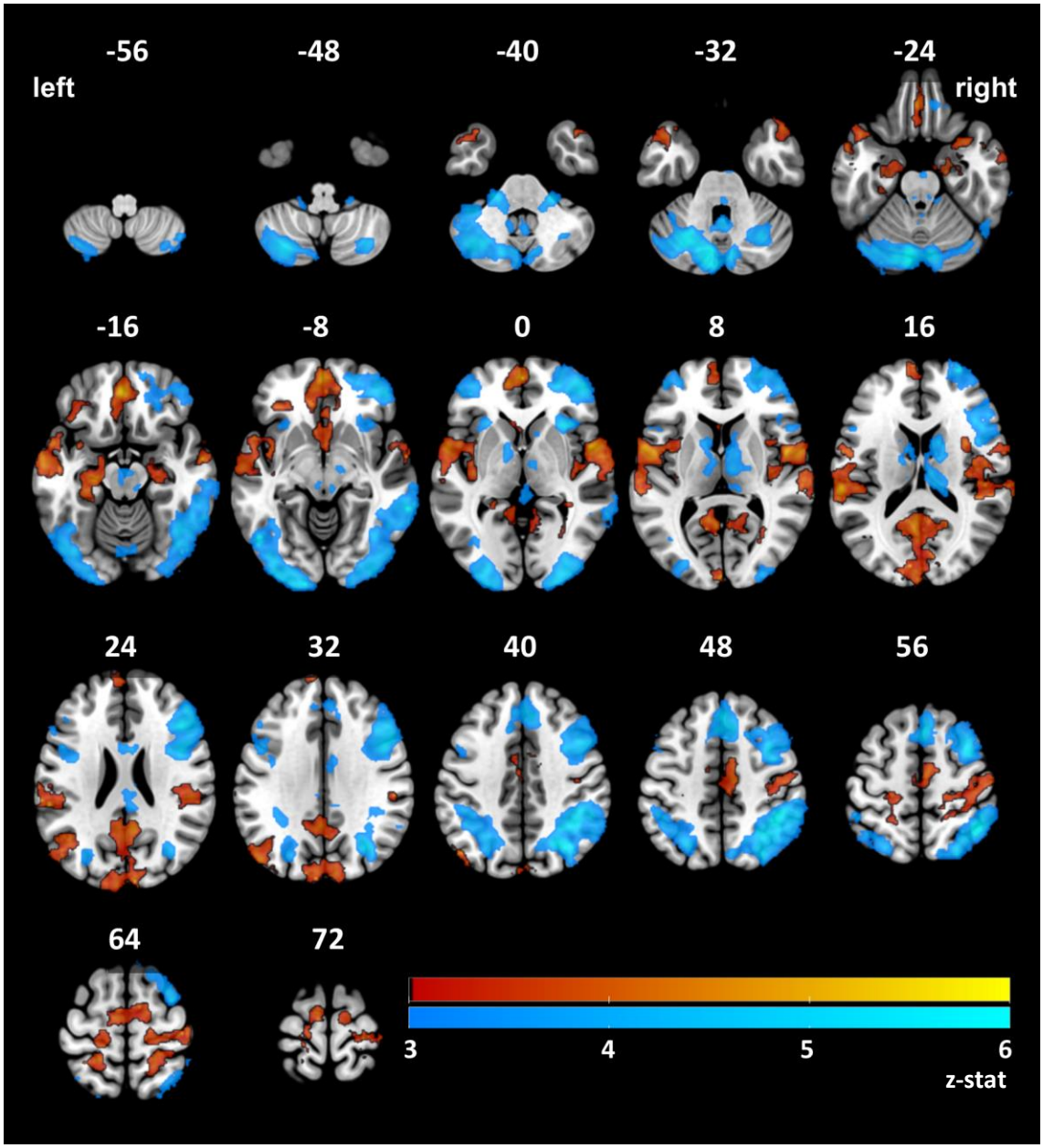


Figure 21: interaction of drug treatment condition with contrasts. Heart > Target – red, Stomach > Target – green, overlap – yellow (all clusters Z>3, $p_{FWE\text{-}corrected}$ <0.05, not shaded by Z value – see tables 26 & 27 for values)

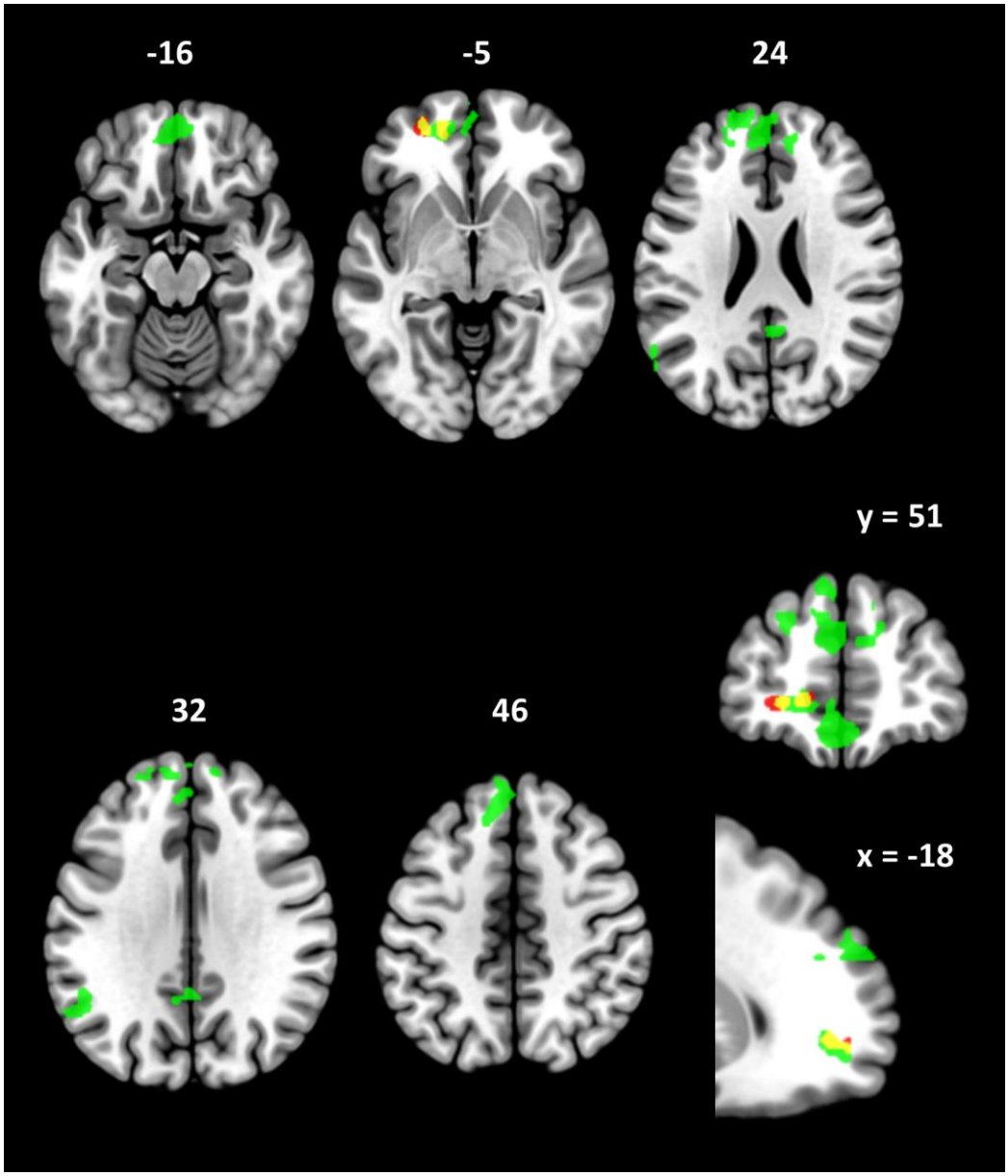


Figure 22 shows percentage changes in BOLD signals of the contrast-drug treatment interactions in the significant clusters from Figure 21. Activation in interoceptive conditions tended to show reduction under placebo compared to citalopram, and the opposite pattern was shown in the exteroceptive condition. Charts exclude one participant with signal change exceeding ±2.5 SD from mean.

Figure 22: BOLD signal change in lateral orbitofrontal cortex cluster for each parameter estimate in interaction (Heart > Target) × (Citalopram > Placebo). Error bars are within-subjects 95% CI

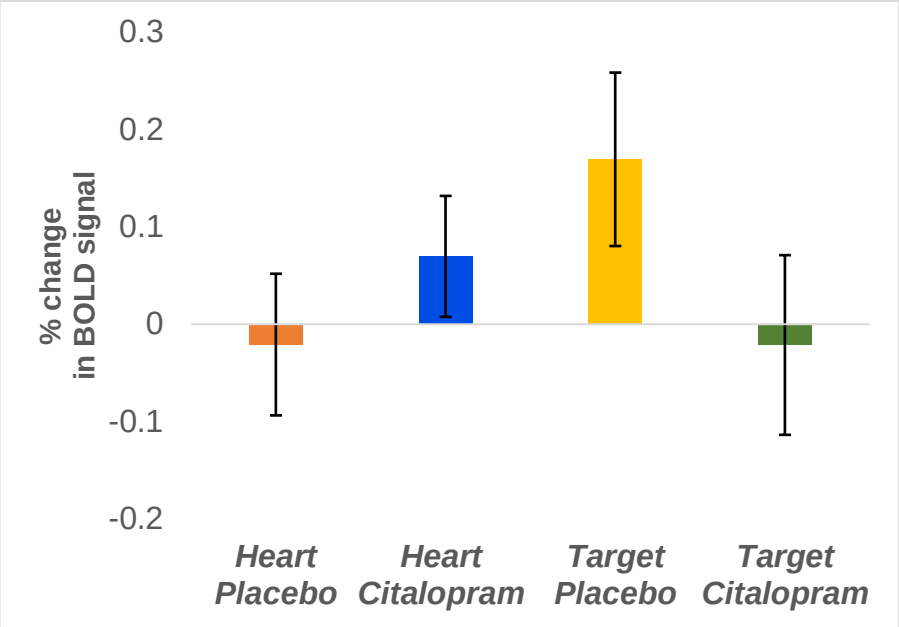


Figure 23: BOLD signal change in clusters for each parameter estimate in interaction (Stomach > Target) × (Citalopram > Placebo). PCC – Posterior Cingulate Cortex, IPL – Inferior Parietal Lobule, lOFC – lateral orbitofrontal cortex, SFG – Superior Frontal Gyrus. Error bars are within-subjects 95% CI

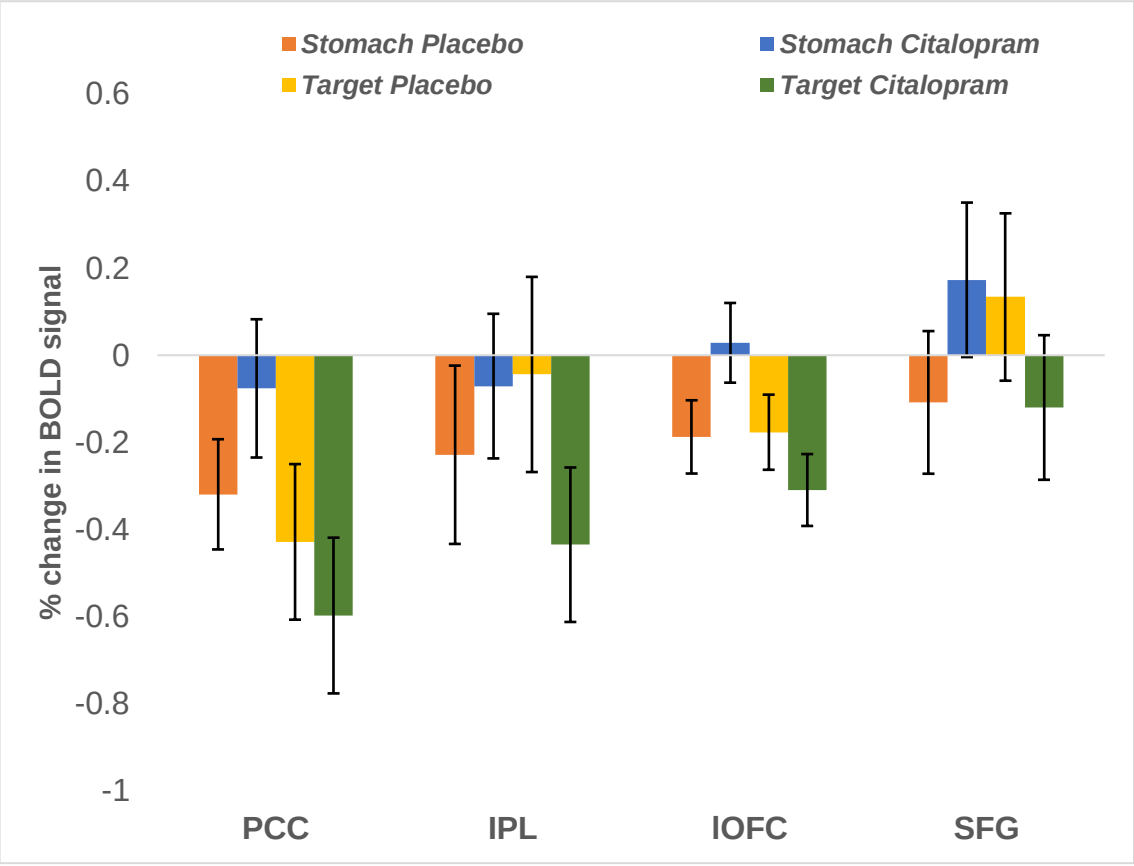


Table 28 shows the conjunction of *Heart vs Target* and *Stomach vs Target* contrasts, showing largely overlapping areas between the two sets of contrasts. The interactions with drug treatment did not show significant clusters in the conjunction.

Table 28: significant clusters in conjunction of Heart vs Target and Stomach vs Target condition, mean across citalopram and placebo treatments. STG – Superior Temporal Gyrus, SFG – Superior Frontal Gyrus, SPL – Superior Parietal Lobule

					Z	MNI coordinates		
					score	(at max Z)		
Contrast	**Hemis.**	**Region**	**BA**	**# voxels**	**(max)**	**x**	**y**	**z**
(Heart >	R	Postcentral Gyrus	3	3157	4.58	44	-20	58
Target) ∩	L	Precentral Gyrus	6	2716	5.16	-58	2	10
(Stomach >	R	Cuneus	18	2485	4.77	18	-86	22
Target)	R	Precentral Gyrus	44	2067	5.24	50	2	6
	L	SFG	9	1443	5.54	-8	64	30
	L	Middle Temporal Gyrus	39	504	4.8	-44	-74	28
	L	Parahippocampal Gyrus	36	343	4.33	-16	-8	-18
	L	STG	38	218	4.02	-44	14	-30
	R	Parahippocampal Gyrus	36	208	5.88	18	-14	-20
	R	Superior Temporal Gyrus	38	143	4.05	42	18	-38
(Target >	L	Cerebellum	-	5510	5.85	-10	-80	-34
Heart) ∩	R	Middle Frontal Gyrus	6	4763	5.62	34	6	60
(Target >	R	SPL	40	3709	8.08	42	-52	54
Stomach)	R	Inferior Temporal Gyrus	20	3299	5.67	60	-48	-14
	L	Intraparietal Sulcus	40	1194	5.19	-46	-42	44
	R	SFG	8	841	5.33	4	24	52
	R	Caudate	36	396	4.26	10	6	16
	R	Cerebellum	-	197	4.38	2	-50	-34
	L	Middle Frontal Gyrus	36	178	3.98	-38	48	0
	L	Cerebellum	-	166	4.71	-18	-34	-42
	L	Cerebellum	36	162	4.23	-40	-44	-40
	R	Insula	-	161	5.16	36	20	-2

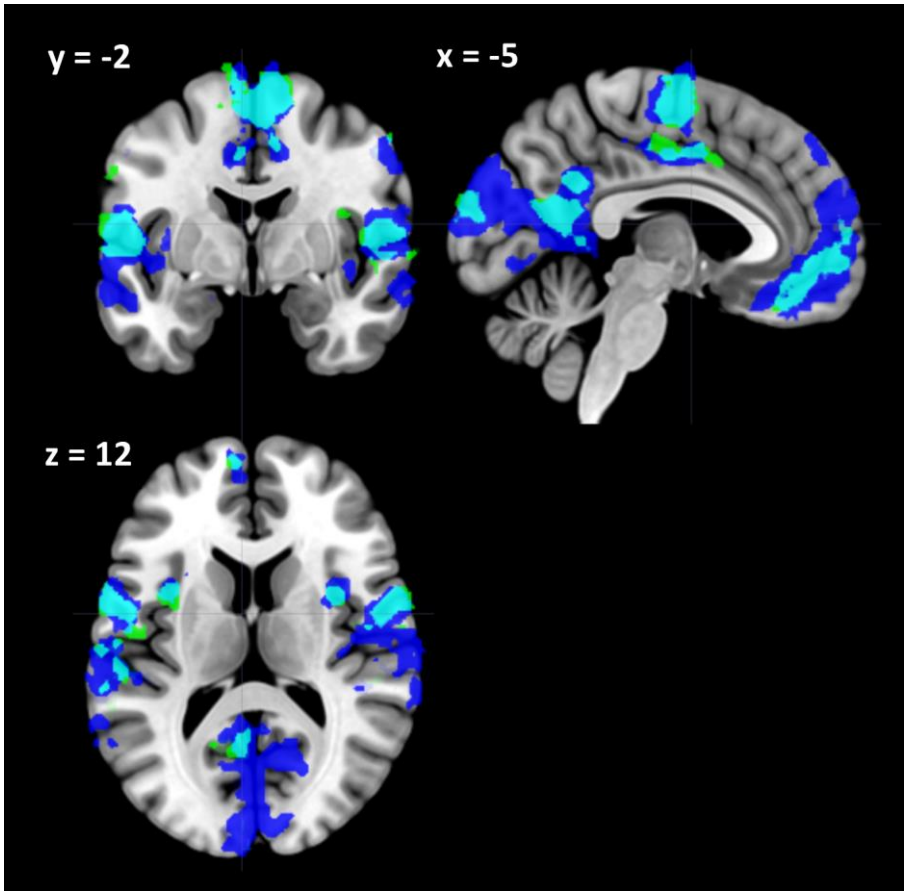
-

Table 29 shows the number of significant voxels in GLM contrasts carried out on citalopram and placebo conditions separately. Figure 24 shows this contrast for *Heart > Target.* A far greater extent of significant voxels was shown under citalopram than placebo for interoceptive conditions, and for placebo than citalopram for the exteroceptive control condition.

Table 29: total number of significant non-overlapping voxels comprising suprathreshold clusters in separate GLM analyses of drug and placebo condition for given contrasts

Contrast	Placebo condition	Citalopram condition
Heart > Target	46051	202294
Target > Heart	230279	92978
Stomach > Target	57520	165333
Target > Stomach	343056	93516

Figure 24: comparison of Heart > Target contrast, separate GLMs of citalopram and placebo conditions. Blue – citalopram, green – placebo, cyan – overlap



The group-level analysis using drug-placebo changes in interoceptive awareness showed no significant clusters.

5.5 DISCUSSION

We first sought to confirm and contrast our results with previous studies. We found several areas across the cortex which were differentially recruited during interoceptive focus, many of which were in common with those found in previous literature. This gave us confidence that our task was successfully measuring interoceptive focus. There was considerable overlap between *Heart* and *Stomach* conditions. This provides further evidence of common regions in the brain processing information across interoceptive modalities (Avery et al., 2015), where both have important roles in maintaining homeostasis and both receive input from afferent vagus nerve fibres (K. N. Browning et al., 2017; Koizumi et al., 1985). Additionally, citalopram differentially shifted the relation of neural correlates between interoceptive and exteroceptive focus. However, there was no observed neural effect of citalopram specific to interoceptive processing on its own.

Areas showing significant clusters in the average of citalopram and placebo conditions for both *Heart > Target* and *Stomach > Target* included superior and middle temporal gyri, cuneus, parahippocampal gyrus and lingual gyrus. Specific to the area of focus, heart-focussed interoception showed clusters in precuneus, precentral and postcentral gyri, superior frontal gyrus and cerebellum, while stomach focus showed activations in medial and inferior frontal gyri and inferior parietal lobule. The peak voxel of the cluster in precentral gyrus was within 6.5mm of the peak of a similar cluster identified in the Schulz (2016) meta-analysis and of two component studies (Critchley et al., 2004; Wiebking & Northoff, 2015), while three other studies also found recruitment of this area (Caseras et al., 2013; Kuehn et al., 2016; Simmons et al., 2013). Other areas in common with previous studies included, from *Heart* contrast: postcentral gyrus (Caseras et al., 2013;

Critchley et al., 2004; Simmons et al., 2013; Wiebking & Northoff, 2015), superior temporal gyrus (STG; Caseras et al., 2013; Critchley et al., 2004; Wiebking & Northoff, 2015), middle temporal gyrus (Pollatos et al., 2007), lingual gyrus (Terasawa, Shibata, et al., 2013), precuneus and cuneus (Simmons et al., 2013). From the *Stomach* contrast, we additionally found commonalities with heart-focussed areas of previous research in inferior parietal lobule (Critchley et al., 2004; Pollatos et al., 2007; Wiebking & Northoff, 2015), inferior frontal gyrus (Critchley et al., 2004; Zaki et al., 2012) and medial frontal gyrus (Caseras et al., 2013; Pollatos et al., 2007; Wiebking & Northoff, 2015).

For the interaction of drug treatment and focus condition, both interoceptive conditions (against target) showed a positive interaction with citalopram (against placebo) in lateral orbitofrontal cortex (lOFC), with stomach focus additionally showing positive interactions in superior frontal gyrus (SFG), inferior parietal lobule (IPL) and posterior cingulate (PCC). These were seen in the absence of any significant changes on resting cerebral blood flow as measured by ASL, with the only trend cluster distant to the significant BOLD interactions. Inspection of signal change for each condition revealed that for all clusters, citalopram was associated with higher BOLD signals than placebo, with the reverse for *Target* condition. This demonstrates a citalopram effect driven by both higher activity in interoceptive focus and lower activity in exteroceptive focus compared with placebo. All clusters were in areas shown to have involvement with interoceptive focus (Critchley et al., 2004; Pollatos et al., 2007; Simmons et al., 2013; Wiebking & Northoff, 2015; Zaki et al., 2012) or self-related processing (Henseler et al., 2011; Ruby & Decety, 2003) in previous literature. Three clusters were in areas previously associated with default mode activity: PCC (Greicius, Krasnow, Reiss, & Menon, 2003; Meindl et al., 2010; Spreng & Grady, 2009), SFG (Scheibner, Bogler, Gleich, Haynes, & Bermpohl, 2017; Yan et al., 2009) and IPL (Passow et al., 2015; Ryan, Sheu, & Gianaros, 2011; Vatansever, Menon, Manktelow, Sahakian, & Stamatakis, 2015). As noted in the introduction, both PCC and IPL have been associated with bodily aspects of selfhood, with the latter also specifically showing cardiac modulation in

conscious experience. Thus the finding that they are more active during interoceptive focus under citalopram suggests serotonergic contributions to awareness of the self, which are dysfunctional in serotonergic disorders that show deficits of interoception and emotional self-awareness (B. M. Herbert et al., 2011; Paulus & Stein, 2010).

The lOFC change during heart focus, with increased activity under citalopram, shows an intriguing parallel with the Avery et al. (2014) study. They reported higher bilateral lOFC activity in healthy compared to depressed populations during heart-focussed interoceptive vs exteroceptive attention. Our findings suggest that manipulation with a serotonergic antidepressant similarly alters activity in this region in healthy populations. Acute citalopram has been shown to increase processing of positive stimuli without concomitant changes in mood (M. Browning et al., 2007; C. J. Harmer, Bhagwagar, et al., 2003), so this finding is in keeping with early changes in processing that reflect its longer-term antidepressant effect. Both Browning et al. and Harmer et al. also showed increases in response to fearful stimuli, with the former also showing potentiation of the startle response. Anxiogenesis in early-stage SSRI treatment is a common side effect (Harada et al., 2014; Kent et al., 1998), which we postulated in the previous chapter may be linked to enhanced access to interoceptive signals prior to longer-term attenuating of biases that interpret these signals as threatening. Our findings point to potential neural correlates of this enhanced access. Contrary to our predictions, we did not find evidence of changes in insula activity under citalopram, nor in precentral or medial frontal gyri. However as previously noted, multiple areas support interoceptive focus, and our findings were in keeping with previous research indicating interaction between lOFC activity in interoception and serotonergic dysfunction.

It should be noted that overall, changes in BOLD signal were larger for the target condition than those of the interoceptive conditions. We did not have any specific hypotheses relating to changes in processing of visual stimuli under citalopram. However, there are parallels from findings of deactivation of DMN nodes during task performance

(Fornito et al., 2012), which is impaired in depression (Anticevic et al., 2012). Moreover in healthy participants, specific effects of serotonergic function in reducing both DMN connectivity during rest (Ven et al., 2013) and activity during task performance (Scharinger et al., 2014) have been shown. Thus, the effect of serotonergic challenge on areas known to be involved with self-referential processing appears to shift neural responses in different directions, reducing it when task demands are external and increasing it when internal focus is required.

As well as the clusters identified that showed significant differences between drug and placebo for each contrast, there were marked differences in the cluster extent between conditions. The drug condition showed a far greater extent of voxels meeting cluster significance thresholds for both interoceptive conditions than target condition, with the reverse pattern seen under placebo. While this should be seen as indicative, it suggests that citalopram is associated with a shift of activation patterns towards greater recruitment for internally-directed processes. Order effects were only seen in two clusters on the (*Stomach > Target) × (Citalopram > Placebo)* interaction (see Supplemental for details).

We did not see significant clusters in the group analysis using changes in interoceptive awareness as a covariate. Our task was designed to measures effects of interoceptive focus. Previous authors (Garfinkel et al., 2015; Schulz, 2016) have noted that different processes may be elicited by judgements requiring accuracy (heartbeat counting or synchrony judgements), awareness judgements of own interoceptive ability, and general interoceptive focus without these requirements. We also did not see significant shifts in self-reported intensity. While caution should be taken in interpreting these null results, they are in line with dissociation of interoceptive processes seen behaviourally at different levels (accuracy, awareness and self-reported confidence; Garfinkel et al., 2016, 2015). This suggests that SSRI effects seen here may have different substrates from those of processes involving higher-order awareness of interoception.

Areas responding more to target than interoceptive focus included areas related to visual attention such as fusiform gyrus and intraparietal sulcus (Bédard et al., 2011; Mukai et al., 2007; Sestieri et al., 2006), and frontal areas associated with externally-orientated task attention and cognitive control (Cole & Schneider, 2007; Liang Wang et al., 2010), as well as cerebellum. Additionally, a cluster in right anterior insula showed higher activation in target than heart focus. As noted in the introduction, this replicates Kerr et al's (2016) finding in healthy controls. This may be in line with anterior insula's role within the salience network mediating stimulus-driven attentional control (Uddin, 2015), with the salient changes in target intensity combined with instructions to monitor these changes rendering the insula more active than during the unchanging stimulus cueing interoceptive focus. Another potential explanation is that, as target phase was designed to mimic heart rate (though uncorrelated to the actual participant's heart), participants may have inferred or attempted to infer that the target phase reflected their heart rate, recruiting insula for this cross-modal comparison.

This study has shown that a single dose of citalopram increases activity in brain areas linked with selfhood and interoceptive processes, adding to a serotonergic theory of interoception. We showed changes in line with serotonergic enhancement of activity in brain areas previously shown be attenuated in depression during interoceptive focus and at rest. This shows parallels in acute neural and behavioural changes seen in other studies of SSRI effects in healthy volunteers. Future work would be needed to determine whether these changes correlate with clinical outcomes in patients receiving SSRI treatment, to gain an understanding of whether altered serotonergic function and interoceptive deficits are linked in mood disorders.

5.5.1 Limitations

While *Target* condition showed lower activation under citalopram in clusters shown in the interaction, post hoc testing on the parametrically modulated target response

revealed changes in the opposite direction – where the shift between shades was greater, activity under citalopram was also greater (figures in Supplemental). This potentially confounds inferences on exteroceptive conditions, although these did not form part of the hypothesised effects. Speculatively, a lower intensity shift (and hence less contrast between black and grey shades of target) may have prompted more attention on the target because shifts became more difficult to detect. Further research using another control target manipulation without changing intensity could provide further insight into why these changes would be seen under serotonergic influences.

5.6 SUPPLEMENTAL RESULTS

Table 30: results of behavioural VAS

scale	Placebo	Citalopram	*t*-stat (df=21)	*p*
nausea	3.50 (5.36)	7.85 (9.43)	-1.79	.089
headache	12.8 (16.3)	11.2 (14.7)	0.627	.54
dizziness	9.71 (10.5)	12.1 (12.8)	-1.01	.33
alert – drowsy	50.3 (15.6)	54.6 (17.2)	-0.916	.37
stimulated – sedated	49.6 (15.9)	51.6 (14.9)	-0.456	.65
restless – peaceful	60.6 (23.3)	60.6 (18.7)	0.003	1.0
irritable – good-humoured	64.7 (18.8)	64.7 (15.8)	-0.012	.99
anxious – calm	71.3 (20.6)	66.3 (17.5)	1.53	.14

Table 31: order effects

Contrast	Hemis.	Region	BA	# voxels	Z score (max)	MNI coordinates (at max Z) x	y	z
(Stomach > Target) × (Citalopram > Placebo) × Order	L	Superior Temporal Gyrus	41	206	4.06	-44	-32	10
	R	Auditory Cortex	41	141	4.14	40	-20	10

Figure 25: BOLD signal change in lateral orbitofrontal cortex cluster for parametric variable of target intensity from interaction (Heart > Target) × (Citalopram > Placebo). Error bars are within-subjects 95% CI

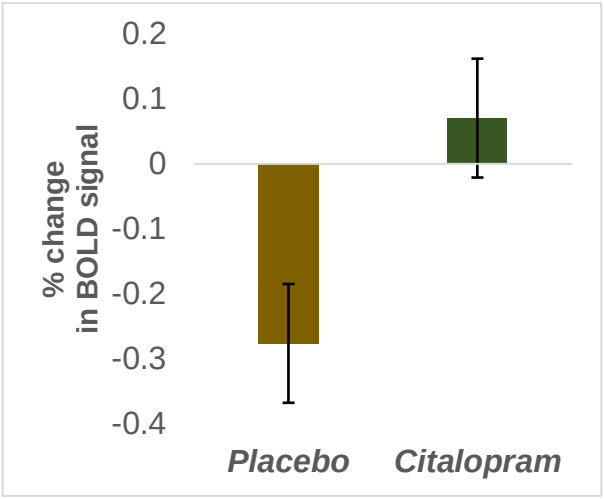
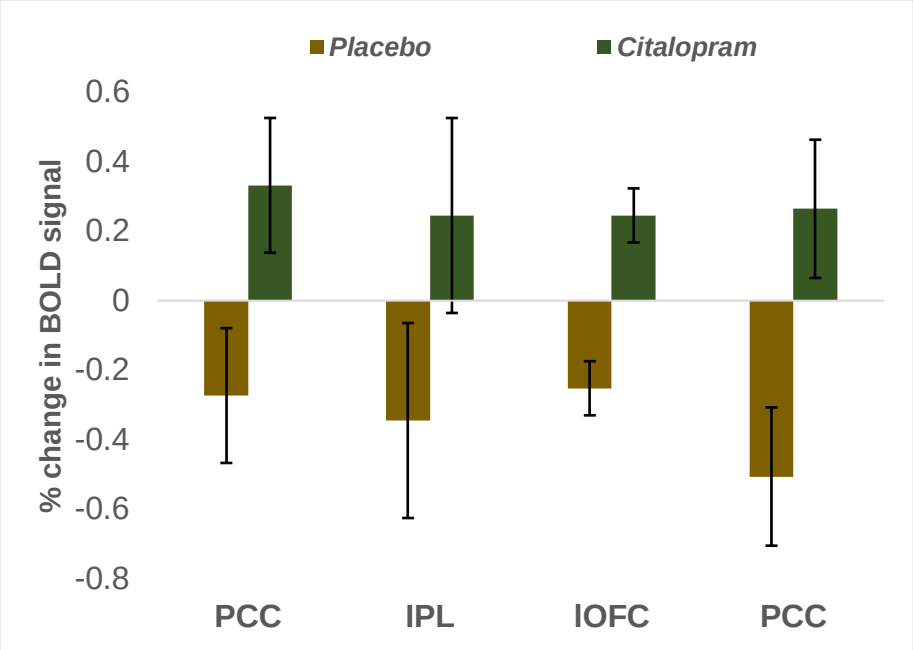


Figure 26: BOLD signal change in clusters for parametric variable of target intensity in interaction (Stomach > Target) × (Citalopram > Placebo). SFG – Superior Frontal Gyrus, lOFC – lateral orbitofrontal cortex, IPL – Inferior Parietal Lobule, PCC – Posterior Cingulate Cortex. Error bars are within-subjects 95% CI



6 EMOTIONAL SELF- AND OTHER-DIRECTED APPRAISAL: NEURAL SUBSTRATES AND EFFECTS OF CITALOPRAM CHALLENGE

6.1 ABSTRACT

Understanding one's own emotions and the emotions of others is important for healthy functioning, and deficits are seen in disorders that present with serotonin dysfunction. Previous research also links this functioning to interoceptive processes, or processes directed at internal physiological states. In previous chapters we demonstrated serotonergic effects on interoception and its neural correlates. Here we examine the effect of acute citalopram on the neural correlates of emotional appraisal directed at the self and others. We used a double-blind placebo crossover design with 22 healthy participants on a task that elicited focus on either the participant's emotional state, or that of a face with a neutral or fearful expression. Across both drug conditions, we showed differential activity when focussed on the self in several areas associated with default mode function including anterior cingulate/medial prefrontal cortex and temporoparietal junction, as well as other frontal and temporal regions including superior frontal gyrus. In the conjunction of self-focus and interoceptive focus, medial prefrontal cortex, insula, angular gyrus and cuneus showed significant activation, in line with previous research. We also showed a drug effect on connectivity between superior frontal gyrus and parahippocampal gyrus: connectivity was reduced in self-focus under citalopram but increased in other-focus. This demonstrates that serotonergic challenge can influence emotional appraisal, and points to another potential avenue of serotonergic effect on perception of the social environment.

6.2 INTRODUCTION

Forming an accurate understanding of the emotions of oneself and of others is closely linked to healthy functioning, and ties into interoceptive processes. Deficits in emotional self-recognition and empathic theory of mind are linked with depression and anxiety (Berthoz et al., 1999; Hezel & McNally, 2014; Honkalampi et al., 2000; Marchesi et al., 2000; Washburn et al., 2016; Weightman et al., 2014). Both are linked also to interoceptive processes. Alexithymia (lack of understanding of one's own emotional state) is associated with interoceptive deficit (Brewer et al., 2016; B. M. Herbert et al., 2011; Longarzo et al., 2015; J. Murphy et al., 2018), in line with putative connections between emotional experience and bodily states (Craig, 2009; Seth, 2013). In the brain, alexithymia manifests in attenuation of insula response to valenced stimuli (van der Velde et al., 2013), an area repeatedly linked with interoceptive processes (Schulz, 2016). Social cognition and interoceptive processes also share a set of neural substrates in orbitofrontal and inferior frontal gyri, amygdala, and mid temporal lobe (Adolfi et al., 2017), as well as insula (Uddin et al., 2017). With well-established links between the serotonin system and mood disorders (Akimova et al., 2009; M. J. Owens & Nemeroff, 1994; Schinka et al., 2004), and the links between the serotonin system and interoception established in earlier chapters, an understanding of the serotonergic contributions of affective appraisal of self and others and their interplay with interoception's neural substrates may offer insight into how these processes become dysfunctional.

Converging evidence shows common neural substrates of affective self-appraisal and appraisal of the emotions of others. In particular, a set of regions known as 'cortical midline structures' (CMS; Northoff & Bermpohl, 2004; Uddin, Iacoboni, Lange, & Keenan, 2007) have been implicated in both processes. These areas, including medial prefrontal cortex (MPFC), anterior cingulate cortex (ACC) and precuneus, constitute parts of the brain's default-mode network (DMN) which are also active during periods of rest (Greicius

et al., 2003; Gusnard & Raichle, 2001). Amodio & Frith's (2006) extensive review of medial frontal areas examined evidence that anterior rostral regions of the MPFC have roles in both affective self-monitoring (Gusnard et al., 2001; Lane et al., 1997) and mentalising (representing another person's perspective; Frith & Frith, 2003). Various authors (e.g. Mahy, Moses, & Pfeifer, 2014; Mars et al., 2012) note that this overlap is consistent with a view of theory of mind (TOM) as a simulation of another's behaviour by imagining one's own response in their situation, i.e. simulation theory (Gallese & Goldman, 1998). Areas such as the right temporoparietal junction (Saxe & Kanwisher, 2003) have been shown to be involved in mentalising, perspective-taking and empathy, often act in concert with these medial frontal areas (Aichhorn et al., 2006; Spengler et al., 2009; Völlm et al., 2006).

Nonetheless, distinctive levels and areas of activity between self and other-directed tasks have been shown from comparisons of rest, self-directed and other-directed inferences (meta-analysis in Wicker, Ruby, Royet, & Fonlupt, 2003). Greater MPFC activity is shown in self- than other-directed tasks, with greatest activity at rest. Ochsner et al. (2004) showed that affective judgements of self and other both increase activity in MPFC, superior temporal gyrus (STG) and posterior cingulate/precuneus, but self-directed judgements show greater activity in subdivisions of the MPFC and in the middle temporal gyrus. Other-directed judgements in the same study were associated with greater activity in inferior frontal gyrus and cuneus.

For both emotional experience and social cognition, overlapping subdivisions of the insular cortex have been shown as important correlates (reviewed in Uddin, Nomi, Hebert-Seropian, Ghaziri, & Boucher, 2017). Its central role in interoceptive processes give it a key place in embodied theories of emotion (Seth, 2013), where visceral responses to external stimuli are modelled as affective states, with the anterior portion in particular singled out as essential for emotional inferences on somatic states (Craig, 2009; Garfinkel & Critchley, 2013). Within social processes, empathy in particular has been linked with the insula. A meta-analysis by Fan, Duncan, de Greck, & Northoff (2011) showed common neural

substrates across affective states of pain, fear, happiness, disgust, and anxiety, with left insula activity recruited in both cognitive empathy (explicit appraisal of another's feelings) and affective empathy (automatic response to another's feelings), right insula only in affective empathy, and left anterior midcingulate cortex activity only in cognitive empathy.

Affective disorders present with biases in both self- and other-related processing, and corresponding effects in regions identified as involved in those processes. Depression is associated with abnormally high activity in cortical midline structures during self-referential processing (reviewed in Nejad, Fossati, & Lemogne, 2013), where rumination (repetitive thinking about negative mood states and their causes and consequences) is posited to be a mediating factor. In social cognition, bias toward mood-congruent interpretation of others' emotions have been observed, as well as deficits in cognitive theory of mind (Weightman et al., 2014), possibly mediated by executive function (Knight & Baune, 2019). In the brain, enhanced activity in emotion-related areas such as amygdala and reduction of activity in frontal areas has been observed during social cognition (Cusi et al., 2012).

The well-established serotonin hypothesis provides an important theoretical background in depressive disorders (M. J. Owens & Nemeroff, 1994); consequently considerable psychopharmacological research has looked at links between serotonergic challenge and emotional processing in a social context. Research using acute selective serotonin reuptake inhibitor (SSRI) has shown effects in related areas, though a clear pattern is yet to emerge, with some studies showing increased bilateral amygdala activity under citalopram challenge to presentations of negative facial expressions (Bigos et al., 2008; Selvaraj et al., 2018), and some showing attenuated activity in right amygdala (Del-Ben et al., 2005; S. E. Murphy, Norbury, et al., 2009). In self- and other-directed trait appraisal based on whether valenced adjectives applied to the participant or their best friend, Matthews et al. (2010) found that chronic escitalopram treatment resulted in attenuated posterior cingulate activity for self-appraisal. However, no previous research

has looked at self- and other-directed affective state appraisal under acute challenge, nor compared this with serotonergic changes during interoceptive focus. With connections established between emotional state and interoceptive processes, an understanding of the serotonergic relationship with appraisal of own and others' emotions may provide insight into the aetiology and characteristic patterns of related disorders. Patterns of interoceptive (Furman et al., 2013; Pollatos et al., 2009) and social cognitive (Knight & Baune, 2019) deficits in depression, along with high comorbidity of alexithymia (Honkalampi et al., 2000) are suggestive of a connection.

This study was designed to test this connection with pharmacological challenge using the SSRI citalopram. We tested healthy volunteers in a double-blind placebo crossover factorial design, with a task that involved focussing on own and others' mental states while viewing faces depicting neutral and fearful expressions. We then used conjunction analysis to compare patterns of activity between this task and an interoception task (a full discussion of the latter task is given in the previous chapter).

We hypothesised that CMS and parietal structures previously associated with self-reflection and mentalising, such as MPFC, ACC, precuneus and TPJ would be differentially recruited for self vs other processing, to confirm earlier reports. We also hypothesized that that self-related areas in particular would show overlapping substrates with interoception (in areas shown as differentially recruited in interoceptive over exteroceptive focussing). Finally, we hypothesised that citalopram would increase activity in the insula and CMS areas during self-focus, and in areas such as the amygdala during presentation of fearful compared with neutral facial expressions, in line with previous research (M. Browning et al., 2007; C. J. Harmer, Bhagwagar, et al., 2003). We hypothesised that connectivity patterns would be enhanced under citalopram, although as the drug could affect both self- and other-related processes we remained agnostic about which would show greater increases. Having shown citalopram-related changes in interoceptive processing, with areas of increased activity and greater extent of significant voxels during interoceptive focus under citalopram,

we also hypothesised that there would be overlapping areas of activity increase between these areas and clusters showing drug-modulated increases during self-focus.

6.3 METHODS AND MATERIALS

6.3.1 Participants

Ethical permission was granted by University of Sussex Sciences & Technology C-REC (ER/JL332/9). Potential participants were screened with a health questionnaire (see appendix) and the Mini-International Neuropsychiatric Interview (Sheehan et al., 1998). Exclusion criteria included: age under 18 or over 35 years; history of psychiatric disorder (including anxiety disorder, depression, eating disorder, psychosis and substance abuse disorder); presence of significant ongoing medical condition (including migraine, diabetes, epilepsy, glaucoma and hypertension); pregnancy or breastfeeding; currently taking any medication (excluding contraceptive pill); first-degree family history of bipolar disorder; Mini International Neuropsychiatric Interview (MINI) indication of: major depressive episode, manic episode, panic disorder, social phobia, OCD, PTSD, alcohol dependence, substance dependence, mood disorder with psychotic features, psychotic disorder, anorexia nervosa, bulimia nervosa, generalised anxiety disorder, or antisocial personality disorder; or scanner contraindications (e.g. metallic implants). They were also instructed to abstain from alcohol or caffeine in the preceding 12 hours before the start of test sessions.

24 healthy participants aged 18-35 were recruited for this study. Neuroimaging data from two participants were excluded due to excessive motion (>6% of volumes identified as motion outliers for either session scan; see section 6.3.4.3 for details), leaving 22 participants (7 males, age $M = 24.0, SD = 3.29$).

6.3.2 Study procedure

We used a repeated-measures design with separate drug and placebo sessions. held at least 7 days apart (days between sessions $M = 10.3$, $SD = 6.87$). Assignment to treatment order was double-blind and counterbalanced, with the drug treatment administered in one session and the placebo in another. Doses in the drug treatment conditions consisted of 20mg citalopram. Drug and placebo doses were delivered in gelatine capsules, indistinguishable from one another, with the capsule filled with microcrystalline cellulose (in addition to the active drug in the drug conditions). Drug and placebo doses were all manufactured according to good manufacturing practice (GMP) guidelines.

Participants were also given eight visual analogue scales at four timepoints – immediately following the dose, preceding the start of the scan, following the end of the scan, and at the end of the test session. Scales (from 0-100) were given to assess three somatic effects (nausea, headache and dizziness) and five emotion/arousal related effects (pairs of antonyms: alert–drowsy, stimulated–sedated, restless–peaceful, irritable–good-humoured, anxious–calm) to measure whether the drug was affecting these measures.

Following the dose but prior to testing, participants were given instructions and practice on the task, and were shown every stimulus picture that would be used in the scanner in that session. To allow for drug levels to reach peak absorption (Milne & Goa, 1991; Sauer et al., 2005), participants commenced behavioural testing (not listed in this chapter) after 3 hours after drug/placebo administration, with neuroimaging starting approximately 45 minutes later. The neuroimaging tasks included the emotional appraisal task described in the following section, and an interoception task, the full results of which are given in the previous chapter.

6.3.3 Task design

Figure 27: order of stimulus presentation

Prior to the experimental trials, participants were given the following instructions: *'When you see the word 'SELF' focus your attention on how you feel. Maintain this focus as long as the picture is shown. When you see the word 'OTHER', focus your attention on how the person in the picture feels. Maintain this focus as long as the picture is shown. When a CROSS is shown on the screen, you can rest. Please keep your eyes open and try not to think of anything in particular. You will be asked to rate how either you or the person in the picture feels using the buttons shown. Please make sure that you breathe evenly throughout the task. You must NOT hold your breath at any time.'*

Figure 18 shows the order of stimulus presentation. Stimuli were taken from the Karolinska Directed Emotional Faces database (Calvo & Lundqvist, 2008), and the task was programmed in Matlab (version 2018a, Mathworks, Natick, MA, USA) using Psychtoolbox-3 (Kleiner et al., 2007). 5 male and 5 female faces were used in each session, with two pictures of each individual displaying a fearful and a neutral facial expression, and different individuals used for each session. Each picture was displayed twice at different times in the session for self-focus (*Self*) and other-focus (*Other*) conditions, for a total of 40 trials per session. Order of presentation was pseudorandomly varied, with a fixation cross displayed for a variable time (3-12s) between trials. After each trial, participants were asked to rate

how they felt (for *Self*) or how the person in the picture felt (*Other*) from negative to positive (valence rating), and then the intensity of the feeling from low to high intensity (intensity rating), both on a scale of 1-4.

6.3.4 Neuroimaging

6.3.4.1 Neuroimaging procedure

Testing took place at the Clinical Imaging Sciences Centre (CISC), Brighton and Sussex Medical School. MRI data were acquired using a Siemens Magnetom Prisma 3T scanner (Siemens Healthcare GmbH, Erlangen, Germany) and a 32 channel head coil. Stimuli were displayed using NNL's (NordicNeuroLab, Bergen, Norway) MRI-safe 40-inch monitor (1920×1080 resolution) at the head of the bore and a mirrored headset. Responses for participants on behavioural measures were recorded using two NNL MRI-safe 2 button response boxes affixed together, operated with their right hand.

6.3.4.2 Image acquisition

Functional images were acquired using T2*-weighted echo-planar imaging (multiband factor 4, echo time 37ms, repetition time 1500ms, voxel size 2.2 × 2.2 × 2mm, 104 × 104 voxels per slice, 72 slices, FOV 205 × 205mm^2, flip angle 52°), with a varying number of volumes according to the speed of task completion (M = 580, SD = 36.8). Prior to the task fMRI sequence, pairs of phase-encode reversed images were acquired for distortion correction (FOV 228 × 228mm^2 104 × 104 voxels per slice, echo spacing 0.54ms). On one session, a structural T1-weighted magnetization-prepared rapid acquisition gradient echo (MPRAGE) image was acquired (echo time 2.2ms, repetition time 2400ms, 0.8mm isotropic voxels, 300 × 320 voxels per slice, 208 slices, FOV 256 × 256mm^2, flip angle 8°).

To test for changes in resting perfusion which may accompany the pharmacological challenge and confound task-related changes in BOLD signal (Detre et al., 2012; D. J. J. Wang et al., 2011), pulsed arterial spin labelling (ASL) images were acquired using a FAIR-QII

sequence (4 label-control image pairs, echo time 16ms, repetition time 4600ms, voxel size $1.5 \times 1.5 \times 3$mm, 126×128 voxels per slice, 40 slices, interleaved slices, FOV $192 \times 192\text{mm}^2$, flip angle 180°, inversion time 1990ms, bolus duration 700ms).

6.3.4.3 Pre-processing

Functional MRI images were pre-processed with a standard FSL pipeline including motion correction using the middle volume of the timeseries as reference volume, identification of motion outlying volumes using FSL Motion Outliers (root mean squared intensity difference of adjacent volumes method), distortion correction using phase-encode reversed image pairs, high-pass filtering at 80 seconds, brain extraction and co-registration to subject structural images. Structural images were co-registered to standard space (Montreal Neurological Institute (MNI) 152 stereotactic template). Independent Component Analysis (ICA) decomposition was carried out using FSL-MELODIC (Beckmann & Smith, 2004) on unsmoothed data, and denoising was carried out manually, with noise components identified and removed using criteria laid out in Griffanti et al. (2017). Following spatial smoothing with 5mm full-width at half maximum (FWHM) Gaussian kernel, a second denoising step was carried out to remove residual noise. ASL images were processed using Bayesian Inference for Arterial Spin Labelling (BASIL) to produce perfusion-weighted images.

6.3.4.4 Analyses

Functional images were all analysed using FSL FEAT v6.0.0 (Woolrich et al., 2001, 2004). We employed a three-level generalised linear model (GLM), with the first level as individual sessions using events corresponding to trial condition (self/other, negative/neutral) along with regressors of no interest for fixation cross, button presses and outlying motion volumes. Second level models were within-subjects contrasts between drug and placebo sessions for participants, and third level was the group-wise analysis across participant session contrasts using FMRIB's Local Analysis of Mixed Effects (FLAME)

1 & 2. First-level contrasts were employed on *Self-Other* focus conditions and *Negative-Neutral* stimuli valence, as well as their interaction. Z (Gaussianised T/F) statistic images were thresholded non-parametrically using clusters determined by $Z > 3$, chosen to maintain Type-I error rates below the .05 level (Eklund et al., 2016), and a (family-wise error rate corrected) cluster significance threshold of $p < .05$. Group level analyses contrasted drug effects, employing covariates for order and gender.

Following this, we performed a conjunction analysis on the *Self > Other* contrasts from citalopram and placebo conditions (Nichols et al., 2005) to determine common brain areas of activity across drug conditions. Cerebral clusters identified were then used as seed regions in a set of psychophysiological interaction (PPI) analyses, to model functional connectivity of these areas (O'Reilly et al., 2012). Each of these analyses comprised an event for *Self > Other* contrast, the timeseries of activity extracted from the relevant cluster (one cluster per analysis), the interaction of contrast event and timeseries, plus the same regressors of no interest as the main GLM. Significant clusters on this interaction would mean that the time course of activity in these voxels is correlated with the seed region, indicating functional connectivity.

Additionally, we performed conjunction analysis of the *Self vs Other* contrasts with the *Heart vs Target* contrast from the interoception task described in the previous chapter, using the mean effects across both drug conditions. We tested all combinations of these contrasts (*Heart > Target ∩ Self > Other, Heart > Target ∩ Other > Self, Target > Heart ∩ Self > Other* and *Target > Heart ∩ Other > Self*) to determine whether clusters identified in conjunctions were specific to the hypothesised effect of overlapping substrates of interoceptive and emotional self-focus.

Perfusion-weighted images were analysed at the group level in a paired t-test of drug and placebo sessions using Statistical Parametric Mapping (SPM12; https://www.fil.ion.ucl.ac.uk/spm/software/spm12/), with cluster-wise family-wise error

rate of $p < .05$, and with an additional exploratory threshold of $p < .001$ with extent > 50 voxels per cluster.

Behavioural analyses using ANOVAs were carried out on the valence and intensity ratings to test the effects of the stimulus and drug manipulations, and drug-placebo differences on VAS scales were analysed using paired t-tests.

6.4 RESULTS

6.4.1 Behavioural effects

The results from the $2 \times 2 \times 2$ ANOVAs carried out on the subjective ratings of valence and intensity against stimulus valence, self/other focus and drug condition are shown in Table 32. The ANOVAs showed that the subjective ratings of both valence and intensity were significantly affected by the valence of the picture displayed and by whether participants were instructed to focus on themselves or the face in the stimulus. Valence ratings were lower (more negative) for negatively valenced stimuli and lower for other than self. Intensity ratings showed the opposite pattern, reflecting a relatively low-intensity and valence-neutral state of participants (see Table 33). In addition, there were significant interactions of stimulus valence and order on rated valence, stimulus valence and self/other focus on intensity and valence, and self/other focus and drug condition on intensity. No post-hoc comparisons (performed with Holm corrections for multiple comparisons) showed any drug effect. There were no effects of drug on self-reported psychological or somatic variables (see Supplemental).

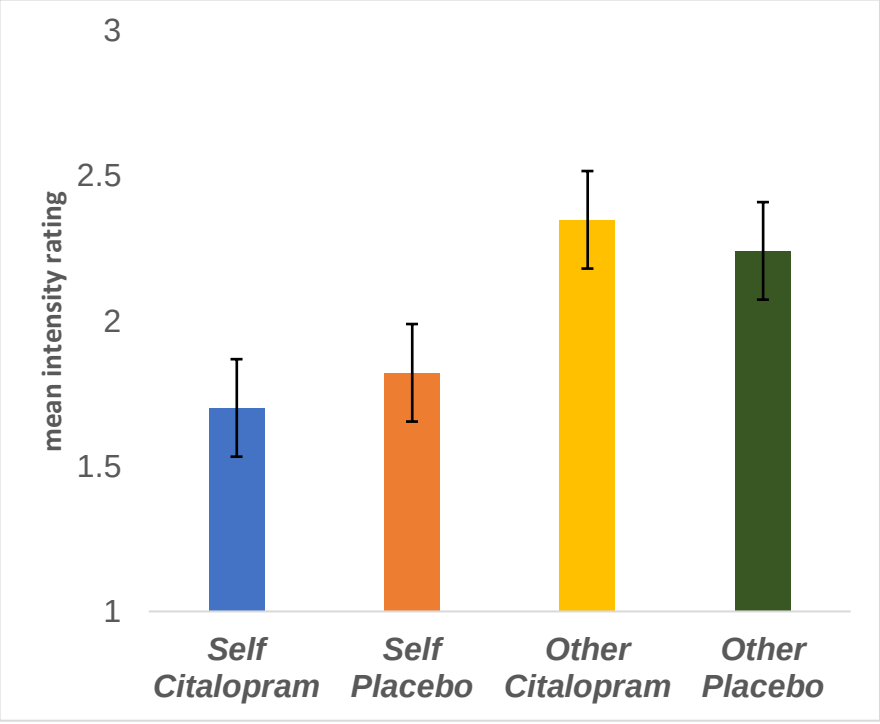
Table 32: ANOVAs of behavioural measures. stimV – valence of stimulus picture (negative or neutral), S vs O – self vs other condition, C vs P – citalopram vs placebo treatment, SS – sum of squares, MS – mean square

		INTENSITY MEASURE				**VALENCE MEASURE**			
predictor	**df**	**SS**	**MS**	***F***	***p***	**SS**	**MS**	***F***	***p***
stimV	1	35.5	35.5	149.4	**< .001**	16.0	16.0	163.4	**<.001**
*(stimV *order)*	1	0.27	0.268	1.13	.30	0.489	0.489	5.0	**.037**
resid.	20	4.75	0.237			1.958	0.098		
S vs O	1	12.14	12.1	34.7	**<.001**	2.90	2.90	21.5	**<.001**
*(S vs O *order)*	1	0.31	0.308	0.88	.36	0.004	0.004	0.027	.87
resid.	20	7.01	0.35			2.70	0.135		
C vs P	1	0.002	0.002	0.021	.89	0.017	0.017	0.199	.66
*(C vs P *order)*	1	0.77	0.765	6.437	**.020**	0.012	0.012	0.131	.72
resid.	20	2.38	0.119			1.76	0.088		
(S vs O *stimV)	1	13.2	13.15	95.7	**<.001**	2.59	2.59	36.5	**<.001**
*(S vs O * stimV *order)*	1	0.10	0.101	0.734	.40	0.065	0.065	0.914	.35
resid.	20	2.75	0.137			1.415	0.071		
(stimV *C vs P)	1	0.07	0.065	1.07	.31	0.004	0.004	0.08	.78
*(stimV *C vs P *order)*	1	0.0003	0.0003	0.005	.95	0.027	0.027	0.612	.44
resid.	20	1.22	0.061			0.881	0.044		
(S vs O *C vs P)	1	0.55	0.553	8.47	**.009**	0.07	0.07	1.21	.29
*(S vs O *C vs P *order)*	1	0.14	0.14	2.14	.16	0.008	0.008	0.14	.71
resid.	20	1.31	0.065			1.159	0.058		

Table 33: marginal means for Self vs Other within-subjects interactions on intensity measure

					95% CI	
Interaction	**Factor 1**	**Factor 2**	**Marginal Mean**	**SE**	**Lower**	**Upper**
(S vs O * stimV)	*Self*	*Negative*	1.94	0.090	1.76	2.12
		Neutral	1.58	0.090	1.40	1.76
	Other	*Negative*	3.03	0.090	2.85	3.21
		Neutral	1.56	0.090	1.38	1.74
(S vs O * C vs P)	*Self*	*Citalopram*	1.70	0.084	1.53	1.87
		Placebo	1.82	0.084	1.65	1.99
	Other	*Citalopram*	2.35	0.084	2.18	2.52
		Placebo	2.24	0.084	2.07	2.41

Figure 28: marginal means of (Self vs Other) × (Citalopram vs Placebo) intensity ratings



6.4.2 ASL

The analysis of ASL image pairs showed no significant clusters at the familywise error rate, suggesting that citalopram compared to placebo did not affect cerebral blood flow in general, and that any effects on BOLD responses were not mediated by general

effects on blood flow. For the uncorrected threshold, a single significant cluster was shown in the occipital cortex (peak voxel MNI coordinate: (18, -96, 14), $p_{uncorrected} = .001$, $p_{FWE} = .10$).

6.4.3 Task fMRI results: main effects and interaction

Table 34: significant clusters in contrasts of Self vs Other condition, mean across citalopram and placebo conditions. IPL – Inferior Parietal Lobule, TPJ – temporoparietal junction, SMA – supplementary motor area

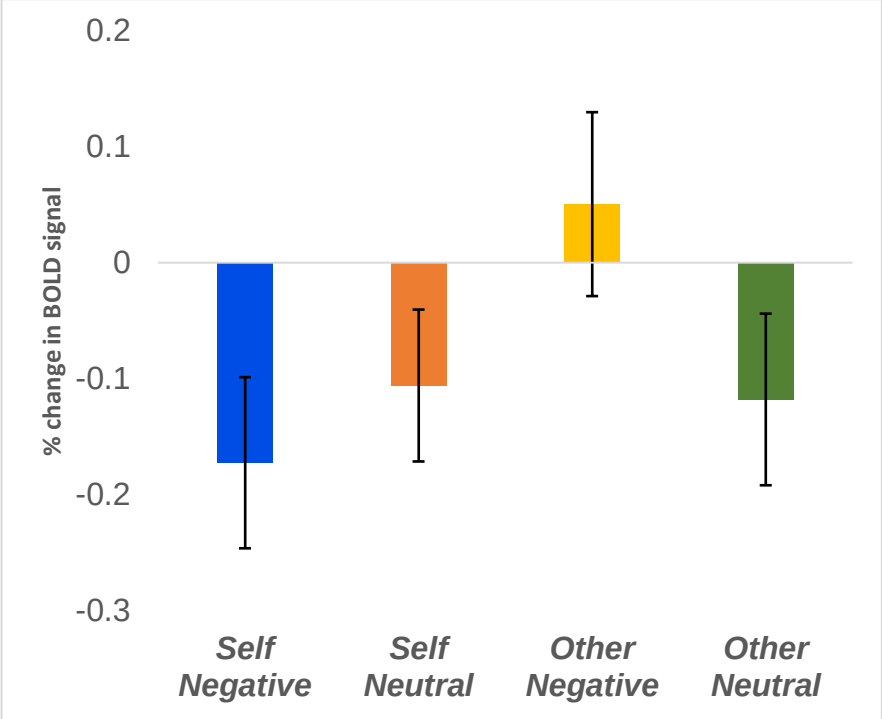
Contrast	Hemis.	Region	BA	# voxels	Z score (max)	MNI coordinates (at max Z)		
						x	y	z
Self > Other	L	Cingulate Gyrus	24	3864	6.65	0	-16	44
	L	Anterior Cingulate / MPFC	10/32	2084	5.92	-2	50	2
	L	IPL / TPJ	39/40	1610	6.97	-58	-46	34
	L	Superior Frontal Gyrus	9/10	1342	5.08	-26	50	28
	L	Cerebellum	-	1159	5.29	-48	-64	-32
	R	Cerebellum	-	768	4.95	40	-58	-42
	R	Insula	13	654	5.97	48	6	0
	R	Supramarginal Gyrus / TPJ	39/40	612	6.45	60	-46	28
	R	SMA	6	563	5.27	14	10	70
	R	Middle Frontal Gyrus	9	556	6.15	30	50	32
	L	Cerebellum	-	392	4.83	-12	-74	-12
	L	Insula	-	364	5.23	-36	8	2
	R	Middle Temporal Gyrus	21/38	263	4.52	58	6	-28
	R	Superior Temporal Gyrus	22	229	5.02	54	-22	-6
	L	Middle Frontal Gyrus	6	220	5.1	-44	12	50
	R	Cerebellum	-	194	4.36	26	-78	-34
Other > Self	L	Caudate	48	124	3.92	-14	26	6

Table 34 shows significant clusters in the contrasts of *Self* vs *Other* conditions. Table 35 shows significant clusters in the contrasts of *Negative vs Neutral* conditions (where only *Negative > Neutral* showed significant clusters) and the interaction between the two sets of contrasts. There were no significantly different clusters in this analysis for the interaction of drug and placebo with either contrast. A single cluster in superior parietal lobule showed the interaction *(Self > Other)* × *(Negative > Neutral)* (Figure 29).

Table 35: significant clusters in contrast of Negative > Neutral condition, mean across citalopram and placebo conditions (no significant clusters for Neutral > Negative)

Contrast	Hemis.	Region	BA	# voxels	Z score (max)	MNI coordinates (at max Z)		
						x	y	z
Negative > Neutral	L	Middle Temporal Gyrus	39	13933	6.33	-60	-62	14
	R	Middle Temporal Gyrus	39	1438	5.91	62	-56	14
	R	Inferior Frontal Gyrus	45	1132	6.29	40	30	0
	R	Fusiform Gyrus	19	215	4.61	24	-86	-14
	L	Cerebellum	-	204	4.13	-36	-60	-52
	R	Cerebellum	-	192	4.27	24	-74	-44
	R	Cerebellum	-	186	4.27	22	-56	-16
	L	Temporal Lobe Sub-Gyral	21	165	4.52	-38	-10	-12
	L	Precuneus	7	159	4.24	2	-50	52
	R	Cuneus	17	149	5.27	12	-82	12
	R	Medial Frontal Gyrus	9	146	4.33	24	36	26
(Self > Other) × *(Negative > Neutral)*	R	Superior Parietal Lobule	19	121	4.06	40	-74	48

Figure 29: BOLD signal change in superior parietal lobule cluster for each parameter estimate in interaction



6.4.4 Connectivity analysis

To test hypotheses of differential effects of citalopram vs placebo on functional connectivity between regions during performance on the task, we first tried to identify suitable seed regions. To do this, we carried out a conjunction analysis between the *Self > Other* contrast under placebo and the same contrast under citalopram. Results are shown in Table 37 and Figure 33. PPI analyses were carried out on all seeds from the cerebrum, testing the mean across drug and placebo conditions for *Self > Other, Other > Self,* and the interaction with drug treatment.

Table 36: significant clusters in conjunction of citalopram and placebo treatments for contrasts of Self > Other condition, (no significant clusters were found for the conjunction of Other > Self). TPJ – Temporoparietal Junction

Contrast	Hemis.	Region	BA	# voxels	Z score (max)	MNI coordinates (at max Z)		
						x	y	z
Self >	L	Angular Gyrus / TPJ	40	499	5.05	-56	-54	40
Other	L	Frontal Pole	9	408	4.59	-26	50	30
	L	Precuneus	31	242	4.22	-12	-48	30
	R	Superior Frontal Gyrus	6	224	4.44	16	8	64
	R	Frontal Pole	9	207	4.15	32	52	28
	L	Cerebellum	-	175	3.8	-30	-60	-50
	L	Anterior Cingulate	32	131	4.08	0	34	-4

Table 37: results of PPI analysis for Self > Other contrast

Cluster	Hemis.	Region	BA	# voxels	*Z* score (max)	MNI coordinates (at max Z)		
						x	y	z
Precuneus	R	Inferior Frontal Gyrus	45	475	4.21	56	24	14
	L	Superior Frontal Gyrus	8	287	4.16	-2	22	52
	L	Cerebellum	-	207	4.16	-10	-78	-28
	L	Cerebellum	-	150	3.84	-36	-60	-30
	R	Cerebellum	-	144	4.17	38	-62	-30
L Frontal Pole	R	Superior Frontal Gyrus	-	112	4.15	-2	12	52
Angular Gyrus	L	Inferior Frontal Gyrus	45	700	4.11	-48	22	4
	L	Superior Frontal Gyrus	6	306	4.31	-2	12	52
	R	Middle Frontal Gyrus	9	124	4.05	56	22	30
	R	Cerebellum	-	113	4.06	38	-68	-24

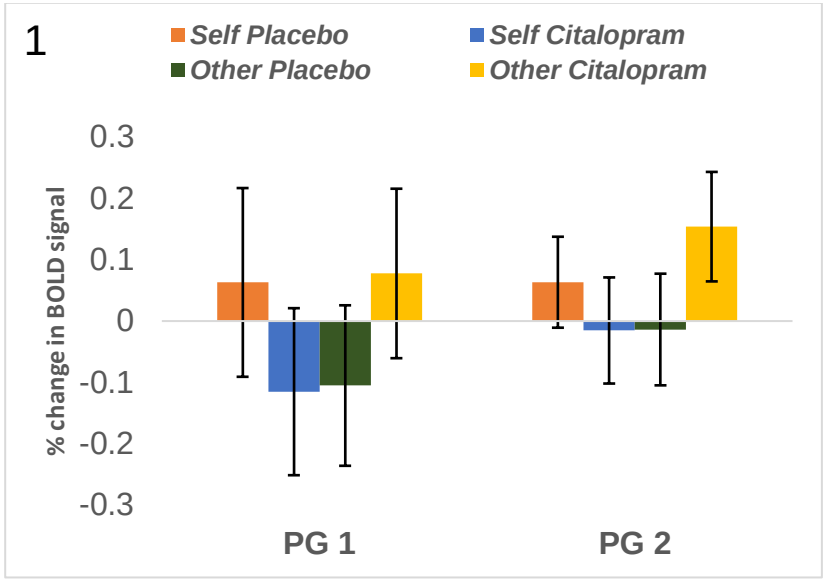
Table 38: results of PPI analysis for Other > Self contrast

						MNI coordinates (at max Z)		
Cluster	**Hemis.**	**Region**	**BA**	**# voxels**	***Z* score (max)**	**x**	**y**	**z**
Angular Gyrus	L	Fusiform Gyrus	37	508	4.95	-26	-50	-8
	L	Cuneus	18	496	4.09	0	-90	18
	L	Middle Temporal Gyrus	39	248	4.7	-40	-78	16
	R	Anterior Cingulate	32	139	3.87	4	40	-14
Superior Frontal Gyrus	R	Parahippocampal Gyrus	19	121	4.47	26	-56	-6
	R	Posterior Cingulate	31	97	4.14	4	-66	20
	R	Anterior Cingulate	32	679	4.17	12	40	16
	R	Posterior Cingulate	30	486	3.97	8	-56	8
	L	Fusiform Gyrus	37	450	4.59	-24	-52	-10
	R	Cingulate Gyrus	31	382	4.71	10	-38	38
	R	Middle Temporal Gyrus	20	280	4.24	64	-44	-12
	L	Fusiform Gyrus	37	255	4.32	-46	-64	-10
	R	Parahippocampal Gyrus	-	115	4.05	30	-32	-10
	L	Cuneus	18	101	3.96	-28	-72	22
	L	Cingulate Gyrus	24	91	3.93	0	-2	30
Anterior Cingulate	R	Supramarginal Gyrus	40	252	3.96	54	-50	20
	L	Precuneus	31	174	3.76	-6	-50	40
	L	Cingulate Gyrus	31	98	3.65	-6	-24	42
	R	Superior Temporal Gyrus	21	77	3.84	56	-24	-8
	R	Middle Temporal Gyrus	21	75	4.11	54	4	-22

Table 39: results of PPI analysis for interaction (Other > Self) × (Citalopram > Placebo)

Cluster	Hemis.	Region	BA	# voxels	*Z* score (max)	MNI coordinates (at max Z) x	y	z
Superior Frontal Gyrus	R	Parahippocampal Gyrus	36	164	3.94	30	-34	-14
	R	Parahippocampal Gyrus	18	147	4.12	12	-50	2

Figure 30: (1) connectivity change with superior frontal gyrus seed in each parahippocampal gyrus (PG) cluster for each parameter estimate in interaction, (2) relationship between interactions of subjective intensity rating and connectivity in PG 2 cluster, where scores are (Self CIT–PLAC) – (Other CIT–PLAC)



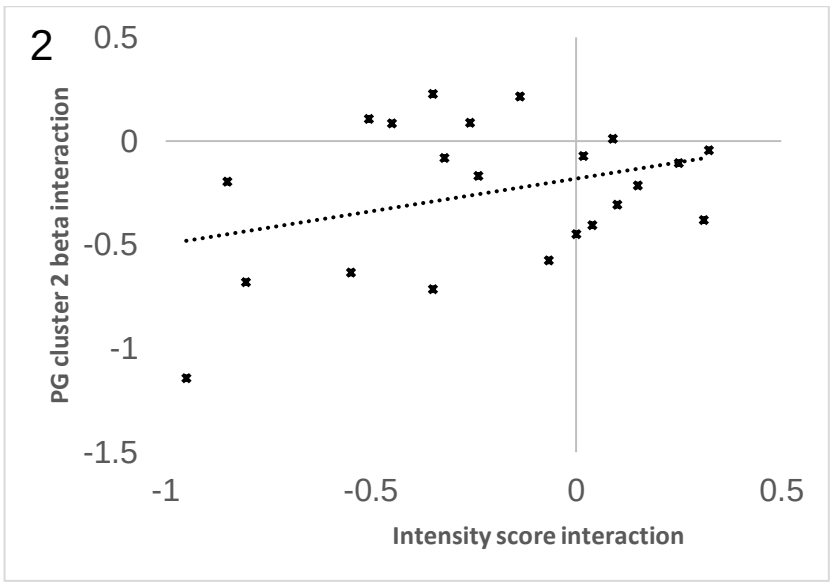


Figure 31: results of PPI analyses, with seed regions shown in green, mean across citalopram and placebo conditions. Self > Other contrast – red-yellow, Other > Self – dark blue-light blue. Number shows z coordinate of slice in MNI space, render shows location of slices in relation to the whole brain

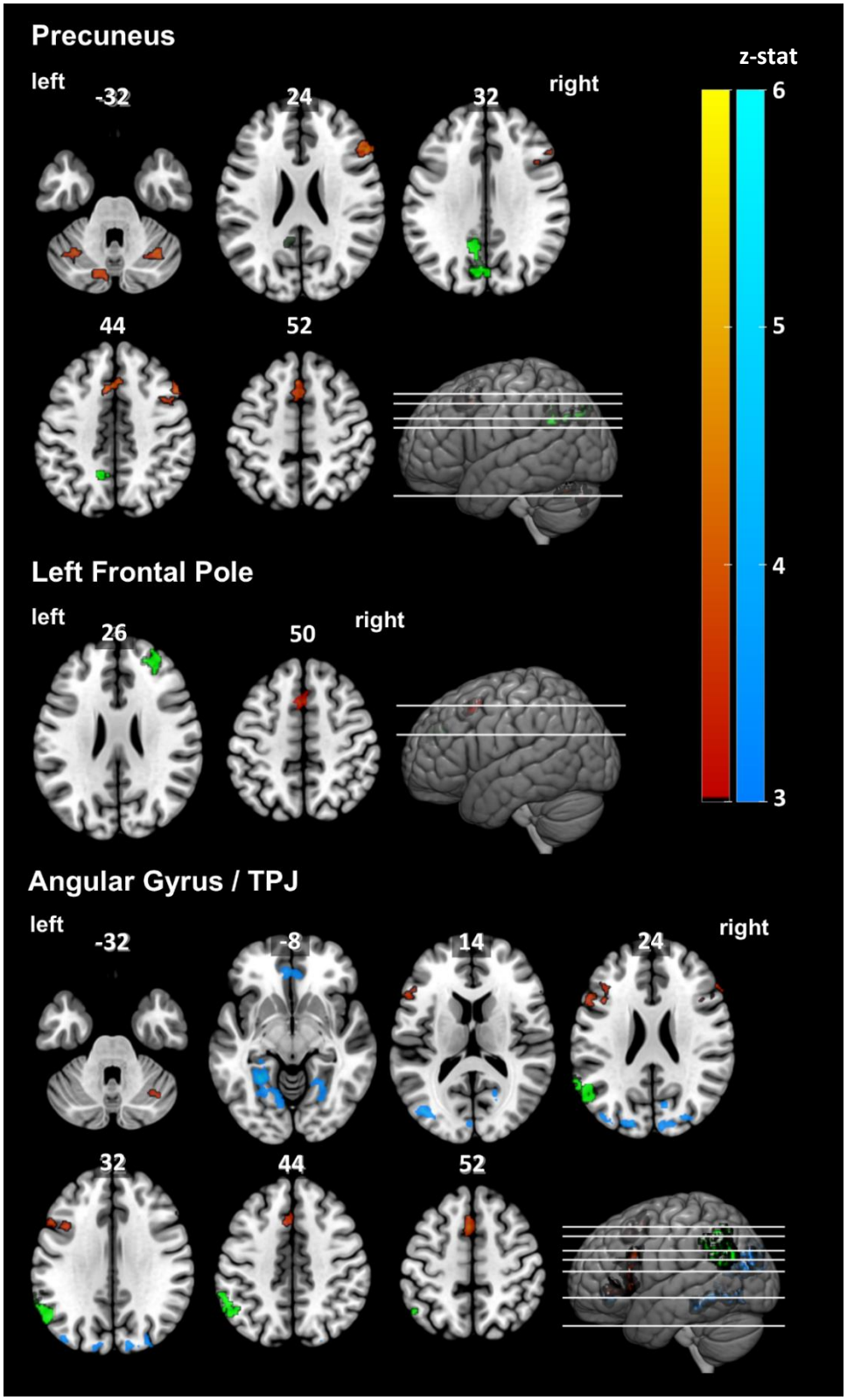
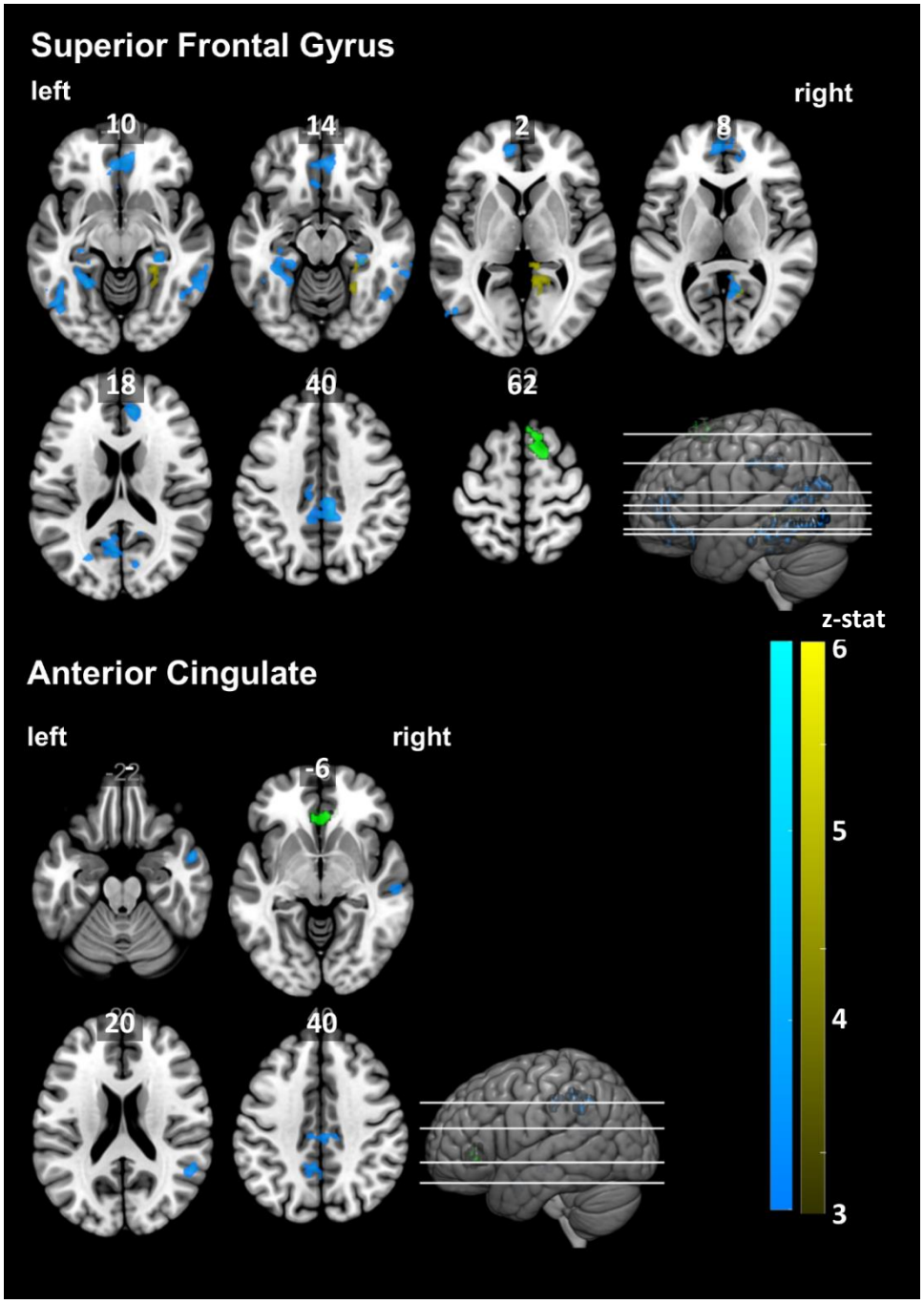


Figure 32: results of PPI analyses, with seed regions shown in green. Other > Self mean across citalopram and placebo conditions – dark blue-light blue, Other > Self × Citalopram > Placebo – dark gold-light gold. Number shows z coordinate of slice in MNI space, render shows location of slices in relation to the whole brain



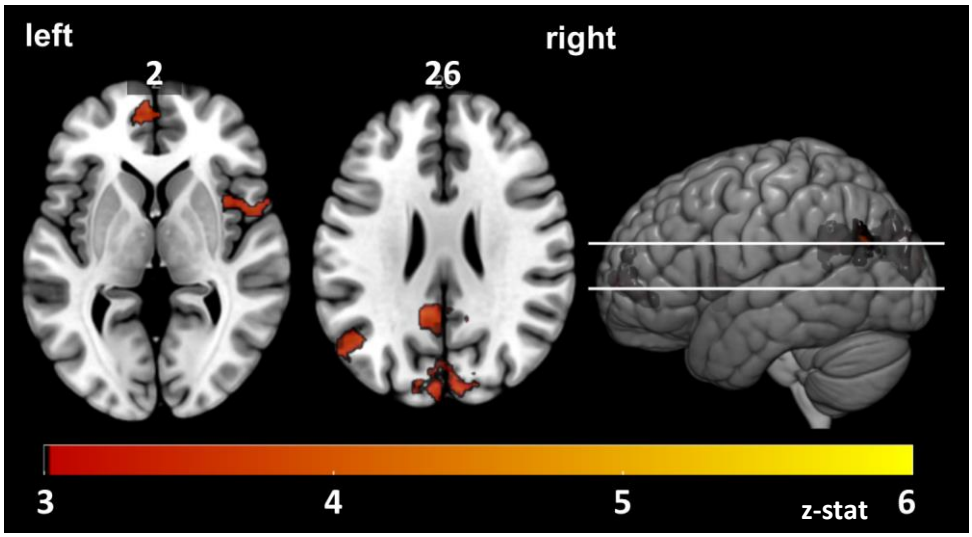
PPI analyses showed significant connectivity between seed regions in precuneus, angular gyrus / TPJ, superior frontal gyrus, anterior cingulate and left frontal pole (Table 37-39 and Figure 31-32). No significant clusters were seen in connectivity analyses of right frontal pole.

6.4.5 Conjunction of self- and interoceptive focus

Table 40: significant clusters in conjunction Self > Other ∩ Heart > Target conditions, mean of citalopram and placebo. MPFC – Medial Prefrontal Cortex

					MNI coordinates (at max Z)		
Hemis.	**Region**	**BA**	**# voxels**	***Z* score (max)**	**x**	**y**	**z**
R	Cuneus	19	751	4.63	16	-84	40
L	Posterior Cingulate	31	340	4.3	-12	-50	26
L	MPFC	9	246	4.43	0	56	18
R	Insula	13	243	4.26	44	10	-6
L	Angular Gyrus	39	225	4.33	-44	-74	28

Figure 33: results of conjunction analysis of Self > Other and Heart > Target contrast, mean across citalopram and placebo conditions. Number shows z coordinate of slice in MNI space, render shows location of slices in relation to the whole brain



Significant clusters were seen for the conjunction of *Self > Other* and *Heart > Target* conditions. There were no significant clusters for the three other combinations of these contrasts (*Target > Heart ∩ Other > Self* etc.) nor for any *Citalopram – Placebo* contrast conjunctions.

No effects of gender were seen in any analysis, but some order effects were observed (see supplemental for details).

6.5 DISCUSSION

Behavioural results and fMRI findings from the main GLM analysis

Behavioural results showed that the manipulation worked as intended, with participants perceiving the stimuli faces in the intended valence, but that their own emotion states were relatively neutral. The drug manipulation did not show main effects, but an order effect was evident in the effect of drug on perceived intensity, although post hoc paired tests were not significant. The most notable interaction was between drug effect and self vs other focus on the intensity rating. Ratings for intensity in self-focus were higher on placebo than drug, but those for other focus were lower, suggesting that participants' own emotional intensity was attenuated by citalopram but perceptions of others' intensity was increased.

The main effect of the *Self > Other* contrast showed several clusters identified in areas in common with previous research. Consistent with functional overlaps between bodily states and processing of emotional aspects of the self and others, we found clusters cingulate gyrus near to those found active in interoceptive aspects of selfhood (Helder F. Araujo et al., 2015), positively-valenced feedback directed at the self (Korn et al., 2012), emotional regulation in response to phobic images (Hermann et al., 2009), and with empathy-related theory of mind (Kanske et al., 2015). The findings of stronger activation during self-related processing is also consistent with previous findings of stronger activity for self- over other-directed processes in areas with shared neural substrates (Ochsner et al., 2004; Wicker et al., 2003). Similarly, strong links with self- and other-directed appraisals have been shown in inferior parietal lobule / temporoparietal junction (reviewed in Kestemont et al., 2015), anterior cingulate (meta-analysis: Helder Filipe Araujo, Kaplan, & Damasio, 2013; review: Ochsner et al., 2005) and superior temporal gyrus (Modinos et al., 2009). The latter two areas showed the same pattern of stronger activity in self vs other processing. The cerebellum was activated significantly more in *Self* than *Other* condition, as

well as showing increased connectivity with precuneus and angular gyrus. The largest cluster overlapped with an area of the cerebellum associated with emotional processing in a previous meta-analysis (Stoodley & Schmahmann, 2009; further review in Adamaszek et al., 2017). A single cluster in the caudate was more active in other than self-processing, previously linked to learning from experience of both self- and other (Canessa et al., 2011).

Areas most strongly activated in the contrast of negative vs neutral faces included bilateral middle temporal gyrus, linked to theory of mind processes (Kestemont et al., 2015; Schlaffke et al., 2015); fusiform gyrus, linked with an array of facial recognition processes (Carré et al., 2014; Katanoda et al., 2000) but also emotional self-appraisal (Terasawa, Fukushima, et al., 2013); inferior frontal gyrus, linked with recognition of emotion in faces (Loughead et al., 2008) and in a location bordering the anterior insula. There were no significant clusters in the inverse contrast. These differences may be interpreted in line with negative faces providing a more salient emotional stimulus than those with a neutral expression, leading to greater activity of areas involved in emotion recognition.

There was also an interaction with *Self vs Other* in a single cluster in the superior parietal lobule, near an area previously associated with recognition of faces (Leveroni et al., 2000). Further investigation showed increases only during *Other Negative* trials. Potentially, this combination of focussing on the other and a negative facial expression made these facial stimuli particularly salient.

Findings from the conjunction and connectivity analyses

The conjunction analysis between citalopram and placebo conditions for the *Self > Other* contrast identified several clusters activated across both conditions. Among them were nodes of the DMN: precuneus, bilateral frontal pole, angular gyrus and anterior cingulate (Greicius et al., 2003; Gusnard et al., 2001; Seghier, 2013; Utevsky et al., 2014). Additionally. a cluster in superior frontal gyrus was identified, previously seen involved in emotional regulation (Silvers, Wager, et al., 2015)

While there were no main effects of the drug manipulation, we did show effects in the connectivity analysis. There was an interaction between drug treatment and *Self vs Other* contrasts, with differential connectivity between the seed region in superior frontal gyrus (SFG) and two clusters in parahippocampal gyrus. This area has been associated with attribution of social motives and inferences on others' decisions (Hartwright et al., 2015; Lombardo et al., 2009; Shane et al., 2009; Vanderwal et al., 2008). Placebo was associated with higher connectivity in *Self* compared with citalopram, and the reverse effect seen for *Other*. Tellingly, this mirrors the behavioural interaction on self-reported intensity, raising the possibility that connectivity in these areas is related to intensity of the emotional appraisal and that citalopram differentially affects this, increasing response to perceived emotional intensity of others. This interpretation is consistent with previous findings that acute citalopram can increase aversion to harming others (Crockett et al., 2010) and perception of others' emotional expressions (M. Browning et al., 2007). The ASL finding of no significant changes in perfusion (with the sole trending change being in a different area) suggests that true changes in task-related activity rather than general drug-related perfusion occurred in this area. Additionally, the seed region in superior frontal gyrus showed increased connectivity in the *Other > Self* contrast in the mean across citalopram and placebo sessions, with midline areas including parahippocampal gyrus, anterior and posterior cingulate, and the fusiform gyrus.

The role of the precuneus in this study revealed interesting features, in line with its functions in a variety of self-related processes. With the seed region in ventral precuneus identified in the conjunction of drug and placebo sessions, several areas showed increased functional connectivity under *Self > Other* but not the inverse contrast. Increased connectivity was shown with a more dorsomedial subarea of the SFG (compared with the SFG seed region). This shows parallels with patterns shown in resting state research (S. Zhang & Li, 2012). The SFG has also previously been linked to emotional self- and self-/other-appraisal (Silvers, Weber, et al., 2015; van der Heiden et al., 2013), and Terasawa,

Fukushima, & Umeda (2013) also showed a similarly increased coupling of precuneus and medial frontal areas during evaluation of current emotional state. They identified this coupling as relevant to the integration of interoceptive and self-related information in emotional self-appraisal. Correspondingly, the precuneus seed region was ~4.5mm (between peak voxels) from the cluster identified in the posterior cingulate in the conjunction analysis of interoceptive and self-directed emotional focus (discussed in the next subsection). The inferior frontal gyrus was the other cerebral area to show increased precuneus connectivity, previously linked to self-related inferences (Doerig et al., 2014; Morel et al., 2014; Wilson-Mendenhall et al., 2013). Interestingly, in the *Other > Self* contrast we found increased connectivity between the anterior cingulate seed and a more dorsal region of the precuneus. This parallels a finding by Atique, Erb, Gharabaghi, Grodd, & Anders (2011) in an emotion mentalising task showing similar increases in this connectivity, and speaks to the dual role of midline structures in mentalising and self-appraisal, with different subregions specialised for each.

The identified seed region in angular gyrus / TPJ showed commonalities with the precuneus for self- vs other-appraisal, with both seeds showing increased connectivity superior and inferior frontal gyri. The angular gyrus has also been shown to exhibit roles in default mode (Gusnard & Raichle, 2001; Mazoyer et al., 2001; Seghier, 2013; Yang et al., 2010) and self-related processing (Hughes & Beer, 2012; Lei Zhu et al., 2013), but also with TOM processing (Schurz et al., 2015; van der Meer et al., 2011). This seed was the only one found to have increases of connectivity in both *Self > Other* and *Other > Self*, which is consistent with research showing cross-modal connectivity across the region subtending various processes within the social environment (Seghier, 2013). Notable among the increased connectivity in the *Other > Self* contrast was with fusiform gyrus, which has been linked previously to perception of others' emotions (Jastorff et al., 2015; Meffert et al., 2015).

Shared areas of self-appraisal and interoceptive focus

The conjunction of *Self > Other* and *Heart > Target* contrasts showed areas involved in self-related processing, including DMN structures in posterior cingulate, medial prefrontal cortex and angular gyrus. The involvement of the insula has also been observed in previous studies on self-appraisal (Qin & Northoff, 2011), in similar subareas. Conjunction analysis of interoceptive and self-directed tasks showed right mid-insula activity in both, consistent with shared substrates of interoceptive and emotional processing. The cuneus also has previously been shown to have increased connectivity during interoceptive focus (Kuehn et al., 2016) and higher activity in self-appraisal (Deming et al., 2018). This provides further evidence of common neural substrates of affective self-processing and interoceptive focus. No other tested combination of contrasts across the two studies reached significance, indicating that these results do not appear to be an artefact of general task demands.

Order effects

There were several effects of treatment order seen from the group-level order covariate (see supplemental for details), including on drug-placebo contrasts that did not show significance otherwise. Previous psychopharmacology research has highlighted the impact of order effects in repeated-measures designs, including in serotonergic studies (S. B. Park et al., 1994; Wood et al., 2006). The sample size of this study was too small to perform reliable between-group analyses on individual sessions, which is a limitation of the study design. The lack of drug effects on the main GLM or on PPI analyses with seeds other than the superior frontal gyrus should be interpreted with caution in light of this. We also did not replicate previous findings on amygdala activity in response to negatively valenced facial expressions under citalopram. However, all of these studies used decision tasks based on gender of face (Del-Ben et al., 2005; S. E. Murphy, Norbury, et al., 2009; Selvaraj et al., 2018) or matching the face to a target (Bigos et al., 2008). The emotion of the face was designed to be incidental, whereas our task sought emotional appraisal. Previous work showed that labelling the affective states of presented facial expressions diminished

amygdala responses to them (Lieberman et al., 2007), suggesting that a similar process may have occurred here. As our main hypotheses were around changes relating to emotional appraisal, this did not affect our main findings.

Limitations

We did not include a control condition with display of facial expressions in the absence of affective judgements. We showed overlap in areas differentially recruited by self-focus and those with greater connectivity in both self and other focus, but this limited our ability to disentangle self- and other-directed focus from facial stimulus processing without emotional appraisal. Future work could include this as a control condition, in order to differentiate serotonergic effects on affective appraisal with this control.

Conclusion

This study showed that self- and other-directed emotional appraisal caused differential activity in CMS structures and lateral areas previously identified with mentalising. Moreover, these showed several significant areas in common with interoceptive focus, lending further weight to theories suggesting a strong functional overlap in neural areas of interoceptive and emotional processing. We also showed a citalopram effect on connectivity between superior frontal gyrus and parahippocampal gyrus, reducing connectivity in self-focus under citalopram but increasing it in other-focus. This demonstrates that serotonergic challenge can influence emotional appraisal, and points to another potential avenue of serotonergic effect on perception of the social environment.

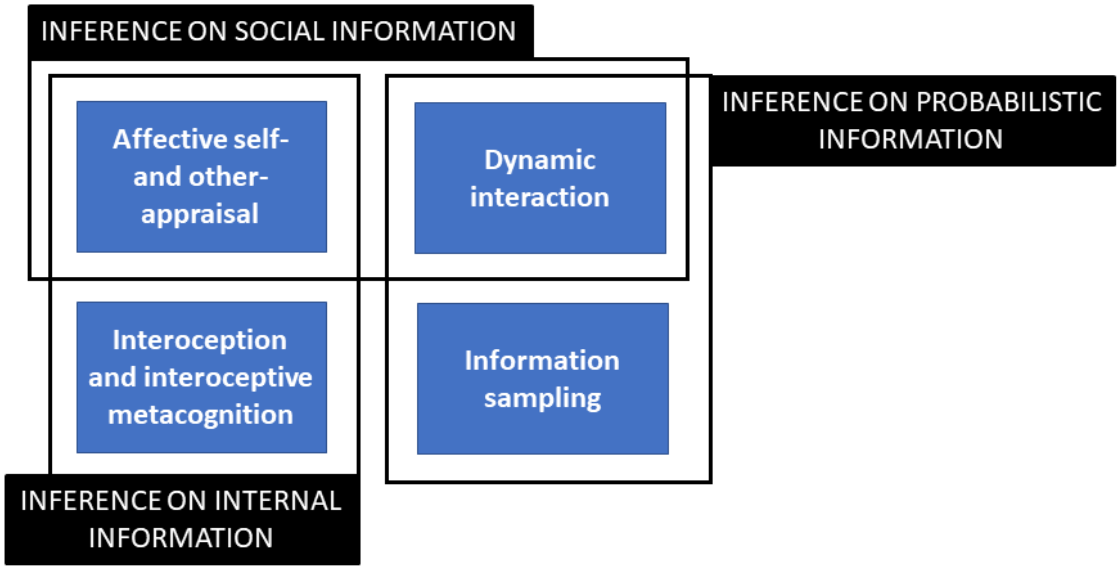
6.6 SUPPLEMENTAL RESULTS

Table 41: results of behavioural VAS

scale	Placebo	Citalopram	*t*-stat (df=21)	*p*
nausea	3.50 (5.36)	7.85 (9.43)	-1.79	.089
headache	12.8 (16.3)	11.2 (14.7)	0.627	.54
dizziness	9.71 (10.5)	12.1 (12.8)	-1.01	.33
alert – drowsy	50.3 (15.6)	54.6 (17.2)	-0.916	.37
stimulated – sedated	49.6 (15.9)	51.6 (14.9)	-0.456	.65
restless – peaceful	60.6 (23.3)	60.6 (18.7)	0.003	1.0
irritable – good-humoured	64.7 (18.8)	64.7 (15.8)	-0.012	.99
anxious – calm	71.3 (20.6)	66.3 (17.5)	1.53	.14

7 GENERAL DISCUSSION

7.1 SUMMARY OF FINDINGS – REVISITING THE MAIN AIMS



This thesis aimed to address four key gaps in the research literature concerning serotonergic effects on internal, social and probabilistic information. From the empirical work detailed in chapters 2-4, I will review our contributions to the understanding of each in turn.

1. *Previous research into decisions to sample probabilistic information under serotonergic challenge had not been able to quantify normative effects through modelling characteristic decision patterns under risk. This gave limited explanatory power to understand whether alterations in patterns of information sampling could also change the effectiveness of decisions in terms of value to the decision maker.*

Chapter 2 showed that acute citalopram administration led to a fall in both probabilities of correct decisions and of expected utility of these decisions given the participant's own risk preferences. By contrast, atomoxetine showed neither effect, despite self-reported side effects being not significantly different between the drugs. This suggests a degree of specificity to serotonergic activity rather than a non-specific effect of the drugs used.

This could be interpreted in line with a postsynaptic increase in serotonergic activity, as our findings were in the opposite direction from a study using acute tryptophan depletion on a similar task (Crockett, Clark, Smillie, et al., 2012). Alternatively, as serotonergic enhancement through optogenetics was shown to increase exploration in mice (Lottem et al., 2018) whereas we found decreased exploration under citalopram, these findings could be consistent with a presynaptic autoreceptor-mediated transient lowering of serotonergic activity. Further research using acute tryptophan depletion with our task variant and modelling technique could help to disentangle these explanations. Additionally, a study with chronic SSRI administration could unpack the time course of changes in decision making ability and determine whether these would persist following neural adaptations in longer-term changes in serotonergic activity.

2. *Despite indications of important links between serotonergic systems and social decision-making, prior research did not incorporate dynamic modelling of interaction – where interacting individuals' choices affect one another's behaviour, and thus repeated interaction leads to changes in incentives to cooperate and compete. This gave them limited generalisability to actual decisions in a social environment.*

Chapter 3 did not show specific effects of either citalopram or atomoxetine on outcomes or learning models on choices in dynamic strategic interaction. However, we showed that that better outcomes were achieved when both participants were in the same

condition, whether drug or placebo, than when they were in different conditions. This applied to both drugs. Learning models on behaviour indicated that when in the same condition, participant actions were reinforced more when they benefited the other party as well. The results suggest that coordination is more easily achieved when pair members are in the same pharmacologically-induced state, with important implications for how real-life interactions achieve coordinated outcomes. If coordinated action requires accurate mutual prediction, resulting in the convergence of generative models of one another's behaviour as suggested by computational theories (Friston & Frith, 2015), then pharmacological challenge to one pair member in isolation may disrupt this process.

Further research could probe whether these effects generalise to other pharmacological treatments. A follow-up study could also redress the limitations of this study by specifically assigning the participants pairs to the same or different conditions to achieve more power through equal group sizes and testing, and test each participant in the two pair conditions. Additionally, using chronic treatment of the two drugs, research could understand whether this dynamic mismatch of states persists once adaptation to drug effects has taken place. Another potential approach would be to look at different levels of endogenous neurotransmitters and genotypes with effects on these levels. Serotonergic approaches could look at levels of serotonin metabolites such as 5-HIAA as a proxy for serotonergic activity (Williams et al., 2001) or polymorphisms of the SERT gene, while prefrontal dopaminergic activity could be proxied by examining COMT genotype (Set et al., 2014). In both cases we would hypothesise that similar neurotransmitter levels between pairs, rather than levels of individual members, might predict higher outcomes. This would have the advantage of greater generalisability. Finally, a fuller picture of underlying neural correlates of this effect could be achieved using neuroimaging with hyperscanning: simultaneous imaging of interactors. There is evidence of synchronised patterns of neural activity during social interaction (reviewed in Koike, Tanabe, & Sadato, 2015). By manipulating pairs of interactors with pharmacological challenge or placebo, this technique

could determine whether neural synchronisation is better predicted by being in the same treatment condition than by treatment condition of the individual, and where these effects might be localised. We would hypothesise increased neural synchronisation in areas involved in social cognition such as medial prefrontal cortex and temporoparietal junction (Amodio & Frith, 2006), as well as areas linked with computation of value (social and non-social) in ventromedial/orbitofrontal cortex (Knutson et al., 2005; Levy & Glimcher, 2012).

3. *The neurochemical mechanisms behind interoception had not been identified despite its pivotal importance to affective processes. There are strong reasons to suspect serotonergic involvement, but to date no studies have tested this hypothesis. Additionally, while research showed that interoceptive decisions (like decisions in other perceptual modalities) demonstrate dissociations between first-order and metacognitive levels, neurotransmitter contributions at different levels had not been examined, nor had their interactions with metacognitive processes in other domains.*

Chapter 4 demonstrated that acute citalopram could enhance cardiac interoceptive awareness on the discrimination task and accuracy on the tracking task, controlling for underlying changes in cardiac measures. This is strongly suggestive of serotonergic involvement with interoceptive processes in the central nervous system. Tests with a separate visual metacognition task showed no drug effect nor mediation of the effect on interoceptive awareness. This indicated that serotonergic effects on interoceptive metacognition were independent of a general effect on metacognitive processes.

Chapter 5 showed that serotonergic effects on interoception, relative to exteroceptive focus, were manifested in altered BOLD response in the absence of global resting cerebral blood flow change. This effect came with the caveat that response was driven in part by enhanced exteroceptive attention. This is consistent with an attenuation of externally-focussed attention and heightening of internal focus. Notably, although there

was no effect on subjective psychological reports, our findings on interoception suggest opposite effects to bluntened interoception seen in depression both behaviourally and in the brain (Avery et al., 2014; Ehlers, 1993; Furman et al., 2013; Pollatos et al., 2009). They were in the same direction to heightened interoception seen in anxiety-related conditions (Ehlers, 1993; Köteles & Doering, 2016; Pollatos et al., 2009). This is in line with findings of SSRI treatment resulting in relatively early improvement in some depressive symptoms (Taylor et al., 2006) but anxiogenic effects (Nutt, 2005).

Further work with clinical populations could explore how changes to interoceptive processes may mediate recovery from mood disorders characterised by serotonin dysfunction. Research with chronic treatment could determine the time course of neural adaptation to these challenges in healthy and disordered populations. Additionally, research using other methods of serotonergic manipulation such as tryptophan depletion or enhancement could be undertaken to provide further tests of the hypothesis of serotonergic involvement.

4. *Despite the relevance of self- and other-related affective processing to mood disorders that present serotonergic dysfunction and are linked to interoceptive processes, their neural correlates and overlap with interoceptive processes in the brain had not been mapped out under serotonergic challenge.*

Chapter 6 showed a set of areas were commonly recruited in the conjunction of emotional self-appraisal (against other-appraisal) and interoceptive focus (against exteroceptive). These included nodes of the default mode network (DMN) in medial prefrontal cortex, posterior cingulate and angular gyrus, plus insula and cuneus. These areas have been previously shown to play roles in emotional and interoceptive processing (Babo-Rebelo, Wolpert, et al., 2016; Gusnard et al., 2001; Lane et al., 1997; Uddin et al., 2017). Additionally, we found increased connectivity between seeds in precuneus, frontal pole and angular gyrus with areas of the frontal gyrus during focus on one's own emotions.

During focus on another person's emotions, seeds in angular gyrus, superior frontal gyrus (SFG) and anterior cingulate showed increased connectivity with a set of cortical midline structures including anterior and posterior cingulate and precuneus, as well as superior and middle temporal gyrus, replicating previous findings on shared substrates of self-appraisal and mentalising (Amodio & Frith, 2006; Mahy et al., 2014; Mars et al., 2012)

We found that citalopram altered connectivity between SFG and parahippocampal gyrus. The drug effect lowered connectivity in self-focus but increased connectivity in other-focus, in line with acute SSRI challenge increasing the salience of social information (M. Browning et al., 2007; C. J. Harmer, Bhagwagar, et al., 2003). They also add to evidence of shared substrates for appraisal of own emotion and interoceptive focus (Terasawa, Fukushima, et al., 2013), with DMN and insula correlates. They highlight a common set of areas involved in processing own emotions and the emotions of others (Kanske et al., 2015; Mahy et al., 2014; Mars et al., 2012), with other-directed processing showing increased connectivity to midline structures.

We did not include a control condition with display of facial expressions in the absence of affective judgements. This limited our ability to disentangle self- and other-directed focus from facial stimulus processing without emotional appraisal. Future work could include this as a control condition in order to differentiate serotonergic effects on affective appraisal and contrast each appraisal condition with this control. Additionally, as with the other aims, research with chronic treatment could help to understand how changes in salience of social information may alter following longer-term alterations in serotonergic activity.

7.2 LINKING INTEROCEPTIVE PROCESSES TO DECISION MAKING

As noted in the introduction, the somatic marker hypothesis (Bechara & Damasio, 2005) links decision making to somatic and affective states, and there is evidence that interoceptive processes may mediate this link (Kandasamy et al., 2016; Werner et al., 2009). However, the resulting impact on the effectiveness of decisions may be a function of existing biases (Dunn, Galton, et al., 2010), so enhanced interoception may not necessarily lead to improved decision making. As these biases may develop over the course of a lifetime and have associated feedback mechanisms (e.g. better decisions leading to better life outcomes and vice versa), it is difficult to disentangle these on the basis simply of innate interoceptive ability.

Under serotonergic challenge, we demonstrated in chapter 2 that utility of decision-making fell, whereas chapter 4 showed that interoceptive awareness increased. While this is not a formal statistical test and these findings are based on (partially) different samples[4], it is indicative that serotonergic enhancement of interoception alone does not promote better decision making. Neuroimaging findings in chapter 5 showed that citalopram increased activity during interoceptive focus in clusters near (<6mm between peak voxels) to those identified in previous work with key decision making attributes. These include experienced value of monetary reward in OFC (Li et al., 2015; Mullett & Tunney, 2013), expected value of reward in PCC (Knutson et al., 2005; Wright et al., 2013), and cognitive control in response to delayed reward in supramarginal gyrus (Hutcherson et al., 2012; Massar et al., 2015). While this is only suggestive, as we did not include decision making tasks in neuroimaging, it is nonetheless telling that three of the four identified areas that showed increased interoceptive activity under citalopram also showed these common substrates. Further research would be needed (for example, with chronic serotonergic treatment) to determine whether longitudinal changes in interoception would take place

[4] See chapter 1 for details of the samples

and would automatically translate into improved decision making, or if the influence of decision making experience is a necessary condition for this to occur.

7.3 LEARNING FROM INFORMATION IN THE SOCIAL AND NON-SOCIAL ENVIRONMENT

Both information sampling and strategic interaction tasks had learning elements. In the IST, developing a model of the generative processes of the environment allowed for more accurate inferences on the probabilities of outcomes. In the strategic interaction, updating the values of actions based on the reward they obtain for both the participant and the other pair member led to higher outcomes. However, there were important differences in the decision problems. The environment in the IST was purely determined by existing contingencies; contingencies neither changed over time nor in response to the actions of the participant. By contrast, the other pair member in the strategic task could modify their strategies over time and in reaction to the participant's behaviour, leading to a more dynamic environment on which to decide. Furthermore, the bonus structure meant a continual changing of incentives. While absent of rapidly changing contingencies, the IST had a potentially greater information space to process. Ideal inferences required monitoring of both current information on the task board (with both revealed colours and an inference on the unrevealed ones) and the history of previous feedback – which due to the static contingencies remained relevant for the duration of the task. Also, the social nature of the information from the strategic interaction task framed decision problems in a different context, with additional considerations of inclination towards the other pair member's reward.

We saw different effects of citalopram on the two tasks, with citalopram appearing to hinder sampling decisions without an accompanying effect on strategic interaction. Work

in non-social domains revealed effects of serotonin-related genotype (den Ouden et al., 2013) and citalopram (Chamberlain et al., 2006) on probabilistic reversal learning (i.e. learning in a non-stationary contingency environment), which suggests that the social framing of our strategic task may have led to differential effects of the drug manipulation. However, we did not replicate existing findings on specific effects on cooperative behaviour under serotonergic manipulation (Tse & Bond, 2002; Wood et al., 2006), with the caveat that our task was designed to evoke a more natural dynamic environment of strategic behaviour.

Our findings on the emotion appraisal task showed areas of parahippocampal gyrus to exhibit more connectivity with superior frontal gyrus during other-directed focus on citalopram compared to placebo. Both areas are linked with inference on social others (Hartwright et al., 2015; Lombardo et al., 2009; Shane et al., 2009; van der Heiden et al., 2013; Vanderwal et al., 2008). Previous work has also shown that acute citalopram can increase aversion to harming others (Crockett et al., 2010) and perception of others' emotional expressions (M. Browning et al., 2007). This raises the possibility that in general, the effect of acute citalopram may increase the salience of socially-mediated information, but that dynamic effects may mitigate this in the acute phase, if more salience cannot lead to better predictions because of a mismatch between the states of interactors. This idea could be tested with neuroimaging work with the strategic interaction task, to determine whether the patterns of connectivity changes from emotional appraisal are mirrored in social decision making. As previous work cited above has shown effects in both affective and cognitive inferences on others in these areas, this suggests that there may be overlapping substrates.

7.4 DIRECTION OF SEROTONERGIC EFFECT

As noted, citalopram has effects on both presynaptic and postsynaptic neurons, with no clear consensus around the net effect in acute challenge. Research in animals has shown dose dependency (David et al., 2003), with the caveat that dose response may be dependent on the type of animal, and that animal studies have tended to use significantly larger dose per kg body weight. Differential effects based on brain area have been observed, with generally greater effects in dorsal raphe nucleus (DRN; Invernizzi, Belli, & Samanin, 1992) and hippocampus (Invernizzi et al., 1997) than prefrontal cortex. As the DRN is densely innervated with serotonergic neurons projecting to cortical and subcortical regions, concentrations there are associated with somatodendritic autoreceptor activation and consequent inhibition of neural firing (Invernizzi et al., 1992; Ohno, 2010). A further complication is from the descending serotonergic projections from prefrontal cortex which exhibit both excitatory and inhibitory control on further DRN activity (Celada et al., 2002). In humans, PET imaging of clinically-relevant doses of the active S-enantiomer escitalopram showed a trend increase in raphe serotonin levels and trend decreases in serotonin levels across the cortex, but which only reached significance in occipital and temporal cortex (Nord et al., 2013).

Our studies cannot determine the net effect of the citalopram treatment on serotonergic activity, nor whether differential serotonergic effects are seen in different brain areas. Findings from the information sampling study could be interpreted in either direction, while the interoception studies are in line with reduction of characteristic interoceptive deficits in depression but enhanced interoception in anxiety disorders. Follow up work using PET with a serotonin receptor radioligand could help to disentangle this. To date only one study has used SSRI manipulation in this experimental setup (Nord et al., 2013) using a relatively high dose of escitalopram (20mg, i.e. containing the same amount of the active S-enantiomer as 40mg citalopram). No study thus far has examined the time course of neural serotonin concentrations in chronic human treatment. Insight into this

could help to determine whether early antidepressant and anxiogenic effects of SSRI treatment are the results of differential changes in serotonin concentrations in brain areas.

7.5 CONCLUSION

In this work, we found a complex pattern of effects of serotonergic manipulation on various cognitive, affective and metacognitive processes, with several parallels to characteristic differences seen in disorders showing serotonergic dysfunction. These must however be seen in the context of healthy volunteers showing quite different baseline serotonergic function and cognitive measures than patients with these disorders. This motivates future clinical work on clinical populations to understand if these changes are manifested with altered baselines.

Limitations notwithstanding, we leveraged the advantages of using single-dose administration of drugs with full experimental control. This allowed us to isolate neurotransmitter effects from confounds of individual life history, which are particularly pertinent in clinical research, and in some studies employ a dual-drug design to test specificity of neurotransmitter contributions. Our designs used the power gained by within-subjects, placebo-controlled designs while controlling for treatment order effects in analysis. In neuroimaging we also tested for altered resting cerebral blood flow to rule out drug-related, task-independent changes in perfusion. We also did not restrict studies to one gender, while controlling for this factor as well, to increase generalisability of findings. We determined that our manipulations exerted their effects in the absence of self-reported changes in mood, furthering work on early mood-independent cognitive changes from serotonergic challenge. We have therefore added to the literature on serotonergic function in the human brain, and its contributions to decisions and inferences on internal, social and probabilistic information.

REFERENCES

aan het Rot, M., Moskowitz, D. S., Pinard, G., & Young, S. N. (2006). Social behaviour and mood in everyday life: The effects of tryptophan in quarrelsome individuals. *Journal of Psychiatry and Neuroscience*, *31*(4), 253–262.

Adamaszek, M., D'Agata, F., Ferrucci, R., Habas, C., Keulen, S., Kirkby, K. C., Leggio, M., Mariën, P., Molinari, M., Moulton, E., Orsi, L., Van Overwalle, F., Papadelis, C., Priori, A., Sacchetti, B., Schutter, D. J., Styliadis, C., & Verhoeven, J. (2017). Consensus Paper: Cerebellum and Emotion. *The Cerebellum*, *16*(2), 552–576. https://doi.org/10.1007/s12311-016-0815-8

Adolfi, F., Couto, B., Richter, F., Decety, J., Lopez, J., Sigman, M., Manes, F., & Ibáñez, A. (2017). Convergence of interoception, emotion, and social cognition: A twofold fMRI meta-analysis and lesion approach. *Cortex*, *88*, 124–142. https://doi.org/10.1016/j.cortex.2016.12.019

Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Do visual perspective tasks need theory of mind? *NeuroImage*, *30*(3), 1059–1068. https://doi.org/10.1016/j.neuroimage.2005.10.026

Akimova, E., Lanzenberger, R., & Kasper, S. (2009). The Serotonin-1A Receptor in Anxiety Disorders. *Biological Psychiatry*, *66*(7), 627–635. https://doi.org/10.1016/j.biopsych.2009.03.012

Aleksandrov, V. G., & Aleksandrova, N. P. (2015). The role of the insular cortex in the control of visceral functions. *Human Physiology*, *41*(5), 553–561. https://doi.org/10.1134/S0362119715050023

Alex, K. D., & Pehek, E. A. (2007). Pharmacologic mechanisms of serotonergic regulation of dopamine neurotransmission. *Pharmacology & Therapeutics*, *113*(2), 296–320. https://doi.org/10.1016/j.pharmthera.2006.08.004

Allen, M., & Tsakiris, M. (2018). *The body as first prior: Interoceptive predictive processing and the primacy of self-models*. Oxford University Press.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(4), 268–277. https://doi.org/10.1038/nrn1884

Anderson, G. M. (2004). Peripheral and central neurochemical effects of the selective serotonin reuptake inhibitors (SSRIs) in humans and nonhuman primates: Assessing bioeffect and

mechanisms of action. *International Journal of Developmental Neuroscience*, *22*(5), 397–404. https://doi.org/10.1016/j.ijdevneu.2004.06.006

Anticevic, A., Cole, M. W., Murray, J. D., Corlett, P. R., Wang, X.-J., & Krystal, J. H. (2012). The role of default network deactivation in cognition and disease. *Trends in Cognitive Sciences*, *16*(12), 584–592. https://doi.org/10.1016/j.tics.2012.10.008

Araujo, Helder F., Kaplan, J., Damasio, H., & Damasio, A. (2015). Neural correlates of different self domains. *Brain and Behavior*, *5*(12). https://doi.org/10.1002/brb3.409

Araujo, Helder Filipe, Kaplan, J., & Damasio, A. (2013). Cortical Midline Structures and Autobiographical-Self Processes: An Activation-Likelihood Estimation Meta-Analysis. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00548

Arce, E., Simmons, A. N., Lovero, K. L., Stein, M. B., & Paulus, M. P. (2008). Escitalopram effects on insula and amygdala BOLD activation during emotional processing. *Psychopharmacology*, *196*(4), 661–672. https://doi.org/10.1007/s00213-007-1004-8

Arnsten, A. F. T. (2009). The Emerging Neurobiology of Attention Deficit Hyperactivity Disorder: The Key Role of the Prefrontal Association Cortex. *The Journal of Pediatrics*, *154*(5), I-S43. https://doi.org/10.1016/j.jpeds.2009.01.018

Arrow, K. J. (1971). The theory of risk aversion. *Essays in the Theory of Risk-Bearing*, 90–120.

Atique, B., Erb, M., Gharabaghi, A., Grodd, W., & Anders, S. (2011). Task-specific activity and connectivity within the mentalizing network during emotion and intention mentalizing. *NeuroImage*, *55*(4), 1899–1911. https://doi.org/10.1016/j.neuroimage.2010.12.036

Averbeck, B. B. (2015). Theory of choice in bandit, information sampling and foraging tasks. *PLoS Computational Biology*, *11*(3), e1004164. https://doi.org/10.1371/journal.pcbi.1004164

Avery, J. A., Drevets, W. C., Moseman, S. E., Bodurka, J., Barcalow, J. C., & Simmons, W. K. (2014). Major depressive disorder is associated with abnormal interoceptive activity and functional connectivity in the insula. *Biological Psychiatry*, *76*(3), 258–266. https://doi.org/10.1016/j.biopsych.2013.11.027

Avery, J. A., Kerr, K. L., Ingeholm, J. E., Burrows, K., Bodurka, J., & Simmons, W. K. (2015). A common gustatory and interoceptive representation in the human mid-insula. *Human Brain Mapping*, *36*(8), 2996–3006. https://doi.org/10.1002/hbm.22823

Axelsen, M. C., Jepsen, J. R. M., & Bak, N. (2018). The Choice of Prior in Bayesian Modeling of the Information Sampling Task. *Biological Psychiatry*, *83*(12), e59–e60. https://doi.org/10.1016/J.BIOPSYCH.2017.04.021

Babo-Rebelo, M., Richter, C. G., & Tallon-Baudry, C. (2016). Neural Responses to Heartbeats in the Default Network Encode the Self in Spontaneous Thoughts. *Journal of Neuroscience*, *36*(30), 7829–7840. https://doi.org/10.1523/JNEUROSCI.0262-16.2016

Babo-Rebelo, M., Wolpert, N., Adam, C., Hasboun, D., & Tallon-Baudry, C. (2016). Is the cardiac monitoring function related to the self in both the default network and right anterior insula? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160004. https://doi.org/10.1098/rstb.2016.0004

Ballesteros, J., & Callado, L. F. (2004). Effectiveness of pindolol plus serotonin uptake inhibitors in depression: A meta-analysis of early and late outcomes from randomised controlled trials. *Journal of Affective Disorders*, *79*(1), 137–147. https://doi.org/10.1016/S0165-0327(02)00404-4

Banca, P., Lange, I., Worbe, Y., Howell, N. A., Irvine, M., Harrison, N. A., Moutoussis, M., & Voon, V. (2016). Reflection impulsivity in binge drinking: Behavioural and volumetric correlates. *Addiction Biology*, *21*(2), 504–515. https://doi.org/10.1111/adb.12227

Barnes, N. M., & Sharp, T. (1999). A review of central 5-HT receptors and their function. *Neuropharmacology*, *38*(8), 1083–1152. https://doi.org/10.1016/S0028-3908(99)00010-6

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beacher, F. D. C. C., Gray, M. A., Minati, L., Whale, R., Harrison, N. A., & Critchley, H. D. (2011). Acute tryptophan depletion attenuates conscious appraisal of social emotional signals in healthy female volunteers. *Psychopharmacology*, *213*(2), 603–613. https://doi.org/10.1007/s00213-010-1897-5

Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior*, *52*(2), 336–372. https://doi.org/10.1016/j.geb.2004.06.010

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, *23*(2), 137–152. https://doi.org/10.1109/TMI.2003.822821

Bédard, P., Wu, M., & Sanes, J. N. (2011). Brain Activation Related to Combinations of Gaze Position, Visual Input, and Goal-Directed Hand Movements. *Cerebral Cortex*, *21*(6), 1273–1282. https://doi.org/10.1093/cercor/bhq205

Benkelfat, C., Ellenbogen, M. A., Dean, P., Palmour, R. M., & Young, S. N. (1994). Mood-Lowering Effect of Tryptophan Depletion: Enhanced Susceptibility in Young Men at Genetic Risk for Major Affective Disorders. *Archives of General Psychiatry*, *51*(9), 687–697. https://doi.org/10.1001/archpsyc.1994.03950090019003

Bennett, D., Oldham, S., Dawson, A., Parkes, L., Murawski, C., & Yu¨cel, M. (2016). Systematic Overestimation of Reflection Impulsivity in the Information Sampling Task. In *Biological Psychiatry*. https://doi.org/10.1016/j.biopsych.2016.05.027

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

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

Bigos, K. L., Pollock, B. G., Aizenstein, H. J., Fisher, P. M., Bies, R. R., & Hariri, A. R. (2008). Acute 5-HT Reuptake Blockade Potentiates Human Amygdala Reactivity. *Neuropsychopharmacology*, *33*(13), 3221–3225. https://doi.org/10.1038/npp.2008.52

Bizot, J.-C., David, S., & Trovero, F. (2011). Effects of atomoxetine, desipramine, d-amphetamine and methylphenidate on impulsivity in juvenile rats, measured in a T-maze procedure. *Neuroscience Letters*, *489*(1), 20–24. https://doi.org/10.1016/J.NEULET.2010.11.058

Blanke, O., Ortigue, S., Landis, T., & Seeck, M. (2002). Stimulating illusory own-body perceptions. *Nature*, *419*(6904), 269–270. https://doi.org/10.1038/419269a

Booij, L., Does, W. V. der, Benkelfat, C., Bremner, J. D., Cowen, P. J., Fava, M., Gillin, C., Leyton, M., Moore, P., Smith, K. A., & Kloot, W. A. V. der. (2002). Predictors of Mood Response to Acute Tryptophan Depletion: A Reanalysis. *Neuropsychopharmacology*, *27*(5), 852–861. https://doi.org/10.1016/S0893-133X(02)00361-5

Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. *American Journal of Psychiatry*, *176*(5), 376–387. https://doi.org/10.1176/appi.ajp.2018.18070881

Bosker, F. J., Winter, T. Y. C. E. D., Klompmakers, A. A., & Westenberg, H. G. M. (1996). Flesinoxan Dose-Dependently Reduces Extracellular 5-Hydroxytryptamine (5-HT) in Rat Median Raphe and Dorsal Hippocampus Through Activation of 5-HT1A Receptors. *Journal of Neurochemistry*, *66*(6), 2546–2555. https://doi.org/10.1046/j.1471-4159.1996.66062546.x

Bourke, C., Douglas, K., & Porter, R. (2010). Processing of Facial Emotion Expression in Major Depression: A Review. *Australian & New Zealand Journal of Psychiatry*, *44*(8), 681–696. https://doi.org/10.3109/00048674.2010.496359

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

Brooks, S., Gelman, A., Jones, G., & Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. CRC press.

Browning, K. N., Verheijden, S., & Boeckxstaens, G. E. (2017). The Vagus Nerve in Appetite Regulation, Mood, and Intestinal Inflammation. *Gastroenterology*, *152*(4), 730–744. https://doi.org/10.1053/j.gastro.2016.10.046

Browning, M., Reid, C., Cowen, P. J., Goodwin, G. M., & Harmer, C. J. (2007). A single dose of citalopram increases fear recognition in healthy subjects. *Journal of Psychopharmacology*, *21*(7), 684–690. https://doi.org/10.1177/0269881106074062

(Bud) Craig, A. D. (2009). How do you feel—Now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, *10*(1), 59–70. https://doi.org/10.1038/nrn2555

Burki, N. K., & Lee, L.-Y. (2010). Mechanisms of dyspnea. *Chest*, *138*(5), 1196–1201. https://doi.org/10.1378/chest.10-0534

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed.). Springer-Verlag. https://www.springer.com/gp/book/9780387953649

Bymaster, F. P., Katner, J. S., Nelson, D. L., Hemrick-Luecke, S. K., Threlkeld, P. G., Heiligenstein, J. H., Morin, S. M., Gehlert, D. R., & Perry, K. W. (2002). Atomoxetine Increases Extracellular Levels of Norepinephrine and Dopamine in Prefrontal Cortex of Rat: A Potential Mechanism for Efficacy in Attention Deficit/Hyperactivity Disorder. *Neuropsychopharmacology*, *27*(5), 699–711. https://doi.org/10.1016/S0893-133X(02)00346-9

Caballero, J., & Nahata, M. C. (2003). Atomoxetine hydrochloride for the treatment of attention-deficit/hyperactivity disorder. *Clinical Therapeutics*, *25*(12), 3065–3083. https://doi.org/10.1016/S0149-2918(03)90092-0

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior Research Methods*, *40*(1), 109–115. https://doi.org/10.3758/BRM.40.1.109

Camerer, C. F., & Hare, T. A. (2014). The Neural Basis of Strategic Choice. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics (Second Edition)* (pp. 479–492). Academic Press. https://doi.org/10.1016/B978-0-12-416008-8.00025-5

Camerer, C., & Ho, T. H. (1999). Experience-weighted Attraction Learning in Normal Form Games. *Econometrica*, *67*(4), 827–874. https://doi.org/10.1111/1468-0262.00054

Cameron, O. G. (2002). *Visceral Sensory Neuroscience*. Oxford University Press.

Campbell-Meiklejohn, D., Wakeley, J., Herbert, V., Cook, J., Scollo, P., Ray, M. K., Selvaraj, S., Passingham, R. E., Cowen, P., & Rogers, R. D. (2011). Serotonin and Dopamine Play Complementary Roles in Gambling to Recover Losses. *Neuropsychopharmacology*, *36*(2), 402–410. https://doi.org/10.1038/npp.2010.170

Canessa, N., Motterlini, M., Alemanno, F., Perani, D., & Cappa, S. F. (2011). Learning from other people's experience: A neuroimaging study of decisional interactive-learning. *NeuroImage*, *55*(1), 353–362. https://doi.org/10.1016/j.neuroimage.2010.11.065

Canli, T., & Lesch, K.-P. (2007). Long story short: The serotonin transporter in emotion regulation and social cognition. *Nature Neuroscience*, *10*(9), 1103–1109. https://doi.org/10.1038/nn1964

Carré, A., Gierski, F., Lemogne, C., Tran, E., Raucher-Chéné, D., Béra-Potelle, C., Portefaix, C., Kaladjian, A., Pierot, L., Besche-Richard, C., & Limosin, F. (2014). Linear association

between social anxiety symptoms and neural activations to angry faces: From subclinical to clinical levels. *Social Cognitive and Affective Neuroscience*, *9*(6), 880–886. https://doi.org/10.1093/scan/nst061

Caseras, X., Murphy, K., Mataix-Cols, D., López-Solà, M., Soriano-Mas, C., Ortriz, H., Pujol, J., & Torrubia, R. (2013). Anatomical and functional overlap within the insula and anterior cingulate cortex during interoception and phobic symptom provocation. *Human Brain Mapping*, *34*(5), 1220–1229. https://doi.org/10.1002/hbm.21503

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A., & Poulton, R. (2003). Influence of Life Stress on Depression: Moderation by a Polymorphism in the 5-HTT Gene. *Science*, *301*(5631), 386–389. https://doi.org/10.1126/science.1083968

Caswell, A. J., Bond, R., Duka, T., & Morgan, M. J. (2015). Further evidence of the heterogeneous nature of impulsivity. *Personality and Individual Differences*, *76*, 68–74. https://doi.org/10.1016/J.PAID.2014.11.059

Cechetto, D. F., & Saper, C. B. (1987). Evidence for a viscerotopic sensory representation in the cortex and thalamus in the rat. *Journal of Comparative Neurology*, *262*(1), 27–45. https://doi.org/10.1002/cne.902620104

Celada, P., Puig, M. V., Casanovas, J. M., Guillazo, G., & Artigas, F. (2001). Control of Dorsal Raphe Serotonergic Neurons by the Medial Prefrontal Cortex: Involvement of Serotonin-1A, GABAA, and Glutamate Receptors. *Journal of Neuroscience*, *21*(24), 9917–9929. https://doi.org/10.1523/JNEUROSCI.21-24-09917.2001

Celada, P., Puig, M. V., Martín-Ruiz, R., Casanovas, J. M., & Artigas, F. (2002). Control of the serotonergic system by the medial prefrontal cortex: Potential role in the etiology of PTSD and depressive disorders. *Neurotoxicity Research*, *4*(5), 409–419. https://doi.org/10.1080/10298420290030550

Chamberlain, S. R., Fineberg, N. A., Blackwell, A. D., Clark, L., Robbins, T. W., & Sahakian, B. J. (2007). A neuropsychological comparison of obsessive–compulsive disorder and trichotillomania. *Neuropsychologia*, *45*(4), 654–662. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2006.07.016

Chamberlain, S. R., Hampshire, A., Müller, U., Rubia, K., del Campo, N., Craig, K., Regenthal, R., Suckling, J., Roiser, J. P., Grant, J. E., Bullmore, E. T., Robbins, T. W., & Sahakian, B. J. (2009). Atomoxetine Modulates Right Inferior Frontal Activation During Inhibitory Control: A Pharmacological Functional Magnetic Resonance Imaging Study. *Biological Psychiatry*, *65*(7), 550–555. https://doi.org/10.1016/j.biopsych.2008.10.014

Chamberlain, S. R., Müller, U., Blackwell, A. D., Clark, L., Robbins, T. W., & Sahakian, B. J. (2006). Neurochemical modulation of response inhibition and probabilistic learning in humans. *Science (New York, N.Y.)*, *311*(5762), 861–863. https://doi.org/10.1126/science.1121218

Chaput, Y., de Montigny, C., & Blier, P. (1986). Effects of a selective 5-HT reuptake blocker, citalopram, on the sensitivity of 5-HT autoreceptors: Electrophysiological studies in the rat brain. *Naunyn-Schmiedeberg's Archives of Pharmacology*, *333*(4), 342–348.

Charney, D. S., Heninger, G. R., & Jatlow, P. I. (1985). Increased Anxiogenic Effects of Caffeine in Panic Disorders. *Archives of General Psychiatry*, *42*(3), 233–243. https://doi.org/10.1001/archpsyc.1985.01790260027003

Cheng, Y., Xu, J., Arnone, D., Nie, B., Yu, H., Jiang, H., Bai, Y., Luo, C., Campbell, R. a. A., Shan, B., Xu, L., & Xu, X. (2017). Resting-state brain alteration after a single dose of SSRI administration predicts 8-week remission of patients with major depressive disorder. *Psychological Medicine*, *47*(3), 438–450. https://doi.org/10.1017/S0033291716002440

Chiappori, P.-A., & Paiella, M. (2011). RELATIVE RISK AVERSION IS CONSTANT: EVIDENCE FROM PANEL DATA. *Journal of the European Economic Association*, *9*(6), 1021–1052. https://doi.org/10.1111/j.1542-4774.2011.01046.x

Ciranna, L. (2006). Serotonin as a modulator of glutamate- and GABA-mediated neurotransmission: Implications in physiological functions and in pathology. *Current Neuropharmacology*, *4*(2), 101–114. https://doi.org/10.2174/157015906776359540

Clark, L., Robbins, T. W., Ersche, K. D., & Sahakian, B. J. (2006). Reflection Impulsivity in Current and Former Substance Users. *Biological Psychiatry*, *60*(5), 515–522. https://doi.org/10.1016/j.biopsych.2005.11.007

Clark, L., Roiser, J., Robbins, T., & Sahakian, B. (2009). Disrupted `reflection' impulsivity in cannabis users but not current or former ecstasy users. *Journal of Psychopharmacology*, *23*(1), 14–22. https://doi.org/10.1177/0269881108089587

Clark, Luke, Studer, B., Bruss, J., Tranel, D., & Bechara, A. (2014). Damage to insula abolishes cognitive distortions during simulated gambling. *Proceedings of the National Academy of Sciences*, *111*(16), 6098–6103. https://doi.org/10.1073/pnas.1322295111

Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage*, *37*(1), 343–360. https://doi.org/10.1016/j.neuroimage.2007.03.071

Cools, R., Calder, A. J., Lawrence, A. D., Clark, L., Bullmore, E., & Robbins, T. W. (2005). Individual differences in threat sensitivity predict serotonergic modulation of amygdala response to fearful faces. *Psychopharmacology*, *180*(4), 670–679. https://doi.org/10.1007/s00213-005-2215-5

Cools, R., Nakamura, K., & Daw, N. D. (2011). Serotonin and dopamine: Unifying affective, activational, and decision functions. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *36*(1), 98–113. https://doi.org/10.1038/npp.2010.121

Cools, R., Roberts, A. C., & Robbins, T. W. (2008). Serotoninergic regulation of emotional and behavioural control processes. *Trends in Cognitive Sciences*, *12*(1), 31–40. https://doi.org/10.1016/j.tics.2007.10.011

Coppen, A. (1967). The Biochemistry of Affective Disorders. *The British Journal of Psychiatry*, *113*(504), 1237–1264. https://doi.org/10.1192/bjp.113.504.1237

Craig, A. D. (2009). How do you feel—Now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, *10*(1), 59–70. https://doi.org/10.1038/nrn2555

Critchley, H. D., & Garfinkel, S. N. (2017). Interoception and emotion. *Current Opinion in Psychology*, *17*, 7–14. https://doi.org/10.1016/j.copsyc.2017.04.020

Critchley, H. D., & Harrison, N. A. (2013). Visceral Influences on Brain and Behavior. *Neuron*, *77*(4), 624–638. https://doi.org/10.1016/j.neuron.2013.02.008

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

Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the*

National Academy of Sciences, *107*(40), 17433–17438. https://doi.org/10.1073/pnas.1009396107

Crockett, M. J., Clark, L., Roiser, J. P., Robinson, O. J., Cools, R., Chase, H. W., den Ouden, H., Apergis-Schoute, A., Campbell-Meikeljohn, D., Seymour, B., Sahakian, B. J., Rogers, R. D., & Robbins, T. W. (2012). Converging evidence for central 5-HT effects in acute tryptophan depletion. *Molecular Psychiatry*, *17*(2), 121–123. https://doi.org/10.1038/mp.2011.106

Crockett, M. J., Clark, L., Smillie, L. D., & Robbins, T. W. (2012). The effects of acute tryptophan depletion on costly information sampling: Impulsivity or aversive processing? *Psychopharmacology*, *219*(2), 587–597. https://doi.org/10.1007/s00213-011-2577-9

Crockett, M. J., Clark, L., Tabibnia, G., Lieberman, M. D., & Robbins, T. W. (2008). Serotonin Modulates Behavioral Reactions to Unfairness. *Science*, *320*(5884), 1739–1739. https://doi.org/10.1126/science.1155577

Crockett, M. J., & Fehr, E. (2014). Social brains on drugs: Tools for neuromodulation in social neuroscience. *Social Cognitive and Affective Neuroscience*, *9*(2), 250–254. https://doi.org/10.1093/scan/nst113

Culverhouse, R. C., Saccone, N. L., Horton, A. C., Ma, Y., Anstey, K. J., Banaschewski, T., Burmeister, M., Cohen-Woods, S., Etain, B., Fisher, H. L., Goldman, N., Guillaume, S., Horwood, J., Juhasz, G., Lester, K. J., Mandelli, L., Middeldorp, C. M., Olié, E., Villafuerte, S., … Bierut, L. J. (2018). Collaborative meta-analysis finds no evidence of a strong interaction between stress and 5-HTTLPR genotype contributing to the development of depression. *Molecular Psychiatry*, *23*(1), 133–142. https://doi.org/10.1038/mp.2017.44

Cusi, A. M., Nazarov, A., Holshausen, K., MacQueen, G. M., & McKinnon, M. C. (2012). Systematic review of the neural basis of social cognition in patients with mood disorders. *Journal of Psychiatry & Neuroscience : JPN*, *37*(3), 154–169. https://doi.org/10.1503/jpn.100179

Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2013). Chapter Two—The Self-Organization of Human Interaction. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 59, pp. 43–95). Academic Press. https://doi.org/10.1016/B978-0-12-407187-2.00002-2

Dalley, J. W., & Roiser, J. P. (2012). Dopamine, serotonin and impulsivity. *Neuroscience*, *215*, 42–58. https://doi.org/10.1016/J.NEUROSCIENCE.2012.03.065

David, D. J. P., Bourin, M., Jego, G., Przybylski, C., Jolliet, P., & Gardier, A. M. (2003). Effects of acute treatment with paroxetine, citalopram and venlafaxine in vivo on noradrenaline and serotonin outflow: A microdialysis study in Swiss mice. *British Journal of Pharmacology*, *140*(6), 1128–1136. https://doi.org/10.1038/sj.bjp.0705538

Daw, N. D., Kakade, S., & Dayan, P. (2002). Opponent interactions between serotonin and dopamine. *Neural Networks*, *15*(4), 603–616. https://doi.org/10.1016/S0893-6080(02)00052-7

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

Dayan, P., & Huys, Q. J. M. (2008). Serotonin, Inhibition, and Negative Mood. *PLOS Computational Biology*, *4*(2), e4. https://doi.org/10.1371/journal.pcbi.0040004

Dayan, P., & Huys, Q. J. M. (2009). Serotonin in Affective Control. *Annual Review of Neuroscience*, *32*(1), 95–126. https://doi.org/10.1146/annurev.neuro.051508.135607

Delazer, M., Högl, B., Zamarian, L., Wenter, J., Gschliesser, V., Ehrmann, L., Brandauer, E., Cevikkol, Z., & Frauscher, B. (2011). Executive functions, information sampling, and decision making in narcolepsy with cataplexy. *Neuropsychology*, *25*(4), 477–487. https://doi.org/10.1037/a0022357

Del-Ben, C. M., Deakin, J. F. W., Mckie, S., Delvai, N. A., Williams, S. R., Elliott, R., Dolan, M., & Anderson, I. M. (2005). The Effect of Citalopram Pretreatment on Neuronal Responses to Neuropsychological Tasks in Normal Volunteers: An fMRI Study. *Neuropsychopharmacology*, *30*(9), 1724–1734. https://doi.org/10.1038/sj.npp.1300728

Deming, P., Philippi, C. L., Wolf, R. C., Dargis, M., Kiehl, K. A., & Koenigs, M. (2018). Psychopathic traits linked to alterations in neural activity during personality judgments of self and others. *NeuroImage: Clinical*, *18*, 575–581. https://doi.org/10.1016/j.nicl.2018.02.029

den Ouden, H. E. M., Daw, N. D., Fernandez, G., Elshout, J. A., Rijpkema, M., Hoogman, M., Franke, B., & Cools, R. (2013). Dissociable effects of dopamine and serotonin on reversal learning. *Neuron*, *80*(4), 1090–1100. https://doi.org/10.1016/j.neuron.2013.08.030

Detre, J. A., Rao, H., Wang, D. J. J., Chen, Y. F., & Wang, Z. (2012). Applications of arterial spin labeled MRI in the brain. *Journal of Magnetic Resonance Imaging*, *35*(5), 1026–1037. https://doi.org/10.1002/jmri.23581

Di Matteo, V., Di Giovanni, G., Pierucci, M., & Esposito, E. (2008). Serotonin control of central dopaminergic function: Focus on in vivo microdialysis studies. In G. Di Giovann, V. Di Matteo, & E. Esposito (Eds.), *Progress in Brain Research* (Vol. 172, pp. 7–44). Elsevier. https://doi.org/10.1016/S0079-6123(08)00902-3

Dick, D. M., Agrawal, A., Keller, M. C., Adkins, A., Aliev, F., Monroe, S., Hewitt, J. K., Kendler, K. S., & Sher, K. J. (2015). Candidate Gene–Environment Interaction Research: Reflections and Recommendations. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691614556682

Djamshidian, A., Sanotsky, Y., Matviyenko, Y., O'Sullivan, S. S., Sharman, S., Selikhova, M., Fedoryshyn, L., Filts, Y., Bearn, J., Lees, A. J., & Averbeck, B. B. (2013). Increased reflection impulsivity in patients with ephedrone-induced Parkinsonism. *Addiction*, *108*(4), 771–779. https://doi.org/10.1111/add.12080

Doerig, N., Schlumpf, Y., Spinelli, S., Späti, J., Brakowski, J., Quednow, B. B., Seifritz, E., & Grosse Holtforth, M. (2014). Neural representation and clinically relevant moderators of individualised self-criticism in healthy subjects. *Social Cognitive and Affective Neuroscience*, *9*(9), 1333–1340. https://doi.org/10.1093/scan/nst123

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). INDIVIDUAL RISK ATTITUDES: MEASUREMENT, DETERMINANTS, AND BEHAVIORAL CONSEQUENCES. *Journal of the European Economic Association*, *9*(3), 522–550. https://doi.org/10.1111/j.1542-4774.2011.01015.x

Domschke, K., Stevens, S., Pfleiderer, B., & Gerlach, A. L. (2010). Interoceptive sensitivity in anxiety and anxiety disorders: An overview and integration of neurobiological findings. *Clinical Psychology Review*, *30*(1), 1–11. https://doi.org/10.1016/j.cpr.2009.08.008

Drevets, W. C., Frank, E., Price, J. C., Kupfer, D. J., Holt, D., Greer, P. J., Huang, Y., Gautier, C., & Mathis, C. (1999). PET imaging of serotonin 1A receptor binding in depression. *Biological Psychiatry*, *46*(10), 1375–1387. https://doi.org/10.1016/S0006-3223(99)00189-4

DuBois, D., Ameis, S. H., Lai, M.-C., Casanova, M. F., & Desarkar, P. (2016). Interoception in Autism Spectrum Disorder: A review. *International Journal of Developmental Neuroscience*, *52*, 104–111. https://doi.org/10.1016/j.ijdevneu.2016.05.001

Dudley, R., Taylor, P., Wickham, S., & Hutton, P. (2016). Psychosis, Delusions and the 'Jumping to Conclusions' Reasoning Bias: A Systematic Review and Meta-analysis. *Schizophrenia Bulletin*, *42*(3), 652–665. https://doi.org/10.1093/schbul/sbv150

Dunn, B. D., Galton, H. C., Morgan, R., Evans, D., Oliver, C., Meyer, M., Cusack, R., Lawrence, A. D., & Dalgleish, T. (2010). Listening to your heart. How interoception shapes emotion experience and intuitive decision making. *Psychological Science*, *21*(12), 1835–1844. https://doi.org/10.1177/0956797610389191

Dunn, B. D., Stefanovitch, I., Evans, D., Oliver, C., Hawkins, A., & Dalgleish, T. (2010). Can you feel the beat? Interoceptive awareness is an interactive function of anxiety- and depression-specific symptom dimensions. *Behaviour Research and Therapy*, *48*(11), 1133–1138. https://doi.org/10.1016/J.BRAT.2010.07.006

Dutta, A., McKie, S., Downey, D., Thomas, E., Juhasz, G., Arnone, D., Elliott, R., Williams, S., Deakin, J. F. W., & Anderson, I. M. (2019). Regional default mode network connectivity in major depressive disorder: Modulation by acute intravenous citalopram. *Translational Psychiatry*, *9*(1), 1–9. https://doi.org/10.1038/s41398-019-0447-0

Dutta, P. K. (1999). *Strategies and Games: Theory and Practice* (1st ed., Vol. 1). The MIT Press. https://EconPapers.repec.org/RePEc:mtp:titles:0262041693

Eckel, C. C., & Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior*, *23*(4), 281–295. https://doi.org/10.1016/S1090-5138(02)00097-1

Eckel, C. C., & Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, *68*(1), 1–17. https://doi.org/10.1016/J.JEBO.2008.04.006

Ehlers, A. (1993). Interoception and panic disorder. *Advances in Behaviour Research and Therapy*, *15*(1), 3–21. https://doi.org/10.1016/0146-6402(93)90001-I

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of*

the United States of America, *113*(28), 7900–7905. https://doi.org/10.1073/pnas.1602413113

Ekselius, L., von Knorring, L., & Eberhard, G. (1997). A double-blind multicenter trial comparing sertraline and citalopram in patients with major depression treated in general practice. *International Clinical Psychopharmacology*, *12*(6), 323–331.

El Mansari, M., Sánchez, C., Chouvet, G., Renaud, B., & Haddjeri, N. (2005). Effects of Acute and Long-Term Administration of Escitalopram and Citalopram on Serotonin Neurotransmission: An In Vivo Electrophysiological Study in Rat Brain. *Neuropsychopharmacology*, *30*(7), 1269–1277. https://doi.org/10.1038/sj.npp.1300686

Ellenbogen, M. A., Young, S. N., Dean, P., Palmour, R. M., & Benkelfat, C. (1996). Mood Response to Acute Tryptophan Depletion in Healthy Volunteers: Sex Differences and Temporal Stability. *Neuropsychopharmacology*, *15*(5), 465–474. https://doi.org/10.1016/S0893-133X(96)00056-5

Evers, E. A. T., Cools, R., Clark, L., Veen, F. M. van der, Jolles, J., Sahakian, B. J., & Robbins, T. W. (2005). Serotonergic Modulation of Prefrontal Cortex during Negative Feedback in Probabilistic Reversal Learning. *Neuropsychopharmacology*, *30*(6), 1138–1147. https://doi.org/10.1038/sj.npp.1300663

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

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, *13*(1), 1–25. https://doi.org/10.1007/s12110-002-1012-7

Fehr, E., & Krajbich, I. (2014). Social Preferences and the Brain. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics (Second Edition)* (pp. 193–218). Academic Press. https://doi.org/10.1016/B978-0-12-416008-8.00011-5

FitzGerald, T. H. B., Schwartenbeck, P., Moutoussis, M., Dolan, R. J., & Friston, K. (2015). Active Inference, Evidence Accumulation, and the Urn Task. *Neural Computation*, *27*(2), 306–328. https://doi.org/10.1162/NECO_a_00699

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1594), 1338–1349. https://doi.org/10.1098/rstb.2011.0417

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00443

Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain : A Journal of Neurology*, *137*(Pt 10), 2811–2822. https://doi.org/10.1093/brain/awu221

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science (New York, N.Y.)*, *329*(5998), 1541–1543. https://doi.org/10.1126/science.1191883

Fornito, A., Harrison, B. J., Zalesky, A., & Simons, J. S. (2012). Competitive and cooperative dynamics of large-scale brain functional networks supporting recollection. *Proceedings of the National Academy of Sciences*, *109*(31), 12788–12793. https://doi.org/10.1073/pnas.1204185109

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

Frazer, A., & Benmansour, S. (2002). Delayed pharmacological effects of antidepressants. *Molecular Psychiatry*, *7*(S1), S23–S28. https://doi.org/10.1038/sj.mp.4001015

Friston, K. J., & Frith, C. D. (2015). Active inference, communication and hermeneutics. *Cortex*, *68*, 129–143. https://doi.org/10.1016/j.cortex.2015.03.025

Frith, U., & Frith, C. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 459–473. https://doi.org/10.1098/rstb.2002.1218

Furman, D. J., Waugh, C. E., Bhattacharjee, K., Thompson, R. J., & Gotlib, I. H. (2013). Interoceptive awareness, positive affect, and decision making in Major Depressive Disorder. *Journal of Affective Disorders*, *151*(2), 780–785. https://doi.org/10.1016/j.jad.2013.06.044

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*(12), 493–501. https://doi.org/10.1016/S1364-6613(98)01262-5

Gamo, N. J., Wang, M., & Arnsten, A. F. T. (2010). Methylphenidate and Atomoxetine Enhance Prefrontal Function Through α2-Adrenergic and Dopamine D1 Receptors. *Journal of the American Academy of Child and Adolescent Psychiatry*, *49*(10), 1011–1023. https://doi.org/10.1016/j.jaac.2010.06.015

Garfinkel, S. N., & Critchley, H. D. (2013). Interoception, emotion and brain: New insights link internal physiology to social behaviour. Commentary on:"Anterior insular cortex mediates bodily sensibility and social anxiety" by Terasawa et al. (2012). *Social Cognitive and Affective Neuroscience*, *8*(3), 231–234. https://doi.org/10.1093/scan/nss140

Garfinkel, S. N., Manassei Miranda F., Hamilton-Fletcher Giles, In den Bosch Yvo, Critchley Hugo D., & Engels Miriam. (2016). Interoceptive dimensions across cardiac and respiratory axes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160014. https://doi.org/10.1098/rstb.2016.0014

Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., & Critchley, H. D. (2015). Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biological Psychology*, *104*, 65–74. https://doi.org/10.1016/J.BIOPSYCHO.2014.11.004

Garfinkel, S. N., Tiley, C., O'Keeffe, S., Harrison, N. A., Seth, A. K., & Critchley, H. D. (2016). Discrepancies between dimensions of interoception in autism: Implications for emotion and anxiety. *Biological Psychology*, *114*, 117–126. https://doi.org/10.1016/j.biopsycho.2015.12.003

Garrod, S., & Pickering, M. J. (2009). Joint Action, Interactive Alignment, and Dialog. *Topics in Cognitive Science*, *1*(2), 292–304. https://doi.org/10.1111/j.1756-8765.2009.01020.x

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, *100*(1), 253–258. https://doi.org/10.1073/pnas.0135058100

Griffanti, L., Douaud, G., Bijsterbosch, J., Evangelisti, S., Alfaro-Almagro, F., Glasser, M. F., Duff, E. P., Fitzgibbon, S., Westphal, R., Carone, D., Beckmann, C. F., & Smith, S. M. (2017). Hand

classification of fMRI ICA noise components. *NeuroImage*, *154*, 188–205. https://doi.org/10.1016/J.NEUROIMAGE.2016.12.036

Grillon, C., Levenson, J., & Pine, D. S. (2007). A Single Dose of the Selective Serotonin Reuptake Inhibitor Citalopram Exacerbates Anxiety in Humans: A Fear-Potentiated Startle Study. *Neuropsychopharmacology*, *32*(1), 225–231. https://doi.org/10.1038/sj.npp.1301204

Grynberg, D., & Pollatos, O. (2015). Perceiving one's body shapes empathy. *Physiology & Behavior*, *140*, 54–60. https://doi.org/10.1016/j.physbeh.2014.12.026

Gurguis, G. N. M., Vitton, B. J., & Uhde, T. W. (1997). Behavioral, sympathetic and adrenocortical responses to yohimbine in panic disorder patients and normal controls. *Psychiatry Research*, *71*(1), 27–39. https://doi.org/10.1016/S0165-1781(97)00041-3

Gusnard, D. A., Akbudak, E., Shulman, G. L., & Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*(7), 4259–4264. https://doi.org/10.1073/pnas.071043098

Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews Neuroscience*, *2*(10), 685–694. https://doi.org/10.1038/35094500

Guterstam, A., Björnsdotter, M., Gentile, G., & Ehrsson, H. H. (2015). Posterior Cingulate Cortex Integrates the Senses of Self-Location and Body Ownership. *Current Biology*, *25*(11), 1416–1425. https://doi.org/10.1016/j.cub.2015.03.059

Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian Journal of Internal Medicine*, *4*(2), 627–635.

Hamilton, J. P., Etkin, A., Furman, D. J., Lemus, M. G., Johnson, R. F., & Gotlib, I. H. (2012). Functional Neuroimaging of Major Depressive Disorder: A Meta-Analysis and New Integration of Baseline Activation and Neural Response Data. *American Journal of Psychiatry*, *169*(7), 693–703. https://doi.org/10.1176/appi.ajp.2012.11071105

Harada, T., Inada, K., Yamada, K., Sakamoto, K., & Ishigooka, J. (2014). A prospective naturalistic study of antidepressant-induced jitteriness/anxiety syndrome. *Neuropsychiatric Disease and Treatment*, *10*, 2115–2121. https://doi.org/10.2147/NDT.S70637

Harinck, F., Van Dijk, E., Van Beest, I., & Mersmann, P. (2007). When Gains Loom Larger Than Losses. *Psychological Science*, *18*(12), 1099–1105. https://doi.org/10.1111/j.1467-9280.2007.02031.x

Hariri, A. R., Mattay, V. S., Tessitore, A., Kolachana, B., Fera, F., Goldman, D., Egan, M. F., & Weinberger, D. R. (2002). Serotonin Transporter Genetic Variation and the Response of the Human Amygdala. *Science*, *297*(5580), 400–403. https://doi.org/10.1126/science.1071829

Harmer, C. J., Bhagwagar, Z., Perrett, D. I., Völlm, B. A., Cowen, P. J., & Goodwin, G. M. (2003). Acute SSRI Administration Affects the Processing of Social Cues in Healthy Volunteers. *Neuropsychopharmacology*, *28*(1), 148–152. https://doi.org/10.1038/sj.npp.1300004

Harmer, C. J., Rogers, R. D., Tunbridge, E., Cowen, P. J., & Goodwin, G. M. (2003). Tryptophan depletion decreases the recognition of fear in female volunteers. *Psychopharmacology*, *167*(4), 411–417. https://doi.org/10.1007/s00213-003-1401-6

Harmer, Catherine J. (2008). Serotonin and emotional processing: Does it help explain antidepressant drug action? *Neuropharmacology*, *55*(6), 1023–1028. https://doi.org/10.1016/j.neuropharm.2008.06.036

Harmer Catherine J., & Cowen Philip J. (2013). 'It's the way that you look at it'—A cognitive neuropsychological account of SSRI action in depression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1615), 20120407. https://doi.org/10.1098/rstb.2012.0407

Harmer, Catherine J., Shelley, N. C., Cowen, P. J., & Goodwin, G. M. (2004). Increased positive versus negative affective perception and memory in healthy volunteers following selective serotonin and norepinephrine reuptake inhibition. *The American Journal of Psychiatry*, *161*(7), 1256–1263. https://doi.org/10.1176/appi.ajp.161.7.1256

Harsanyi, J. C. (1967). Games with Incomplete Information Played by "Bayesian" Players, I–III Part I. The Basic Model. *Management Science*. https://doi.org/10.1287/mnsc.14.3.159

Hart, N., McGowan, J., Minati, L., & Critchley, H. D. (2013). Emotional regulation and bodily sensation: Interoceptive awareness is intact in borderline personality disorder. *Journal of Personality Disorders*, *27*(4), 506–518. https://doi.org/10.1521/pedi_2012_26_049

Hartwright, C. E., Apperly, I. A., & Hansen, P. C. (2015). The special case of self-perspective inhibition in mental, but not non-mental, representation. *Neuropsychologia*, *67*, 183–192. https://doi.org/10.1016/j.neuropsychologia.2014.12.015

Hauser, T. U., Allen, M., Purg, N., Moutoussis, M., Rees, G., & Dolan, R. J. (2017). Noradrenaline blockade specifically enhances metacognitive performance. *ELife*, *6*, e24901. https://doi.org/10.7554/eLife.24901

Hauser, T. U., Allen, M., Rees, G., & Dolan, R. J. (2017). Metacognitive impairments extend perceptual decision making weaknesses in compulsivity. *Scientific Reports*, *7*(1), 6614. https://doi.org/10.1038/s41598-017-06116-z

Hauser, T. U., Moutoussis, M., Dayan, P., & Dolan, R. J. (2017). Increased decision thresholds trigger extended information gathering across the compulsivity spectrum. *Translational Psychiatry*, *7*(12), 1296. https://doi.org/10.1038/s41398-017-0040-3

Hauser, T. U., Moutoussis, M., Iannaccone, R., Brem, S., Walitza, S., Drechsler, R., Dayan, P., & Dolan, R. J. (2017). Increased decision thresholds enhance information gathering performance in juvenile Obsessive-Compulsive Disorder (OCD). *PLOS Computational Biology*, *13*(4), e1005440. https://doi.org/10.1371/journal.pcbi.1005440

Henseler, I., Krüger, S., Dechent, P., & Gruber, O. (2011). A gateway system in rostral PFC? Evidence from biasing attention to perceptual information and internal representations. *NeuroImage*, *56*(3), 1666–1676. https://doi.org/10.1016/j.neuroimage.2011.02.056

Herbert, B. M., Herbert, C., & Pollatos, O. (2011). On the Relationship Between Interoceptive Awareness and Alexithymia: Is Interoceptive Awareness Related to Emotional Awareness? *Journal of Personality*, *79*(5), 1149–1175. https://doi.org/10.1111/j.1467-6494.2011.00717.x

Herbert, B., & Pollatos, O. (2019). The relevance of interoception for eating behavior and eating disorders. In *The Interoceptive Mind* (pp. 165–186). Oxford University Press.

Herman, A. M., Rae, C. L., Critchley, H. D., & Duka, T. (2019). Interoceptive accuracy predicts nonplanning trait impulsivity. *Psychophysiology*, *56*(6), e13339. https://doi.org/10.1111/psyp.13339

Hermann, A., Schäfer, A., Walter, B., Stark, R., Vaitl, D., & Schienle, A. (2009). Emotion regulation in spider phobia: Role of the medial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, *4*(3), 257–267. https://doi.org/10.1093/scan/nsp013

Hezel, D. M., & McNally, R. J. (2014). Theory of Mind Impairments in Social Anxiety Disorder. *Behavior Therapy*, *45*(4), 530–540. https://doi.org/10.1016/j.beth.2014.02.010

Hirschfeld, R. M. A., Montgomery, S. A., Keller, M. B., Kasper, S., Schatzberg, A. F., Moller, H.-J., Healy, D., Baldwin, D., Humble, M., & Versiani, M. (2000). Social Functioning in Depression: A Review. *The Journal of Clinical Psychiatry*, *61*(4), 268–275.

Hjorth, S. (1993). Serotonin 5-HT1A autoreceptor blockade potentiates the ability of the 5-HT reuptake inhibitor citalopram to increase nerve terminal output of 5-HT in vivo: A microdialysis study. *Journal of Neurochemistry*, *60*(2), 776–779. https://doi.org/10.1111/j.1471-4159.1993.tb03217.x

Hjorth, S., Bengtsson, H. J., Kullberg, A., Carlzon, D., Peilot, H., & Auerbach, S. B. (2000). Serotonin autoreceptor function and antidepressant drug action. *Journal of Psychopharmacology*, *14*(2), 177–185. https://doi.org/10.1177/026988110001400208

Holt, C. A., & Laury, S. K. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, *92*(5), 1644–1655. https://doi.org/10.1257/000282802762024700

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

Horn, A. S. (1973). Structure-activity relations for the inhibition of catecholamine uptake into synaptosomes from noradrenaline and dopaminergic neurones in rat brain homogenates. *British Journal of Pharmacology*, *47*(2), 332–338.

Hughes, B. L., & Beer, J. S. (2012). Medial orbitofrontal cortex is associated with shifting decision thresholds in self-serving cognition. *NeuroImage*, *61*(4), 889–898. https://doi.org/10.1016/j.neuroimage.2012.03.011

Hutcherson, C. A., Plassmann, H., Gross, J. J., & Rangel, A. (2012). Cognitive Regulation during Decision Making Shifts Behavioral Control between Ventromedial and Dorsolateral Prefrontal Value Systems. *Journal of Neuroscience*, *32*(39), 13543–13554. https://doi.org/10.1523/JNEUROSCI.6387-11.2012

Hyttel, J., Bøgesø, K. P., Perregaard, J., & Sánchez, C. (1992). The pharmacological effect of citalopram resides in the (S)-(+)-enantiomer. *Journal of Neural Transmission / General Section JNT*, *88*(2), 157–160. https://doi.org/10.1007/BF01244820

Iigaya, K., Fonseca, M. S., Murakami, M., Mainen, Z. F., & Dayan, P. (2018). An effect of serotonergic stimulation on learning rates for rewards apparent after long intertrial intervals. *Nature Communications*, *9*(1), 2477. https://doi.org/10.1038/s41467-018-04840-2

Invernizzi, R., Belli, S., & Samanin, R. (1992). Citalopram's ability to increase the extracellular concentrations of serotonin in the dorsal raphe prevents the drug's effect in the frontal cortex. *Brain Research*, *584*(1), 322–324. https://doi.org/10.1016/0006-8993(92)90914-U

Invernizzi, R., Velasco, C., Bramante, M., Longo, A., & Samanin, R. (1997). Effect of 5-HT1A Receptor Antagonists on Citalopram-induced Increase in Extracellular Serotonin in the Frontal Cortex, Striatum and Dorsal Hippocampus. *Neuropharmacology*, *36*(4), 467–473. https://doi.org/10.1016/S0028-3908(97)00060-9

Ionta, S., Heydrich, L., Lenggenhager, B., Mouthon, M., Fornari, E., Chapuis, D., Gassert, R., & Blanke, O. (2011). Multisensory Mechanisms in Temporo-Parietal Cortex Support Self-Location and First-Person Perspective. *Neuron*, *70*(2), 363–374. https://doi.org/10.1016/j.neuron.2011.03.009

Jastorff, J., Huang, Y.-A., Giese, M. A., & Vandenbulcke, M. (2015). Common neural correlates of emotion perception in humans. *Human Brain Mapping*, *36*(10), 4184–4201. https://doi.org/10.1002/hbm.22910

Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-subject designs. *Psychological Methods*, *6*(2), 115–134.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263. https://doi.org/10.2307/1914185

Kandasamy, N., Garfinkel, S. N., Page, L., Hardy, B., Critchley, H. D., Gurnell, M., & Coates, J. M. (2016). Interoceptive Ability Predicts Survival on a London Trading Floor. *Scientific Reports*, *6*, 32986. https://doi.org/10.1038/srep32986

Kandel, E. R. (2013). *Principles of Neural Science* (5th ed.). McGraw-Hill.

Kanske, P., Böckler, A., Trautwein, F.-M., & Singer, T. (2015). Dissecting the social brain: Introducing the EmpaToM to reveal distinct neural networks and brain–behavior relations for empathy

and Theory of Mind. *NeuroImage*, *122*, 6–19. https://doi.org/10.1016/j.neuroimage.2015.07.082

Katanoda, K., Yoshikawa, K., & Sugishita, M. (2000). Neural substrates for the recognition of newly learned faces: A functional MRI study. *Neuropsychologia*, *38*(12), 1616–1625. https://doi.org/10.1016/s0028-3932(00)00069-5

Katkin, E. S., Morell, M. A., Goldband, S., Bernstein, G. L., & Wise, J. A. (1982). Individual Differences in Heartbeat Discrimination. *Psychophysiology*, *19*(2), 160–166. https://doi.org/10.1111/j.1469-8986.1982.tb02538.x

Kennedy, N., Foy, K., Sherazi, R., McDonough, M., & McKeon, P. (2007). Long-term social functioning after depression treated by psychiatrists: A review. *Bipolar Disorders*, *9*(1–2), 25–37. https://doi.org/10.1111/j.1399-5618.2007.00326.x

Kent, J. M., Coplan, J. D., & Gorman, J. M. (1998). Clinical utility of the selective serotonin reuptake inhibitors in the spectrum of anxiety. *Biological Psychiatry*, *44*(9), 812–824. https://doi.org/10.1016/S0006-3223(98)00210-8

Kerr, K. L., Moseman, S. E., Avery, J. A., Bodurka, J., Zucker, N. L., & Simmons, W. K. (2016). Altered Insula Activity during Visceral Interoception in Weight-Restored Patients with Anorexia Nervosa. *Neuropsychopharmacology*, *41*(2), 521–528. https://doi.org/10.1038/npp.2015.174

Kestemont, J., Ma, N., Baetens, K., Clément, N., Van Overwalle, F., & Vandekerckhove, M. (2015). Neural correlates of attributing causes to the self, another person and the situation. *Social Cognitive and Affective Neuroscience*, *10*(1), 114–121. https://doi.org/10.1093/scan/nsu030

Khalsa, S. S., & Feinstein, J. S. (2019). The somatic marker hypothesis of anxiety. In *The Interoceptive Mind* (pp. 144–164). Oxford University Press.

Khoury, N. M., Lutz, J., & Schuman-Olivier, Z. (2018). Interoception in Psychiatric Disorders: A Review of Randomized, Controlled Trials with Interoception-Based Interventions. *Harvard Review of Psychiatry*, *26*(5), 250–263. https://doi.org/10.1097/HRP.0000000000000170

Kiser, D., Steemers, B., Branchi, I., & Homberg, J. R. (2012). The reciprocal interaction between serotonin and social behaviour. *Neuroscience & Biobehavioral Reviews*, *36*(2), 786–798. https://doi.org/10.1016/j.neubiorev.2011.12.009

Klabunde, M., Acheson, D. T., Boutelle, K. N., Matthews, S. C., & Kaye, W. H. (2013). Interoceptive sensitivity deficits in women recovered from bulimia nervosa. *Eating Behaviors*, *14*(4), 488–492. https://doi.org/10.1016/j.eatbeh.2013.08.002

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1–16.

Klumpp, H., Fitzgerald, D. A., Cook, E., Shankman, S. A., Angstadt, M., & Phan, K. L. (2014). Serotonin transporter gene alters insula activity to threat in social anxiety disorder. *Neuroreport*, *25*(12), 926–931. https://doi.org/10.1097/WNR.0000000000000210

Knight, M. J., & Baune, B. T. (2019). The Direct and Indirect Relationship Between Social Cognition and Psychosocial Dysfunction in Major Depressive Disorder. *Frontiers in Psychiatry*, *10*. https://doi.org/10.3389/fpsyt.2019.00347

Knoll, J. F., & Hodapp, V. (1992). A comparison between two methods for assessing heartbeat perception. *Psychophysiology*, *29*(2), 218–222.

Knuth, D. E. (1969). Seminumerical algorithms. In *The Art of Computer Programming. 2. Reading, MA: Addison–Wesley.* (2nd ed., pp. 139–140). Reading, MA: Addison–Wesley.

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., & Glover, G. (2005). Distributed Neural Representation of Expected Value. *Journal of Neuroscience*, *25*(19), 4806–4812. https://doi.org/10.1523/JNEUROSCI.0642-05.2005

Koda, K., Ago, Y., Cong, Y., Kita, Y., Takuma, K., & Matsuda, T. (2010). Effects of acute and chronic administration of atomoxetine and methylphenidate on extracellular levels of noradrenaline, dopamine and serotonin in the prefrontal cortex and striatum of mice. *Journal of Neurochemistry*, *114*(1), 259–270. https://doi.org/10.1111/j.1471-4159.2010.06750.x

Koike, T., Tanabe, H. C., & Sadato, N. (2015). Hyperscanning neuroimaging technique to reveal the "two-in-one" system in social interactions. *Neuroscience Research*, *90*, 25–32. https://doi.org/10.1016/j.neures.2014.11.006

Koizumi, K., Terui, N., & Kollai, M. (1985). Effect of cardiac vagal and sympathetic nerve activity on heart rate in rhythmic fluctuations. *Journal of the Autonomic Nervous System*, *12*(2), 251–259. https://doi.org/10.1016/0165-1838(85)90065-7

Koot, S., Zoratto, F., Cassano, T., Colangeli, R., Laviola, G., van den Bos, R., & Adriani, W. (2012). Compromised decision-making and increased gambling proneness following dietary serotonin depletion in rats. *Neuropharmacology*, *62*(4), 1640–1650. https://doi.org/10.1016/J.NEUROPHARM.2011.11.002

Korn, C. W., Prehn, K., Park, S. Q., Walter, H., & Heekeren, H. R. (2012). Positively Biased Processing of Self-Relevant Social Feedback. *Journal of Neuroscience*, *32*(47), 16832–16844. https://doi.org/10.1523/JNEUROSCI.3016-12.2012

Köteles, F., & Doering, B. K. (2016). The many faces of somatosensory amplification: The relative contribution of body awareness, symptom labeling, and anxiety. *Journal of Health Psychology*, *21*(12), 2903–2911. https://doi.org/10.1177/1359105315588216

Kuehn, E., Mueller, K., Lohmann, G., & Schuetz-Bosbach, S. (2016). Interoceptive awareness changes the posterior insula functional connectivity profile. *Brain Structure and Function*, *221*(3), 1555–1571. https://doi.org/10.1007/s00429-015-0989-8

Kupfer, D. J., Frank, E., & Phillips, M. L. (2012). Major depressive disorder: New clinical, neurobiological, and treatment perspectives. *The Lancet*, *379*(9820), 1045–1055. https://doi.org/10.1016/S0140-6736(11)60602-8

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lane, R., Fink, G., Chau, P., & Dolan, R. (1997). Neural activation during selective attention to subjective emotional responses. *Neuroreport*, *8*(18), 3969–3972.

Lanzenberger, R. R., Mitterhauser, M., Spindelegger, C., Wadsak, W., Klein, N., Mien, L.-K., Holik, A., Attarbaschi, T., Mossaheb, N., Sacher, J., Geiss-Granadia, T., Kletter, K., Kasper, S., & Tauscher, J. (2007). Reduced Serotonin-1A Receptor Binding in Social Anxiety Disorder. *Biological Psychiatry*, *61*(9), 1081–1089. https://doi.org/10.1016/j.biopsych.2006.05.022

Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*(8), 365–373. https://doi.org/10.1016/j.tics.2011.05.009

Lenggenhager, B., Azevedo, R. T., Mancini, A., & Aglioti, S. M. (2013). Listening to your heart and feeling yourself: Effects of exposure to interoceptive signals during the ultimatum game.

Experimental Brain Research, *230*(2), 233–241. https://doi.org/10.1007/s00221-013-3647-5

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and Decision Making. *Annual Review of Psychology*, *66*(1), 799–823. https://doi.org/10.1146/annurev-psych-010213-115043

Lesch, K.-P., Bengel, D., Heils, A., Sabol, S. Z., Greenberg, B. D., Petri, S., Benjamin, J., Müller, C. R., Hamer, D. H., & Murphy, D. L. (1996). Association of Anxiety-Related Traits with a Polymorphism in the Serotonin Transporter Gene Regulatory Region. *Science*, *274*(5292), 1527–1531. https://doi.org/10.1126/science.274.5292.1527

Leveroni, C. L., Seidenberg, M., Mayer, A. R., Mead, L. A., Binder, J. R., & Rao, S. M. (2000). Neural Systems Underlying the Recognition of Familiar and Newly Learned Faces. *Journal of Neuroscience*, *20*(2), 878–886. https://doi.org/10.1523/JNEUROSCI.20-02-00878.2000

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology*, *22*(6), 1027–1038. https://doi.org/10.1016/j.conb.2012.06.001

Li, Y., Sescousse, G., Amiez, C., & Dreher, J.-C. (2015). Local Morphology Predicts Functional Organization of Experienced Value Signals in the Human Orbitofrontal Cortex. *Journal of Neuroscience*, *35*(4), 1648–1658. https://doi.org/10.1523/JNEUROSCI.3058-14.2015

Lichters, M., Brunnlieb, C., Nave, G., Sarstedt, M., & Vogt, B. (2016). The Influence of Serotonin Deficiency on Choice Deferral and the Compromise Effect. *Journal of Marketing Research*, *53*(2), 183–198. https://doi.org/10.1509/jmr.14.0482

Lieberman, M. D., Eisenberger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., & Way, B. M. (2007). Putting feelings into words: Affect labeling disrupts amygdala activity in response to affective stimuli. *Psychological Science*, *18*(5), 421–428. https://doi.org/10.1111/j.1467-9280.2007.01916.x

Linnér, L., Endersz, H., Öhman, D., Bengtsson, F., Schalling, M., & Svensson, T. H. (2001). Reboxetine Modulates the Firing Pattern of Dopamine Cells in the Ventral Tegmental Area and Selectively Increases Dopamine Availability in the Prefrontal Cortex. *Journal of Pharmacology and Experimental Therapeutics*, *297*(2), 540–546.

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Wheelwright, S. J., Sadek, S. A., Suckling, J., & Baron-Cohen, S. (2009). Shared Neural Circuits for Mentalizing about the Self and Others. *Journal of Cognitive Neuroscience*, *22*(7), 1623–1635. https://doi.org/10.1162/jocn.2009.21287

Long, A. B., Kuhn, C. M., & Platt, M. L. (2009). Serotonin shapes risky decision making in monkeys. *Social Cognitive and Affective Neuroscience*, *4*(4), 346–356. https://doi.org/10.1093/scan/nsp020

Longarzo, M., D'Olimpio, F., Chiavazzo, A., Santangelo, G., Trojano, L., & Grossi, D. (2015). The relationships between interoception and alexithymic trait. The Self-Awareness Questionnaire in healthy subjects. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01149

Lottem, E., Banerjee, D., Vertechi, P., Sarra, D., Lohuis, M. oude, & Mainen, Z. F. (2018). Activation of serotonin neurons promotes active persistence in a probabilistic foraging task. *Nature Communications*, *9*(1), 1000. https://doi.org/10.1038/s41467-018-03438-y

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

Lundberg, J., Christophersen, J. S., Petersen, K. B., Loft, H., Halldin, C., & Farde, L. (2007). PET measurement of serotonin transporter occupancy: A comparison of escitalopram and citalopram. *International Journal of Neuropsychopharmacology*, *10*(6), 777–785. https://doi.org/10.1017/S1461145706007486

Luscher, B., Shen, Q., & Sahir, N. (2011). The GABAergic deficit hypothesis of major depressive disorder. *Molecular Psychiatry*, *16*(4), 383–406. https://doi.org/10.1038/mp.2010.120

Mahy, C. E. V., Moses, L. J., & Pfeifer, J. H. (2014). How and where: Theory-of-mind in the brain. *Developmental Cognitive Neuroscience*, *9*, 68–81. https://doi.org/10.1016/j.dcn.2014.01.002

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

Marquand, A. F., De Simoni, S., O'Daly, O. G., Williams, S. C. R., Mourão-Miranda, J., & Mehta, M. A. (2011). Pattern classification of working memory networks reveals differential effects of methylphenidate, atomoxetine, and placebo in healthy volunteers. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *36*(6), 1237–1247. https://doi.org/10.1038/npp.2011.9

Mars, R. B., Neubert, F.-X., Noonan, M. P., Sallet, J., Toni, I., & Rushworth, M. F. S. (2012). On the relationship between the "default mode network" and the "social brain". *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00189

Marsh, D. M., Dougherty, D. M., Moeller, F. G., Swann, A. C., & Spiga, R. (2002). Laboratory-Measured Aggressive Behavior of Women: Acute Tryptophan Depletion and Augmentation. *Neuropsychopharmacology*, *26*(5), 660–671. https://doi.org/10.1016/S0893-133X(01)00369-4

Massar, S. A. A., Libedinsky, C., Weiyan, C., Huettel, S. A., & Chee, M. W. L. (2015). Separate and overlapping brain areas encode subjective value during delay and effort discounting. *NeuroImage*, *120*, 104–113. https://doi.org/10.1016/j.neuroimage.2015.06.080

Matthews, S. C., Simmons, A. N., Strigo, I. A., Arce, E., Stein, M. B., & Paulus, M. P. (2010). Escitalopram attenuates posterior cingulate activity during self-evaluation in healthy volunteers. *Psychiatry Research*, *182*(2), 81–87. https://doi.org/10.1016/j.pscychresns.2010.02.003

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Mazoyer, B., Zago, L., Mellet, E., Bricogne, S., Etard, O., Houdé, O., Crivello, F., Joliot, M., Petit, L., & Tzourio-Mazoyer, N. (2001). Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain Research Bulletin*, *54*(3), 287–298. https://doi.org/10.1016/S0361-9230(00)00437-8

Meffert, H., Brislin, S. J., White, S. F., & Blair, J. R. (2015). Prediction errors to emotional expressions: The roles of the amygdala in social referencing. *Social Cognitive and Affective Neuroscience*, *10*(4), 537–544. https://doi.org/10.1093/scan/nsu085

Meindl, T., Teipel, S., Elmouden, R., Mueller, S., Koch, W., Dietrich, O., Coates, U., Reiser, M., & Glaser, C. (2010). Test–retest reproducibility of the default-mode network in healthy individuals. *Human Brain Mapping*, *31*(2), 237–246. https://doi.org/10.1002/hbm.20860

Meyer, J. H., Wilson, A. A., Sagrati, S., Hussey, D., Carella, A., Potter, W. Z., Ginovart, N., Spencer, E. P., Cheok, A., & Houle, S. (2004). Serotonin transporter occupancy of five selective serotonin reuptake inhibitors at different doses: An [11C]DASB positron emission tomography study. *The American Journal of Psychiatry*, *161*(5), 826–835. https://doi.org/10.1176/appi.ajp.161.5.826

Michelson, D., Adler, L., Spencer, T., Reimherr, F. W., West, S. A., Allen, A. J., Kelsey, D., Wernicke, J., Dietrich, A., & Milton, D. (2003). Atomoxetine in adults with ADHD: Two randomized, placebo-controlled studies. *Biological Psychiatry*, *53*(2), 112–120. https://doi.org/10.1016/S0006-3223(02)01671-2

Milne, R. J., & Goa, K. L. (1991). Citalopram. A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic potential in depressive illness. *Drugs*, *41*(3), 450–477.

Modinos, G., Ormel, J., & Aleman, A. (2009). Activation of Anterior Insula during Self-Reflection. *PLOS ONE*, *4*(2), e4618. https://doi.org/10.1371/journal.pone.0004618

Morel, N., Villain, N., Rauchs, G., Gaubert, M., Piolino, P., Landeau, B., Mézenge, F., Desgranges, B., Eustache, F., & Chételat, G. (2014). Brain Activity and Functional Coupling Changes Associated with Self-Reference Effect during Both Encoding and Retrieval. *PLOS ONE*, *9*(3), e90488. https://doi.org/10.1371/journal.pone.0090488

Moret, C., & Briley, M. (1996). Effects of acute and repeated administration of citalopram on extracellular levels of serotonin in rat brain. *European Journal of Pharmacology*, *295*(2), 189–197. https://doi.org/10.1016/0014-2999(95)00646-X

Morón, J. A., Brockington, A., Wise, R. A., Rocha, B. A., & Hope, B. T. (2002). Dopamine Uptake through the Norepinephrine Transporter in Brain Regions with Low Levels of the Dopamine Transporter: Evidence from Knock-Out Mouse Lines. *Journal of Neuroscience*, *22*(2), 389–395. https://doi.org/10.1523/JNEUROSCI.22-02-00389.2002

Moskowitz, D. S., Pinard, G., Zuroff, D. C., Annable, L., & Young, S. N. (2001). The Effect of Tryptophan on Social Interaction in Everyday Life: A Placebo-Controlled Study.

Neuropsychopharmacology, *25*(2), 277–289. https://doi.org/10.1016/S0893-133X(01)00219-6

Mueller, E. M., Evers, E. A., Wacker, J., & Van Der Veen, F. (2012). Acute tryptophan depletion attenuates brain-heart coupling following external feedback. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00077

Mukai, I., Kim, D., Fukunaga, M., Japee, S., Marrett, S., & Ungerleider, L. G. (2007). Activations in visual and attention-related areas predict and correlate with the degree of perceptual learning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*(42), 11401–11411. https://doi.org/10.1523/JNEUROSCI.3002-07.2007

Mukherjee, S., Sahay, A., Chandrasekhar Pammi, V. S., & Srinivasan, N. (2017). Is loss-aversion magnitude-dependent? Measuring prospective affective judgments regarding gains and losses. In *Judgment and Decision Making* (Vol. 12, Issue 1).

Mul, C., Stagg, S. D., Herbelin, B., & Aspell, J. E. (2018). The Feeling of Me Feeling for You: Interoception, Alexithymia and Empathy in Autism. *Journal of Autism and Developmental Disorders*, *48*(9), 2953–2967. https://doi.org/10.1007/s10803-018-3564-3

Mullett, T. L., & Tunney, R. J. (2013). Value representations by rank order in a distributed network of varying context dependency. *Brain and Cognition*, *82*(1), 76–83. https://doi.org/10.1016/j.bandc.2013.02.010

Murphy, J., Brewer, R., Hobson, H., Catmur, C., & Bird, G. (2018). Is alexithymia characterised by impaired interoception? Further evidence, the importance of control variables, and the problems with the Heartbeat Counting Task. *Biological Psychology*, *136*, 189–197. https://doi.org/10.1016/j.biopsycho.2018.05.010

Murphy, S. E., Longhitano, C., Ayres, R. E., Cowen, P. J., & Harmer, C. J. (2006). Tryptophan supplementation induces a positive bias in the processing of emotional material in healthy female volunteers. *Psychopharmacology*, *187*(1), 121–130. https://doi.org/10.1007/s00213-006-0401-8

Murphy, S. E., Longhitano, C., Ayres, R. E., Cowen, P. J., Harmer, C. J., & Rogers, R. D. (2009). The Role of Serotonin in Nonnormative Risky Choice: The Effects of Tryptophan Supplements on the "Reflection Effect" in Healthy Adult Volunteers. *Journal of Cognitive Neuroscience*, *21*(9), 1709–1719. https://doi.org/10.1162/jocn.2009.21122

Murphy, S. E., Norbury, R., O'Sullivan, U., Cowen, P. J., & Harmer, C. J. (2009). Effect of a single dose of citalopram on amygdala response to emotional faces. *The British Journal of Psychiatry*, *194*(6), 535–540. https://doi.org/10.1192/bjp.bp.108.056093

Nakamura, K. (2013). The role of the dorsal raphé nucleus in reward-seeking behavior. *Frontiers in Integrative Neuroscience*, *7*. https://doi.org/10.3389/fnint.2013.00060

Nandam, L. S., Hester, R., & Bellgrove, M. A. (2014). Dissociable and common effects of methylphenidate, atomoxetine and citalopram on response inhibition neural networks. *Neuropsychologia*, *56*, 263–270. https://doi.org/10.1016/j.neuropsychologia.2014.01.023

Nandam, L. S., Hester, R., Wagner, J., Cummins, T. D. R., Garner, K., Dean, A. J., Kim, B. N., Nathan, P. J., Mattingley, J. B., & Bellgrove, M. A. (2011). Methylphenidate But Not Atomoxetine or Citalopram Modulates Inhibitory Control and Response Time Variability. *Biological Psychiatry*, *69*(9), 902–904. https://doi.org/10.1016/j.biopsych.2010.11.014

Nejad, A. B., Fossati, P., & Lemogne, C. (2013). Self-Referential Processing, Rumination, and Cortical Midline Structures in Major Depression. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00666

NICE. (2019a). *Atomoxetine—National Institute for Health and Care Excellence / British National Formulary* [Clinical Guidelines]. https://bnf.nice.org.uk/drug/atomoxetine.html

NICE. (2019b). *Citalopram—National Institute for Health and Care Excellence / British National Formulary* [Clinical Guidelines]. https://bnf.nice.org.uk/drug/citalopram.html

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J.-B. (2005). Valid conjunction inference with the minimum statistic. *NeuroImage*, *25*(3), 653–660. https://doi.org/10.1016/j.neuroimage.2004.12.005

Nocedal, J., & Wright, S. (2006). Sequential Quadratic Programming. In *Numerical Optimization* (2nd ed., pp. 526–572). Springer-Verlag. https://www.springer.com/gp/book/9780387303031

Nord, M., Finnema, S. J., Halldin, C., & Farde, L. (2013). Effect of a single dose of escitalopram on serotonin concentration in the non-human and human primate brain. *International Journal of Neuropsychopharmacology*, *16*(7), 1577–1586. https://doi.org/10.1017/S1461145712001617

Northoff, G., & Bermpohl, F. (2004). Cortical midline structures and the self. *Trends in Cognitive Sciences*, *8*(3), 102–107. https://doi.org/10.1016/j.tics.2004.01.004

Nutt, D. J. (2005). Overview of Diagnosis and Drug Treatments of Anxiety Disorders. *CNS Spectrums*, *10*(1), 49–56. https://doi.org/10.1017/S1092852900009901

Ochsner, K. N., Beer, J. S., Robertson, E. R., Cooper, J. C., Gabrieli, J. D. E., Kihsltrom, J. F., & D'Esposito, M. (2005). The neural correlates of direct and reflected self-knowledge. *NeuroImage*, *28*(4), 797–814. https://doi.org/10.1016/j.neuroimage.2005.06.069

Ochsner, K. N., Knierim, K., Ludlow, D. H., Hanelin, J., Ramachandran, T., Glover, G., & Mackey, S. C. (2004). Reflecting upon feelings: An fMRI study of neural systems supporting the attribution of emotion to self and other. *Journal of Cognitive Neuroscience*, *16*(10), 1746–1772. https://doi.org/10.1162/0898929042947829

Ohno, Y. (2010). New insight into the therapeutic role of 5-HT1A receptors in central nervous system disorders. *Central Nervous System Agents in Medicinal Chemistry*, *10*(2), 148–157.

O'Reilly, J. X., Woolrich, M. W., Behrens, T. E. J., Smith, S. M., & Johansen-Berg, H. (2012). Tools of the trade: Psychophysiological interactions and functional connectivity. *Social Cognitive and Affective Neuroscience*, *7*(5), 604–609. https://doi.org/10.1093/scan/nss055

Owens, M. J., & Nemeroff, C. B. (1994). Role of serotonin in the pathophysiology of depression: Focus on the serotonin transporter. *Clinical Chemistry*, *40*(2), 288–295.

Owens, Michael J, Knight, D. L., & Nemeroff, C. B. (2001). Second-generation SSRIs: Human monoamine transporter binding profile of escitalopram and R-fluoxetine. *Biological Psychiatry*, *50*(5), 345–350. https://doi.org/10.1016/S0006-3223(01)01145-3

Park, H.-D., & Tallon-Baudry, C. (2014). The neural subjective frame: From bodily signals to perceptual consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1641), 20130208–20130208. https://doi.org/10.1098/rstb.2013.0208

Park, Hyeong-Dong, Correia, S., Ducorps, A., & Tallon-Baudry, C. (2014). Spontaneous fluctuations in neural responses to heartbeats predict visual detection. *Nature Neuroscience*, *17*(4), 612–618. https://doi.org/10.1038/nn.3671

Park, S. B., Coull, J. T., McShane, R. H., Young, A. H., Sahakian, B. J., Robbins, T. W., & Cowen, P. J. (1994). Tryptophan depletion in normal volunteers produces selective impairments in learning and memory. *Neuropharmacology*, *33*(3), 575–588. https://doi.org/10.1016/0028-3908(94)90089-2

Passow, S., Specht, K., Adamsen, T. C., Biermann, M., Brekke, N., Craven, A. R., Ersland, L., Grüner, R., Kleven-Madsen, N., Kvernenes, O.-H., Schwarzlmüller, T., Olesen, R. A., & Hugdahl, K. (2015). Default-mode network functional connectivity is closely related to metabolic activity. *Human Brain Mapping*, *36*(6), 2027–2038. https://doi.org/10.1002/hbm.22753

Paulus, M. P., Feinstein, J. S., & Khalsa, S. S. (2019). An Active Inference Approach to Interoceptive Psychopathology. *Annual Review of Clinical Psychology*, *15*(1), 97–122. https://doi.org/10.1146/annurev-clinpsy-050718-095617

Paulus, M. P., & Stein, M. B. (2010). Interoception in anxiety and depression. *Brain Structure and Function*, *214*(5), 451–463. https://doi.org/10.1007/s00429-010-0258-9

Payne, R. A., Symeonides, C. N., Webb, D. J., & Maxwell, S. R. J. (2006). Pulse transit time measured from the ECG: An unreliable marker of beat-to-beat blood pressure. *Journal of Applied Physiology (Bethesda, Md.: 1985)*, *100*(1), 136–141. https://doi.org/10.1152/japplphysiol.00657.2005

Pazos, A., Probst, A., & Palacios, J. M. (1987a). Serotonin receptors in the human brain—III. Autoradiographic mapping of serotonin-1 receptors. *Neuroscience*, *21*(1), 97–122. https://doi.org/10.1016/0306-4522(87)90326-5

Pazos, A., Probst, A., & Palacios, J. M. (1987b). Serotonin receptors in the human brain—IV. Autoradiographic mapping of serotonin-2 receptors. *Neuroscience*, *21*(1), 123–139. https://doi.org/10.1016/0306-4522(87)90327-7

Peyron, C., Petit, J.-M., Rampon, C., Jouvet, M., & Luppi, P.-H. (1997). Forebrain afferents to the rat dorsal raphe nucleus demonstrated by retrograde and anterograde tracing methods. *Neuroscience*, *82*(2), 443–468. https://doi.org/10.1016/S0306-4522(97)00268-6

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional Neuroanatomy of Emotion: A Meta-Analysis of Emotion Activation Studies in PET and fMRI. *NeuroImage*, *16*(2), 331–348. https://doi.org/10.1006/nimg.2002.1087

Phillips, M. L., Drevets, W. C., Rauch, S. L., & Lane, R. (2003). Neurobiology of emotion perception I: The neural basis of normal emotion perception. *Biological Psychiatry*, *54*(5), 504–514. https://doi.org/10.1016/S0006-3223(03)00168-9

Piech, R. M., Strelchuk, D., Knights, J., Hjälmheden, J. V., Olofsson, J. K., & Aspell, J. E. (2017). People with higher interoceptive sensitivity are more altruistic, but improving interoception does

not increase altruism. *Scientific Reports*, *7*(1), 1–5. https://doi.org/10.1038/s41598-017-14318-8

Pohl, R., Yeragani, V. K., Balon, R., Rainey, J. M., Lycaki, H., Ortiz, A., Berchou, R., & Weinberg, P. (1988). Isoproterenol-induced panic attacks. *Biological Psychiatry*, *24*(8), 891–902. https://doi.org/10.1016/0006-3223(88)90224-7

Pollatos, O., Schandry, R., Auer, D. P., & Kaufmann, C. (2007). Brain structures mediating cardiovascular arousal and interoceptive awareness. *Brain Research*, *1141*, 178–187. https://doi.org/10.1016/j.brainres.2007.01.026

Pollatos, O., Traut-Mattausch, E., & Schandry, R. (2009). Differential effects of anxiety and depression on interoceptive accuracy. *Depression and Anxiety*, *26*(2), 167–173. https://doi.org/10.1002/da.20504

Puig, M. V., & Miller, E. K. (2012). The Role of Prefrontal Dopamine D1 Receptors in the Neural Mechanisms of Associative Learning. *Neuron*, *74*(5), 874–886. https://doi.org/10.1016/j.neuron.2012.04.018

Qin, P., & Northoff, G. (2011). How is our self related to midline regions and the default-mode network? *NeuroImage*, *57*(3), 1221–1233. https://doi.org/10.1016/j.neuroimage.2011.05.028

Quadt, L., Critchley, H. D., & Garfinkel, S. N. (2018). The neurobiology of interoception in health and disease. *Annals of the New York Academy of Sciences*, *1428*(1), 112–128. https://doi.org/10.1111/nyas.13915

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rae, C. L., Nombela, C., Rodríguez, P. V., Ye, Z., Hughes, L. E., Jones, P. S., Ham, T., Rittman, T., Coyle-Gilchrist, I., Regenthal, R., Sahakian, B. J., Barker, R. A., Robbins, T. W., & Rowe, J. B. (2016). Atomoxetine restores the response inhibition network in Parkinson's disease. *Brain*, *139*(8), 2235–2248. https://doi.org/10.1093/brain/aww138

Raiteri, M., Del Carmine, R., Bertollini, A., & Levi, G. (1977). Effect of sympathomimetic amines on the synaptosomal transport of noradrenaline, dopamine and 5-hydroxytryptamine. *European Journal of Pharmacology*, *41*(2), 133–143. https://doi.org/10.1016/0014-2999(77)90202-3

Rapoport, A., & Amaldoss, W. (2000). Mixed strategies and iterative elimination of strongly dominated strategies: An experimental investigation of states of knowledge. *Journal of Economic Behavior & Organization*, *42*(4), 483–521. https://doi.org/10.1016/S0167-2681(00)00101-3

Rassovsky, Y., & Kushner, M. G. (2003). Carbon dioxide in the study of panic disorder: Issues of definition, methodology, and outcome. *Journal of Anxiety Disorders*, *17*(1), 1–32. https://doi.org/10.1016/S0887-6185(02)00181-0

Ray, R. S., Corcoran, A. E., Brust, R. D., Kim, J. C., Richerson, G. B., Nattie, E., & Dymecki, S. M. (2011). Impaired Respiratory and Body Temperature Control Upon Acute Serotonergic Neuron Inhibition. *Science*, *333*(6042), 637–642. https://doi.org/10.1126/science.1205295

Reiman, E. M., Lane, R. D., Ahern, G. L., Schwartz, G. E., Davidson, R. J., Friston, K. J., Yun, L. S., & Chen, K. (1997). Neuroanatomical correlates of externally and internally generated human emotion. *The American Journal of Psychiatry*, *154*(7), 918–925. https://doi.org/10.1176/ajp.154.7.918

Reist, C., Mazzanti, C., Vu, R., Tran, D., & Goldman, D. (2001). Serotonin transporter promoter polymorphism is associated with attenuated prolactin response to fenfluramine. *American Journal of Medical Genetics*, *105*(4), 363–368. https://doi.org/10.1002/ajmg.1360

Richelson, E. (1994). Pharmacology of Antidepressants—Characteristics of the Ideal Drug. *Mayo Clinic Proceedings*, *69*(11), 1069–1081. https://doi.org/10.1016/S0025-6196(12)61375-5

Romero, L., Bel, N., Artigas, F., de Montigny, C., & Blier, P. (1996). Effect of pindolol on the function of pre- and postsynaptic 5-HT1A receptors: In vivo microdialysis and electrophysiological studies in the rat brain. *Neuropsychopharmacology*, *15*(4), 349–360. https://doi.org/10.1016/0893-133X(95)00240-E

Rouault, M., Seow, T., Gillan, C. M., & Fleming, S. M. (2018). Psychiatric Symptom Dimensions Are Associated With Dissociable Shifts in Metacognition but Not Task Performance. *Biological Psychiatry*, *84*(6), 443–451. https://doi.org/10.1016/j.biopsych.2017.12.017

Ruby, P., & Decety, J. (2003). What you believe versus what you think they believe: A neuroimaging study of conceptual perspective-taking. *European Journal of Neuroscience*, *17*(11), 2475–2480. https://doi.org/10.1046/j.1460-9568.2003.02673.x

Ryan, J. P., Sheu, L. K., & Gianaros, P. J. (2011). Resting state functional connectivity within the cingulate cortex jointly predicts agreeableness and stressor-evoked cardiovascular reactivity. *NeuroImage*, *55*(1), 363–370. https://doi.org/10.1016/j.neuroimage.2010.11.064

Sacchetti, G., Bernini, M., Bianchetti, A., Parini, S., Invernizzi, R. W., & Samanin, R. (1999). Studies on the acute and chronic effects of reboxetine on extracellular noradrenaline and other monoamines in the rat brain. *British Journal of Pharmacology*, *128*(6), 1332–1338. https://doi.org/10.1038/sj.bjp.0702926

Sáez, I., Zhu, L., Set, E., Kayser, A., & Hsu, M. (2015). Dopamine modulates egalitarian behavior in humans. *Current Biology : CB*, *25*(7), 912–919. https://doi.org/10.1016/j.cub.2015.01.071

Salmon, T. C. (2001). An Evaluation of Econometric Models of Adaptive Learning. *Econometrica*, *69*(6), 1597–1628. JSTOR.

Sánchez, C., Bøgesø, K. P., Ebert, B., Reines, E. H., & Braestrup, C. (2004). Escitalopram versus citalopram: The surprising role of the R-enantiomer. *Psychopharmacology*, *174*(2), 163–176. https://doi.org/10.1007/s00213-004-1865-z

Santangelo, A. M., Sawiak, S. J., Fryer, T., Hong, Y., Shiba, Y., Clarke, H. F., Riss, P. J., Ferrari, V., Tait, R., Suckling, J., Aigbirhio, F. I., & Roberts, A. C. (2019). Insula serotonin 2A receptor binding and gene expression contribute to serotonin transporter polymorphism anxious phenotype in primates. *Proceedings of the National Academy of Sciences*, *116*(29), 14761–14768. https://doi.org/10.1073/pnas.1902087116

Sauer, J.-M., Ring, B. J., & Witcher, J. W. (2005). Clinical Pharmacokinetics of Atomoxetine. *Clinical Pharmacokinetics*, *44*(6), 571–590. https://doi.org/10.2165/00003088-200544060-00002

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in “theory of mind”. *NeuroImage*, *19*(4), 1835–1842. https://doi.org/10.1016/S1053-8119(03)00230-1

Schandry, R. (1981). Heart Beat Perception and Emotional Experience. *Psychophysiology*, *18*(4), 483–488. https://doi.org/10.1111/j.1469-8986.1981.tb02486.x

Scharinger, C., Rabl, U., Kasess, C. H., Meyer, B. M., Hofmaier, T., Diers, K., Bartova, L., Pail, G., Huf, W., Uzelac, Z., Hartinger, B., Kalcher, K., Perkmann, T., Haslacher, H., Meyer-Lindenberg, A., Kasper, S., Freissmuth, M., Windischberger, C., Willeit, M., … Pezawas, L. (2014). Platelet

Serotonin Transporter Function Predicts Default-Mode Network Activity. *PLOS ONE*, *9*(3), e92543. https://doi.org/10.1371/journal.pone.0092543

Scheibner, H. J., Bogler, C., Gleich, T., Haynes, J.-D., & Bermpohl, F. (2017). Internal and external attention and the default mode network. *NeuroImage*, *148*, 381–389. https://doi.org/10.1016/j.neuroimage.2017.01.044

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *The Behavioral and Brain Sciences*, *36*(4), 393–414. https://doi.org/10.1017/S0140525X12000660

Schinka, J. A., Busch, R. M., & Robichaux-Keene, N. (2004). A meta-analysis of the association between the serotonin transporter gene polymorphism (5-HTTLPR) and trait anxiety. *Molecular Psychiatry*, *9*(2), 197–202. https://doi.org/10.1038/sj.mp.4001405

Schlaffke, L., Lissek, S., Lenz, M., Juckel, G., Schultz, T., Tegenthoff, M., Schmidt-Wilcke, T., & Brüne, M. (2015). Shared and nonshared neural networks of cognitive and affective theory-of-mind: A neuroimaging study using cartoon picture stories. *Human Brain Mapping*, *36*(1), 29–39. https://doi.org/10.1002/hbm.22610

Schulz, S. M. (2016). Neural correlates of heart-focused interoception: A functional magnetic resonance imaging meta-analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160018. https://doi.org/10.1098/rstb.2016.0018

Schurz, M., Kronbichler, M., Weissengruber, S., Surtees, A., Samson, D., & Perner, J. (2015). Clarifying the role of theory of mind areas during visual perspective taking: Issues of spontaneity and domain-specificity. *NeuroImage*, *117*, 386–396. https://doi.org/10.1016/j.neuroimage.2015.04.031

Seghier, M. L. (2013). The Angular Gyrus: Multiple Functions and Multiple Subdivisions. *The Neuroscientist*, *19*(1), 43–61. https://doi.org/10.1177/1073858412440596

Selvaraj, S., Walker, C., Arnone, D., Cao, B., Faulkner, P., Cowen, P. J., Roiser, J. P., & Howes, O. (2018). Effect of Citalopram on Emotion Processing in Humans: A Combined 5-HT_{1A} [^{11}C]CUMI-101 PET and Functional MRI Study. *Neuropsychopharmacology*, *43*(3), 655–664. https://doi.org/10.1038/npp.2017.166

Sen, S., Burmeister, M., & Ghosh, D. (2004). Meta-analysis of the association between a serotonin transporter promoter polymorphism (5-HTTLPR) and anxiety-related personality traits.

American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, *127B*(1), 85–89. https://doi.org/10.1002/ajmg.b.20158

Sestieri, C., Di Matteo, R., Ferretti, A., Del Gratta, C., Caulo, M., Tartaro, A., Olivetti Belardinelli, M., & Romani, G. L. (2006). 'What' versus 'where' in the audiovisual domain: An fMRI study. *NeuroImage*, *33*(2), 672–680. https://doi.org/10.1016/j.neuroimage.2006.06.045

Set, E., Saez, I., Zhu, L., Houser, D. E., Myung, N., Zhong, S., Ebstein, R. P., Chew, S. H., & Hsu, M. (2014). Dissociable contribution of prefrontal and striatal dopaminergic genes to learning in economic games. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(26), 9615–9620. https://doi.org/10.1073/pnas.1316259111

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

Seth Anil K., & Friston Karl J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160007. https://doi.org/10.1098/rstb.2016.0007

Shah, P., Catmur, C., & Bird, G. (2017). From heart to mind: Linking interoception, emotion, and theory of mind. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *93*, 220–223. https://doi.org/10.1016/j.cortex.2017.02.010

Shah, P., Hall, R., Catmur, C., & Bird, G. (2016). Alexithymia, not autism, is associated with impaired interoception. *Cortex*, *81*, 215–220. https://doi.org/10.1016/j.cortex.2016.03.021

Shane, M. S., Stevens, M. C., Harenski, C. L., & Kiehl, K. A. (2009). Double dissociation between perspective-taking and empathic-concern as predictors of hemodynamic response to another's mistakes. *Social Cognitive and Affective Neuroscience*, *4*(2), 111–118. https://doi.org/10.1093/scan/nsn043

Sheehan, D. V, Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry*, *59 Suppl 20*, 22-33;quiz 34-57.

Siegel, J. Z., & Crockett, M. J. (2013). How serotonin shapes moral judgment and behavior. *Annals of the New York Academy of Sciences*, *1299*(1), 42–51. https://doi.org/10.1111/nyas.12229

Silvers, J. A., Wager, T. D., Weber, J., & Ochsner, K. N. (2015). The neural bases of uninstructed negative emotion modulation. *Social Cognitive and Affective Neuroscience*, *10*(1), 10–18. https://doi.org/10.1093/scan/nsu016

Silvers, J. A., Weber, J., Wager, T. D., & Ochsner, K. N. (2015). Bad and worse: Neural systems underlying reappraisal of high- and low-intensity negative emotions. *Social Cognitive and Affective Neuroscience*, *10*(2), 172–179. https://doi.org/10.1093/scan/nsu043

Simmons, W. K., Avery, J. A., Barcalow, J. C., Bodurka, J., Drevets, W. C., & Bellgowan, P. (2013). Keeping the body in mind: Insula functional organization and functional connectivity integrate interoceptive, exteroceptive, and emotional awareness. *Human Brain Mapping*, *34*(11), 2944–2958. https://doi.org/10.1002/hbm.22113

Simonsen, A., Scheel-Krüger, J., Jensen, M., Roepstorff, A., Møller, A., Frith, C. D., & Campbell-Meiklejohn, D. (2014). Serotoninergic effects on judgments and social learning of trustworthiness. *Psychopharmacology*, *231*(14), 2759–2769. https://doi.org/10.1007/s00213-014-3444-2

Skandali, N., Rowe, J. B., Voon, V., Deakin, J. B., Cardinal, R. N., Cormack, F., Passamonti, L., Bevan-Jones, W. R., Regenthal, R., Chamberlain, S. R., Robbins, T. W., & Sahakian, B. J. (2018a). Dissociable effects of acute SSRI (escitalopram) on executive, learning and emotional functions in healthy humans. *Neuropsychopharmacology 2018 43:13*, *43*(13), 2645. https://doi.org/10.1038/s41386-018-0229-z

Skandali, N., Rowe, J. B., Voon, V., Deakin, J. B., Cardinal, R. N., Cormack, F., Passamonti, L., Bevan-Jones, W. R., Regenthal, R., Chamberlain, S. R., Robbins, T. W., & Sahakian, B. J. (2018b). Dissociable effects of acute SSRI (escitalopram) on executive, learning and emotional functions in healthy humans. *Neuropsychopharmacology*, *43*(13), 2645. https://doi.org/10.1038/s41386-018-0229-z

Song, C., Kanai, R., Fleming, S. M., Weil, R. S., Schwarzkopf, D. S., & Rees, G. (2011). Relating inter-individual differences in metacognitive performance on different perceptual tasks. *Consciousness and Cognition*, *20*(4), 1787–1792. https://doi.org/10.1016/j.concog.2010.12.011

Spengler, S., Cramon, D. Y. von, & Brass, M. (2009). Control of shared representations relies on key processes involved in mental state attribution. *Human Brain Mapping*, *30*(11), 3704–3718. https://doi.org/10.1002/hbm.20800

Spreng, R. N., & Grady, C. L. (2009). Patterns of Brain Activity Supporting Autobiographical Memory, Prospection, and Theory of Mind, and Their Relationship to the Default Mode Network. *Journal of Cognitive Neuroscience*, *22*(6), 1112–1123. https://doi.org/10.1162/jocn.2009.21282

Sprouse, J. S., & Aghajanian, G. K. (1987). Electrophysiological responses of serotoninergic dorsal raphe neurons to 5-HT1A and 5-HT1B agonists. *Synapse*, *1*(1), 3–9. https://doi.org/10.1002/syn.890010103

Steenen, S. A., van Wijk, A. J., van der Heijden, G. J., van Westrhenen, R., de Lange, J., & de Jongh, A. (2016). Propranolol for the treatment of anxiety disorders: Systematic review and meta-analysis. *Journal of Psychopharmacology (Oxford, England)*, *30*(2), 128–139. https://doi.org/10.1177/0269881115612236

Stoodley, C. J., & Schmahmann, J. D. (2009). Functional topography in the human cerebellum: A meta-analysis of neuroimaging studies. *NeuroImage*, *44*(2), 489–501. https://doi.org/10.1016/j.neuroimage.2008.08.039

Strigo, I. A., & Craig, A. D. (2016). Interoception, homeostatic emotions and sympathovagal balance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160010. https://doi.org/10.1098/rstb.2016.0010

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT Press. http://site.ebrary.com/id/10225275

Swanson, C. J., Perry, K. W., Koch-Krueger, S., Katner, J., Svensson, K. A., & Bymaster, F. P. (2006). Effect of the attention deficit/hyperactivity disorder drug atomoxetine on extracellular concentrations of norepinephrine and dopamine in several brain regions of the rat. *Neuropharmacology*, *50*(6), 755–760. https://doi.org/10.1016/j.neuropharm.2005.11.022

Székely, M. (2000). The vagus nerve in thermoregulation and energy metabolism. *Autonomic Neuroscience: Basic & Clinical*, *85*(1–3), 26–38. https://doi.org/10.1016/S1566-0702(00)00217-4

Tammimaki, A., Aonurm-Helm, A., Kaenmaki, M., & Mannisto, P. T. (2016). Elimination of extracellular dopamine in the medial prefrontal cortex of conscious mice analysed using selective enzyme and uptake inhibitors. *Journal of Physiology and Pharmacology: An Official Journal of the Polish Physiological Society*, *67*(2), 301–309.

Tavares, J. V. T., Clark, L., Cannon, D. M., Erickson, K., Drevets, W. C., & Sahakian, B. J. (2007). Distinct Profiles of Neurocognitive Function in Unmedicated Unipolar Depression and Bipolar II Depression. *Biological Psychiatry*, *62*(8), 917–924. https://doi.org/10.1016/J.BIOPSYCH.2007.05.034

Taylor, M. J., Freemantle, N., Geddes, J. R., & Bhagwagar, Z. (2006). Early Onset of Selective Serotonin Reuptake Inhibitor Antidepressant Action: Systematic Review and Meta-analysis. *Archives of General Psychiatry*, *63*(11), 1217–1223. https://doi.org/10.1001/archpsyc.63.11.1217

Terasawa, Y., Fukushima, H., & Umeda, S. (2013). How does interoceptive awareness interact with the subjective experience of emotion? An fMRI study. *Human Brain Mapping*, *34*(3), 598–612. https://doi.org/10.1002/hbm.21458

Terasawa, Y., Moriguchi, Y., Tochizawa, S., & Umeda, S. (2014). Interoceptive sensitivity predicts sensitivity to the emotions of others. *Cognition and Emotion*, *28*(8), 1435–1448. https://doi.org/10.1080/02699931.2014.888988

Terasawa, Y., Shibata, M., Moriguchi, Y., & Umeda, S. (2013). Anterior insular cortex mediates bodily sensibility and social anxiety. *Social Cognitive and Affective Neuroscience*, *8*(3), 259–266. https://doi.org/10.1093/scan/nss108

Tsakiris, M. (2017). The multisensory basis of the self: From body to identity to others. *Quarterly Journal of Experimental Psychology*, *70*(4), 597–609. https://doi.org/10.1080/17470218.2016.1181768

Tsakiris, M., & Critchley, H. (2016). Interoception beyond homeostasis: Affect, cognition and mental health. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160002. https://doi.org/10.1098/rstb.2016.0002

Tse, W. S., & Bond, A. J. (2002). Serotonergic intervention affects both social dominance and affiliative behaviour. *Psychopharmacology*, *161*(3), 324–330. https://doi.org/10.1007/s00213-002-1049-7

Tse, W. S., & Bond, A. J. (2003). Reboxetine Promotes Social Bonding in Healthy Volunteers. *Journal of Psychopharmacology*, *17*(2), 189–195. https://doi.org/10.1177/0269881103017002007

Tunbridge, E. M., Bannerman, D. M., Sharp, T., & Harrison, P. J. (2004). Catechol-O-Methyltransferase Inhibition Improves Set-Shifting Performance and Elevates Stimulated Dopamine Release in the Rat Prefrontal Cortex. *Journal of Neuroscience*, *24*(23), 5331–5335. https://doi.org/10.1523/JNEUROSCI.1124-04.2004

Turlejski, K. (1996). Evolutionary ancient roles of serotonin: Long-lasting regulation of activity and development. *Acta Neurobiologiae Experimentalis*, *56*(2), 619–636.

Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience*, *16*(1), 55–61. https://doi.org/10.1038/nrn3857

Uddin, L. Q., Iacoboni, M., Lange, C., & Keenan, J. P. (2007). The self and social cognition: The role of cortical midline structures and mirror neurons. *Trends in Cognitive Sciences*, *11*(4), 153–157. https://doi.org/10.1016/j.tics.2007.01.001

Uddin, L. Q., Nomi, J. S., Hebert-Seropian, B., Ghaziri, J., & Boucher, O. (2017). Structure and function of the human insula. *Journal of Clinical Neurophysiology : Official Publication of the American Electroencephalographic Society*, *34*(4), 300. https://doi.org/10.1097/WNP.0000000000000377

Utevsky, A. V., Smith, D. V., & Huettel, S. A. (2014). Precuneus Is a Functional Core of the Default-Mode Network. *The Journal of Neuroscience*, *34*(3), 932–940. https://doi.org/10.1523/JNEUROSCI.4227-13.2014

Van der Does, A. J. W. (2001). The effects of tryptophan depletion on mood and psychiatric symptoms. *Journal of Affective Disorders*, *64*(2), 107–119. https://doi.org/10.1016/S0165-0327(00)00209-3

van der Heiden, L., Scherpiet, S., Konicar, L., Birbaumer, N., & Veit, R. (2013). Inter-individual differences in successful perspective taking during pain perception mediates emotional responsiveness in self and others: An fMRI study. *NeuroImage*, *65*, 387–394. https://doi.org/10.1016/j.neuroimage.2012.10.003

van der Meer, L., Groenewold, N. A., Nolen, W. A., Pijnenborg, M., & Aleman, A. (2011). Inhibit yourself and understand the other: Neural basis of distinct processes underlying Theory of Mind. *NeuroImage*, *56*(4), 2364–2374. https://doi.org/10.1016/j.neuroimage.2011.03.053

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

Vanderwal, T., Hunyadi, E., Grupe, D. W., Connors, C. M., & Schultz, R. T. (2008). Self, mother and abstract other: An fMRI study of reflective social processing. *NeuroImage*, *41*(4), 1437–1446. https://doi.org/10.1016/j.neuroimage.2008.03.058

Vatansever, D., Menon, D. K., Manktelow, A. E., Sahakian, B. J., & Stamatakis, E. A. (2015). Default mode network connectivity during task execution. *NeuroImage*, *122*, 96–104. https://doi.org/10.1016/j.neuroimage.2015.07.053

Ven, V. van de, Wingen, M., Kuypers, K. P. C., Ramaekers, J. G., & Formisano, E. (2013). Escitalopram Decreases Cross-Regional Functional Connectivity within the Default-Mode Network. *PLOS ONE*, *8*(6), e68355. https://doi.org/10.1371/journal.pone.0068355

Völlm, B. A., Taylor, A. N. W., Richardson, P., Corcoran, R., Stirling, J., McKie, S., Deakin, J. F. W., & Elliott, R. (2006). Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. *NeuroImage*, *29*(1), 90–98. https://doi.org/10.1016/j.neuroimage.2005.07.022

Von Neumann, J., & Morgenstern, O. (1944). Theory of Games and Economic Behavior. *Princeton University Press*. https://doi.org/10.1177/1468795X06065810

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196. https://doi.org/10.3758/BF03206482

Wang, D. J. J., Chen, Y., Fernández-Seara, M. A., & Detre, J. A. (2011). Potentials and Challenges for Arterial Spin Labeling in Pharmacological Magnetic Resonance Imaging. *Journal of Pharmacology and Experimental Therapeutics*, *337*(2), 359–366. https://doi.org/10.1124/jpet.110.172577

Wang, Liang, Liu, X., Guise, K. G., Knight, R. T., Ghajar, J., & Fan, J. (2010). Effective connectivity of the fronto-parietal network during attentional control. *Journal of Cognitive Neuroscience*, *22*(3), 543–553. https://doi.org/10.1162/jocn.2009.21210

Wang, Ling, Zhou, C., Zhu, D., Wang, X., Fang, L., Zhong, J., Mao, Q., Sun, L., Gong, X., Xia, J., Lian, B., & Xie, P. (2016). Serotonin-1A receptor alterations in depression: A meta-analysis of

molecular imaging studies. *BMC Psychiatry*, *16*(1). https://doi.org/10.1186/s12888-016-1025-0

Warden, M. R., Selimbeyoglu, A., Mirzabekov, J. J., Lo, M., Thompson, K. R., Kim, S.-Y., Adhikari, A., Tye, K. M., Frank, L. M., & Deisseroth, K. (2012). A prefrontal cortex–brainstem neuronal projection that controls response to behavioural challenge. *Nature*, *492*(7429), 428–432. https://doi.org/10.1038/nature11617

Warren, C. M., Wilson, R. C., Wee, N. J. van der, Giltay, E. J., Noorden, M. S. van, Cohen, J. D., & Nieuwenhuis, S. (2017). The effect of atomoxetine on random and directed exploration in humans. *PLOS ONE*, *12*(4), e0176034. https://doi.org/10.1371/journal.pone.0176034

Washburn, D., Wilson, G., Roes, M., Rnic, K., & Harkness, K. L. (2016). Theory of mind in social anxiety disorder, depression, and comorbid conditions. *Journal of Anxiety Disorders*, *37*, 71–77. https://doi.org/10.1016/j.janxdis.2015.11.004

Weightman, M. J., Air, T. M., & Baune, B. T. (2014). A Review of the Role of Social Cognition in Major Depressive Disorder. *Frontiers in Psychiatry*, *5*. https://doi.org/10.3389/fpsyt.2014.00179

Werner, N. S., Jung, K., Duschek, S., & Schandry, R. (2009). Enhanced cardiac perception is associated with benefits in decision-making. *Psychophysiology*, *46*(6), 1123–1129. https://doi.org/10.1111/j.1469-8986.2009.00855.x

Wernicke, J. F., Faries, D., Girod, D., Brown, J. W., Gao, H., Kelsey, D., Quintana, H., Lipetz, R., Michelson, D., & Heiligenstein, J. (2003). Cardiovascular Effects of Atomoxetine in Children, Adolescents, and Adults. *Drug Safety*, *26*(10), 729–740. https://doi.org/10.2165/00002018-200326100-00006

Whitehead, W. E., Drescher, V. M., Heiman, P., & Blackwell, B. (1977). Realtion of heart rate control to heartbeat perception. *Biofeedback and Self-Regulation*, *2*(4), 317–392.

WHO. (2017). *World Health Organisation | Depression and Other Common Mental Disorders*. WHO. http://www.who.int/mental_health/management/depression/prevalence_global_health_estimates/en/

Wicker, B., Ruby, P., Royet, J.-P., & Fonlupt, P. (2003). A relation between rest and the self in the brain? *Brain Research Reviews*, *43*(2), 224–230. https://doi.org/10.1016/j.brainresrev.2003.08.003

Wiebking, C., Bauer, A., de Greck, M., Duncan, N. W., Tempelmann, C., & Northoff, G. (2010). Abnormal body perception and neural activity in the insula in depression: An fMRI study of the depressed 'material me'. *The World Journal of Biological Psychiatry: The Official Journal of the World Federation of Societies of Biological Psychiatry*, *11*(3), 538–549. https://doi.org/10.3109/15622970903563794

Wiebking, C., de Greck, M., Duncan, N. W., Tempelmann, C., Bajbouj, M., & Northoff, G. (2015). Interoception in insula subregions as a possible state marker for depression—An exploratory fMRI study investigating healthy, depressed and remitted participants. *Frontiers in Behavioral Neuroscience*, *9*. https://doi.org/10.3389/fnbeh.2015.00082

Wiebking, C., Duncan, N. W., Tiret, B., Hayes, D. J., Marjańska, M., Doyon, J., Bajbouj, M., & Northoff, G. (2014). GABA in the insula—A predictor of the neural response to interoceptive awareness. *NeuroImage*, *86*, 10–18. https://doi.org/10.1016/j.neuroimage.2013.04.042

Wiebking, C., & Northoff, G. (2015). Neural activity during interoceptive awareness and its associations with alexithymia—An fMRI study in major depressive disorder and non-psychiatric controls. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00589

Wiens, S., & Palmer, S. N. (2001). Quadratic trend analysis and heartbeat detection. *Biological Psychology*, *58*(2), 159–175. https://doi.org/10.1016/S0301-0511(01)00110-7

Willeit, M., & Praschak-Rieder, N. (2010). Imaging the effects of genetic polymorphisms on radioligand binding in the living human brain: A review on genetic neuroreceptor imaging of monoaminergic systems in psychiatry. *NeuroImage*, *53*(3), 878–892. https://doi.org/10.1016/j.neuroimage.2010.04.030

Williams, R. B., Marchuk, D. A., Gadde, K. M., Barefoot, J. C., Grichnik, K., Helms, M. J., Kuhn, C. M., Lewis, J. G., Schanberg, S. M., Stafford-Smith, M., Suarez, E. C., Clary, G. L., Svenson, I. K., & Siegler, I. C. (2001). Central Nervous System Serotonin Function and Cardiovascular Responses to Stress. *Psychosomatic Medicine*, *63*(2), 300.

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General*, *143*(6), 2074–2081. http://dx.doi.org/10.1037/a0038199

Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2013). Situating emotional experience. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00764

Wood, R. M., Rilling, J. K., Sanfey, A. G., Bhagwagar, Z., & Rogers, R. D. (2006). Effects of Tryptophan Depletion on the Performance of an Iterated Prisoner's Dilemma Game in Healthy Adults. *Neuropsychopharmacology*, *31*(5), 1075. https://doi.org/10.1038/sj.npp.1300932

Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage*, *21*(4), 1732–1747. https://doi.org/10.1016/j.neuroimage.2003.12.023

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal Autocorrelation in Univariate Linear Modeling of FMRI Data. *NeuroImage*, *14*(6), 1370–1386. https://doi.org/10.1006/nimg.2001.0931

Wright, N. D., Symmonds, M., & Dolan, R. J. (2013). Distinct encoding of risk and value in economic choice between multiple risky options. *NeuroImage*, *81*, 431–440. https://doi.org/10.1016/j.neuroimage.2013.05.023

Yan, C., Liu, D., He, Y., Zou, Q., Zhu, C., Zuo, X., Long, X., & Zang, Y. (2009). Spontaneous Brain Activity in the Default Mode Network Is Sensitive to Different Resting-State Conditions with Limited Cognitive Load. *PLOS ONE*, *4*(5), e5743. https://doi.org/10.1371/journal.pone.0005743

Yang, J., Weng, X., Zang, Y., Xu, M., & Xu, X. (2010). Sustained activity within the default mode network during an implicit memory task. *Cortex*, *46*(3), 354–366. https://doi.org/10.1016/j.cortex.2009.05.002

Ye, Z., Rae, C. L., Nombela, C., Ham, T., Rittman, T., Jones, P. S., Rodríguez, P. V., Coyle-Gilchrist, I., Regenthal, R., Altena, E., Housden, C. R., Maxwell, H., Sahakian, B. J., Barker, R. A., Robbins, T. W., & Rowe, J. B. (2016). Predicting beneficial effects of atomoxetine and citalopram on response inhibition in Parkinson's disease with clinical and neuroimaging measures. *Human Brain Mapping*, *37*(3), 1026–1037. https://doi.org/10.1002/hbm.23087

Young, S. N., Smith, S. E., Pihl, R. O., & Ervin, F. R. (1985). Tryptophan depletion causes a rapid lowering of mood in normal males. *Psychopharmacology*, *87*(2), 173–177. https://doi.org/10.1007/BF00431803

Zaki, J., Davis, J. I., & Ochsner, K. N. (2012). Overlapping activity in anterior insula during interoception and emotional experience. *NeuroImage*, *62*(1), 493–499. https://doi.org/10.1016/j.neuroimage.2012.05.012

Zamariola, G., Maurage, P., Luminet, O., & Corneille, O. (2018). Interoceptive accuracy scores from the heartbeat counting task are problematic: Evidence from simple bivariate correlations. *Biological Psychology*, *137*, 12–17. https://doi.org/10.1016/j.biopsycho.2018.06.006

Zhang, H., Li, L., Wu, M., Chen, Z., Hu, X., Chen, Y., Zhu, H., Jia, Z., & Gong, Q. (2016). Brain gray matter alterations in first episodes of depression: A meta-analysis of whole-brain studies. *Neuroscience & Biobehavioral Reviews*, *60*, 43–50. https://doi.org/10.1016/j.neubiorev.2015.10.011

Zhang, L., Liu, L., Li, X., Song, Y., & Liu, J. (2015). Serotonin transporter gene polymorphism (5-HTTLPR) influences trait anxiety by modulating the functional connectivity between the amygdala and insula in Han Chinese males. *Human Brain Mapping*, *36*(7), 2732–2742. https://doi.org/10.1002/hbm.22803

Zhang, S., & Li, C. R. (2012). Functional connectivity mapping of the human precuneus by resting state fMRI. *NeuroImage*, *59*(4), 3548–3562. https://doi.org/10.1016/j.neuroimage.2011.11.023

Zhu, Lei, Zheng, L., Chen, M., Guo, X., Li, J., Chen, L., & Yang, Z. (2013). The neural correlates of incidental self-processing induced by handwritten negative words. *Experimental Brain Research*, *228*(1), 1–8. https://doi.org/10.1007/s00221-013-3531-3

Zhu, Lusha, Mathewson, K. E., & Hsu, M. (2012). Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(5), 1419–1424. https://doi.org/10.1073/pnas.1116783109

Zimmerman, C. A., Huey, E. L., Ahn, J. S., Beutler, L. R., Tan, C. L., Kosar, S., Bai, L., Chen, Y., Corpuz, T. V., Madisen, L., Zeng, H., & Knight, Z. A. (2019). A gut-to-brain signal of fluid osmolarity controls thirst satiation. *Nature*, *568*(7750), 98–102. https://doi.org/10.1038/s41586-019-1066-x

APPENDIX

A1: Health questionnaire for first and second testing periods

University of Sussex

HR:

BP:

Participant ID: _____

Health Questionnaire

Fill this out with a medical doctor and ask questions if necessary

No	Question	Please initial	
		Yes	No
1	I have a history of a psychiatric disorder (including anxiety disorders, depression, eating disorders, psychosis and substance abuse disorders)		
2	I have a first-degree family member (mother, father, brother/sister) with a history of bipolar disorder		
3	I have diabetes		
4	I have epilepsy or have had a seizure in my lifetime		
5	I have respiratory problems, including asthma		
6	I have had a problem with hypertension, heart disease, angina, irregular heartbeat, channelopathies (disorders caused by the dysfunction of ion channels), cerebral aneurysm, stroke or ANYTHING affected by high blood pressure or accelerated heartbeat.		
7	I sometimes have an abnormal heartbeat or blood pressure		
8	I have low levels of potassium or magnesium (hypokalaemia/hypomagnesaemia)		
9	I get migraines		
10	I have problems with my liver or kidneys		
11	I have had a problem with the blood vessels in my brain (such as a stroke or aneurysm)		
12	I could be pregnant or breastfeeding		
13	I have used antihistamines or painkillers in the past week		
14	I have used a psychoactive drug (a drug that affects the way I think or makes me hyper or drowsy) in the last month		
15	I have taken an MAO inhibitor in the last month		
16	I have had an allergic reaction to medication before		
17	I have a tumour on my adrenal gland		
18	I have had mood swings lately, or abnormally hostile thoughts		
19	I sometimes use tryptophan or St John's Wort		
20	I have another medical condition not covered by this sheet		
	Physician note:		

I certify that all of the information above is to the best of my knowledge and belief true, correct and complete

Subject Signature Subject Name Date:

Physician Signature Physician Name Date:

A2: Health questionnaire for third testing period

brighton and sussex medical school

UNIVERSITY OF SUSSEX

Health Questionnaire

Please read the following questions carefully. If you answer 'yes' to any of the questions, please give details in the space at the end of the form. If you are unsure of any of your answers, please leave the question blank and ask the researcher to explain further

No	Question	Please initial	
		Yes	No
1	I have a history of a psychiatric disorder (including anxiety disorders, depression, eating disorders, psychosis and substance abuse disorders)		
2	I have a first-degree family member (mother, father, brother/sister) with a history of bipolar disorder		
3	I have migraines		
4	I have diabetes		
5	I have epilepsy or have had a seizure at some point in my life		
6	I have glaucoma		
7	I have hypertension		
8	I have reduced kidney or liver function		
9	I have heart abnormalities (e.g. arrhythmia)		
10	I have a bleeding disorder such as haemophilia		
11	I have low levels of potassium or magnesium (hypokalaemia/hypomagnesaemia)		
12	I have impaired kidney or liver function		
13	I have a low resting heart rate (bradycardia)		
14	I have salt depletion or another water balance problem		
15	I could be pregnant or breastfeeding		

16	I have taken medication, either prescription or over the counter (except for birth control pills), in the last 30 days		
17	I have had an allergic reaction to medication before		
18	I have received electroconvulsive therapy		
19	I have another medical condition not covered by this sheet		
If you answered 'yes' to any of the questions above, please give details here:			
Researcher notes:			

I certify that all of the information above is to the best of my knowledge and belief true, correct and complete

Participant Signature	Participant Name	Date
_____	_____	_____

I have reviewed this form in the presence of the participant and given the opportunity for the participant to ask any questions. I am satisfied based on the information available that this person is able to take part in the study

Physician Signature	Physician Name	Date
_____	_____	_____

The MRI scanner uses a powerful magnetic field – we need to ensure that you are safe to enter the scanning room and don't have metal attached to you that can cause artifacts or heating effects.*

*Please remove ALL body piercings and loose metal objects **BEFORE** arriving for your scan*
e.g.: mobile phones, watches, hairclips/extensions, and, if necessary, change into the clothes provided.
Patients / participants with false limbs/callipers: please remove them before entering the scanning room.
Heavy eye make-up (particularly mascara) can distort images. Please do not wear; you may be asked to remove it.

Do you have / ever had any of the following? If yes please include details and dates:			
	No	Yes	Details / Dates
Cardiac pacemaker/defibrillator?			
Heart surgery / valve replacement?			
Stents in any blood vessels?			
Head surgery including that to the eyes or ears (e.g. clips / coils / shunts / cochlear implants)?			
Surgery in the past 6 weeks? Please provide details			
Any other surgery (e.g. pins / plates / screws in any bones / joint replacements)?			
Neurological stimulator or any other implanted electronic medical device (e.g. insulin pump)?			
Implanted contraceptive IUS/IUD (e.g. Mirena or copper coil)? Please provide make / model			
Epilepsy?			
Skin patches? (e.g. HRT, nicotine, pain relief, contraceptive)			
Tattoos or permanent eye makeup? (If yes, where? See overleaf)			
Have you EVER had metal fragments in your eyes or under your skin (e.g. from a car accident / shrapnel injury / welding / grinding / metal sheet worker)?			
Have you had a previous MRI scan? If yes, please indicate at CISC or elsewhere.			
Do you have any orthodontic work and/or any metal within your mouth?			
Women of child-bearing age (12-65): Is there any chance you could be pregnant?			

I confirm that I have answered and understood the above questions/information and that the information I have provided is correct to the best of my knowledge

Participant Name	Participant Signature	Date