Preliminary Investigations of Circular Inference in Bipolar Disorders

Matthew T. Whelan

Master of Science by Research

Institute for Adaptive and Neural Computation
School of Informatics
University of Edinburgh

2021

Abstract

Bipolar disorders are debilitating psychiatric illnesses affecting up to 2.4% of the population, and are one of the leading causes of disabilities worldwide, with up to 19% of patients completing suicide if left untreated. Despite this, current efforts to understand and treat bipolar disorders are still poor, with around 60% of patients relapsing even after proper treatment. New efforts to understand psychiatric illnesses are now being conducted in the growing field of computational psychiatry. Yet even in this field, bipolar disorders are an understudied illness. This study aims to build on the small but significant work being conducted on understanding bipolar disorders through computational modelling techniques. Circular inference is a Bayesian inference model that assumes prior beliefs, and likelihoods, in the form of sensory information, are reverberated in a hierarchical inference paradigm. That is, prior beliefs can be confused and added with likelihoods, whilst likelihoods can be confused and added with priors. This model has successfully been applied to schizophrenia, with results showing impaired circular inferences in schizophrenia compared to healthy controls. Given the genetic and symptomatic overlaps between schizophrenia and bipolar disorders, there is promise in applying this model to bipolar disorders. This study has therefore investigated the potential application of circular inference to bipolar disorders. A model recovery technique was first conducted to test whether the circular inference model could accurately be recovered for the experiment used. Results from this suggests that the circular inference model could not be recovered with any significant accuracy, being recovered in only 10% of the simulations. However, if the other two models used in this study, namely simple Bayes and weighted Bayes, were used to simulate the data, the circular inference model was never the best fitting model. This suggests that if the circular inference model is the winning model, there is a strong chance that the data is best fit by circular inference rather than the other models. A probabilistic decision task was designed, based on previous work, that aims to capture the behaviour of individuals in how they infer decisions given limited prior and sensory information. The behavioural data was fit using the three models. Of the 7 data sets obtained from the experiment, two participants were best fit by circular inference. Of these two participants, one scored the highest for the personality trait *cyclothemia*, a trait closely related to bipolar disorders. Whilst these results do not prove statistically significant, this study has provided a useful framework with which to explore further the potential of circular inference in bipolar disorders, potentially with redesigns of both the model and the experiment.

Acknowledgements

I would like to thank the organisers of the CDT in Biomedical AI, and in particular Dr. Diego Oyarzún and Dr. Ian Simpson, who have both provided an immense amount of support throughout this particularly challenging academic year. Despite the challenges, the MScR program was organised to provide plenty of intellectual freedom and stimulation, and has further provided, and the CDT is continuing to provide, many new skills and opportunities for career and personal growth.

A special thank you, and most probably not the last, to my supervisor Dr. Peggy Seriès, who has been both a supervisor and a mentor to me over the last year. It is hard for me to find much else that could be more meaningful than having a career working in the field of psychiatry. I am therefore incredibly grateful to have the opportunity to continue this line of research into the PhD, and am particularly pleased to be doing so with further supervision from Peggy.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Matthew T. Whelan*)

To my siblings, who are the best kind of siblings one could possibly have asked for.
And in particular to James, whose life and love has motivated much of this work.

Table of Contents

Chapter 1

Background

This chapter will provide a brief overview of the nosology of bipolar disorder (BD), including its symptoms, current treatments, purported causes, and symptomatic and genetic overlap with other psychiatric disorders. A short background on the methods used in the newly emerging field of Computational Psychiatry (CP), and how CP aims to help solve some of the current problems faced in the field of psychiatry, is then provided. The methods applied in this dissertation are common approaches used in CP, and it is the potential gains that CP may provide to the general study of psychiatry, and in this case BD in particular, that provides the motivation for the work conducted in this dissertation.

1.1 Bipolar Disorder

Bipolar disorder is a psychiatric illness broadly described as significant and debilitating changes in mood, such that mood alternates between (hypo)manic states and depressive states, interspersed with euthymic states. It was French psychiatrist Jean-Pierre Falret who is noted to be the first to observe and categorise BD in the 1850s, which he named *la folie circulaire*, or circular insanity (Angst and Sellaro, 2000). The *Diagnostic and Statistical Manual of Mental Disorders, fifth edition* (DSM-5, American Psychiatric Association, 2013), which is the most widely used manual in the United States for classifying mental disorders, now categorises BD into a total of seven subcategories, although the three predominant subcategories are bipolar I disorder, bipolar II disorder, and cyclothymic disorder.

Based on the World Mental Health Survey Initiative, BD has a lifetime and 12-month prevalence rate of 2.4% and 1.5% respectively (Merikangas et al., 2011), and is

the 17th leading cause of disability in the global burden of diseases (Vigo et al., 2016). Left untreated, it is a lethal disease, with up to 60% of sufferers attempting suicide at least once, and up to 19% of sufferers completing suicide (Novick et al., 2010). The total annual cost of BD to the UK, which includes NHS resource use, non-health-care resource use and indirect costs, was $2 billion in 1999/2000 (Gupta and Guest, 2002), or the equivalent of approximately $3.4 billion in 2020 (average inflation at 2.8% per year, ONS, 2021). An estimated 300,000 people in the UK are affected with BD (Gupta and Guest, 2002).

1.1.1 Symptoms of Bipolar Disorder

According to the DSM-5, bipolar I disorder is diagnosed if a patient has at least one lifetime manic episode. Although patients diagnosed with bipolar I disorder often experience hypomanic (a less severe manic episode) and major depressive episodes, these are not necessary for a diagnosis of bipolar I disorder. A manic episode is characterised by an abnormal and persistent period of elevated, expansive, or irritable mood and increased goal-directed activity, which lasts for at least 1 week and is present nearly every day, for most of the day. The mood disturbance is so severe that it has detrimental outcomes on the person's social or occupational functions, and may necessitate hospitalisation to prevent harm to the self or others. It may also include psychotic features. During this period the patient also exhibits behaviours that differ significantly from their usual behaviour. These include behaviours such as: having an inflated self-esteem or grandiosity; a decreased need for sleep, and therefore feeling rested after very little or no sleep at all; increased talkativeness that can be intrusive, including with strangers; an experience of racing thoughts or flight of ideas; increased distractibility, such that attention is constantly drawn to irrelevant external stimuli; increased goal-directed activity, such as at work/school, sexually, or socially; and doing to excess activities that have high potential for negative consequences, such as unrestrained spending sprees accumulating lots of debt, or sexual indiscretions that pay little or no regard to relationship consequences or disease transmission.

Bipolar II disorder meanwhile includes experiencing at least one episode of major depression and one episode of hypomania without ever experiencing full mania. During a state of hypomania, the same behavioural changes as specified in mania above are present, with the difference being that it often does not significantly impair the person's social or occupational functions, or require hospitalisation. Despite this, the

behavioural changes are often noticeable by others. The hypomanic state may also have a shorter timeline to that of mania, lasting only 4 or more days, as opposed to 1 week (although hypomania can last for 1 week or more). Individuals with bipolar II disorder tend to have greater and more chronic periods of depression than those with bipolar I disorder, and depression often does cause impairments. Depressive symptoms can also co-occur within a period of hypomania, and likewise hypomanic symptoms can co-occur during a period of major depression.

A major depressive episode constitutes either a symptom of depressed mood or a symptom loss of interest/pleasure, which lasts for at least 2 weeks. During depressed mood, there might be either a subjective feeling of sadness, emptiness, or hopelessness, or an observation from others that the person appears to be so (for instance tearfulness). During a loss of interest/pleasure, there is a significant reduction in the interest or pleasure found from all, or almost all, activities. Either a depressed mood or loss of interest/pleasure lasts nearly every day, for most of the day. Major depression also also constitutes a significant change in behaviour occurring within the same 2 week period that can cause significant detrimental effects in social or occupational functions. These can include: significant weight loss or gain, and/or a significant increase or decrease in appetite; insomnia or hypersomnia; an observable psychomotor agitation or retardation; fatigue or loss of energy; feelings of worthlessness or guilt that can be delusional; an inability to concentrate or think, or indecisiveness; thoughts of death, including suicide ideation, or a suicide attempt.

In cyclothymic disorder, symptoms of hypomania and depression can be present for at least 2 years, but without fulfilling the full criteria for hypomania and depression (American Psychiatric Association, 2013). The symptoms of hypomania and depression are also distinct from one another, and hence do not co-occur.

1.1.2 Current Treatments for Bipolar Disorder

The treatments for BD focuses on two predominant types, that of pharmacological intervention through drugs, and a second through psychotherapy. In addition, treatments can alter depending on which symptoms require alleviating in the short-term (mania or depression), as well as there being treatments for the long-term maintenance of the illness (Geddes and Miklowitz, 2013).

1.1.2.1 Pharmacotherapy for Bipolar Disorder

In treating acute mania, the most common and effective pharmacological treatments show to be antipsychotic drugs, and in particular risperidone, olanzapine, and haloperidol, which show to be more effective than other antipsychotics (Cipriani et al., 2011). Whilst these drugs are most effective for short-term treatment of mania, lithium has been shown to have better long-term evidence of efficacy when continued drug treatment is used (Geddes and Miklowitz, 2013).

Pharmacological treatment of depressive episodes, also known as bipolar depression, seems to be more challenging than treatment of mania. Evidence suggests that the use of antidepressants for treating bipolar depression is no more effective than placebo, and could potentially lead to the induction of mania or mood instability (Sidor and MacQueen, 2012). Despite this lack of evidence for the effectiveness of antidepressants, estimates suggest that 50-80% of BD patients are nonetheless prescribed antidepressants (Sidor and MacQueen, 2012). More effective treatment options for bipolar depression show promise in atypical, or second generation, antipsychotics, and in particular quetiapine and olanzapine (De Fruyt et al., 2012). Unfortunately, these drugs also have unwanted side effects such as somnolence (or sleepiness), sedation, akathisia (restlessness and inability to remain still), and metabolic issues (such as weight gain).

For long-term maintenance and treatment, lithium appears to remain the best treatment option for BD (Cade, 1949). A meta-analysis of lithium in the long-term treatment of BD showed a reduction in relapse from 61% to 40% following lithium treatment, with greater effectiveness in preventing relapses of mania than of depression, as well as reducing risk of suicide (Geddes et al., 2004). Unfortunately, long-term lithium use is restricted by potentially dangerous adverse side-effects, including disruption to renal function and potential renal failure, risk of congenital malformation in the babies of mothers who take lithium during pregnancy, and increased risk of hyperparathyroidism.

1.1.2.2 Psychotherapy for Bipolar Disorder

Psychotherapy is now considered an important treatment option in conjunction to pharmacotherapy for treating BD (Geddes and Miklowitz, 2013). This is because of evidence that psychosocial stressors, such as excessive family discord and distress, negative life events, disrusption of sleep, or accelerated goal attainment, can provoke relapse and worsen symptoms (Geddes and Miklowitz, 2013).

Unfortunately, it is unlikely that psychotherapy would work sufficiently well in treating patients during acute manic episodes, due to insufficient insight or rejection of help (Geddes and Miklowitz, 2013). Psychotherapy can however help during and following depressive episodes, by reducing recovery time following acute depression, preventing relapse and improving overall life functioning and satisfaction for patients (Miklowitz, 2006; Miklowitz et al., 2007).

There are a number of different forms of psychotherapy, with the main ones being family-focused therapy, cognitive-behavioural therapy, interpersonal and social rhythm therapy, group psychoeducation, functional remediation and systematic care management (Geddes and Miklowitz, 2013). They each work to help address some of the psychosocial factors that can contribute to relapses and worsening of symptoms. For instance, family-focused therapy aims to address the frequently shown association between hostility of caregivers (parents or spouse) and increased risk of relapse, by providing therapy to both patients and caregivers (Hooley, 2007). Cognitive-behavioural therapy meanwhile assumes that pessimistic or overly optimistic thinking in response to life events, as well as dysfunctional beliefs about the self, the world, and the future, determines relapse into either depression or mania (Geddes and Miklowitz, 2013), and attempts to address these dysfunctional beliefs and behaviours. Cognitive-behavioural therapy seems to help prevent depressive episodes significantly more so than manic episodes (Lam et al., 2005). Generally, psychotherapy is provided alongside pharmacotherapy, and meta-analyses suggest that psychotherapy in combination with pharmacotherapy is significantly more successful in preventing relapses than pharmacotherapy alone (Scott et al., 2007).

Despite promising evidence for both pharmacological and psychological treatments of BD, even with treatment, relapse of mania or depression occurs in about 37% of patients within 1 year, and in around 60% of patients within 2 years (Geddes and Miklowitz, 2013). New therapy options for both the short-term (and in particular for bipolar depression) and the long-term treatment of BD are therefore desperately needed.

1.1.3 Purported Aetiology of Bipolar Disorder

There are three factors that are purported to be the main causes of BD. These are genetic factors, biological factors, and psychosocial factors.

The genetic risk factor for BD shows itself in the high rates of heritability, it being one of the most heritable mental disorders (Goodwin and Jamison, 2007). As such,

many theories on the aetiology of BD focus on particular genes or groups of genes (Neves-Pereira et al., 2002). Biological factors meanwhile relate to abnormalities in brain regions, such as the hypothalamic-pituitary-adrenal axis (Furnham and Anthony, 2010), or in differences in neurotransmitter levels, such as norepinephrine, serotonin and dopamine (Young et al., 1994). Psychosocial factors suggest that environmental events, such as stressful life events, childhood trauma, and abuse, could lead to increased risk of BD (Hammen and Gitlin, 1997; Leverich et al., 2002). Furthermore, sleep deprivation and disruption to circadian rhythms has been shown to induce manic episodes (Wehr et al., 1987).

The prevailing view is that it is probably the interaction of all these three factors, the genetic, biological, and psychosocial, that leads to the onset of BD, in what has been termed the 'diathesis-stress' model (Furnham and Anthony, 2010).

1.1.4 Bipolar Disorder in Relation to Other Psychiatric Disorders

Bipolar disorder shares many features, including symptoms, genetic markers, and brain activity abnormalities, with other psychiatric disorders. In particular, bipolar disorder has been found to share many features with schizophrenia and major depressive disorder. Factor structures of symptoms in schizophrenia shows to present with three primary factors, or symptom clusters (American Psychiatric Association, 2013). These are 1) positive symptoms (hallucinations and delusions); 2) negative symptoms (anhedonia, asociality, etc.); and 3) disorganisation (disorganised thinking/speech or abnormal behaviour). Toomey et al. (1998) showed in a factor analysis that these same three factors can be extended and applied to major depression and bipolar disorder. They concluded that there was a continuous measure of psychosis relevant to all three illnesses, and that the similar symptom factors found across these illnesses were reflecting the underlying dimension of psychosis. Furthermore, Reynolds de Sousa et al. (2021) have explored the hypothesis that BD and schzophrenia lie on a continuum, and relate findings that show the same or similar neurotransmitter dysfunctions, genetic similarities, and common abnormalities in EEG and imaging patterns between these two disorders.

Given BD, and particularly BD type II, includes periods of major depression, it should be clear that it shares symptoms with major depressive disorder. As such, bipolar disorder is often misdiagnosed as recurrent major depressive disorder. In fact, it is estimated that potentially up to 21% of patients diagnosed with unipolar depression

may have undiagnosed bipolar disorder (Smith et al., 2011). Misdiagnoses such as this point towards the need for improved diagnostic tools, but again it indicates a shared dimensional element to these illnesses.

BD, in addition to schizophrenia, has also shown to have strong genetic overlap with autism spectrum disorder, with all three illnesses sharing common single-nucleotide polymorphism alleles (Carroll and Owen, 2009). It is not surprising therefore that a family history of both schizophrenia and BD increases the risk factors for autism spectrum disorder, such that first degree relatives of an individual with either schizophrenia or BD have increased risk of autism (Sullivan et al., 2012).

There is also evidence of comorbidity between bipolar disorder and a range of psychiatric illnesses. In particular, there is a strong association between BD and substance abuse, but also associations with anxiety disorders, attention-deficit/hyperactivity disorder, eating disorders, cylcothymia and axis II personality disorders (Krishnan, 2005).

Given such a wide range of overlapping symptoms, shared genetic variances, and comorbidities between BD and other psychiatric disorders, these all indicate that a dimensional approach to diagnosing and understanding mental illness may be more sensible than the current approach. Indeed, the methods of modelling behaviour that we have taken as part of this study could be more readily applied to a dimensional understanding of mental illness, such that a persons ability to perform inference is used as a marker of illness, as opposed to, or perhaps in combination with, symptoms only.

1.2 A Computational Approach to Psychiatry

The traditional approach to psychiatry, as outlined above, has been one of symptom clustering used to classify illnesses, with many of these symptoms being subjective evaluations from the patient. However, due to the heterogeneous nature of symptoms within an illness, and the sharing of symptoms across different illnesses, classifications can often be difficult and even incorrect. One such common occurrence of this is that bipolar disorder is often misclassified as unipolar depression (Smith et al., 2011). In addition, the best treatments have tended to come from trial-and-error methods, particularly pharmacotherapy treatments, which are clearly inefficient methods for effective drug discovery. Stephan et al. (2016a,b) have laid out a non-exhaustive series of some of the major challenges faced currently in psychiatry.

Computational approaches to psychiatry now aim to improve upon the current state

of psychiatry, by using two complementary methods of computation: data-driven and theory-driven computation (Huys et al., 2016; Series, 2020). Data-driven computation employs machine learning and high-dimensional data in order to more effectively classify illnesses, predict treatment outcomes or to guide treatment selection. However, data-driven methods do not necessarily provide much insight into the underlying mechanisms of these illnesses. Theory-driven methods meanwhile build mechanistic models at multiple levels of abstraction. These models instantiate hypotheses of the underlying mechanisms, and can be tested using behavioural experimental designs. These two approaches are often complementary. For instance, parameter estimation in a theory-driven model can be performed using maximum likelihood estimation (as has been performed in this study, Section 3.6).

Computational methods applied to psychiatry has spawned the new field of Computational Psychiatry (CP), and CP is now attempting to address a large subset of those major challenges laid out by Stephan et al. (2016a,b). This includes a broad list of challenges, such as (again, non-exhaustive): the role and problems of using symptoms in disease diagnosis (the current practice), and the need to integrate a dimensional approach to mental illnesses; the higher order structure of the mechanisms relevant for diagnostics – that is, to understand the cluster of mechanisms that are specific to patients and that may be shared across subgroups of patients; the development of computational assays, or generative models that characterise the (aberrant) computations performed by the brain, for psychiatric nosology; and understanding learning dysfunctions in relation to psychiatric illnesses.

This study aims to build on the promising methods of CP, and employs a theory-driven approach to understanding BD, which is a comparatively understudied psychiatric disorder in CP.

1.3 Motivation for this Study

BD is a relatively understudied psychiatric disorder in CP (Series, 2020). However, there are significant overlaps between BD and schizophrenia. There is promise therefore in applying some of the same or similar experiments and models that have been applied to schizophrenia for bipolar disorder. One such promising computational model that has been successfully applied to schizophrenia is 'circular inference' (Jardri and Denève, 2013; Jardri et al., 2017). In a recent extension of circular inference in schizophrenia, Simonsen et al. (2021) successfully showed deficits in circular infer-

ence using an experiment that included social factors that were not included in the original experiment by Jardri et al. (2017). In this study, we aim to adopt the experiment by Simonsen et al. (2021) and investigate whether it could be applied to bipolar disorders.

Prior to applying it to patients however, it is first beneficial to assess the general effectiveness of the modelling approach. This study will use model evaluation techniques in order to assess how reliable the circular inference model is in describing the behavioural data extracted from the study. In addition, given this experiment is still very new, this study will evaluate the subjective experience of taking the experiment, by providing the participants with a set of evaluation questions after having conducted the experiment. Furthermore, a preliminary analysis on whether mood might be correlated with behaviour in the experimental task is conducted by assessing the participants using a non-clinical questionnaire known as TEMPS-A. Finally, given this experiment includes a social factor, and despite it not being the primary objective of this study, it might also be interesting to investigate whether autistic traits are correlated with behaviour in the experiment. Hence, a second questionnaire is provided to the participants, known as the autism spectrum quotient, to assess traits related to autism.

The work presented in this dissertation will therefore assess the quality of both the model and experiment, as well as exploring any potential correlations between mood temperaments, autistic traits, and the model parameters fitted to the experimental data. As the experimental design, model design, and analyses have been conducted using a small, preliminary set of participants, this work in effect provides a ready-made framework that can later be more easily applied to larger groups of patients.

Chapter 2

Computational Modelling of Bipolar Disorder

This Chapter will first review briefly the current approaches to computational modelling of bipolar disorder. It will then move on to describe Bayesian methods of modelling behaviour and psychiatric illnesses, which is an approach not yet seen in modelling approaches to bipolar disorder. Three Bayesian methods are described, each one building on the previous, and all three being used as part of this study: Simple Bayes, Weighted Bayes, and Circular Inference. An example of the application of these Bayesian methods for modelling psychiatric illnesses is provided with regards to schizophrenia, which has arguably been modelled through this approach more than any other illness.

2.1 Current Computational Models of Bipolar Disorder

Bipolar disorder remains a poorly understood illness, and is perhaps one of the least studied major psychiatric disorders in computational psychiatry (Series, 2020). As such, there is only a limited number of theory-driven computational models available. Broadly, the literature on theory-driven computational models of BD falls under two main approaches. The first is non-linear dynamical systems modelling, which was arguably the first attempt at applying mathematical modelling to BD. And it was Daugherty et al. (2009) who were the first to attempt to model BD as a dynamical oscillator. They made the assumption that mood in BD type II fluctuates periodically and worsens over time if left untreated. A negatively damped harmonic oscillator was therefore used to model mood as,

$$\ddot{m} - \alpha\dot{m} + \omega^2 m = 0, \tag{2.1}$$

where m represents mood and $\alpha > 0$ and ω are parameters. However, given this system would go towards infinity, and yet mood, even if left untreated, cannot realistically become infinitely severe, an additional term can be added to Eqtn 2.1 allowing the system to come to a stable limit cycle,

$$\ddot{m} - \alpha\dot{m} + \omega^2 m - \beta m^2 \dot{m} = 0, \tag{2.2}$$

where β is an additional parameter. This system is also known as the van der Pol oscillator. Whilst this system can grow over time, representing a worsening of BD severity, it eventually reaches a stable limit cycle. Autonomous forcing can be included in order to represent the aggregate effects of treatment of BD (the combined effect of antidepressants, antipsychotics, mood stabilisers, psychotherapy, etc.), by including an additional forcing function $g(m,\dot{m})$,

$$\ddot{m} - \alpha\dot{m} + \omega^2 m - \beta m^2 \dot{m} = g(m,\dot{m}). \tag{2.3}$$

Daugherty et al. (2009) use the forcing function $g(m,\dot{m}) = \gamma m^4 \dot{m} + \delta m^2 \dot{m}$ in their analysis. Applying the forcing function when the van der Pol limit cycle is severe reduces the amplitude of the oscillations to less severe levels. The above model is an example of the dynamical approach to BD, and further models have extended on this (Steinacher and Wright, 2013; Goldbeter, 2011; Chang and Chou, 2018). A review conducted by Cochran et al. (2017) offers a useful comparison of the dynamical modelling approaches to BD.

The second main body of theoretical modelling of BD comes from Reinforcement Learning (RL). A recent computational framework described by Mason et al. (2017) suggests that mood swings are a result of potentially hypersensitivity and hyposensitivity to rewarding stimuli during states of high mood and low mood, respectively. During a manic cycle for instance, an increased biased perception to rewarding stimuli as a result of hypersensitivity leads to increased positive surprises, thus increasing mood further which increases the biased perception even further still (Figure 2.1). Over time however, as reality becomes increasingly discrepant with the perceived rewards, there is an eventual extreme negative surprise. This begins a depressive cycle, which works in the same manner as the manic cycle but with decreased biased perceptions and negative surprises. In this case, over time, reality eventually becomes much more

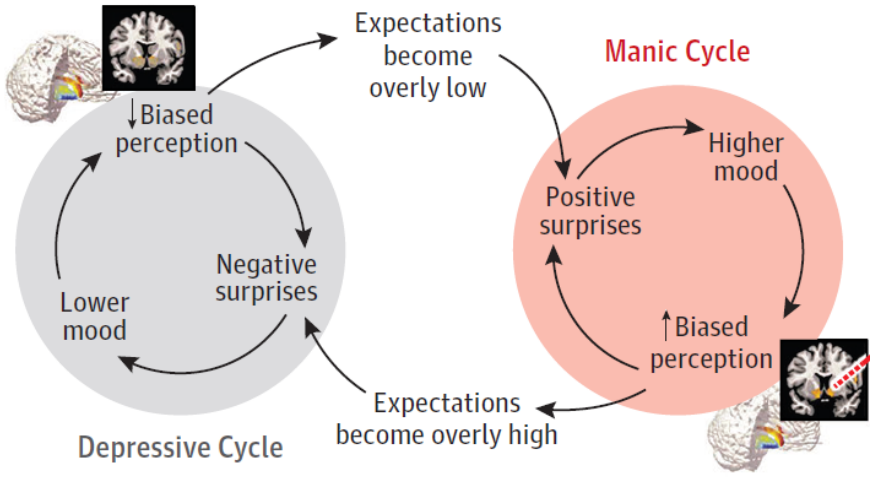

Figure 2.1: Model of mood instability proposed by Mason et al. (2017). A manic cycle occurs when perception of rewards are much higher than reality (left cycle). However, once expectations of rewards become overly high and unrealistic, a large negative surprise is elicited. This begins a depressive cycle in which perception of rewards become much lower than reality (right cycle). Figure adapted from Mason et al. (2017).

rewarding than the biased perception suggests, which then causes an extreme positive surprise and hence another manic cycle. Linke et al. (2020) have suggested that it could be a case of reduced learning rates for negative surprises leads to overly optimistic expectations, hence initiating the bipolar cycle as indicated in Figure 2.1. They proposed a simplistic RL rule that has two separate learning rates each for positive and negative rewards,

$$Q_i(t+1) = Q_i(t) + \alpha_+[r(t) - Q_i(t)]_+ + \alpha_-[r(t) - Q_i(t)]_- \tag{2.4}$$

where $Q_i(t)$ is the value for a stimulus i on a trial t, $r(t)$ is a reward received during trial t, and α_+ and α_- are the positive and negative learning rates for rewards that are better or worse than expected, respectively. Linke et al. (2020) tested such a model on first-degree relative of patients with BD in a binary decision making task. They hypothesised that an aberrancy in the negative learning rate might also be observed in first-degree relatives of patients with BD. Their results did not strongly support this however, and further investigations still need to be performed to test whether this is even true of BD patients themselves.

2.2 Introduction to Bayes' Theorem in Modelling Behaviour

A large area of computational modelling in CP has now gone into exploring whether there are aberrancies in inference in patients with various psychiatric disorders, but in particular in schizophrenia (Garety et al., 1991; Adams et al., 2016). In the Bayesian inference framework, a belief would arise from an individual's integration of prior beliefs with new sensory information. It has been suggested that in schizophrenia, sensory information is given more weight than prior beliefs (Adams et al., 2016), which could explain the surprising finding that patients with schizophrenia have a reduced susceptibility to optical illusions (Notredame et al., 2014).

The experimental setup used to test inference in this study, which has been adapted from Simonsen et al. (2021), can be used to test the Bayesian inference hypothesis of behaviour. In the experiment (discussed more fully in Section 3.3), there are a range of jars each containing different proportions of red and green beads. One of these jars is chosen at random with the contents of the jar hidden from the participants. 8 beads are then randomly drawn from the jar, and the participant must guess the colour of the 9th bead before it is drawn. Four simulated agents also draw beads from the same jar before the participant does the same, and make guesses as to the colour of the next bead. The participant is shown with the guesses of the four agents. This information therefore represents the prior belief for the colour of the 9th bead. The 8 randomly drawn beads then represent the participant's sensory evidence. Formally, this can be presented as follows,

$$P(R|S) = \frac{P(S|R)P(R)}{P(S)} \tag{2.5}$$

where $P(R|S)$ is the posterior probability of the next bead being red (R) given the sensory evidence (S). $P(R)$ is then the prior probability of the next bead being red, $P(S)$ the marginal likelihood of the sensory evidence, and $P(S|R)$ the likelihood of the sensory evidence given the next bead will be red. This formulation is known as the standard, or simple, Bayes model, and is one such example of Bayesian modelling of behaviour that assumes the participant is 'Bayes optimal'. For a situation in which there is a binary choice (such as between a red and green bead in this instance), and the likelihood and priors for each choice are mutually exclusive and sum to 1, so that $P(S|R) = 1 - P(S|G)$ and $P(R) = 1 - P(G)$, then Eqtn 2.5 can be simplified to a log

odds representation as follows (Jardri et al., 2017; Simonsen et al., 2021),

$$L_r = L_s + L_o \tag{2.6}$$

where $L_r = \log\left(\frac{P(R|S)}{1-P(R|S)}\right)$, $L_s = \log\left(\frac{P(S|R)}{1-P(S|R)}\right)$ and $L_o = \log\left(\frac{P(R)}{1-P(R)}\right)$. This is the form used by Jardri et al. (2017) and Simonsen et al. (2021). Once L_r had been found, it would require only a simple inversion to obtain the posterior,

$$P(R|S) = \frac{e^{L_r}}{1+e^{L_r}}. \tag{2.7}$$

2.2.1 Weighted Bayes

In practice it is unlikely that a participant would be Bayes optimal, and may instead over- or under-count either the sensory evidence or their prior beliefs. To account for this, there is a further extension on simple Bayes known as weighted Bayes (WB), which adds weights to the prior and likelihood (Simonsen et al., 2021),

$$L_r = F(L_s, w_s) + F(L_o, w_o), \tag{2.8}$$

where

$$F(L, w) = \log\left(\frac{we^L + 1 - w}{(1-w)e^L + w}\right). \tag{2.9}$$

The weights for the likelihood and prior, w_s and w_o respectively, represent how much 'trust' the participant has in either the sensory evidence or the prior. When $w_s = w_o = 1$ then the WB model reduces to the SB model.

2.2.2 Circular Inference

Jardri and Denève (2013) first layed out the circular inference hypothesis in schizophrenia. Originally proposed in a neural network framework, their later work (Jardri et al., 2017) simplified the circular inference model down from the neural network description to being one that is an extension on the WB described above. This latter version of the CI model is given as follows,

$$L_r = F(L_s + I, w_s) + F(L_o + I, w_o) \tag{2.10}$$

where

$$I = F(\alpha_s L_s, w_s) + F(\alpha_o L_o, w_o). \tag{2.11}$$

As can be seen in Eqtn 2.10, the log odds for each of the prior and likelihood has an equal term added on to them, I. This additional term, described by Eqtn 2.11, represents an additional increase in the number of times the likelihood and prior are included in the computations. This addition is also described as the 'reverberation' of the likelihood and prior, as they appear to be reverberated within the computation of the posterior log odds (Jardri and Denève, 2013). There are an additional two terms, α_s and α_o, that dictate how many times each of the likelihood and prior are reverberated, respectively. Should $\alpha_s = \alpha_o = 0$, then CI reduces down to the WB case.

This forms the basic understanding of the CI model itself. But it is perhaps worth a closer look at the theory behind this model.

2.2.3 Theory of Circular Inference

Jardri and Denève (2013) first proposed the circular inference model as a mechanism to explain positive symptoms (such as hallucinations and delusions) in schizophrenia. They based it on empirical findings showing impairments in GABA transmission and/or NMDA receptor plasticity in schizophrenia (Stephan et al., 2009). Impairment of GABA, an inhibitory neurotransmitter, and NMDA, being a receptor for the excitatory neurotransmitter glutamate, would indicate an imbalance in excitation and inhibition in the brain (O'Donnell, 2011).

The Bayesian inference hypothesis of the brain posits that a basic function of the brain is to generate percepts via Bayesian inference (Doya et al., 2007). To do so, a hierarchical Bayesian inference framework would be performed via hierarchical networks in the brain (Friston, 2008). These hierarchical networks pass on sensory observations via a bottom-up path, whilst predictions based on prior beliefs would be passed through via a top-down path. An optimal combination of bottom-up sensory signals with top-down predictions would then result in optimal Bayesian inference (Figure 2.2).

However, there is a problematic outcome in having a hierarchical network pass information both ways through the network – how can sensory information that is passed on from the bottom-up pathway not be reverberated back down the network and therefore be misinterpreted as prior knowledge? And likewise, how does prior knowledge not get reverberated back up the network and be misinterpreted as sensory informa-

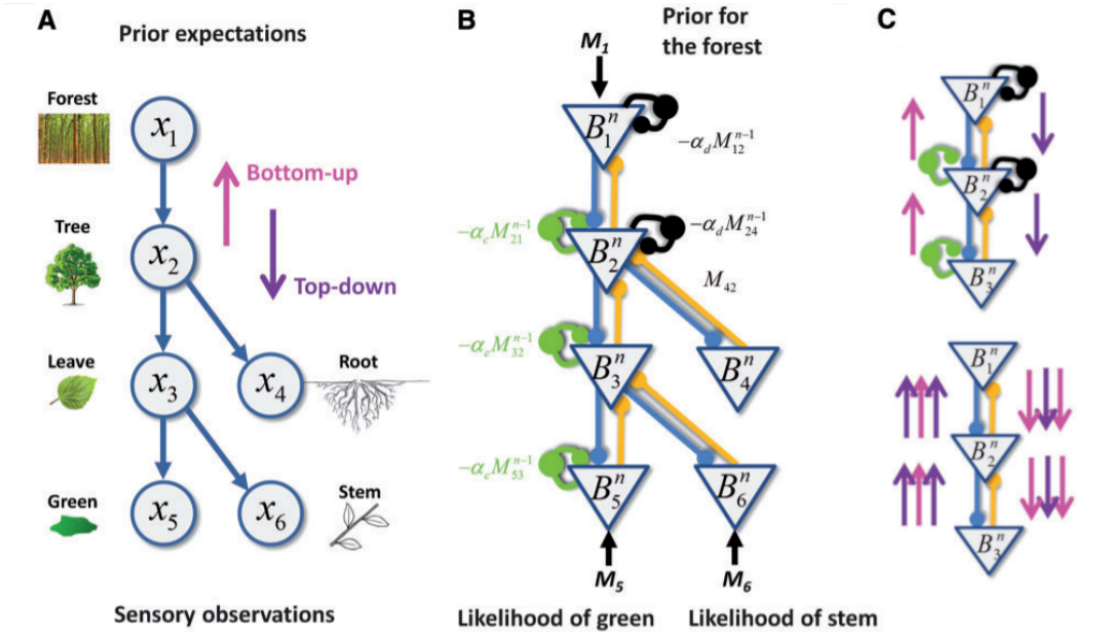

Figure 2.2: Neural network model of circular inference. A) a hypothetical scenario of hierarchical and probabilistic causations. Sensory information is received from the bottom, whilst prior beliefs are located at the top of the hierarchy. B) A neural network implementation of belief propagation. Messages are sent between the nodes, each of which represents a belief about the probability of the object at that node being present. Recurrent inhibitory connections aim to cancel out messages that would otherwise be reverberated throughout the network. C) An example of a balanced inhibition-excitation network allowing messages to be propagated only once throughout the network (top) versus an imbalanced network without appropriate inhibition, such that messages are passed multiple times throughout the network, resulting in circular belief propagation. Figure adapted from Jardri and Denève (2013).

tion? In the hierarchical network proposed by Jardri and Denève (2013), they suggest that inhibitory loops cancel out reverberated signals, thus avoiding these problems. In such a framework, if inhibition were reduced, or if there was a case of over-excitation in the network, then these signals would indeed be reverberated, resulting in 'circular belief propagation'.

Figure 2.2 gives a toy example for how such a hierarchical network might compute inferences, as first described by Jardri and Denève (2013). In this hypothetical case, bottom up sensory evidence, in the way of the colour green for instance, is passed up the network. Meanwhile, prior beliefs, such as the prior knowledge that one is in a forest (or not) is passed down the network. A combination of these bottom-up and top-

down signals then combine to produce inferences at different levels of the hierarchy. For instance, the presence of the colour green increases the probability of there being a leaf, which in turn increases the probability of there being a tree. Likewise, the prior knowledge that one is in a forest increases the probability of seeing a tree, and therefore also a leaf. Note that there is a causal relationship in the network that travels down the top-down path, such that the presence of a forest causes the presence of trees, the presence of trees causes the presence of leaves, and the presence of leaves causes the colour green (as represented by the downward pointed blue arrows in Figure 2.2A).

Each node in the network therefore represents the probability of the presence of that object. A neural network implementation of this could be as is shown in Figure 2.2B. Here, messages sent from node i to node j are given as M_{ij}. Feedforward and feedback connections are then passed between each of the nodes, whilst the nodes themselves each have inhibitory loops which aim to counteract the excitatory loops in the network. Should there be a balance between excitation and inhibition, then messages are sent only once throughout the network, resulting in optimal inferences (Figure 2.2C), top). However, should there be reduced inhibition so that there is over-excitation in the network, then messages will be passed on multiple times throughout the network (Figure 2.2C, bottom).

A set of equations that act recursively over the course of discrete time steps would look like the following,

$$M_{ij}^{n+1} = W_{ij}(B_i^n - \alpha M_{ji}^n) \tag{2.12}$$

$$B_i^{n+1} = \sum_j M_{ji}^{n+1} \tag{2.13}$$

Here, n indicates the time step, and W_{ij} is the strength of connection from node i onto node j. The term B_i indicates the belief for the presence of the particular object at node i, and is represented as a log odds ratio. That is, the belief at node i is given by,

$$B_i = \log\left(\frac{p_i}{1-p_i}\right) \tag{2.14}$$

where p_i is the probability of the presence of the object, and $1-p_i$ the probability of the absence of the object. For instance, a belief of $B_{leaf} = 4$ would be a strong belief in favour of a leaf being present, whilst a belief of $B_{leaf} = 0$ would equate to total uncertainty about whether a leaf is present or not. A negative belief, say $B_{leaf} = -4$, would be a strong belief against the presence of a leaf.

The negative term in Equation 2.12, $-\alpha M_{ji}^{n}$, is the inhibitory portion that aims to cancel the message that node j sends in return to node i. A value of 1 for α would denote full inhibition, whereas a value of 0 would indicate no inhibition and would therefore generate circular inferences (Figure 2.2C). Different values for α can be given for top-down inhibitory loops versus bottom-up loops. In the example shown in Figure 2.2, for messages sent down the network, for instance if j was above i in Eqtn 2.12, then α is denoted α_d for 'downward' messages, whereas the reverse direction would be α_c for 'climbing' messages.

Jardri and Denève (2013) used this model to show, via simulations, how impairing the network by reducing inhibition leads to aberrant beliefs. As stated above, this is the neural network version of the model. For simplicity however, an algorithmic approach was later used by Jardri et al. (2017) and Simonsen et al. (2021), and is the version used in this study too. Additionally, this model has only been tested in schizophrenia (Jardri et al., 2017; Simonsen et al., 2021) and autism (Chrysaitis et al., 2021), but in no other cases. This study therefore aims to apply the model of circular inference to BD, which as stated in Section 1.1.4, shares many features with schizophrenia.

Chapter 3

Methods

3.1 Participants Recruitment

Participants were recruited through email circulation to colleagues and friends of the author. A total of 14 participants signed up for the study. Questionnaire scores for all 14 participants have been recorded, as well as evaluation feedback from 11 participants. However, due to a technical error for which time unfortunately did not allow for the resolution of, the experimental data for only 7 of the participants have been recorded. The details regarding this technical error are discussed in Chapter 5. No demographic details of the participants have been recorded, such as age, sex, education, known psychiatric disorders, etc. The only information that can be assumed is that a large proportion of the participants were students of the University of Edinburgh.

3.2 Questionnaires

Two questionnaires were used to assess participants personality traits in relation to model parameters and experiment performance. The first was a shortened version of the Temperament Evaluation of Memphis, Pisa, Paris and San Diego-autoquestionnaire, otherwise known as TEMPS-A (Akiskal et al., 2005). This version of the TEMPS-A questionnaire aims to measure 5 temperaments, or factors: cyclothymic (fluctuation of high and low levels of mood), depressive (abnormally low levels of mood), irritable, hyperthymi (abnormally high levels of mood), and anxious. Scores along these factors aim to capture emotional, cognitive, psychomotor and circadian traits that could predispose someone to mood disorders (Akiskal et al., 2005). The questionnaire is a 'yes-or-no' type questionnaire. A 'yes' to a question increases the score by 1. Ques-

tions 1-12 measure the cyclothymic factor, questions 13-20 measure the depressive factor, questions 21-28 measure the irritable factor, questions 29-36 measure the hyperthymi factor, and questions 37-39 measure the anxious factor (see Appendix for the full questionnaire). Scores for each of the five factors can therefore be taken independently, as well as an overall score be given for all questions. The final scores for each factor and the overall score are then normalised (by dividing the score by the total number of questions for each category or by the total number of questions in the whole questionnaire).

The second questionnaire used was the Autism Spectrum Quotient, or AQ (Baron-Cohen et al., 2001). This aims to measure the degree to which participants have traits associated with the autistic spectrum. It is a 50 question, agree/disagree type questionnaire (see Appendix for full questionnaire). This questionnaire was included in this study for two reasons. The first is that, as was described in Section 1.1.4, there is a strong overlap between autism, bipolar disorder, and schizophrenia. The second is that, whilst circular inference has been applied to schizophrenia with positive results, a recent application of circular inference in relation to autistic traits showed no correlation (Chrysaitis et al., 2021). However, the key difference between the study of Chrysaitis et al. (2021) and this one is that this includes a social element (the 4 agents, see next Section). Social deficits in autism are a prominent symptom of the condition (American Psychiatric Association, 2013), and it could be that this experiment captures better the behaviour of inference in social contexts relevant to autism. This would leave open further research avenues that could later be explored.

These questionnaires were designed and hosted on Qualtrics (www.qualtrics.com). The participants would first answer both of these questionnaires, and then upon completion of these they were redirected to the behavioural decision making task.

3.3 Behavioural Experiment – Modified Beads Task

The modified beads task experiment follows the same procedure as that of (Simonsen et al., 2021). In this experiment there are 9 different jars each containing varying proportions of red and green beads (Figure 3.1). One jar is selected at random and the contents are hidden from the participant. 8 beads are then randomly drawn from this jar. The participants are then told that a 9th bead will be drawn, and that they must guess what colour, red or green, this 9th bead will be. In addition to the participant, 4 other simulated agents will perform the same task, drawing 8 beads from the same jar

and guessing the colour of the 9th bead. The agents will also indicate how confident they are in their guess (Figure 3.2). The participant will have access to the agent's guesses and confidences before they then see their own 8 drawn beads. It is important to note that the agents, once having drawn their beads and made their guesses, then replace the beads back into the jar before the participant takes their turn. Therefore the proportions of red and green beads are the same for all agents and the participant when drawing the beads.

The participants first perform 5 practice trials without any other agents, in order to gain an understanding of the experimental procedure. In this practice session, the participant is first described to them the set of instructions as described above, and are shown the 9 jars each containing different proportions of red and green beads. On each trial, the participant is told that a new jar has been chosen and then sees 8 beads that have been drawn from this jar. They then must guess the colour of the next bead as well as provide a confidence for their guess. By performing these practice trials, it should be clear to the participant how the agents will then make their guesses. The procedure for the practice experiment is shown in Figure 3.1.

After completing the first set of practice trials, the participants are told that they will now perform the task with four other agents, as described above. These four agents appear as neutral faces on the screen. The figures for these faces were obtained with consent from the Radboud Faces Database (Langner et al., 2010). This time however, whilst the agents will be providing a confidence, the participants do not – they only need to guess the colour of the 9th bead. The participants first perform 10 practice trials at a slow pace in order to properly understand the procedure, before performing 105 main trials at a quicker pace. The full procedure for the main trials, including times between transitions in the experiment, are shown in Figure 3.2.

For each trial a random set of agents faces are selected. All possible 7 scenarios for the number of red beads, ranging from 1 red bead up to 7 red beads, are randomly presented to the participant. In addition, all 5 possible colour combinations for the agents' decisions, from 0 agents choosing red up to all 4 choosing red, are presented to the participant. For simplicity, agents choosing the same colour have the same confidence. Should the agents be split about the choice of colour, one set of agents will always be confident whilst the other(s) will always be unconfident. This gives a total of 35 trial combinations. This combination is then repeated randomly 3 times, making 105 trials in total.

The behavioural task was hosted on Pavlovia (https://pavlovia.org) and developed

using the PsychoPy software package (Peirce et al., 2019). Upon completion of the behavioural task, the participants were then redirected back to Qualtrics in order to answer a few short questions that aimed to evaluate the behavioural experiment.

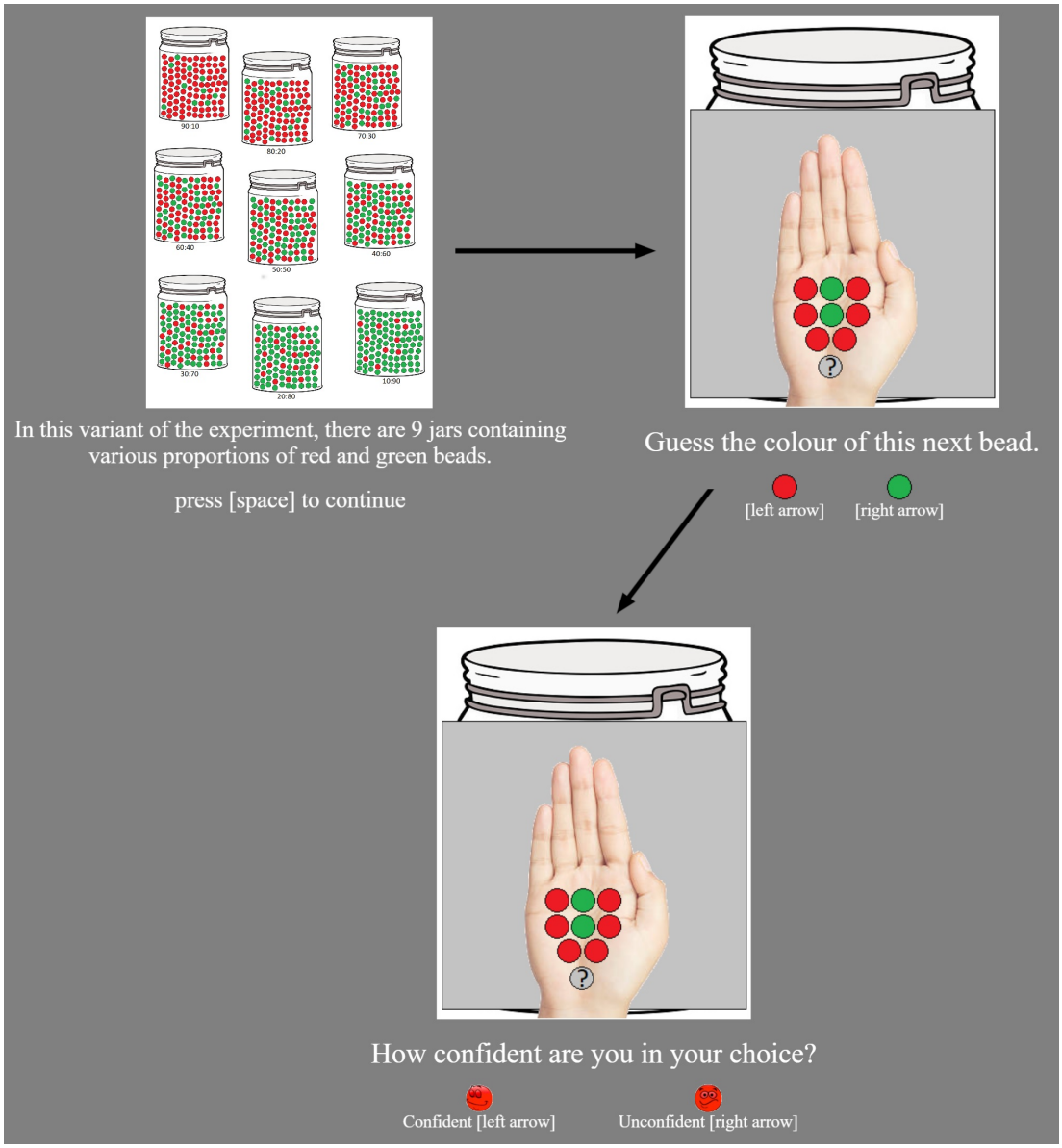

Figure 3.1: In the initial set of practice trials the participants are first presented with instructions for the task, including a presentation of the 9 jars (top left). They then perform 5 practice trials. For each trial a new jar is presented along with 8 beads from the jar, where the participant is then told to guess the colour of a 9th bead (top right). Upon guessing a particular colour, the participant is then asked to provide a confidence for their guess (bottom). Arrows indicate ordering of the frames displayed to the participant.

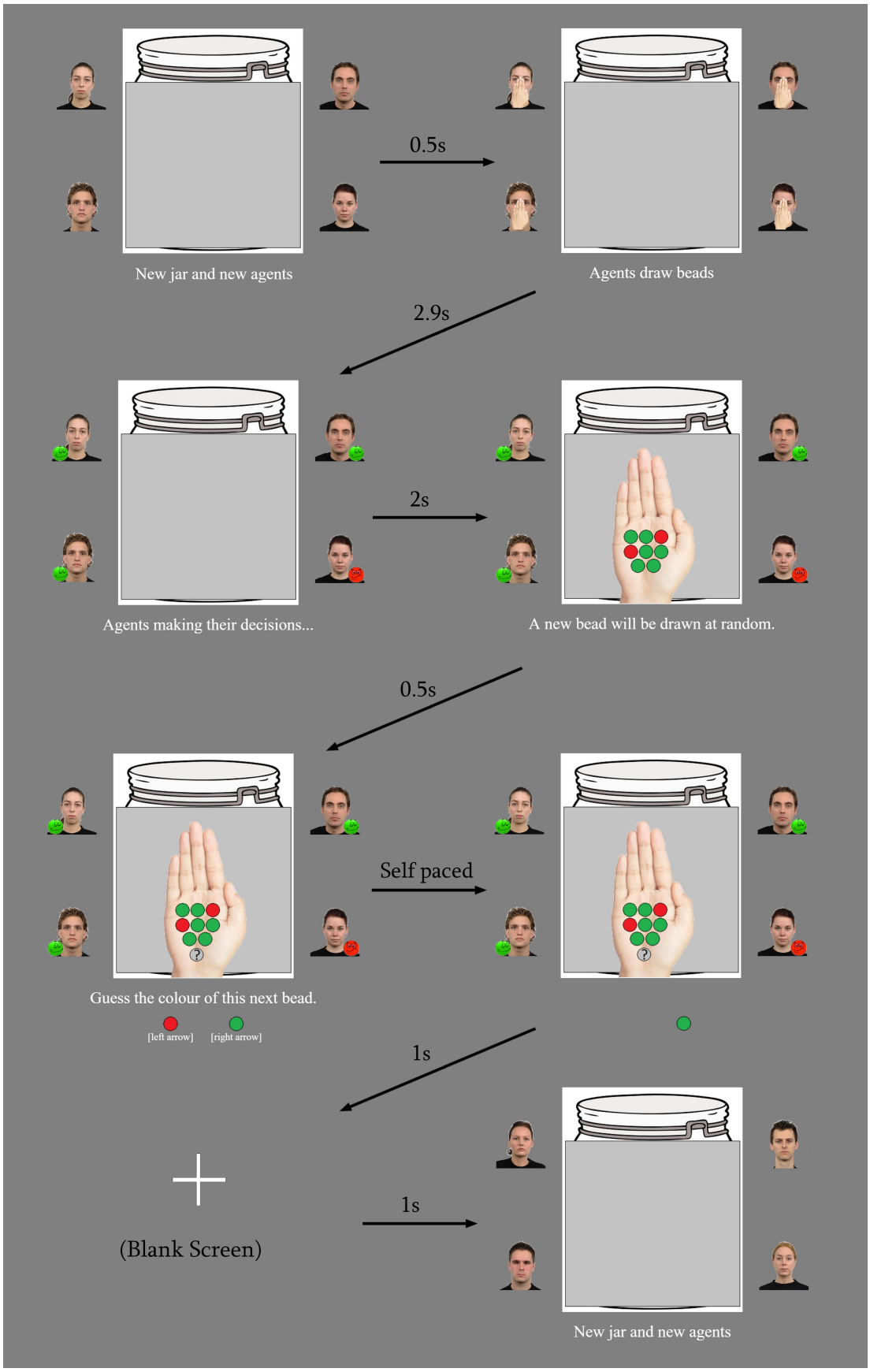
0.5s
New jar and new agents
Agents draw beads
2.9s
2s
Agents making their decisions...
A new bead will be drawn at random.
0.5s
Self paced
Guess the colour of this next bead.
[left arrow]
[right arrow]
1s
1s
(Blank Screen)
New jar and new agents

Figure 3.2: (Previous page.) The full procedure for 1 trial of the behavioural experiment. Times indicate how long the previous frame is shown for before switching to the next frame. At the start of each trial a new jar and set of agents are shown to the participant for 0.5s. The agents each draw beads, taking up 2.9s to make their decisions. Uncertain choices take slightly longer to appear than certain ones. The participant is given 2s to observe the agents decisions before being presented with their own set of beads. The participant then guesses the colour of the next bead, after which their selection remains onscreen for 1s. A blank screen is then presented for 1s before the next trial begins. See Figure 3.2 for the full graphical procedure of the experiment.

3.4 Evaluation Form

As the modified beads task described above, and first introduced by Simonsen et al. (2021), is a relatively novel experiment, an initial evaluation of the experiment was deemed important. This was in order to ensure that the participants were engaged and retained a consistent level of focus throughout, that all task instructions were clear, and to assess whether the design of the task, such as the choices of the agents being independent of the participant's beads, affected how the participants performed.

Five open-ended questions were asked to the participant. These included:

1. How did you feel whilst performing the experiment? (E.g., excited, bored, fatigued, curious, etc.)

2. Did you have any issues whilst performing the experiment? Were the instructions clear? Is there anything that would have made the experience better?

3. How long did you (approximately) spend on the entire experiment? Did it feel long, short, or fine? Did this affect how you behaved in the experiment as it went on?

4. The decisions of the agents were independent of the selected jar. Did you notice this? If so, did it change how you made your decisions?

5. Is there any other feedback you would like to provide? (optional)

Each of the questions attempted to gather information on how the participants felt and performed during the experiment, such that this feedback could be used to gauge qualitatively how effective and engaging the experiment was.

3.5 Models Development

For computational modelling of the experimental data, three models were implemented: Simple Bayes (SB), Weighted Bayes (WB) and Circular Inference (CI). These were each introduced in Section 2.2, but will be reproduced here for convenience. The development of these three models for the specific case of the beads task presented here is taken from Simonsen et al. (2021).

3.5.1 Simple Bayes

SB is computed, in log odds form, as a simple summation of the likelihood log odds and the prior log odds,

$$L_r = L_s + L_o \tag{3.1}$$

where L_s is the likelihood (sensory information) and L_o the prior. Recall that L_r is the posterior log odds for the probability of the next bead being *red*, and is computed as $L_r = \log\left(\frac{p_r}{1-p_r}\right)$, where p_r is the posterior probability of the next bead being red given the number of red beads that were drawn from the jar. The likelihood is determined from the number of red and green beads shown to the participant, and is computed as the log of the ratio of the number of red beads to the number of green beads,

$$L_s = \log\left(\frac{n_r}{n_g}\right) \tag{3.2}$$

where n_r is the number of red beads and n_g the number of green beads. The prior is determined from the guesses and confidence of the agents, and is computed as the summation of the agents confidences and guesses,

$$L_o = \sum_{k=1}^{4} C_k O_k. \tag{3.3}$$

The term C_k is the confidence of the k^{th} agent, and is equal to 0.5 if the agent is 'unconfident', and equal to 1 if the agent is 'confident'. The term O_k represents the choice of the agent, and is equal to a log odds of 0.9 if the agent chose 'red', and equal to a log odds of 0.1 if the agent chose 'green'. I.e., $O_k = \log\left(\frac{0.9}{0.1}\right)$ if the choice is red, and $O_k = \log\left(\frac{0.1}{0.9}\right)$ if the choice is green.

3.5.2 Weighted Bayes

Recall that the equation for the WB case is,

$$L_r = F(L_s, w_s) + F(L_o, w_o) \tag{3.4}$$

where the function $F(\cdot)$ is described as,

$$F(L, w) = \log\left(\frac{we^L + 1 - w}{(1-w)e^L + w}\right). \tag{3.5}$$

The terms L_s and L_o are again computed according to Eqtns 3.2 and 3.3, whereas the weights, w_s and w_o, are constants representing how much weight is given to the likelihood and prior, and both values should fall in the range [0.5, 1], as in (Simonsen et al., 2021). Note that Eqtn 3.4 is the original model used by Jardri et al. (2017), and the one used in this study, however, Simonsen et al. (2021) modified this slightly. More will be said on this in the Discussion.

3.5.3 Circular Inference

Recall that the model for circular inference was given by,

$$L_r = F(L_s + I, w_s) + F(L_o + I, w_o) \tag{3.6}$$

where

$$I = F(\alpha_s L_s, w_s) + F(\alpha_o L_o, w_o). \tag{3.7}$$

As before, both L_s and L_o are computed according to Eqtns 3.2 and 3.3, and the weights are constants as in the WB case. The additional terms α_s and α_o are constants that indicate the number of reverberations of the likelihood and prior respectively, and range from 0 up to 6.

3.5.4 Taking Actions as Bernoulli Trials

Once the log odds of the posteriors have been calculated for either of the three models, the action of choosing the red bead in each model is modelled as a Bernoulli trial,

$$a_r | p_r \sim \text{Bernoulli}(p_r) \tag{3.8}$$

where p_r is the posterior probability of selecting a red bead, and can be computed from the posterior log odds ratio, L_r, as,

$$p_r = \frac{e^{L_r}}{1 + e^{L_r}}. \tag{3.9}$$

3.6 Model Fitting

Fitting each of the models to the behavioural data is done by maximum likelihood estimation, or equivalently by minimising the negative log likelihood (NLL) (Myung, 2003). The likelihood (separate to the likelihood of L_s) is set using the probability distribution for selecting the read bead in each trial. This is given by the Bernoulli distribution as per Eqtn 3.8, which is a distribution parameterised by the posterior probability, p_r. The likelihood function can therefore be represented as,

$$\begin{aligned}\mathcal{L}(\mathbf{p}_r|\mathbf{a}_r) &= p(\mathbf{a}_r|\mathbf{p}_r) \\ &= \prod_{i=1}^{n} p_{r,i}^{a_{r,i}}(1-p_{r,i})^{(1-a_{r,i})}.\end{aligned} \tag{3.10}$$

The terms $\mathbf{p}_r$ and $\mathbf{a}_r$ are emboldened to indicate that these are vectors of posteriors and actions for a whole sequence of trials. In the second line of Eqtn 3.10, the total number of trials is given by n, and each individual trial is indexed by i. $p_{r,i}$ is therefore the posterior probability for choosing a red bead in trial i, and $a_{r,i}$ is the action of choosing a red bead in trial i, where $a_{r,i} = 1$ if the red bead is chosen, and $a_{r,i} = 0$ if the green bead is chosen. The second line of Eqtn 3.10 is therefore the product of n probability mass functions of the Bernoulli distribution for selecting a red bead in each trial, assuming that each choice in each trial is independent of the other trials.

Taking the log of Eqtn 3.10 and negating it will then produce the NLL function,

$$NLL(\mathcal{L}) = -\sum_{i=1}^{n} \left[a_{r,i}\log(p_{r,i}) + (1-a_{r,i})\log(1-p_{r,i})\right]. \tag{3.11}$$

One could then replace $p_{r,i}$ in each trial with the definition of p_r from Eqtn 3.9, as computed from either SB, WB or CI. Since SB is actually Bayes optimal there is no fitting that can be done, though computing the NLL would give an indication as to how close to being Bayes optimal a participant was. Fitting for WB and CI would equate to fitting the parameters (w_s and w_o in the case of WB, and w_s, w_o α_s, and α_o in the case of CI), which can be done by minimising Eqtn 3.11 with respect to the parameters. The parameters are optimized with bounds, such that the weights w_s and w_o are bounded to limits of $[0,1]$ and the CI parameters α_s and α_o are bounded to the limits $[0,6]$. To perform this minimisation, the 'trust-constr' optimisation method found in the optimiser package of SciPy (Python 3.8) was used. This optimisation algorithm is based on the constrained optimiser of Lalee et al. (1998).

3.7 Model Evaluation

An important method in evaluating the effectiveness of each model is to perform parameter and model recovery (Wilson and Collins, 2019). These methods check that the correct parameters and model can accurately be recovered. The basic principle is that one first simulates fake data sets using a pre-selected model and parameter values, and then using these sets fit the models and parameters and check how close the best fits are to the original values. This way can one be more confident that the fitting procedure outlines above gives meaningful results.

3.7.1 Parameter Recovery

In parameter recovery, a set of parameters are initially chosen for a particular model, and using these parameters, a fake data set is then simulated. For this study, a series of actions can be simulated using Bernoulli trials as per Eqtn 3.8. The task of parameter recovery is to then reverse the process and, using the model method described in Section 3.6, check whether the original parameters can successfully be recovered from the simulated data set.

For each of the WB and CI models this process was performed 100 times using 100 randomly sampled parameter values. Each of the weights, w_s and w_o, were sampled from a Uniform distribution with lower and upper bounds of 0 and 1, respectively. The CI parameters, α_s and α_o, were sampled from a Uniform distribution with lower and upper bounds of 0 and 6, respectively.

3.7.2 Model Recovery

The task in model recovery is to again simulate a fake data set using a set of pre-selected parameter values, but here the task is to fit all models and test which model then fits the data set best. So for instance, one could simulate data using the WB model, and then fit the WB and CI model to this data set. The NLL score can be used for each of the three models (SB, WB and CI) indicating which model fits the data best. A simple method for assessing model fit is the Bayesian Information Criterion, or BIC (Wilson and Collins, 2019). The BIC penalises the number of free parameters in the model, which therefore not only measures how good the fit is, but also how simple (and therefore least prone to overfitting) the model is. BIC is given as,

$$BIC = 2N\hat{L}L + k_m \log(n). \tag{3.12}$$

The term $N\hat{L}L$ is the NLL for the best fitting parameters of the model, whilst the term k_m is the number of parameters that model m has. n is the number of trials. The best fitting model is the one with the smallest BIC score.

For each of the three models 100 data sets were simulated using 100 parameters determined using the same Uniform distributions as was used in parameter recovery. This amounted to a total of 300 data sets (100 for each model being simulated). For each simulated data set the WB and CI models were then fit, and a BIC score was then computed for all three models. The model with the smallest BIC was then deemed the winning model.

3.8 Code Availability

All code and anonymised data is available at the following Github repository: https://github.com/MatthewTWhelan/circular_inference_BD.

Chapter 4

Results

4.1 Questionnaires

There was a total of 6 factors that were measured via the two questionnaires. The TEMPS-A questionnaire measured 5 factors (cyclothemia, depressive, irritable, hyperthermi, anxious), whilst the AQ questionnaire measured the autistic traits factor. Normalised scores across all 6 factors are shown in Table 4.1. The distributions of the normalised scores for these 6 factors are shown in Figure 4.1. In general, there appears to be a relatively even distribution of scores for all participants across all 6 factors. A full set of the scores for all participants is provided in the Appendix.

The focus of this study is on BD, which could be most closely related to the cyclothemia factor. It is interesting to note potential relationships between cyclothemia

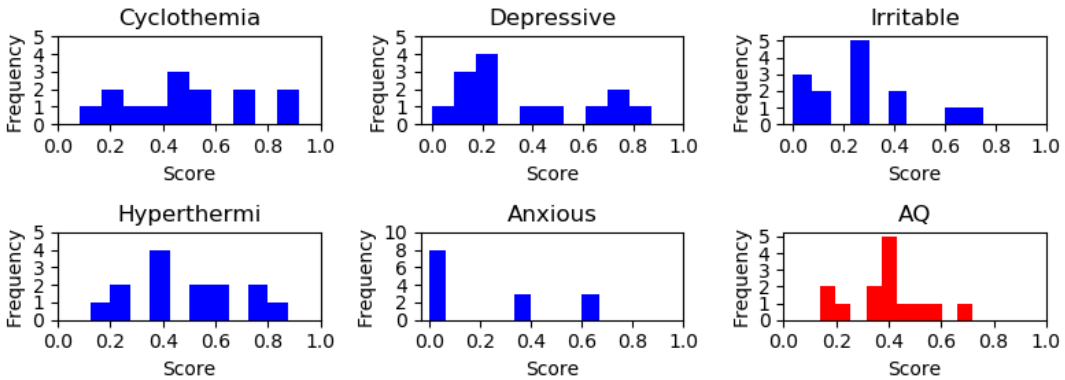

Figure 4.1: Frequencies of the questionnaire scores for all participants. The 5 temperament factors are coloured in blue, and the Autism-Spectrum Quotient is coloured in red.

Participant	C	D	I	H	A	AQ
P1	0.50	0.13	0.00	0.38	0.00	0.40
P2	0.92	0.75	0.75	0.38	0.33	0.46
P3	0.42	0.25	0.00	0.88	0.00	0.40
P4	0.33	0.38	0.13	0.38	1.00	0.52
P5	0.42	0.75	0.25	0.13	1.00	0.72
P6	0.17	0.25	0.25	0.13	0.00	0.60
P7	0.17	0.25	0.63	0.75	0.00	0.24
P8	0.08	0.13	0.13	0.75	0.00	0.18
P9	0.50	0.50	0.00	0.25	0.00	0.32
P10	0.42	0.25	0.25	0.38	0.67	0.38
P11	0.67	0.00	0.25	0.38	0.67	0.14
P12	0.67	0.88	0.38	0.63	0.33	0.40
P13	0.25	0.13	0.38	0.38	0.00	0.40
P14	0.83	0.63	0.25	0.63	0.00	0.36

Table 4.1: Normalised scores for all participants in the questionnaires. C = Cyclothemia; D = Depressive; I = Irritable; H = Hyperthermi; A = Anxious; AQ = Autism-Spectrum Quotient.

and the other temperament factors however. Figure 4.2 plots the cyclothemia score against the other 4 temperament. The depressive factor has a significant positive correlation with cyclothemia (Pearson correlation coefficient (r) $= 0.554$; p-value $= 0.040$). The other correlations are not significant, but worth noting. Cyclothemia with: hyperthermi ($r = -0.090$; p-value $= 0.760$); irritable ($r = 0.239$; p-value $= 0.410$); anxious ($r = 0.278$; p-value $= 0.336$).

Another set of relationships of interest are the temperament scores with the AQ scores. Figure 4.3 plots the AQ scores against all 5 of the temperament scores. Whilst none of these relationships are significant, there are two that are worth noting as being near significant. There is a near significant positive correlation between the depressive factor and AQ ($r = 0.489$; p-value $= 0.076$), and a near significant negative correlation between hyperthermi and AQ ($r = -0.482$; p-value $= 0.081$). The other correlations were not significant, but again worth noting. AQ with: cyclothemia ($r = -0.004$; p-value $= 0.988$); irritable ($r = 0.007$; p-value $= 0.980$); anxious ($r = 0.267$; p-value $= 0.356$).

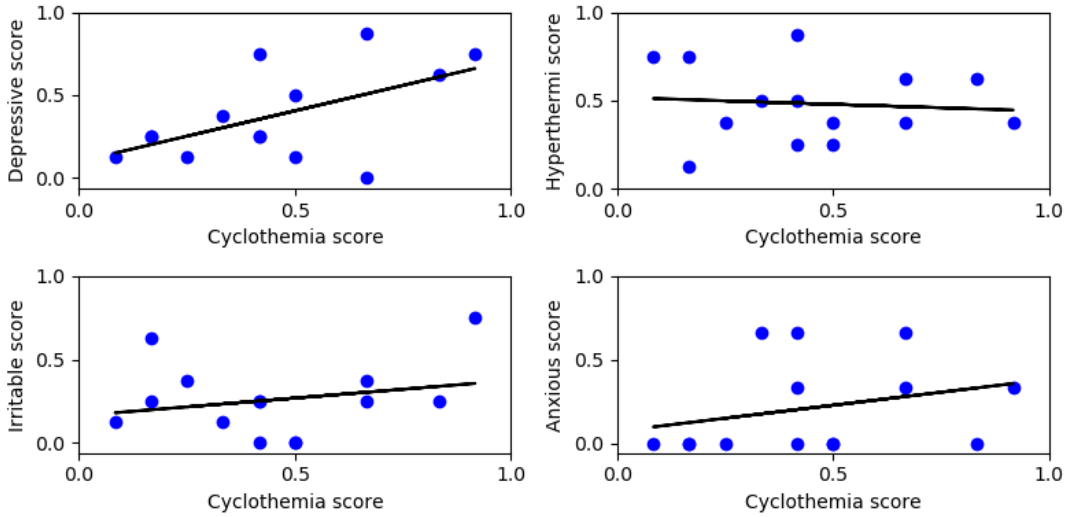

Figure 4.2: Plots of the cyclothemia factor against the other temperament factors for all participants.

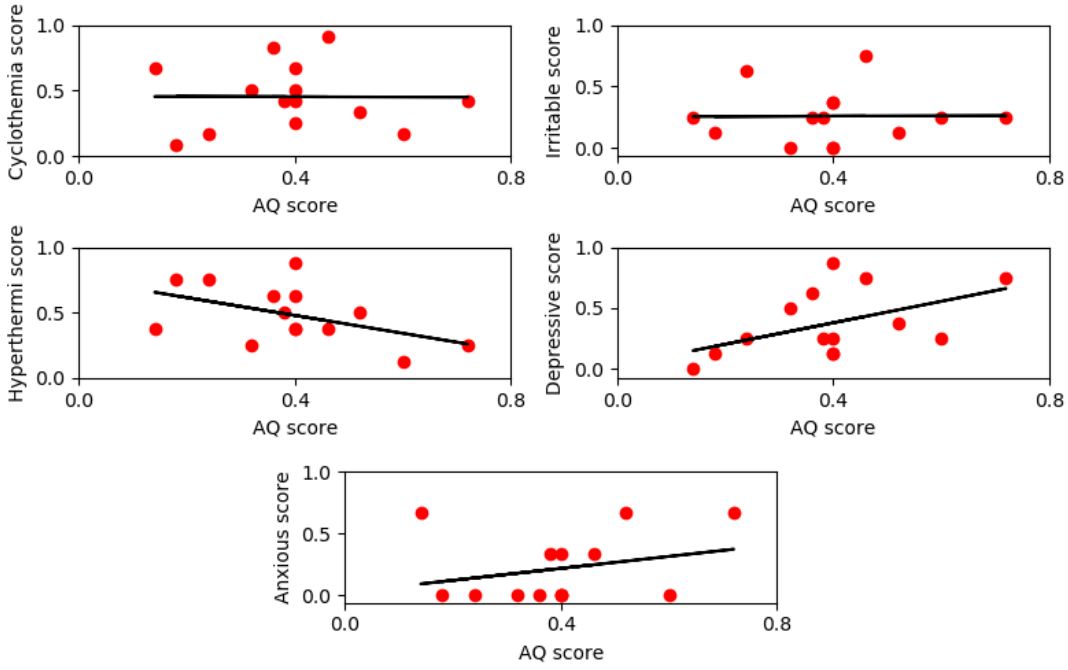

Figure 4.3: Plots of the Autism-Spectrum Quotient score against all 5 of the temperament scores for all participants.

4.2 Model Evaluation

4.2.1 Parameter Recovery

The results of performing parameter recovery with the WB and CI models are shown in Figures 4.4 and 4.5, respectively. The plots show the parameters used to simulate the data against the parameters fit against that data. In the case of WB, there is good parameter recovery for w_s ($r = 0.913$; p-value = 6e-40), and an even better recovery for w_o ($r = 0.985$; p-value = 2e-76). For CI the parameter recovery is not as strong, in particular for the CI reverberation weights α. Parameter recovery for w_s is mediocre ($r = 0.506$; p-value = 7.91e-8), whilst for w_o it is slightly better ($r = 0.797$; p-value = 3.62e-23). Parameter recovery for α_s is very poor ($r = 0.011$; p-value = 0.916), whilst it is somewhat better for α_o, though still poor ($r = 0.106$; p-value = 0.295).

4.2.2 Model Recovery

The results of performing 300 model recovery simulations is shown in the confusion matrix of Figure 4.6. The confusion matrix values are normalised, showing the proportion of occasions that each model won given the model that simulated the data. When SB is used to simulate the data, the SB model successfully wins on all occasions. When

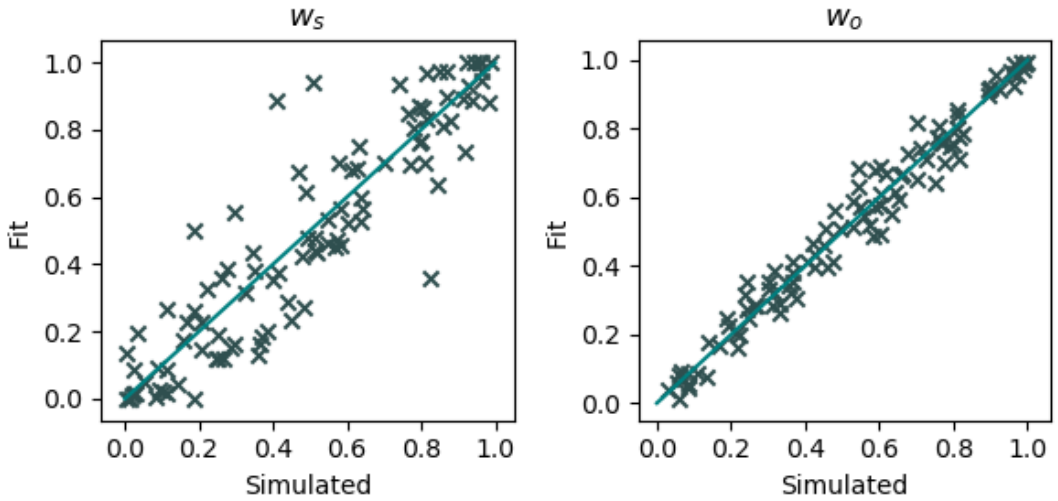

Figure 4.4: Parameter recovery for WB. 100 data sets were simulated using 100 randomly sampled parameters. The parameters used to simulate the data are plotted against the parameters recovered from fitting the model. Centre lines indicate perfect recovery.

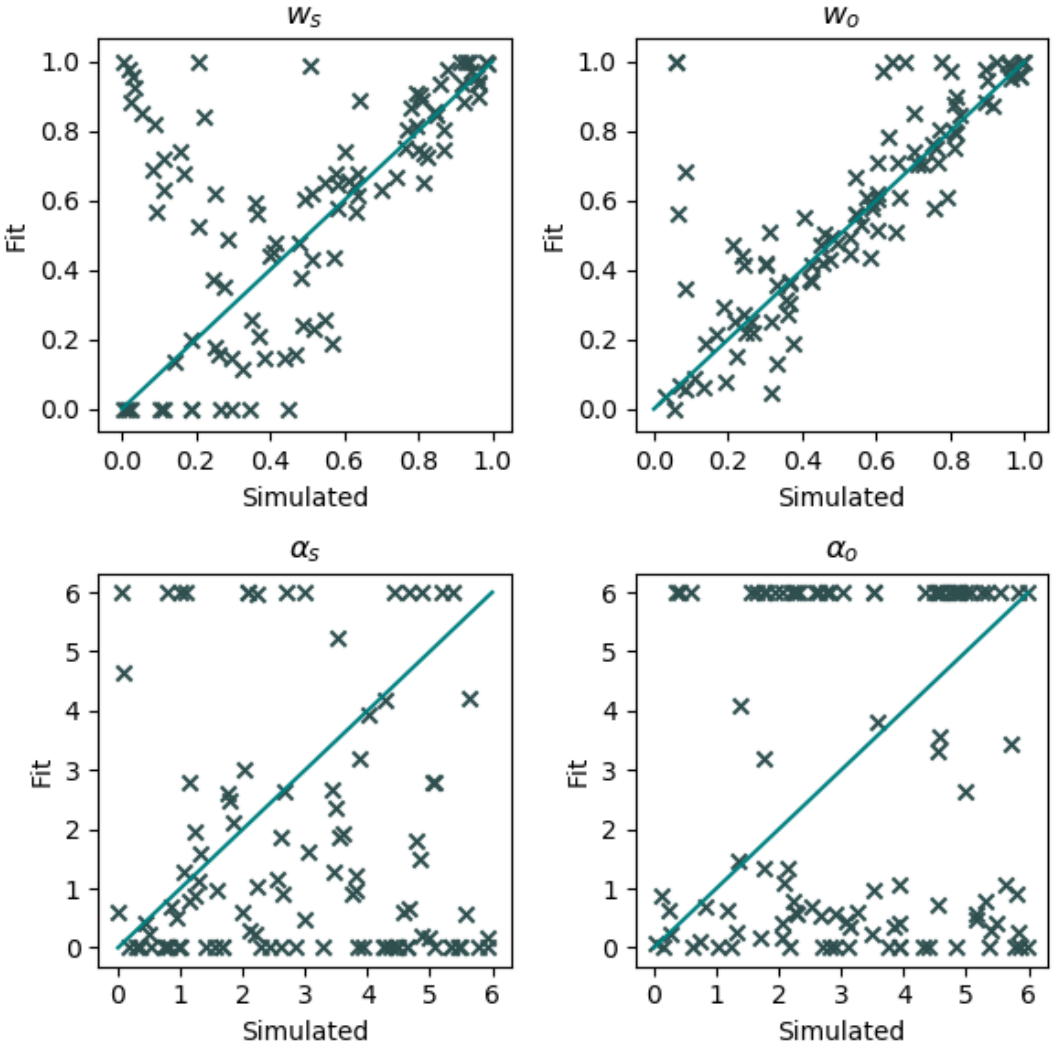

Figure 4.5: Parameter recovery for CI. 100 data sets were simulated using 100 randomly sampled parameters. The parameters used to simulate the data are plotted against the parameters recovered from fitting the model. Centre lines indicate perfect recovery.

WB is used to simulate the data, WB still achieves the correct recovery in 91% of the occasions. When CI is used to simulate the data however, successful recovery only occurs in 10% of the occasions. At 76%, WB is most often the winning model when CI is used to model the data.

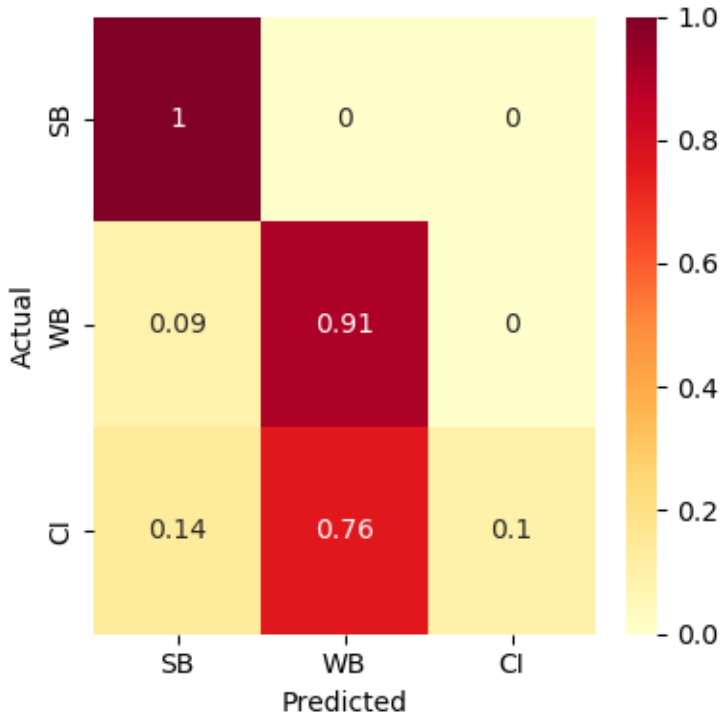

Figure 4.6: Confusion matrix from the model recovery. 100 data sets were simulated for each of the three models, after which each model was fit to the data sets. The proportion of occasions for which each model won, computed using the BIC scores (x-axis), given the actual model that generated that data set (y-axis) are shown here.

4.3 Model Fitting

The WB and CI models were fit to the behavioural data of the 7 participants for which data was gathered. A summary of the fit parameters are shown in Table 4.2.

	WB		CI			
Participant	w_s	w_o	w_s	w_o	α_s	α_o
P1	0.50	1.00	0.98	1.00	0.41	2.65
P2	1.00	0.48	1.00	0.42	5.83	0.00
P5	1.00	1.00	1.00	1.00	0.13	0.00
P8	1.00	0.57	1.00	0.80	5.14	0.00
P9	0.33	0.47	0.28	0.46	0.62	0.29
P10	0.54	0.52	0.54	0.52	0.00	6.00
P14	0.00	0.36	0.00	0.25	0.00	0.18

Table 4.2: A summary of the fitted parameters for the WB and CI models for each participant.

The NLL and BIC scores for each of the models per participant are shown in Table 4.3, with the winning models highlighted in bold for each participant. SB was the winning model in 2 of the participants, WB was the winning model in 3 of the participants, and CI was the winning model in 2 of the participants.

	SB		WB		CI	
Participant	$\hat{NLL}$	*BIC*	$\hat{NLL}$	*BIC*	$\hat{NLL}$	*BIC*
P1	4.11	**8.23**	2.39	14.09	0.00	18.62
P2	284.17	568.35	37.58	84.46	14.56	**47.74**
P5	9.01	**18.02**	8.96	27.23	8.85	36.31
P8	236.42	472.85	37.43	84.17	14.83	**48.29**
P9	350.27	700.54	71.10	**151.51**	71.07	160.75
P10	331.48	662.97	72.65	**154.61**	72.65	163.92
P14	436.87	873.74	40.46	**90.24**	40.34	99.30

Table 4.3: A summary of the negative log-likelihoods for the fitted models and the BIC scores. The BIC scores in bold represent the winning model for that participant.

4.4 Experiment Evaluation

The evaluation of the experiment was largely subjective, and based on feedback provided by the participants at the end of the experiment. Only 11 participants answered the evaluation form. Despite this, there were some responses that were frequently stated by a number of the participants. Some notable features from the evaluation are provided for each question.

4.4.1 Question 1

The overwhelming response to how the participants felt during the experiment was one of 'bored', which appeared in 5 of the participants' answers. This feeling of boredom went along with feedback that the task was 'repetitive' and 'laborious'. One participant in particular mentioned how 'pushing through the 105 trials was a slog'.

4.4.2 Question 2

In response to whether there were any particular issues or whether the instructions were clear, the majority mentioned there were none and that the instructions were indeed clear. One participant did mention that they would 'like to know the reasoning behind the jars and beads part', suggesting a clearer description of how the beads task relates to the scientific hypothesis proposed may be a valuable inclusion.

4.4.3 Question 3

For the participants that provided a time, the range of times that it took them to complete the experiment were between 25 minutes and 35 minutes. 5 of the participants mentioned that it felt too long, with one participant even becoming 'a bit disinterested'. One participant even became 'impatient', and had to 'resist the urge to rush'. Despite this, a number of the participants (3) said it did not affect their behaviour throughout the experiment, whilst 1 mentioned that it may have affected their answer as they became less inclined to think about their answer due to the repetitive nature of the task.

4.4.4 Question 4

In response to whether the participants noticed that the agents chose independently, and whether this affected their behaviour, there was quite a mixed response. 4 of the

participants said they noticed that the agents were choosing independently, whilst 2 did not notice it. 2 of the participants decided to go with their own decisions, and paid little attention to what the agents chose, whilst 1 participant based their choices 'solely on what the agents were saying'. 1 participant didn't understand what was going on, and so was 'just picking the colours at random'.

4.4.5 Question 5

The final question asked the participants if they had any further comments. 3 participants had mentioned their observation that the agents would never agree with the same confidence if they chose different colours. 1 participant mentioned that 'it could've been more realistic if the agents acted independently' from one another.

Chapter 5

Discussion

This dissertation has presented a preliminary analysis on the potential application of the circular inference (CI) model to bipolar disorder (BD). This preliminary analysis did not explicitly include BD patients, but instead was conducted from a small sample of colleagues of the author. The CI model developed by Jardri and Denève (2013); Jardri et al. (2017) has been evaluated using a new experimental paradigm of probabilistic decision making designed by Simonsen et al. (2021). To evaluate the circular inference model, parameter recovery and model recovery analyses were conducted alongside the simple Bayes (SB) and weighted Bayes (WB) models. Each of these three models, SB, WB and CI, were fit to the behavioural data gathered from the participants and BIC scores were computed to assess the best fitting model for each participant. Additionally, two questionnaires were used to assess the participants with regards to temperament traits and autistic traits. The temperament traits were composed of 5 factors, including one of *cyclothemia*, which is closely related to BD (American Psychiatric Association, 2013). Correlations between each of the temperament traits were explored, alongside an exploration of these traits with respect to autistic traits.

The parameter recovery analysis was conducted for both the WB and CI models. In the case of WB parameter recovery was very good, giving confidence that fitting the WB model to behavioural data with the assumption that a WB model produced that data would provide a sufficiently accurate fitted model. Unfortunately, the recovery was weak for CI. This weakness of recovery in the CI model has been shown before (Chrysaitis et al., 2021), although it is particularly weak in this study. The same weakness of recovery for CI was found in the model recovery analysis too. Whereas good recovery was seen in the SB and WB models, CI was only able to recover the correct

model in 10% of the simulations. This raises some questions as to the effectiveness of the CI model.

Of the three models explored in this study, the SB model was taken from Simonsen et al. (2021). Their model gave the likelihood and prior terms as functions of the number of red beads and the decisions/confidences of the 4 agents. However, given these likelihood and prior terms, Simonsen et al. (2021) then altered the WB and CI models slightly from those originally designed by Jardri et al. (2017). For instance, Jardri et al. (2017) used the calculation $L_r = F(L_s, w_s) + F(L_o, w_o)$ to compute the posterior. And this was the form used in this study too, as shown in Eqtn 3.4. Simonsen et al. (2021) meanwhile used a slightly modified version in which the agent confidences were included in the w_o portion of the above Eqtn, rather than in the L_o portion. To be more accurate, their expression was something closer to $L_r = F(L_s, w_s) + \sum_k F(O_k, w_o + \beta C_k)$, where O_k is the choice of agent k, and C_k is the confidence of agent k. They also included therefore an additional term, β, into their computation. And the CI model is also modified to be different to that used by Jardri et al. (2017). It could be then that these modifications were necessary to overcome the shortcomings of using the original WB and CI models. For instance, by including the β term, it is accounting for the possibility that each of the participants might weight the confidences of the agents differently. Despite these modifications, the model recovery analysis performed by Simonsen et al. (2021) gave 60% accuracy for the CI model. So although significantly better than the model recovery found in this study, it still leaves possible questions over the accuracy of the CI model.

Despite these doubts, it is perhaps important to note that when either the SB or WB model was used to simulate data sets, the CI was never inferred as the best model during model recovery. This is significant, as 2 of the participants were best fit by the CI model. This result would suggest that the participants were indeed behaving in line with CI, and not according to SB or WB. It is further interesting to note that of these 2 participants, one of them scored the highest cyclothemia score amongst all participants (P2). Whilst there is not enough data available to infer anything of significance, this may be somewhat indicative. Indeed, the participant with the second highest cyclothemia score (P14) performed worse than all other participants in comparison to the optimal strategy of SB, indicating an aberration in performance away from SB and further evidence of increased weightings or interactions in the likelihood and/or prior.

One possible source of inaccuracy with using the WB and CI models of Jardri et al. (2017) with the SB terms of Simonsen et al. (2021) is that the prior might be over-

counted for. This is due to their being 4 agents, with each agent offering a maximum contribution of $\log(0.9/0.1)$ to the prior term. Compared to the maximum contribution of the likelihood, computed as $\log(7)$, the prior has a maximum value that is around 4.5 times larger than the likelihood. Rationally, this seems to make sense. If a participant trusts that the agents are guessing to the same degree of accuracy as the participant themselves, then the guesses of 4 agents should provide a greater deal of evidence than the single participant alone. However, due to the design of this experiment, some participants recognised that the agents were acting independently and chose to ignore the agents, going only with their own decisions. Participant P14 adopted this strategy, as documented in their evaluation form, and indeed performed worse compared to SB. Participant P1 on the other hand chose to put more trust in the agents than in their own sensory information, again as documented in the evaluation form. And for this participant, they had the lowest (and therefore best) BIC score for SB. The closer the participant is to performing according to the SB strategy, the less likely they are to be best modelled by WB or CI with regards to the BIC score. But due to the biased nature of the prior versus the likelihood, greater trust in the agents takes one closer to an optimal, SB strategy, than does having greater trust in one's own sensory information.

An additional point to raise is the range for the α parameters in the CI model. The range of values that was chosen for this was between 0 and 6. However, the range that was chosen by Jardri et al. (2017) was between 0 and 60. There is no particular reason for choosing an upper limit of 6, other than selecting 60 gave somewhat worse performance in the parameter recovery (data not provided). Given time constraints a more thorough exploration of the effects of using different upper limits was not conducted. However, it could be useful to do such an analysis, and to, for instance, sample a range of upper limits by perhaps performing a grid search over those parameters, and using this to estimate the most promising range for the α terms.

Despite some of the potential drawbacks of this study, it does raise some interesting questions. For instance, does the number of agents matter to the decision of the participant? If there were 4 agents versus 10 agents, would the participant place equal weighting in the decision of all agents combined, or does the weighting increase as the number of participants increases? And if so, how would this affect the modelling choices? Including agents in this task produces a similar effect to the 'conformity versus independence' question raised in the Asch conformity studies (Friend et al., 1990). In addition, research has shown that participants may behave differently when being asked to respond publicly in front of the group compared to when responding privately

(Bond, 2005). This again raises the question as to whether participants would behave differently in front of 4 other people face-to-face, rather than doing so through the screen of a computer, and particularly doing so in front of simulated agents. A number of the participants were able to recognise that the agents acted independently of the actual jar chosen. One potential method to work round this could be to use the practice portions of the experiment for which the participants took part in at the start of the experiment, where they were made to give confidences. This data could be used for other experiments, in replace of the simulated agents choices. This would remove the independent nature of the agent guesses, making for a potentially more realistic experiment.

Finally, it is worth noting the technical error that occurred with participants' experimental data. The experiment itself was designed using PsychoPy (Peirce et al., 2019). Unfortunately, in its current version, PsychoPy does not allow for questionnaire type forms to be designed, hence why the external site Qualtrics was instead used. But this meant editing the JavaScript code that had been generated by PsychoPy in order to automatically redirect the participant away from the experiment and to the evaluation form on Qualtrics. It is believed that this could have caused an issue in some browsers, since the line of code that was used to redirect the participant was placed just prior to PsychoPy's termination code. Unfortunately, placing it after the termination code would result in the redirection not working. For future experiments then, this should be carefully tested, ideally by testing the experiment on all the main browsers. Ideally the experiment could be designed from scratch, rather than using an external package like PsychoPy, for better control over the design of the experiment.

5.1 Future Work

There are two main directions for future work. One direction would be in improving the modelling technique, whilst the other in designing an improved experiment.

On the modelling approach, it would appear sensible to adopt the modified though slightly more complex model by Simonsen et al. (2021) for the WB and CI inference models. Whilst they do have additional parameters compared to using the simpler approach by Jardri et al. (2017), these seem to prove necessary for adequate parameter and model recovery. It may also prove useful to perform a search over the parameters when performing parameter recovery, in order to determine sensible upper and lower bounds and the region in which best recovery is possible. This is because if the recov-

ery is not reliable for a given set of bounds, it is not possible to place much confidence on the fitted parameters for a given data set.

In terms of the experimental design, there are a number of modifications that could be made based upon the feedback received from the participants. An important modification would be to make the agents decisions more realistic. The current approach was to use 35 combinations beads and agent decisions, such that agent decisions are independent of the beads, as well as give agents who chose the same colour to have the same confidence, with the other agents having a different colour and confidence. One means to avoid this could that, for a randomly selected jar, to sample a number of beads for each agent, and use the sample to produce a colour guess and confidence. For instance, if there are 1 or 2 reds, an agent could make the decision 'confident green'. For 3 reds, 'unconfident green'. For 4 reds, a random decisions with 50% probability of either red or green coupled with being 'unconfident', etc. This may produce more realistic decision making in the agents in order to ensure the participants do not ignore the decisions of the agents. It may also be useful to test, via simulation, the minimum number of trials needed to accurately fit the models. One could use parameter and model recovery using a variety of different trial numbers, in order to test how many trials is necessary before recovery accuracy reaches an asymptote. Ideally, the less number of trials needed the better, given a large number of the participants found the experiment to be 'long' and 'boring'. Finally, there seems to be no reason not to simplify the experiment, particularly given 2 of the participants did not understand how to do the experiment. For instance, it could make sense to employ only 2 jars rather than the 9 jars used in this experiment. These 2 jars could have opposing proportions of red and green beads, such as a 40:60 split in one and a 60:40 split in the other. Rather than guess the colour of the next bead, the participant could guess which jar has been chosen given a random sample of x number of beads. This is very much like the original beads task (Huq et al., 1988), except with the addition of the social element via the 4 agents.

Appendix A

Questionnaires

A.1 TEMPS-A

The questions taken from TEMPS-A (Akiskal et al., 2005) were as shown on the next 3 pages. For each question the participant answered ‘Yes’, a score of 1 was added.

Personality Questionnaire Part 1 of 2. Please answer 'Yes' for the following items only if they apply to you for much of your life.

	Yes	No
My ability to think varies greatly from sharp to dull for no apparent reason.	○	○
I constantly switch between being lively and sluggish.	○	○
I get sudden shifts in mood and energy.	○	○
The way I see things is sometimes vivid, but at other times lifeless.	○	○
My mood often changes for no reason.	○	○
I go back and forth between being outgoing and being withdrawn from others.	○	○
My moods and energy are either high or low, rarely in between.	○	○
I go back and forth between feeling overconfident and feeling unsure of myself.	○	○

	Yes	No
My need for sleep varies a lot from just a few hours to more than 9 h.	○	○
I sometimes go to bed feeling great, and wake up in the morning feeling life is not worth living.	○	○

I can really like someone a lot, and then completely lose interest in them.	○	○
I am the kind of person who can be sad and happy at the same time.	○	○
People tell me I am unable to see the lighter side of things.	○	○
I'm the kind of person who doubts everything.	○	○
I am a very skeptical person.	○	○
I am by nature a dissatisfied person.	○	○

	Yes	No
I'm a sad, unhappy person.	○	○
I think things often turn out for the worst.	○	○
I give up easily.	○	○
I complain a lot.	○	○
People tell me I blow up out of nowhere.	○	○
I can get so furious I could hurt someone.	○	○
I often get so mad that I will just trash everything.	○	○
When crossed, I could get into a fight.	○	○

	Yes	No
When I disagree with someone, I can get into a heated argument.	○	○
When angry, I snap at people.	○	○
I am known to swear a lot.	○	○
I have been told that I become violent with just a few drinks.	○	○
I have a gift for speech, convincing and inspiring to others.	○	○
I often get many great ideas.	○	○
I love to tackle new projects, even if risky.	○	○
I like telling jokes, people tell me I'm humorous.	○	○

	Yes	No
I have abilities and expertise in many fields.	○	○
I am totally comfortable even with people I hardly know.	○	○
I love to be with a lot of people.	○	○
I am the kind of person who likes to be the boss.	○	○
I am often fearful of someone in my family coming down with a serious disease.	○	○
I'm always thinking someone might break bad news to me about a family member.	○	○
When someone is late coming home, I fear they may have had an accident.	○	○

A.2 Autism-Spectrum Quotient

The questions taken from the autism-spectrum quotient (Baron-Cohen et al., 2001) are as shown on the next several pages. For questions 1, 2, 4, 5, 6, 7, 9, 12, 13, 16, 18, 19, 20, 21, 22, 23, 26, 33, 35, 39, 41, 42, 43, 45, 46, an answer of "Definitely Agree" or "Slightly Agree" scores 1 point. For all other questions, an answer of "Definitely Disagree" or "Slightly Disagree" scores 1 point.

Personality Questionnaire Part 2 of 2.
Please answer how much you agree with each statement. Again, please do not spend too long thinking about each question.

	Definitely Agree	Slightly Agree	Slightly Disagree	Definitely Disagree
I prefer to do things with others rather than on my own.	○	○	○	○
I prefer to do things the same way over and over again.	○	○	○	○
If I try to imagine something, I find it very easy to create a picture in my mind.	○	○	○	○
I frequently get so strongly absorbed in one thing that I lose sight of other things.	○	○	○	○
I often notice small sounds when others do not.	○	○	○	○
I usually notice car number plates or similar strings of information.	○	○	○	○
Other people frequently tell me that what I've said is impolite, even though I think it is polite.	○	○	○	○
When I'm reading a story, I can easily imagine what the characters might look like.	○	○	○	○
I am fascinated by dates.	○	○	○	○

	Definitely Agree	Slightly Agree	Slightly Disagree	Definitely Disagree
In a social group, I can easily keep track of several different people's conversations.	○	○	○	○
I find social situations easy.	○	○	○	○
I tend to notice details that others do not.	○	○	○	○
I would rather go to a library than a party.	○	○	○	○
I find making up stories easy.	○	○	○	○
I find myself drawn more strongly to people than to things.	○	○	○	○
I tend to have very strong interests, which I get upset about if I can't pursue.	○	○	○	○
I enjoy social chit-chat.	○	○	○	○
When I talk, it isn't always easy for others to get a word in edgeways.	○	○	○	○

	Definitely Agree	Slightly Agree	Slightly Disagree	Definitely Disagree
I am fascinated by numbers.	○	○	○	○
When I'm reading a story, I find it difficult to work out the characters' intentions.	○	○	○	○
I don't particularly enjoy reading fiction.	○	○	○	○
I find it hard to make new friends.	○	○	○	○
I notice patterns in things all the time.	○	○	○	○
I would rather go to the theatre than a museum.	○	○	○	○
It does not upset me if my daily routine is disturbed.	○	○	○	○
I frequently find that I don't know how to keep a conversation going.	○	○	○	○
I find it easy to "read between the lines" when someone is talking to me.	○	○	○	○

	Definitely Agree	Slightly Agree	Slightly Disagree	Definitely Disagree
I usually concentrate more on the whole picture, rather than the small details.	○	○	○	○
I am not very good at remembering phone numbers.	○	○	○	○
I don't usually notice small changes in a situation, or a person's appearance.	○	○	○	○
I know how to tell if someone listening to me is getting bored.	○	○	○	○
I find it easy to do more than one thing at once.	○	○	○	○
When I talk on the phone, I'm not sure when it's my turn to speak.	○	○	○	○
I enjoy doing things spontaneously.	○	○	○	○
I am often the last to understand the point of a joke.	○	○	○	○
I find it easy to work out what someone is thinking or feeling just by looking at their face.	○	○	○	○

	Definitely Agree	Slightly Agree	Slightly Disagree	Definitely Disagree
If there is an interruption, I can switch back to what I was doing very quickly.	○	○	○	○
I am good at social chit-chat.	○	○	○	○
People often tell me that I keep going on and on about the same thing.	○	○	○	○
When I was young, I used to enjoy playing games involving pretending with other children.	○	○	○	○
I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant, etc.).	○	○	○	○
I find it difficult to imagine what it would be like to be someone else.	○	○	○	○
I like to plan any activities I participate in carefully.	○	○	○	○
I enjoy social occasions.	○	○	○	○
I find it difficult to work out people's intentions.	○	○	○	○

	Definitely Agree	Slightly Agree	Slightly Disagree	Definitely Disagree
New situations make me anxious.	○	○	○	○
I enjoy meeting new people.	○	○	○	○
I am a good diplomat.	○	○	○	○
I am not very good at remembering people's date of birth.	○	○	○	○
I find it very easy to play games with children that involve pretending.	○	○	○	○

Bibliography

Adams, R. A., Huys, Q. J., and Roiser, J. P. (2016). Computational psychiatry: towards a mathematically informed understanding of mental illness. *Journal of Neurology, Neurosurgery & Psychiatry*, 87(1):53–63.

Akiskal, H. S., Mendlowicz, M. V., Jean-Louis, G., Rapaport, M. H., Kelsoe, J. R., Gillin, J. C., and Smith, T. L. (2005). Temps-a: validation of a short version of a self-rated instrument designed to measure variations in temperament. *Journal of affective disorders*, 85(1-2):45–52.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders*. Washington, DC.

Angst, J. and Sellaro, R. (2000). Historical perspectives and natural history of bipolar disorder. *Biological psychiatry*, 48(6):445–457.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The autism-spectrum quotient (aq): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of autism and developmental disorders*, 31(1):5–17.

Bond, R. (2005). Group size and conformity. *Group processes & intergroup relations*, 8(4):331–354.

Cade, J. F. (1949). Lithium salts in the treatment of psychotic excitement. *Medical Journal of Australia*.

Carroll, L. S. and Owen, M. J. (2009). Genetic overlap between autism, schizophrenia and bipolar disorder. *Genome medicine*, 1(10):1–7.

Chang, S.-S. and Chou, T. (2018). A dynamical bifurcation model of bipolar disorder based on learned expectation and asymmetry in mood sensitivity. *Computational Psychiatry*, pages 205–222.

Chrysaitis, N. A., Jardri, R., Denève, S., and Series, P. (2021). No increased circular inference in autism or autistic traits. *bioRxiv*.

Cipriani, A., Barbui, C., Salanti, G., Rendell, J., Brown, R., Stockton, S., Purgato, M., Spineli, L. M., Goodwin, G. M., and Geddes, J. R. (2011). Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. *The Lancet*, 378(9799):1306–1315.

Cochran, A. L., Schultz, A., McInnis, M. G., and Forger, D. B. (2017). A comparison of mathematical models of mood in bipolar disorder. In *Computational neurology and psychiatry*, pages 315–341. Springer.

Daugherty, D., Roque-Urrea, T., Urrea-Roque, J., Troyer, J., Wirkus, S., and Porter, M. A. (2009). Mathematical models of bipolar disorder. *Communications in Nonlinear Science and Numerical Simulation*, 14(7):2897–2908.

De Fruyt, J., Deschepper, E., Audenaert, K., Constant, E., Floris, M., Pitchot, W., Sienaert, P., Souery, D., and Claes, S. (2012). Second generation antipsychotics in the treatment of bipolar depression: a systematic review and meta-analysis. *Journal of psychopharmacology*, 26(5):603–617.

Doya, K., Ishii, S., Pouget, A., and Rao, R. P. (2007). *Bayesian brain: Probabilistic approaches to neural coding*. MIT press.

Friend, R., Rafferty, Y., and Bramel, D. (1990). A puzzling misinterpretation of the asch 'conformity'study. *European Journal of Social Psychology*, 20(1):29–44.

Friston, K. (2008). Hierarchical models in the brain. *PLoS computational biology*, 4(11):e1000211.

Furnham, A. and Anthony, E. (2010). Lay theories of bipolar disorder: The causes, manifestations and cures for perceived bipolar disorder. *International Journal of Social Psychiatry*, 56(3):255–269.

Garety, P., Hemsley, D., Wessely, S., et al. (1991). Reasoning in deluded schizophrenic and paranoid patients. *Journal of Nervous and Mental Disease*, 179(4):194–201.

Geddes, J. R., Burgess, S., Hawton, K., Jamison, K., and Goodwin, G. M. (2004). Long-term lithium therapy for bipolar disorder: systematic review and meta-analysis of randomized controlled trials. *American Journal of Psychiatry*, 161(2):217–222.

Geddes, J. R. and Miklowitz, D. J. (2013). Treatment of bipolar disorder. *The lancet*, 381(9878):1672–1682.

Goldbeter, A. (2011). A model for the dynamics of bipolar disorders. *Progress in biophysics and molecular biology*, 105(1-2):119–127.

Goodwin, F. K. and Jamison, K. R. (2007). *Manic-depressive illness: bipolar disorders and recurrent depression*, volume 2. Oxford University Press.

Gupta, R. D. and Guest, J. F. (2002). Annual cost of bipolar disorder to uk society. *The British Journal of Psychiatry*, 180(3):227–233.

Hammen, C. and Gitlin, M. (1997). Stress reactivity in bipolar patients and its relation to prior history of disorder. *The American journal of psychiatry*.

Hooley, J. M. (2007). Expressed emotion and relapse of psychopathology. *Annu. Rev. Clin. Psychol.*, 3:329–352.

Huq, S., Garety, P. A., and Hemsley, D. R. (1988). Probabilistic judgements in deluded and non-deluded subjects. *The Quarterly Journal of Experimental Psychology Section A*, 40(4):801–812.

Huys, Q. J., Maia, T. V., and Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature neuroscience*, 19(3):404–413.

Jardri, R. and Denève, S. (2013). Circular inferences in schizophrenia. *Brain*, 136(11):3227–3241.

Jardri, R., Duverne, S., Litvinova, A. S., and Denève, S. (2017). Experimental evidence for circular inference in schizophrenia. *Nature communications*, 8(1):1–13.

Krishnan, K. R. R. (2005). Psychiatric and medical comorbidities of bipolar disorder. *Psychosomatic medicine*, 67(1):1–8.

Lalee, M., Nocedal, J., and Plantenga, T. (1998). On the implementation of an algorithm for large-scale equality constrained optimization. *SIAM Journal on Optimization*, 8(3):682–706.

Lam, D. H., Hayward, P., Watkins, E. R., Wright, K., and Sham, P. (2005). Relapse prevention in patients with bipolar disorder: cognitive therapy outcome after 2 years. *American Journal of Psychiatry*, 162(2):324–329.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., and Van Knippenberg, A. (2010). Presentation and validation of the radboud faces database. *Cognition and emotion*, 24(8):1377–1388.

Leverich, G. S., McElroy, S. L., Suppes, T., Keck Jr, P. E., Denicoff, K. D., Nolen, W. A., Altshuler, L. L., Rush, A. J., Kupka, R., Frye, M. A., et al. (2002). Early physical and sexual abuse associated with an adverse course of bipolar illness. *Biological psychiatry*, 51(4):288–297.

Linke, J. O., Koppe, G., Scholz, V., Kanske, P., Durstewitz, D., and Wessa, M. (2020). Aberrant probabilistic reinforcement learning in first-degree relatives of individuals with bipolar disorder. *Journal of affective disorders*, 264:400–406.

Mason, L., Eldar, E., and Rutledge, R. B. (2017). Mood instability and reward dysregulation—a neurocomputational model of bipolar disorder. *JAMA psychiatry*, 74(12):1275–1276.

Merikangas, K. R., Jin, R., He, J.-P., Kessler, R. C., Lee, S., Sampson, N. A., Viana, M. C., Andrade, L. H., Hu, C., Karam, E. G., et al. (2011). Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Archives of general psychiatry*, 68(3):241–251.

Miklowitz, D. J. (2006). A review of evidence-based psychosocial interventions for bipolar disorder. *Journal of Clinical Psychiatry*, 67:28.

Miklowitz, D. J., Otto, M. W., Frank, E., Reilly-Harrington, N. A., Kogan, J. N., Sachs, G. S., Thase, M. E., Calabrese, J. R., Marangell, L. B., Ostacher, M. J., et al. (2007). Intensive psychosocial intervention enhances functioning in patients with bipolar depression: results from a 9-month randomized controlled trial. *American Journal of Psychiatry*, 164(9):1340–1347.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology*, 47(1):90–100.

Neves-Pereira, M., Mundo, E., Muglia, P., King, N., Macciardi, F., and Kennedy, J. L. (2002). The brain-derived neurotrophic factor gene confers susceptibility to bipolar disorder: evidence from a family-based association study. *The American Journal of Human Genetics*, 71(3):651–655.

Notredame, C.-E., Pins, D., Deneve, S., and Jardri, R. (2014). What visual illusions teach us about schizophrenia. *Frontiers in integrative neuroscience*, 8:63.

Novick, D. M., Swartz, H. A., and Frank, E. (2010). Suicide attempts in bipolar i and bipolar ii disorder: a review and meta-analysis of the evidence. *Bipolar disorders*, 12(1):1–9.

O'Donnell, P. (2011). Adolescent onset of cortical disinhibition in schizophrenia: insights from animal models. *Schizophrenia bulletin*, 37(3):484–492.

ONS, O. f. N. S. (2021). Retail prices index: Long run series: 1947 to 2019. `https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/cdko/mm23`.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., and Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior research methods*, 51(1):195–203.

Reynolds de Sousa, T., Correia, D. T., and Novais, F. (2021). Exploring the hypothesis of a schizophrenia and bipolar disorder continuum: Biological, genetic and pharmacologic data. *CNS & Neurological Disorders - Drug Targets*, 20:1–1.

Scott, J., Colom, F., and Vieta, E. (2007). A meta-analysis of relapse rates with adjunctive psychological therapies compared to usual psychiatric treatment for bipolar disorders. *International Journal of Neuropsychopharmacology*, 10(1):123–129.

Series, P. (2020). *Computational psychiatry: A primer*. MIT Press.

Sidor, M. M. and MacQueen, G. M. (2012). An update on antidepressant use in bipolar depression. *Current psychiatry reports*, 14(6):696–704.

Simonsen, A., Fusaroli, R., Petersen, M. L., Vermillet, A.-Q., Bliksted, V., Mors, O., Roepstorff, A., and Campbell-Meiklejohn, D. (2021). Taking others into account: combining directly experienced and indirect information in schizophrenia. *Brain*, 144(5):1603–1614.

Smith, D. J., Griffiths, E., Kelly, M., Hood, K., Craddock, N., and Simpson, S. A. (2011). Unrecognised bipolar disorder in primary care patients with depression. *The British Journal of Psychiatry*, 199(1):49–56.

Steinacher, A. and Wright, K. A. (2013). Relating the bipolar spectrum to dysregulation of behavioural activation: a perspective from dynamical modelling. *PloS one*, 8(5):e63345.

Stephan, K. E., Bach, D. R., Fletcher, P. C., Flint, J., Frank, M. J., Friston, K. J., Heinz, A., Huys, Q. J., Owen, M. J., Binder, E. B., et al. (2016a). Charting the landscape of priority problems in psychiatry, part 1: classification and diagnosis. *The Lancet Psychiatry*, 3(1):77–83.

Stephan, K. E., Binder, E. B., Breakspear, M., Dayan, P., Johnstone, E. C., Meyer-Lindenberg, A., Schnyder, U., Wang, X.-J., Bach, D. R., Fletcher, P. C., et al. (2016b). Charting the landscape of priority problems in psychiatry, part 2: pathogenesis and aetiology. *The Lancet Psychiatry*, 3(1):84–90.

Stephan, K. E., Friston, K. J., and Frith, C. D. (2009). Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. *Schizophrenia bulletin*, 35(3):509–527.

Sullivan, P. F., Magnusson, C., Reichenberg, A., Boman, M., Dalman, C., Davidson, M., Fruchter, E., Hultman, C. M., Lundberg, M., Långström, N., et al. (2012). Family history of schizophrenia and bipolar disorder as risk factors for autism. *Archives of general psychiatry*, 69(11):1099–1103.

Toomey, R., Faraone, S. V., Simpson, J. C., and Tsuang, M. T. (1998). Negative, positive, and disorganized symptom dimensions in schizophrenia, major depression, and bipolar disorder. *The Journal of nervous and mental disease*, 186(8):470–476.

Vigo, D., Thornicroft, G., and Atun, R. (2016). Estimating the true global burden of mental illness. *The Lancet Psychiatry*, 3(2):171–178.

Wehr, T. A., Sack, D. A., and Rosenthal, N. E. (1987). Sleep reduction as a final common pathway in the genesis of mania. *The American journal of psychiatry*.

Wilson, R. C. and Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *Elife*, 8:e49547.

Young, L. T., Warsh, J. J., Kish, S. J., Shannak, K., and Hornykeiwicz, O. (1994). Reduced brain 5-ht and elevated ne turnover and metabolites in bipolar affective disorder. *Biological psychiatry*, 35(2):121–127.