A Predictive Coding Model of Language Dysfunction in Early Psychosis

Joshua S. LeClair

Department of Psychology, Western University

PSYCH 4850E: Psychology Honours Thesis

Supervisors: Dr. L. K. Palaniyappan, Dr. R. Limongi

April 12, 2020

Abstract

Computational approaches are emerging as a powerful tool for understating psychiatric illness. Predictive coding models, in particular, have been productively applied to investigate various facets of schizophrenia symptomatology. By blending experimental data with hypothesis-driven predictive coding computational modelling, the present study sought to understand if different mechanistic cognitive processes underlie language dysfunction in schizophrenia. In particular, a type of reinforcement learning model, a Q-learning model, was fit to previously collected speech-listening task data. Involving healthy control participants and treatment-naïve individuals experiencing a first episode of psychosis, the speech-listening task targeted features of both low-level language processing and high-level language comprehension. Although multiple aberrations in language functioning have been proposed to underlie the deficits present in schizophrenia, only one modelling assumption was made for this study – namely that, in aggregate, these aberrant processes give rise to differences in *a priori* beliefs in task-competence. On the basis of this simple assumption, a unique pattern of model attributions was found to characterize the clinical group. These results suggest that predictive coding models may help elucidate important unobservable mechanistic differences in cognitive processes.

Acknowledgements

First and foremost, I would like to thank you, Roberto. Your mentorship, patience, and unfaltering kindness made it possible to overcome what (at points) felt like insurmountable challenges. Moreover, your support and guidance made this a fun and immensely rewarding experience. Thank you for showing me it was possible to combine my previous skills with my newfound passion.

Next, I'd like to thank you, Lena for taking a chance on me and welcoming me into your group; little could I have anticipated that our chance encounter over coffee in California would lead to this! Thank you for your quiet but ever-present support this past year.

To my colleagues in the NIMI Group, thank you for providing a sense of community during what was (undoubtedly) a challenging year for all of us. Although in many cases we never had the opportunity to meet in person, thank you – each of you – for never failing to make me feel welcome.

Finally, to my closest friend, Tyler. Thank you, simply, for everything – for your steadfast presence and your unwavering support. Without you, I know I would have neither embarked upon nor finished this crazy adventure. More often than you realize, your encouragement and your kindness have made all the difference.

Statement of Contribution

Development of the research problem and planning of the modelling approach were the result of a collaborative effort between myself (JSL) and Dr. R. Limongi. I wrote the necessary MATLAB code to fit the selected predictive coding model to the experimental data previously collected by Ana-Bianca Popa. Additionally, with guidance from Dr. R. Limongi, I carried out all analyses that were performed on the data.

A Predictive Coding Model of Language Dysfunction in Early Psychosis

Schizophrenia is a complex and devastating brain disorder which affects approximately 1% of the world's adult population (Regier et al., 1993). Individuals with schizophrenia face a significantly reduced life expectancy, and often cope with impairments in social functioning, difficulties with independent living, and problems maintaining employment (Harvey, 2014). Cutting across the heterogeneity of the disorder, language is implicated in many of the major symptoms and perceptual and cognitive abnormalities that characterize schizophrenia. Although many theories have been proposed, the etiological causes and biopsychosocial mechanisms underpinning the disorder remain poorly understood (Valton et al., 2017). Recently, however, generative computational models have emerged as a promising way of rigorously defining and testing hypotheses about mental disorders (Huys et al., 2011). One such family of computational models – rooted in what is known as the predictive coding theory of psychosis – is finding application in elucidating many facets of schizophrenia symptomology (Adams et al., 2013). In particular, language dysfunction in schizophrenia has been explained theoretically within this framework (Brown & Kuperberg, 2015).

Before proceeding, it is instructive to delineate between psychosis, which is a constellation of symptoms associated with a loss of contact with external reality, and schizophrenia itself. Psychosis can involve alterations in cognition, affect, and behaviour and is associated with a multitude of major mental disorders. Schizophrenia, however, is a chronic psychiatric condition in which the presence of psychotic symptoms, such as delusions or hallucinations, are potential – yet neither necessary nor sufficient – diagnostic criteria. In addition to these so-called positive symptoms, schizophrenia is also characterized by negative symptoms (*e.g.*, blunted affect, anhedonia) and disorganized symptoms (*e.g.*, disorganized

thoughts or behaviour). Language disturbances associated with these three categories of symptoms include auditory verbal hallucinations (AVHs), reduced speech output, and disorganized speech (positive thought disorder), respectively.

Research into language abnormalities in schizophrenia has primarily developed along four lines of inquiry (Brown & Kuperberg, 2015). Two of these lines of research have focused on the clinical manifestations of language dysfunction, exploring AVHs (for a review, see Moseley et al., 2013) and the speech associated with thought disorder (for a review, see Kircher et al., 2018). In contrast, however, a separate body of literature has examined the low-level auditory processing deficits (for a review, see Javitt and Freedman, 2015) and high-level sentence comprehension deficits (for a review, see Boudewyn et al., 2012) associated with the perceptual and cognitive abnormalities typical of schizophrenia.

It is these latter two research foci that are most relevant to the present study. Before moving into a deeper discussion of these aspects of language dysfunction in schizophrenia, however, an overview of the Bayesian brain theory and predictive coding theory of psychosis will be given. This will provide context to then explore how both high-level and low-level language deficits may be theoretically explained within a predictive coding framework. Subsequently, the details of the present study will be discussed.

Bayesian Brain Theory

The literature on schizophrenia is replete with many diverging (but not necessarily mutually exclusive) theories of the origin of the disorder. At the level of neurophysiology, deficits associated with schizophrenia are variously attributed to faulty dopaminergic (*e.g.*, Howes & Kapur, 2009), glutamatergic (*e.g.*, Mogahaddam & Javitt, 2012), or GABAergic (*e.g.*, Nakazawa et al., 2013) signalling. Unifying these theories, a fourth hypothesis – the

dysconnection hypothesis – attempts to establish a link between aberrant synaptic neuromodulation (functional connectivity) and the symptoms of schizophrenia (Friston & Frith, 1995). Within this hypothesis, the notion of the Bayesian brain is an emerging computational perspective being used to bridge this explanatory gap (Friston et al., 2016).

The Bayesian Brain theory is a proposed explanation of generalized brain function (Friston, 2005; Knill and Pouget, 2004). According to this theory, the brain is first and foremost an organ of inference. Specifically, it is assumed that cognition can be described in terms of Bayesian inference (Valton et al., 2017). Conceptually speaking, this mathematical formalism offers a way to optimally combine one's previous experiences ("prior beliefs") with real-time sensory evidence (the "likelihood") to yield a probabilistic prediction about the environment (the "posterior"). The relative contribution of each of these sources of information is determined by their certainty ("precision"). Under this view, the brain's primary task is to infer the distal causes which produce the patterns of sensory evidence that it encounters. To accomplish this, the brain leverages its prior experiences to generate hypotheses which are then compared against incoming sensory data; the closer the match between this "simulated data" and the real-world data, the more probable it is that the inferred cause of the sensation is correct (Clark, 2013). In the case of language comprehension, for example, the cause of the sensation (speech) that a listener must infer is the underlying message that a speaker is trying to communicate.

Language comprehension serves as a useful example as it also illustrates how a sensory signal can be encoded at multiple levels of representation. A hierarchical generative view of language (Kuperberg & Jaeger, 2016) suggests that higher-level hypotheses about message-level meaning can inform predictions about upcoming words, which, in turn, can influence expectations about lower-level phonetic features of incoming speech. Any discrepancies between

the top-down predictions and bottom-up sensory inputs – formally, prediction errors – may be used to update high-level beliefs about the meaning of the message being conveyed by the speaker (Brown & Kuperberg, 2015). In general, models in which lower-level phenomena are explained in terms of more abstract higher-level features are known as hierarchical generative models. Predictive coding is the name given to the message-passing scheme wherein the recursive exchange of information between levels generates prediction errors which drive learning (Adams et al., 2013).

Predictive Coding Theory of Psychosis

Although the Bayesian brain theory offers a general account of brain function, it has been productively applied to the study of psychopathology. In particular, this theory has inspired a growing a body of literature investigating the mechanisms underlying psychosis and schizophrenia (Sterzer et al., 2018). If the brain is viewed as an organ of inference, then schizophrenic symptoms may be conceptualized as failures of prediction – a result of faulty inference (Adams et al., 2013; Valton et al., 2017).

Specifically, the predictive coding theory of psychosis proposes that a wide variety of symptoms associated with the disorder may be explained in terms of the aberrant encoding of the precision of prior beliefs relative to the sensory data (Adams et al., 2013; Frith & Friston, 2013). This one computational change has been hypothesized to underlie a variety of symptoms of schizophrenia, including delusions, hallucinations, and other subtle perceptual aberrations. Indeed, most work exploring the explanatory power of the predictive coding account has focused on the emergence of psychotic symptoms (*e.g.*, Corlett et al., 2009; Fletcher & Frith, 2009).

One of the primary strengths of computational models is their capacity to test hypotheses about a main variable of interest, such as a behavioural or perceptual deficit (Colombo & Series,

2012). Mechanistic or descriptive insight may then be gleaned if the associated experimental phenomena are well-captured by the model. With the exception of clusters of studies exploring resistance to visual illusions (*e.g.*, Dima et al., 2010; Silverstein & Keane 2011; Teufel et al., 2015) and formal statistical learning abnormalities (*e.g.*, Averbeck et al., 2011; Garety & Freeman, 2013; Joyce et al., 2013), however, literature blending experimental data with hypothesis-driven predictive coding computational modelling has been scarce until recent times.

Some recent studies have leveraged the hypothesis-driven computational approach to explore predictive coding aberrations in the context of auditory perception. Using a Pavlovian learning task which paired a visual stimulus with an auditory tone, Powers et al. (2017) conditioned auditory hallucinations in a group of individuals diagnosed with psychotic illness, a group of 'psychics', and a healthy control group. Part of their analysis used a hierarchical generative model to capture the behavioural data and explore the computational mechanisms underlying hallucinations. Their results supported the notion that there is an aberrant encoding of prior beliefs in individuals who regularly hallucinate voices. A related study by Cassidy et al. (2018) investigated aberrant prior beliefs in a sample of schizophrenia patients and healthy controls by using a task that measured illusory changes in the perceived duration of a simple auditory tone. Most recently, Haarsma et al. (2020) explored how prior beliefs varied across individuals at various stages of psychotic illness using two auditory perception paradigms. A lower-level perception task based on the McGurk illusion (McGurk & MacDonald, 1976) assessed how lip movements influenced the perception of ambiguous phonemes. In contrast, a higher-level cognitive task relied on learned written word – phoneme associations. Results from this study suggest that alterations in prior beliefs differ across hierarchical levels of sensory perception.

Although each of the aforementioned studies found evidence for predictive coding aberrations in the context of auditory perception, no work has yet explored this phenomenon using more complex speech stimuli. Given that the linguistic phenomena associated with schizophrenia are most clearly manifest at the level of words, sentences, and discourse (Kuperberg, 2010), this is a particularly relevant direction of potential inquiry. The dearth of experimental research exploring how language dysfunction, broadly construed, may be reconciled with the predictive coding account of psychosis is particularly striking in light of a recently proposed theoretical framework endorsing this approach (Brown & Kuperberg, 2015).

Language Dysfunction in Schizophrenia

In a recent theory paper, Brown and Kuperberg (2015) put forth a hierarchical generative framework for understanding language dysfunction in schizophrenia. In part, they elucidate how both the higher-level language comprehension and lower-level language perception abnormalities that characterize schizophrenia may be conceptually unified from the perspective of predictive coding. In the following two subsections, aspects of this framework will be discussed in the context of relevant experimental findings and illustrative research paradigms.

High-level Language Comprehension

At the higher level of language comprehension, experimental evidence suggests that individuals with schizophrenia engage with language in a less predictive and more reactive manner (Brown & Kuperberg, 2015). This observation is supported by numerous studies in which patients demonstrate difficulty in resolving lexically ambiguous homophones (*e.g.*, Chapman et al., 1964) and exhibit a tendency to interpret figurative language overly literally (*e.g.*, Chapman, 1960). Each of these phenomena point towards an impaired ability to use linguistic context in the process of language comprehension.

One way the effects of linguistic context on speech perception have been investigated experimentally is through the use of regular and anomalous speech stimuli. Kuperberg et al. (1998) presented healthy control participants and schizophrenia patients with a combination of spoken sentences; some sentences were entirely coherent whereas other contained pragmatic, syntactic, or semantic violations. When tasked with listening for a target word in such sentences, patients with schizophrenia demonstrated significantly smaller deviations in performance across conditions, indicating a relative insensitivity to linguistic violations. In a follow-up study (Kuperberg et al., 2000), this insensitivity was shown to be more pronounced in patients with more severe thought disorder.

Casting these observations in terms of predictive coding and Bayesian inference, Brown and Kuperberg (2015) suggest that such impairments might be attributed to an underweighting of prior beliefs. That is to say, less precise prior beliefs would bias inference in the direction of the sensory evidence, leading to an over-dependence on bottom-up information processing.

Low-level Language Perception

At the lower level of speech perception, top-down predictions still play an integral role in language processing. In addition to helping anticipate upcoming words, high-level predictions are also informative at the level of acoustic-phonetic features (Brown & Kuperberg, 2015). Most predominantly, this is reflected in the fact that the accurate disambiguation of phonemes in speech is dependant on the use of contextual information (Allen et al., 2004; McMurray & Jongman, 2011). As was the case for high-level language deficits, an overweighting of sensory information may explain why, relative to healthy individuals, schizophrenia patients demonstrate impaired phoneme discrimination (Kugler & Caudrey, 1983).

Uncertainty at the level of perception may be experimentally manipulated through the use of degraded speech stimuli. At the level of word recognition, it has been shown that individuals with schizophrenia are more vulnerable to masking using speech-spectrum noise (Wu et al., 2012). Exacerbated speech processing deficits are also reflected in studies using sentence-length speech stimuli. Alpert (1985) presented schizophrenia patients with brief phrases masked by white noise at various signal-to-noise ratios (SNRs). Although all patients demonstrated impaired performance in reporting the stimulus phrases, the nature of the errors differed depending on the hallucination-proneness of the individual. In a subsequent study (Hoffman et al., 1999), individuals with schizophrenia demonstrated similarly pronounced deficits in perception during a masked speech repetition task.

In contrast to higher level comprehension deficits, Brown and Kuperberg (2015) hypothesize that low-level perceptual prediction errors might drive a rapid learning process to accommodate deviant phoneme pronunciations. It should be noted, however, that there is conflicting evidence in the literature as to whether patients might be able to adjust phoneme representations (*e.g.*, Cienfuegos et al., 1999; Kasai et al., 2002) on the basis of prediction errors.

The Present Study

Using previously collected experimental data, the present study examines if a predictive coding model can capture abnormalities associated with high-level language comprehension and low-level language perception in individuals with schizophrenia. Participants in the aforementioned experiment consisted of a group of individuals experiencing a first episode of psychosis and a matched group of healthy controls. All subjects completed a speech-listening task which involved listening to and immediately recalling and reporting a series of sentences. In this task, two sources of uncertainty were manipulated. First, addressing high-level language

comprehension, some sentences were semantically coherent (from here on, “meaningful” sentences) and others contained numerous semantic violations (from here on, “meaningless” sentences). Within these meaningfulness conditions, sentences were presented at both high and low levels of acoustic degradation, directly affecting low-level language perception. Response data from this task was coded for numerous error measures.

Although Brown and Kuperberg (2015) make specific predictions about the nature of the predictive coding aberrations that may be present in individuals with schizophrenia, it is not possible to directly test these hypotheses on the basis of the available experimental data. Rather, for the present modelling study, it is assumed that, in aggregate, these various language processing aberrations lead participants to approach the experimental trials with potentially differing beliefs (levels of “confidence”) about their abilities to perform the task. Although difficult to grasp phenomenologically, such assumptions about differences in prior beliefs over sets of behaviour are eminently compatible with Bayesian accounts of human cognition (Smith et al., 2021).

One way to formalize these assumptions is via a simple reinforcement learning model (*i.e.*, a Q-learning model). Rather than conceptualizing reinforcement learning as the optimization of expected reward, however, a Bayesian inference perspective defines an optimal policy in terms of the probability distribution of desired (behavioural) states (Friston et al., 2009). In the context of the speech-listening task, these desired states represent participants’ condition-optimal self-perceptions about their performance in each experimental trial. Cast in this light, then, prior beliefs restrict the frequency with which such states are visited. Such a conceptualization of the Q-learning model is suggestive of the main scientific aim of this modelling study

The mathematical analysis of the experimental data presented here asks if different generative processes – formally, differences in model structure – may plausibly underlie differences in between-group performance on a speech-listening task. To investigate this question, a variational Bayesian analysis of previously collected experimental data was conducted. In addition to examining differences in model structure, Q-learning models were used to estimate latent variables at the level of the individual (*e.g.* learning rate). These subject-level parameters were subsequently compared across conditions and between groups.

Method

Experimental Design

Collected as part of the TOPSY (Tracking Outcomes of Psychosis) longitudinal research study, the data used for the present work comes from a previous experiment by Popa (2018).

Participants

The present analysis is based on data from 48 clinical participants and 29 healthy controls. The clinical group consisted of individuals experiencing a first episode of psychosis (FEP) who were recruited from the Prevention and Early Intervention Program for Psychoses (PEPP) at Victoria Hospital in London, Ontario. Recruited via advertisements, the matched control group consisted of healthy individuals from Western University and the broader local (London) community. Demographic information for both groups of participants are summarized in Table 1 (Popa, 2018).

To be included in the study, clinical participants must have been receiving antipsychotic medication for less than two weeks. As such, participants experiencing a first episode of psychosis were recruited within two weeks of their initial assessment at the PEPP clinic. Further inclusion criteria included a capacity to provide informed consent and the ability to participate

safely in the experimental task. In addition to the more general exclusion criteria described below, clinical participants were excluded if they met diagnostic criteria for a major substance use disorder in the past year. Psychiatrists established a consensus diagnosis for 32 of the 48 clinical participants six months post FEP; of these 32, most were diagnosed with schizophrenia (n = 22), schizoaffective disorder (n = 2), or schizophreniform disorder (n = 1). The remaining participants for whom follow-up information is available were diagnosed with bipolar disorder (n = 1), manic depressive disorder (n = 2), brief psychotic disorder (n = 1), clinical high-risk for psychosis (n = 1), or psychosis not otherwise specified (n = 2).

Control participants were statistically equivalent to the clinical group in terms of age, years of education, sex, and socio-economic status. In addition to the general exclusion criteria, control participants were excluded if they had a history of mental illness. General exclusion criteria for all participants included a major head injury, a significant uncontrolled medical illness, or a hearing impairment. The study was approved by Western University's Research Ethics Board (see Appendix A).

Table 1

Summary of Demographic Data

	Clinical Group			Control Group		
	n	Median	Range	*n*	Median	Range
Sex (M:F)	39:9			19:10		
Age	48	22	16 – 39	29	22	16 – 29
Years of Education	47	12	11 – 18	29	14	11 – 16
National Statistics Socio-economic Classification 5-class scale	43	4	1 – 5	28	3	1 – 5

Materials

The experimental stimuli for the speech-listening task consisted of 34 pairs of syntactically matched sentences between 6 and 13 words in length. As detailed by Popa (2018), these sentences were derived from the stimuli produced by Davis et al. (2011) and were chosen based on data from a pilot study. One sentence in each matched pair was semantically meaningful (*e.g.*, "He searched the pack for the ace of hearts.") and the other was semantically meaningless (*e.g.*, "He charged the lap for the niece of wheels."). Each of the 68 sentences was spoken by a female native English speaker and recorded at a sampling rate of 44.1 kHz.

Within each meaningfulness condition, half (*i.e.*, 17) of the sentences were presented at a high level of acoustic degradation (-4 dB SNR), with the remaining half presented at a low level of acoustic degradation (0 dB SNR). Speech-spectrum noise matching the amplitude envelope of the masked sentence was used to degrade each recording. To counterbalance the sentence materials across SNR levels, two versions of the experimental stimuli were prepared: between the two versions, each pair of sentences were presented at both levels of acoustic degradation. Participants were assigned versions at random such that an equal number of individuals in both the clinical and control groups were tested with each version. Within each version, the order of presentation of the sentences remained the same between participants.

Procedure

The speech-listening task was conducted in a quiet room at the Robarts Research Institute in London, Ontario. Clinical measures relevant to the original study by Popa (2018) were collected prior to the commencement of the experiment. Participants were told that the sentences they would be listening to may be challenging to hear; importantly, they were not apprised of the meaningfulness manipulation in advance. After receiving instructions for the task, participants

were seated at a laptop computer running E-Prime 2.0 stimulus presentation software. One at a time, sentences were presented at a pre-set comfortable listening level over a pair of headphones. After each sentence was presented, participants were prompted to report, at their own pace, what they just heard by typing into a response box. As part of the task instructions, participants were told to simply report as much of the sentence as they could recall without concern for completeness, punctuation, or spelling. However, in the event that participants could not discern any words, they were provided with the option to enter "none". Participants completed four practice trials before being presented with the experimental stimuli. Although the 68 sentences were presented one after another, participants were given the option to take a break halfway through. In total, the experimental task took between 20 to 30 minutes to complete, after which participants were debriefed and compensated for their time.

Measures

Multiple measures were extracted from the speech-listening task response data by Popa (2018). As a simple performance measure, the fraction of words each participant correctly recalled for each sentence was computed. Three different error measures were also extracted. Trained research assistants blinded to the group membership of participants scored the response data for morphological variants (*e.g.*, reporting "charged" as "charges"), phonological variants (*e.g.*, reporting "wheels" as "feels"), and intrusion errors (*i.e.*, erroneous words not belonging to either of the aforementioned categories). In each case, errors were computed as a percentage relative to the total number of words in the original sentence. For the purposes of mathematical modelling, however, only the simple performance measure data were used.

Mathematical Modelling

Implemented using the Variational Bayesian Analysis MATLAB toolbox (Danizeau et al., 2014), Q-learning models were fit to the experimental data. In order to do so, however, certain modelling assumptions were made and a correspondence between the experimental task and mathematical structure of the model was established.

Assumptions

In order to investigate if there were between-group differences in the generative (predictive coding) processes that underlay performance in the speech-listening task, it was assumed that participants may approach the experiment with various levels of confidence in their ability to successfully complete the task. Conceivably, such a level of confidence may vary continuously; however, due to computational limitations, only four extreme – albeit illustrative – discrete values were encoded in the Q-learning models. Two of these values reflected a low level of confidence (*i.e.*, beliefs that success was likely on 10 or 20 percent of the trials) whereas the remaining two reflected a high level of confidence (*i.e.*, beliefs that success was likely on 80 or 90 percent of the trials). Mathematically, these levels of confidence correspond to probability mass functions of a (discrete) random variable possessing differing means. These *a priori* levels of confidence, in turn, influence performance on the speech-listening task.

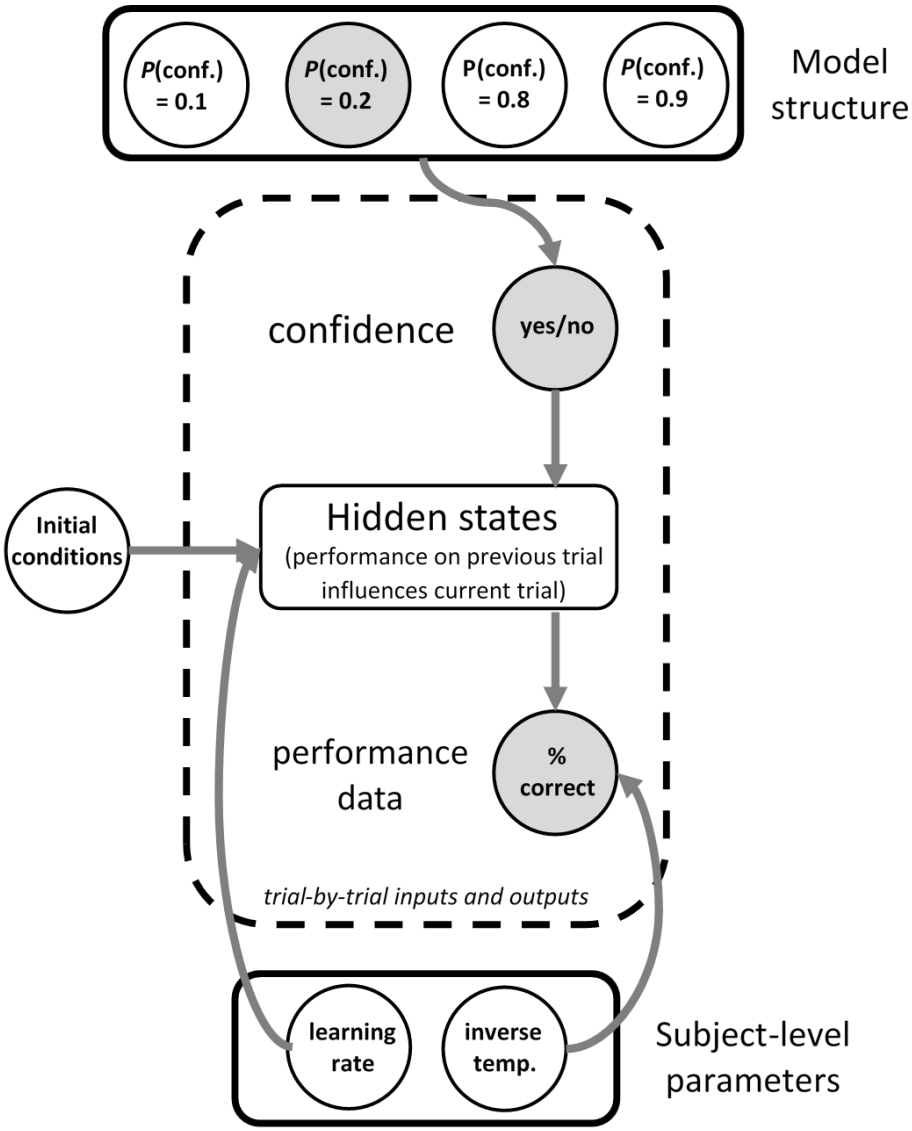
Structure

The mathematical structure of the Q-learning model employed in the analysis is outlined in Figure 1. Corresponding to the modelling assumptions that were made, four different model structures were fit to each participant's data; mathematically, each of the four model structures are described by Bernoulli distributions possessing differing means (*e.g.*, the distribution representing a 10 percent level of confidence has a mean of 0.1) . In turn, probabilistic draws

from these Bernoulli distributions were inputted into the model as binary choices (*i.e.*, representing confidence, or the lack thereof, in one's performance on a specific trial).

Figure 1

Mathematical Structure of Q-learning Model



Note. Grey nodes represent data-driven or assumption-driven inputs into the model. The centre plate indicates repetitions over trials. Arrows denote causal relationships between components of the model.

The process of fitting a model to these assumption-driven inputs involves computing the so-called "hidden states" (formally, the Q-values) which dictate how task performance (specified by the experimental data) changes over the duration of the experiment. In effect, these hidden states are the mechanism by which performance on previous trials influences subsequent performance. The time-evolution of these hidden states are driven by a subject-level parameter which is free to vary between individuals – the learning rate. A second subject-level parameter, the inverse temperature, determines the correspondence between the hidden states and the observed experimental performance data. Conceptually, the inverse temperature captures something of the level of randomness in variations in task performance; higher values of this parameter indicate that certain hidden states map more definitively onto specific levels of task performance.

Finally, since there is a mathematical dependence on performance on previous trials, initial values for the subject-level parameters needed to be estimated in order to fit the model to the first experimental trial for each participant. Uniform priors were chosen for these initial conditions – conceptually, this means that the subject-level parameters were not initially biased or constrained in any particular way.

Procedure

Fitting the Q-learning models to the experimental data was a multi-step process. First, previously scored performance data from Popa (2018) was inputted into the Q-learning models (see Table 2 for descriptive statistics). Next, subject-level inversions for each of the four "confidence" models, for each experimental condition (*i.e.,* 2 levels of meaningfulness x 2 levels of noise) were performed for each participant. Fitting the models in this way produced three important metrics for each combination of model and experimental condition (*i.e.,* 4 models x 4

experimental conditions) for each participant: the learning-rate, the inverse temperature, and the model evidence. However, due to the probabilistic nature of the inputs from the Bernoulli distributions, these metrics needed to be obtained by averaging over a large number of such simulations. Convergence of the subject-level parameters and model evidence was reached after averaging over 150 sets of simulations for each participant (see Appendix B).

Table 2

Sample Descriptives of Raw Performance Data

	Control Group		Clinical Group	
Experimental Condition	*M*	*SD*	*M*	*SD*
Meaningless with high noise	0.24	0.21	0.20	0.21
Meaningless with low noise	0.65	0.25	0.59	0.30
Meaningful with high noise	0.42	0.34	0.40	0.34
Meaningful with low noise	0.91	0.18	0.86	0.25

Note. Here, performance data refers to the simplest metric that was collected – namely, the fraction of words each participant correctly identified in each experimental trial.

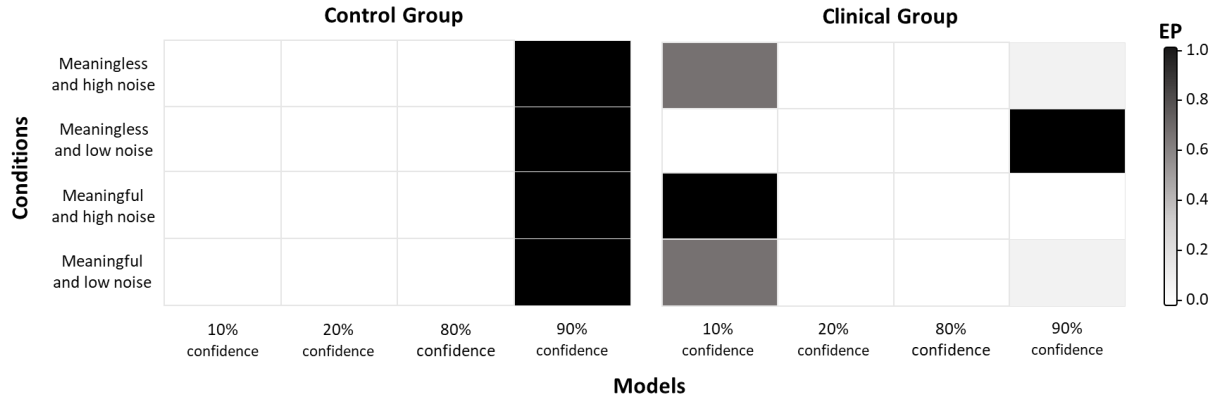
Results

Model Structure

To determine which model structures best represented the experimental data, Bayesian model selection was performed. In a between-condition analysis, the (averaged) model evidence from each ensemble of subject-level inversions was used to generate composite models that described each participant's performance across conditions. These 256 composite models encoded all combinations of candidate models and conditions (*i.e.*, 4 candidate "confidence" models raised to the power of 4 experimental conditions produces 256 possible combinations of confidence and conditions). The resulting per-condition model attributions are displayed in Figure 2.

Figure 2

Per-condition Bayesian Model Selection



Note. The exceedance probability (EP) measures how likely it is that a given model is more frequent (describes the performance of a larger number of participants) than all other models under consideration.

As suggested by the results in Figure 2, a composite model assuming 90 percent confidence throughout each of the experimental conditions best described the performance of the control group participants (EP = 1.00). In contrast, a mixed-confidence composite model best described the performance of the clinical participants. Here, a model assuming 10 percent confidence for three of the experimental conditions and 90 percent confidence for the semantically meaningless low noise condition was found to be most likely (EP = 0.96).

Finally, a between-group random effects analysis was performed to test the hypothesis that these patterns of model attributions differed between the control group and clinical group (H_o: model frequencies are the same for each group; H_A: groups have distinct model frequencies). The posterior probability that the two groups have the same model frequencies, $P(H_o \mid \text{data}) < .001$, supports the rejection of the null hypothesis; that is, model frequencies differ significantly between groups.

Model Parameters

Once the composite models best describing performance in each group were identified, subject-level parameters for each participant were extracted. Descriptive statistics for both of these parameters – the learning rate and the inverse temperature – are summarized in Table 3.

Table 3

Sample Descriptives of Subject-level Parameters

	Control Group				Clinical Group			
	Learning Rate		Inverse Temperature		Learning Rate		Inverse Temperature	
Experimental Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Meaningless with high noise	0.36	0.01	5.24	0.30	0.37	0.02	5.22	0.44
Meaningless with low noise	0.38	0.02	4.49	0.33	0.37	0.02	4.56	0.42
Meaningful with high noise	0.36	0.02	4.84	0.36	0.35	0.02	4.83	0.33
Meaningful with low noise	0.38	0.02	3.77	0.23	0.37	0.02	3.90	0.52

A factorial MANOVA was subsequently conducted to compare main effects and interaction effects of group membership, meaningfulness, and noise on learning rate and inverse temperature. The analysis of this data utilized a 2 (Group: Control and Clinical) x 2 (Meaningfulness: Meaningless and Meaningful) x 2 (Noise: High and Low) design.

Before proceeding, homogeneity of variance was checked for each dependent variable using Levene's test. This test was not significant for learning rate, Levene $F(7, 304) = 0.71$, $p = .665$, but was found to be significant for inverse temperature, Levene $F(7, 304) = 2.26$, $p = .030$. Accordingly, equal variances were not assumed for subsequent analyses involving inverse temperature and an alpha level stricter than .05 was used to evaluate significance. Furthermore, Box's test was significant, $F(21, 189570.67) = 2.44$, $p < .001$, indicating that the homogeneity of covariance matrices assumption was violated. However, given the sample sizes for each group, the MANOVA is robust against such violations (Allen & Bennett, 2008); nonetheless, results were interpreted with caution.

The factorial MANOVA revealed a significant effect of both meaningfulness, Pillai's Trace = .32, $F(2, 303) = 72.57$, $p < .001$, and noise, Pillai's Trace = .55, $F(2, 303) = 186.39$, $p < .001$, on the combined dependent variables. However, group membership was not found to have a significant effect, Pillai's Trace = .01, $F(2, 303) = 0.908$, $p = .404$. Furthermore, a significant interaction effect between meaningfulness and noise on the combined dependent variables was found, Pillai's Trace = .05, $F(2, 303) = 7.90$, $p < .001$. In contrast, no significant interactions were found between group and meaningfulness, Pillai's Trace < .01, $F(2, 303) = 0.11$, $p = .894$, or between group and noise, Pillai's Trace = .01, $F(2, 303) = 1.04$, $p = .356$. Three-way interactions between the independent variables also did not reach the level of significance, Pillai's Trace < .01, $F(2, 303) = 0.22$, $p = .801$.

Follow-up tests of univariate effects indicated that while meaningfulness had a significant main effect (ME) on inverse temperature, $F(1, 304) = 138.79$, $p < .001$, power > .99, its effect on learning rate was not significant, $F(1, 304) = 2.60$, $p = .108$, power = .36. Collapsing across noise levels, inverse temperature was, on average, higher in both groups for trials involving

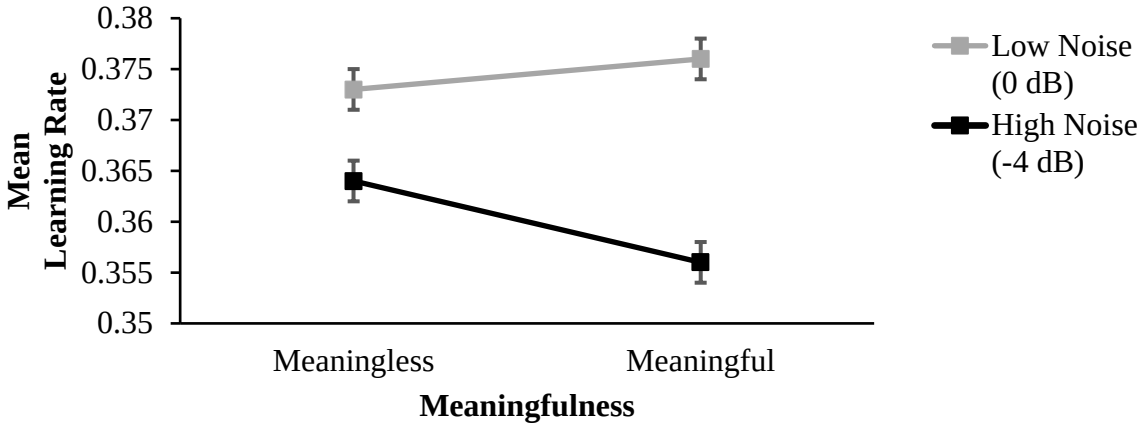
meaningless sentences. This effect was large, $\eta^2 = .31$. In contrast, noise was found to have a significant ME on both learning rate, $F(1, 304) = 49.87$, $p < .001$, power > .99, and inverse temperature, $F(1, 304) = 343.54$, $p < .001$, power > .99. In both groups, learning rates were, on average, higher in the low noise conditions ($\eta^2 = .14$); inverse temperature, however, was lower, on average, in the low noise conditions ($\eta^2 = .53$). As indicated by the partial eta-squared values, the effect of noise on each subject-level parameter was large. However, these effects must be interpreted cautiously in light of potential interactions between meaningfulness and noise on learning rate and inverse temperature.

Indeed, tests of univariate effects revealed a significant interaction between meaningfulness and noise on learning rate, $F(1, 304) = 7.06$, $p = .008$, $\eta^2 = .02$, power = .76. This effect was small. From visual inspection of this interaction (see Figure 3), it is clear that the effect of noise on learning rate varies depending on the level of meaningfulness; differences in learning rate become more pronounced under the meaningful sentence condition. Interestingly, the direction of effect also changes - while learning rate increases across levels of the meaningfulness factor at a high level of noise, it decreases across levels of meaningfulness at a low level of noise.

Finally, a potential interaction between meaningfulness and noise on inverse temperature was investigated, $F(1, 304) = 10.12$, $p = .002$, $\eta^2 = .03$, power > .89. Using the stricter alpha level necessitated by the violation of the homogeneity assumption (*i.e.*, $p < .001$), this interaction is not significant. Indeed, visual inspection of a plot depicting this potential interaction (see Figure 4) supports such a conservative conclusion.

Figure 3

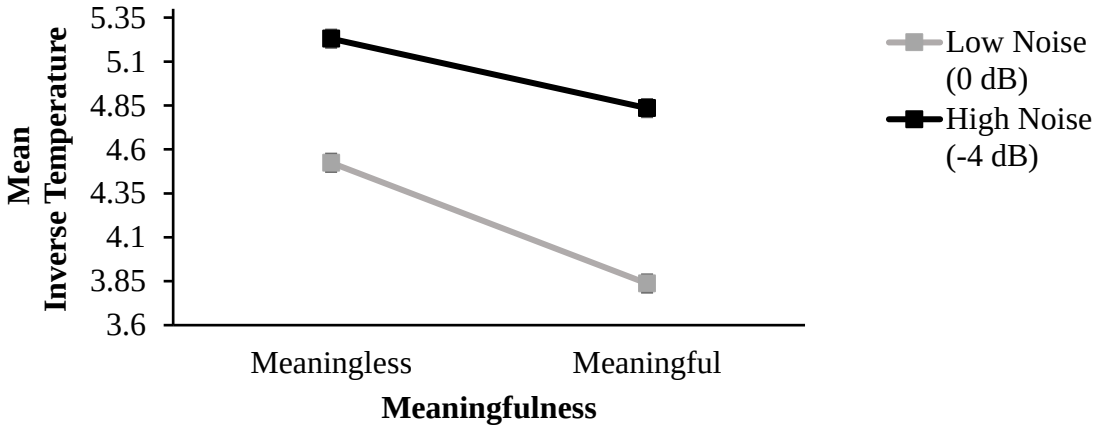
Simple Main Effects of Noise on Learning Rate



Note. Mean learning rates are shown for the two noise levels, plotted by the two levels of meaningfulness. To improve the graphical presentation of the data, a truncated range is used for the ordinate. Error bars represent the standard error.

Figure 4

Simple Main Effects of Noise on Temperature



Note. Mean inverse temperatures are shown for the two noise levels, plotted by the two levels of meaningfulness. To improve the graphical presentation of the data, a truncated range is used for the ordinate. Error bars represent the standard error.

Discussion

Informed by a wealth of experimental literature, it has been hypothesized that a number of predictive coding aberrations may underlie language dysfunction in schizophrenia. In aggregate, such aberrations may influence an individual's confidence in their perceptions of sentence-level auditory stimuli. Premised on this assumption, the present mathematical modelling study sought to investigate if a simple reinforcement learning model can reveal between-group differences in hidden cognitive processes on a speech-listening task. The computational approach employed here is rooted in the belief that important between-group differences may not be self-evident when behavioural observations are simply analyzed using classic inferential statistical procedures. If such explicit differences are to be found in the experimental data, however, they will naturally emerge as variability in the subject-level parameters produced by the generative models.

Implications and Conclusions

Suggesting that distinct generative processes may indeed influence how individuals experiencing a first episode of psychosis engage with language, significant differences at the level of model structure were found. While the performance of the healthy participants was best explained by high levels of confidence across all experimental conditions, a contrasting pattern of attributions was found for the clinical group; across all but one of the conditions, low levels of confidence best accounted for their performance on the experimental task. Given the limitations of the mathematical model (*e.g.*, the models tested here only encoded four discrete levels of confidence), it is difficult to draw more detailed conclusions about the specific nature of these differences. As such, it is not necessarily *how* the groups differed that matters but rather *that* they differed. In this way, these results offer some of the first hypothesis-driven computational

modelling evidence for Brown and Kuperberg's (2015) hierarchical generative view of language dysfunction in schizophrenia.

In contrast to the between-group differences in model structure that were identified, the results – or rather, the lack thereof – from the analysis of the model parameters are more perplexing. Although main effects and interaction effects involving the meaningfulness and noise factors were found, they are not particularly germane to the present inquiry; given that the experimental conditions undoubtedly influenced task performance (and thus, indirectly, the model parameters), such observations are merely reflective of the structure of the speech-listening task itself. More interesting, perhaps, are the null results – neither learning rate nor inverse temperature differed significantly between the clinical group and the control group. Additionally, no significant interactions between group membership and any of the experimental factors were found. How, then, are these null results to be reconciled with the well-established language deficits that characterize schizophrenia?

It is important to note that the exact nature of language impairment in FEP patients remains poorly understood (Roche et al., 2016). Historically, most studies investigating language dysfunction in schizophrenia have relied upon chronic patient populations. As such, a plausible explanation for these null results may be that the aspects of language processing and comprehension targeted by the speech-listening task are not (yet) sufficiently impaired in FEP patients so as to produce directly observable differences in the subject-level parameters that drive performance. It is here that the strength of a generative modelling approach becomes clear. Although no self-evident between-group differences exist within the experimental data, predictive coding models are able to elucidate otherwise *unobservable* mechanistic differences – differences that may well be of both theoretical and clinical importance.

Indeed, from the perspective of translational clinical science, evidence that it may be possible to distinguish between first episode psychosis patients and healthy individuals on the basis of a simple speech-listening task is promising. Although many FEP patients go on to be diagnosed with schizophrenia, the diagnostic boundaries separating schizophrenia from related psychotic disorders – particularly in the early stages of the illness – are not well established (Baldwin et al., 2005). Given the promise shown by the modelling approach taken in this study, it would be interesting to consider if differences in underlying generative processes may be used to differentiate between diagnostic groups within a heterogenous clinical sample.

Limitations and Future Work

To address such open research questions, however, more complex predictive coding models will be required. Improving upon two of the mathematical limitations of the present model, in particular, offer clear avenues for future work. First, the model space employed in the analysis presented here was artificially constrained; only four discrete confidence values were encoded in the Q-learning models. Conceivably, however, beliefs in one's own abilities may vary continuously. Accordingly, future modelling efforts should allow structural parameters (*e.g.*, "confidence") to vary freely. Although this requires a significant increase in model complexity, such an improvement would make it possible to examine how the structural parameters are distributed within the sample populations under consideration. As such, it would then be possible to draw clearer conclusions about the exact nature and magnitude of any between-group differences in model attributions. Second, although a (relatively) simple Q-learning model was appropriate for the exploratory nature of this study, it may not be the best model to describe the structure of the speech-listening task. A more complex predictive coding model, such as a Bayesian associative learning model (Mathys et al., 2011), may be better suited

to capture both forms of uncertainty inherent in the experimental manipulation (*i.e.*, environmental volatility and perceptual uncertainty). Despite these limitations, this modelling study represents a promising first step towards understanding and quantifying the aberrant processes which may underlie language dysfunction in schizophrenia.

References

Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., & Friston, K. J. (2013). The computational anatomy of psychosis. *Frontiers in Psychiatry*, *4*, 47. https://doi.org/10.3389/fpsyt.2013.00047

Allen, P. J., & Bennett, K. (2008). *SPSS for the health and behavioural sciences*. Thomson.

Allen, P. P., Johns, L. C., Fu, C. H., Broome, M. R., Vythelingum, G. N., & McGuire, P. K. (2004). Misattribution of external speech in patients with hallucinations and delusions. *Schizophrenia Research*, *69*(2-3), 277-287. https://doi.org/10.1016/j.schres.2003.09.008

Alpert. (1985). The signs and symptoms of schizophrenia. *Comprehensive Psychiatry.*, *26*(2), 103–112. https://doi.org/10.1016/0010-440X(85)90030-6

Averbeck, B. B., Evans, S., Chouhan, V., Bristow, E., & Shergill, S. S. (2011). Probabilistic learning and inference in schizophrenia. *Schizophrenia Research, 127*(1-3), 115-122. https://doi.org/10.1016/j.schres.2010.08.009

Baldwin, P., Browne, D., Scully, P. J., Quinn, J. F., Morgan, M. G., Kinsella, A., ... & Waddington, J. L. (2005). Epidemiology of first-episode psychosis: illustrating the challenges across diagnostic boundaries through the Cavan-Monaghan study at 8 years. *Schizophrenia Bulletin*, *31*(3), 624-638. https://doi.org/10.1093/schbul/sbi025

Boudewyn, M. A., Carter, C. S., & Swaab, T. Y. (2012). Cognitive control and discourse comprehension in schizophrenia. *Schizophrenia Research and Treatment, 2012*, 484502–484507. https://doi.org/10.1155/2012/484502

Brown, M., & Kuperberg, G. R. (2015). A hierarchical generative framework of language processing: Linking language perception, interpretation, and production abnormalities in schizophrenia. *Frontiers in Human Neuroscience*, *9*, 643. https://doi.org/10.3389/fnhum.2015.00643

Cassidy, C. M., Balsam, P. D., Weinstein, J. J., Rosengard, R. J., Slifstein, M., Daw, N. D., ... & Horga, G. (2018). A perceptual inference mechanism for hallucinations linked to striatal dopamine. *Current Biology*, *28*(4), 503-514. https://doi.org/10.1016/j.cub.2017.12.059

Chapman, L. J. (1960). Confusion of figurative and literal usages of words by schizophrenics and brain damaged patients. *The Journal of Abnormal and Social Psychology*, *60*(3), 412–416. https://doi.org/10.1037/h0043371

Chapman, L. J., Chapman, J. P., & Miller, G. A. (1964). A theory of verbal behavior in schizophrenia. *Progress in Experimental Personality Research*, *72*, 49.

Cienfuegos, A., March, L., Shelley, A. M., & Javitt, D. C. (1999). Impaired categorical perception of synthetic speech sounds in schizophrenia. *Biological Psychiatry*, *45*(1), 82-88. https://doi.org/10.1016/S0006-3223(98)00064-X

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181-204. https://doi.org/10.1017/S0140525X12000477

Colombo, M., & Seriès, P. (2012). Bayes in the brain—on Bayesian modelling in neuroscience. *The British Journal for the Philosophy of Science*, *63*(3), 697-723. https://doi.org/10.1093/bjps/axr043

Corlett, P. R., Frith, C. D., & Fletcher, P. C. (2009). From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology*, *206*(4), 515-530. https://doi.org/10.1007/s00213-009-1561-0

Daunizeau, J., Adam, V., & Rigoux, L. (2014). VBA: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. *PLoS Comput Biol*, *10*(1). https://doi.org/10.1371/journal.pcbi.1003441

Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through "top-down" processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, *23*(12), 3914–32. https://doi.org/10.1162/jocn_a_00084

Dima, D., Roiser, J. P., Dietrich, D. E., Bonnemann, C., Lanfermann, H., Emrich, H. M., & Dillo, W. (2009). Understanding why patients with schizophrenia do not perceive the hollow-mask illusion using dynamic causal modelling. *Neuroimage*, *46*(4), 1180-1186. https://doi.org/10.1016/j.neuroimage.2009.03.033

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, *10*(1), 48-58. https://doi.org/10.1038/nrn2536

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815-836. https://doi.org/10.1098/rstb.2005.1622

Friston, K., Brown, H. R., Siemerkus, J., & Stephan, K. E. (2016). The dysconnection hypothesis (2016). *Schizophrenia Research*, *176*(2-3), 83-94. https://doi.org/10.1016/j.schres.2016.07.014

Friston, K. J., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement learning or active inference?. *PloS One*, *4*(7). https://doi.org/10.1371/journal.pone.0006421

Friston, K. J., & Frith, C. D. (1995). Schizophrenia: a disconnection syndrome?. *Clinical Neuroscience (New York, N.Y.)*, *3*(2), 89–97.

Frith, C., & Friston, K. J. (2013). False perceptions and false beliefs: understanding schizophrenia. *Neurosciences and the Human Person: New Perspectives on Human Activities*, *121*, 1-15.

Garety, P. A., & Freeman, D. (2013). The past and future of delusions research: from the inexplicable to the treatable. *The British Journal of Psychiatry*, *203*(5), 327-333. https://doi.org/10.1192/bjp.bp.113.126953

Haarsma, J., Knolle, F., Griffin, J. D., Taverne, H., Mada, M., Goodyer, I. M., ... & Nspn Consortium. (2020). Influence of prior beliefs on perception in early psychosis: Effects of illness stage and hierarchical level of belief. *Journal of Abnormal Psychology*, *129*(6), 581. http://dx.doi.org/10.1037/abn0000494

Harvey, P. D. (2014). Assessing disability in schizophrenia: tools and contributors. *J. Clin. Psychiatry*, *75*, e27. https://doi.org/10.4088/JCP.13049tx5c

Hoffman, R. E., Rapaport, J., Mazure, C. M., & Quinlan, D. M. (1999). Selective speech perception alterations in schizophrenic patients reporting hallucinated “voices”. *American Journal of Psychiatry*, *156*(3), 393-399. https://doi.org/10.1176/ajp.156.3.393Howes, O. D., & Kapur, S. (2009). The dopamine hypothesis of schizophrenia: version III—the final common pathway. *Schizophrenia Bulletin*, *35*(3), 549-562. https://doi.org/10.1093/schbul/sbp006

Huys, Q. J., Moutoussis, M., & Williams, J. (2011). Are computational models of any use to psychiatry?. *Neural Networks*, *24*(6), 544-551. https://doi.org/10.1016/j.neunet.2011.03.001

Javitt, D. C., & Freedman, R. (2015). Sensory processing dysfunction in the personal experience and neuronal machinery of schizophrenia. *American Journal of Psychiatry*, *172*(1), 17-31. https://doi.org/10.1176/appi.ajp.2014.13121691

Joyce, D. W., Averbeck, B. B., Frith, C. D., & Shergill, S. S. (2013). Examining belief and confidence in schizophrenia. *Psychological Medicine*, *43*(11), 2327. https://doi.org/10.1017/S0033291713000263

Kasai, K., Nakagome, K., Itoh, K., Koshida, I., Hata, A., Iwanami, A., ... & Kato, N. (2002). Impaired cortical network for preattentive detection of change in speech sounds in schizophrenia: a high-resolution event-related potential study. *American Journal of Psychiatry*, *159*(4), 546-553. https://doi.org/10.1176/appi.ajp.159.4.546

Kircher, T., Bröhl, H., Meier, F., & Engelen, J. (2018). Formal thought disorders: from phenomenology to neurobiology. *The Lancet Psychiatry*, *5*(6), 515-526. https://doi.org/10.1016/S2215-0366(18)30059-2

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, *27*(12), 712-719. https://doi.org/10.1016/j.tins.2004.10.007

Kugler, B. T., & Caudrey, D. J. (1983). Phoneme discrimination in schizophrenia. *The British Journal of Psychiatry*, *142*(1), 53-59. https://doi.org/10.1192/bjp.142.1.53

Kuperberg, G. R. (2010). Language in schizophrenia part 1: an introduction. *Language and Linguistics Compass*, *4*(8), 576-589. https://doi.org/10.1111/j.1749-818X.2010.00216.x

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?. *Language, Cognition and Neuroscience*, *31*(1), 32-59. https://doi.org/10.1080/23273798.2015.1102299

Kuperberg, G. R., McGuire, P. K., & David, A. S. (1998). Reduced sensitivity to linguistic context in schizophrenic thought disorder: evidence from on-line monitoring for words in linguistically anomalous sentences. *Journal of Abnormal Psychology*, *107*(3), 423. https://doi.org/10.1037/0021-843X.107.3.423

Kuperberg, G. R., McGuire, P. K., & David, A. S. (2000). Sensitivity to linguistic anomalies in spoken sentences: a case study approach to understanding thought disorder in schizophrenia. *Psychological Medicine*, *30*(2), 345-357. https://doi.org/10.1017/S0033291700001744

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, *5*, 39. https://doi.org/10.3389/fnhum.2011.00039

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746-748. https://doi.org/10.1038/264746a0

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, *118*(2), 219. https://doi.org/10.1037/a0022325

Moghaddam, B., & Javitt, D. (2012). From revolution to evolution: the glutamate hypothesis of schizophrenia and its implication for treatment. *Neuropsychopharmacology*, *37*(1), 4-15. https://doi.org/10.1038/npp.2011.181

Moseley, P., Fernyhough, C., & Ellison, A. (2013). Auditory verbal hallucinations as atypical inner speech monitoring, and the potential of neurostimulation as a treatment option. *Neuroscience & Biobehavioral Reviews*, *37*(10), 2794-2805. https://doi.org/10.1016/j.neubiorev.2013.10.001

Nakazawa, K., Zsiros, V., Jiang, Z., Nakao, K., Kolata, S., Zhang, S., & Belforte, J. E. (2012). GABAergic interneuron origin of schizophrenia pathophysiology. *Neuropharmacology*, *62*(3), 1574-1583. https://doi.org/10.1016/j.neuropharm.2011.01.022

Popa, A. (2018). *Inviting hallucinatory percepts during speech-listening to detect cognitive changes in early psychosis* (Publication No. 5854). [Master's thesis, Western University]. Electronic Thesis and Dissertation Repository. https://ir.lib.uwo.ca/etd/5854

Powers, A. R., Mathys, C., & Corlett, P. R. (2017). Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. *Science*, *357*(6351), 596-600. https://doi.org/10.1126/science.aan3458

Regier, D. A., Narrow, W. E., Rae, D. S., Manderscheid, R. W., Locke, B. Z., & Goodwin, F. K. (1993). The de facto US mental and addictive disorders service system: Epidemiologic Catchment Area prospective 1-year prevalence rates of disorders and services. *Archives of General Psychiatry*, *50*(2), 85-94. https://doi.org/10.1001/archpsyc.1993.01820140007001

Roche, E., Segurado, R., Renwick, L., McClenaghan, A., Sexton, S., Frawley, T., ... & Clarke, M. (2016). Language disturbance and functioning in first episode psychosis. *Psychiatry Research*, *235*, 29-37. https://doi.org/10.1016/j.psychres.2015.12.008

Silverstein, S. M., & Keane, B. P. (2011). Perceptual organization impairment in schizophrenia and associated brain mechanisms: review of research from 2005 to 2010. *Schizophrenia Bulletin*, *37*(4), 690-699. https://doi.org/10.1093/schbul/sbr052

Smith, R., Friston, K., & Whyte, C. (2021, January 2). A Step-by-Step Tutorial on Active Inference and its Application to Empirical Data. https://doi.org/10.31234/osf.io/b4jm6

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., ... & Corlett, P. R. (2018). The predictive coding account of psychosis. *Biological Psychiatry*, *84*(9), 634-643. https://doi.org/10.1016/j.biopsych.2018.05.015

Teufel, C., Subramaniam, N., Dobler, V., Perez, J., Finnemann, J., Mehta, P. R., ... & Fletcher, P. C. (2015). Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proceedings of the National Academy of Sciences*, *112*(43), 13401-13406. https://doi.org/10.1073/pnas.1503916112

Valton, V., Romaniuk, L., Steele, J. D., Lawrie, S., & Seriès, P. (2017). Comprehensive review: computational modelling of schizophrenia. *Neuroscience & Biobehavioral Reviews*, *83*, 631-646. https://doi.org/10.1016/j.neubiorev.2017.08.022

Wu, C., Cao, S., Zhou, F., Wang, C., Wu, X., & Li, L. (2012). Masking of speech in people with first-episode schizophrenia and people with chronic schizophrenia. *Schizophrenia Research*, *134*(1), 33-41. https://doi.org/10.1016/j.schres.2011.09.019

Appendix A

Documentation of Ethics Approval

Western Research

Date: 30 September 2020

To: Dr. Lena Palaniyappan

Project ID: 108268

Study Title: The Pathophysiology of Thought Disorder in Psychosis (TOPSY)

Application Type: Continuing Ethics Review (CER) Form

Review Type: Delegated

REB Meeting Date: 20/Oct/2020

Date Approval Issued: 30/Sep/2020

REB Approval Expiry Date: 24/Oct/2021

Dear Dr. Lena Palaniyappan,

The Western University Research Ethics Board has reviewed the application. This study, including all currently approved documents, has been re-approved until the expiry date noted above.

REB members involved in the research project do not participate in the review, discussion or decision.

Western University REB operates in compliance with, and is constituted in accordance with, the requirements of the TriCouncil Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2); the International Conference on Harmonisation Good Clinical Practice Consolidated Guideline (ICH GCP); Part C, Division 5 of the Food and Drug Regulations; Part 4 of the Natural Health Products Regulations; Part 3 of the Medical Devices Regulations and the provisions of the Ontario Personal Health Information Protection Act (PHIPA 2004) and its applicable regulations. The REB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Please do not hesitate to contact us if you have any questions.

Sincerely,

The Office of Human Research Ethics

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Appendix B

Figure B1

Sample Learning Rate Convergence Plot for a Control Subject

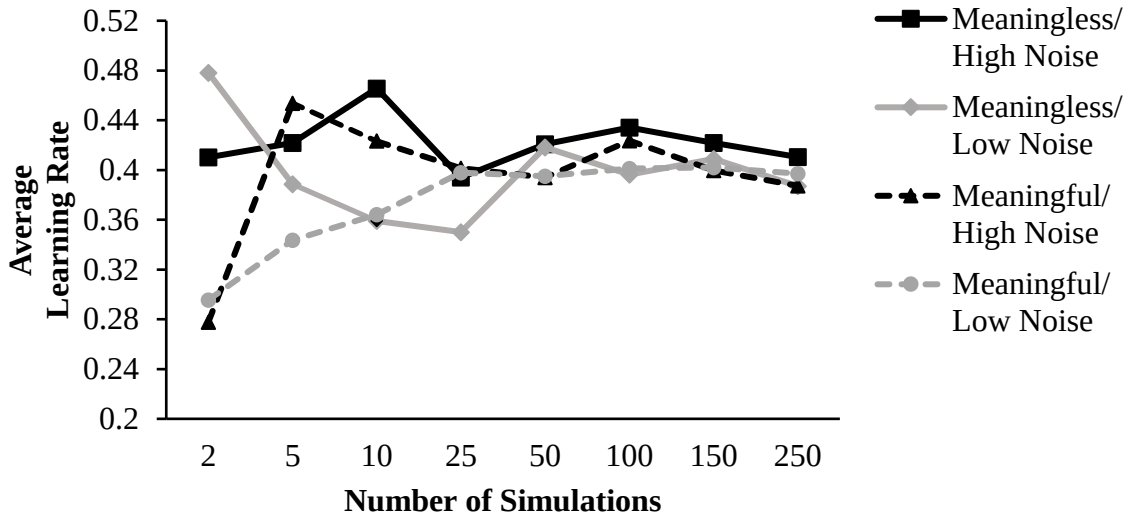


Figure B2

Sample Inverse Temperature Convergence Plot for a Control Subject

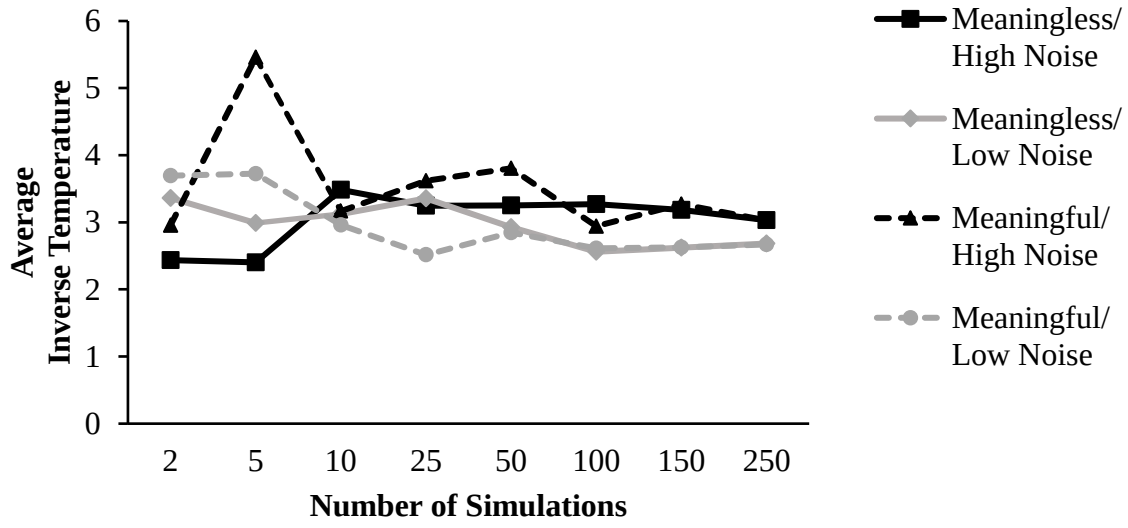


Figure B3

Sample Model Evidence Convergence Plot for a Control Subject

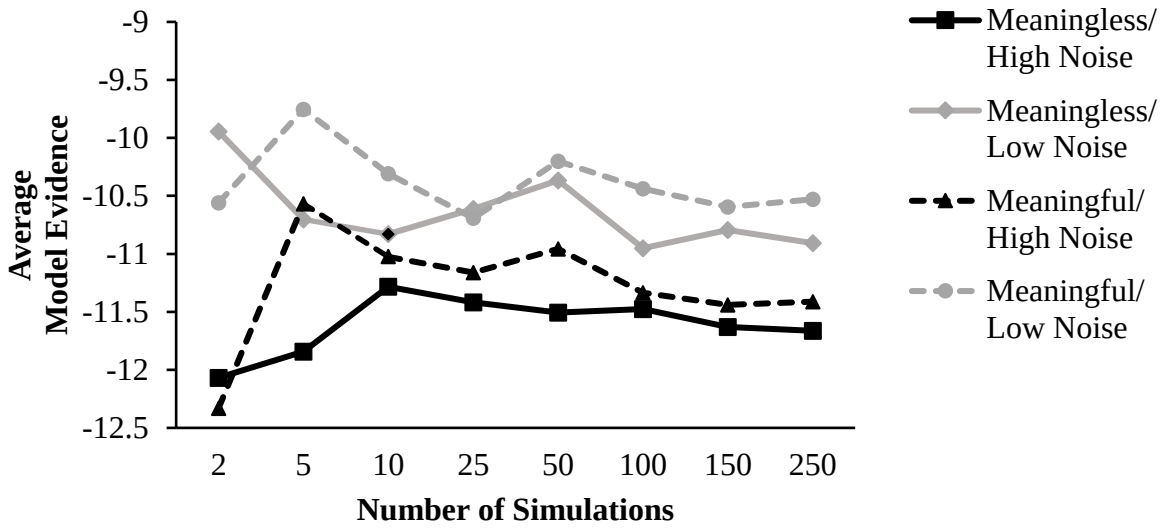


Figure B4

Sample Learning Rate Convergence Plot for a Clinical Subject

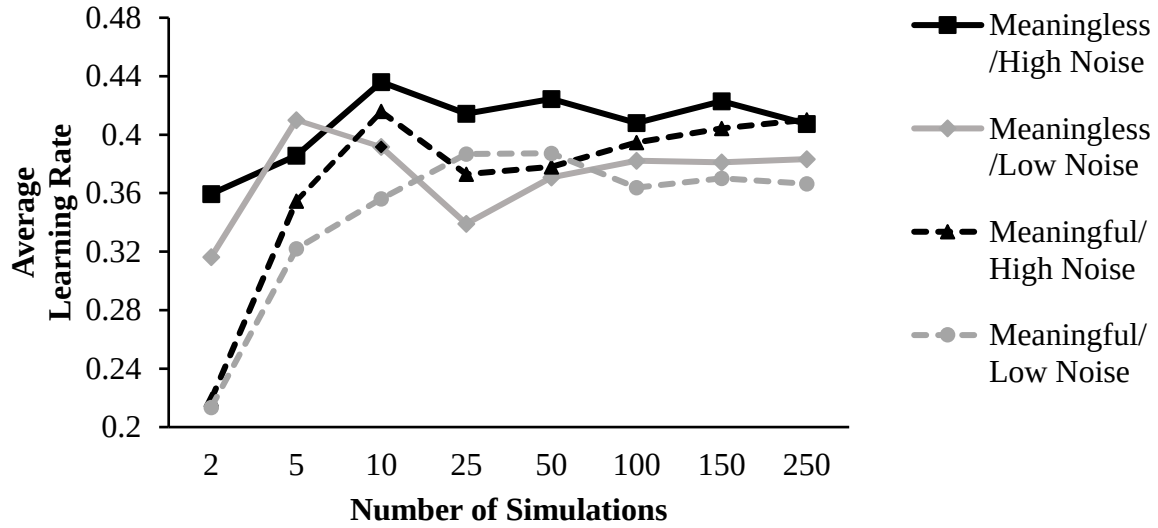


Figure B5

Sample Inverse Temperature Convergence Plot for a Clinical Subject

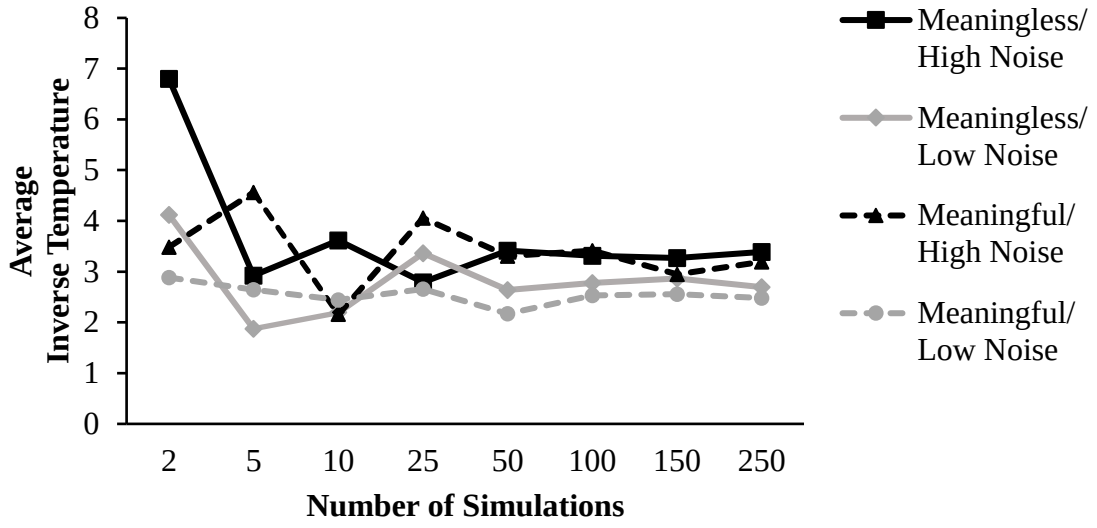


Figure B6

Sample Model Evidence Convergence Plot for a Clinical Subject