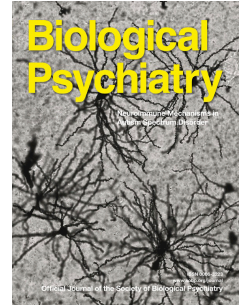


Journal Pre-proof

Computational phenotyping of aberrant belief updating in individuals with schizotypal traits and schizophrenia

Nace Mikus, Claus Lamm, Christoph Mathys



PII: S0006-3223(24)01554-3

DOI: <https://doi.org/10.1016/j.biopsych.2024.08.021>

Reference: BPS 15582

To appear in: *Biological Psychiatry*

Received Date: 8 December 2023

Revised Date: 10 August 2024

Accepted Date: 13 August 2024

Please cite this article as: Mikus N., Lamm C. & Mathys C., Computational phenotyping of aberrant belief updating in individuals with schizotypal traits and schizophrenia, *Biological Psychiatry* (2024), doi: <https://doi.org/10.1016/j.biopsych.2024.08.021>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Published by Elsevier Inc on behalf of Society of Biological Psychiatry.

Computational phenotyping of aberrant belief updating in individuals with schizotypal traits and schizophrenia

Nace Mikus^{1,2}, Claus Lamm^{1,*} & Christoph Mathys^{2,3,4,*}

¹ Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Austria;

² Interacting Minds Centre, Aarhus University, Denmark;

³ Translational Neuromodeling Unit, University of Zurich and ETH Zurich, Zurich, Switzerland;

⁴ Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy

* Shared senior authors

Corresponding author: Nace Mikus, nace.mikus@univie.ac.at, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Austria, Liebiggasse 5, 1010 Vienna, Austria

Keywords: Predictive processing, Bayesian inference, Psychotic symptoms, Test-retest reliability, Symptom differentiation, Predictive inference task

Short title: Computational phenotyping of Psychotic Experiences

Abstract

Background

Psychotic experiences are thought to emerge from various interrelated patterns of disrupted belief updating, such as overestimating the reliability of sensory information and misjudging task volatility. Yet, these substrates have never been jointly addressed under one computational framework and it is not clear to what degree they reflect trait-like computational patterns.

Methods

We introduced a novel hierarchical Bayesian model that describes how individuals simultaneously update their beliefs about the task volatility and noise in observation. We applied this model to data from a modified Predictive inference task in a test-retest study with healthy volunteers (N=45, 4

sessions) and examined the relationship between model parameters and schizotypal traits in a larger online sample (N = 437) and in a cohort of patients with schizophrenia (N = 100).

Results

The interclass correlations were moderate to high for model parameters and excellent for averaged belief trajectories and precision-weighted learning rates estimated through hierarchical Bayesian inference. We found that uncertainty about the task volatility was related to schizotypal traits and to positive symptoms in patients, when learning to gain rewards. In contrast, negative symptoms in patients were associated with more rigid beliefs about observational noise, when learning to avoid losses.

Conclusion

These findings suggest that individuals with schizotypal traits across the psychosis continuum are less likely to learn or utilize higher-order statistical regularities of the environment and showcase the potential of clinically relevant computational phenotypes for differentiating symptom groups in a transdiagnostic manner.

Introduction

Computational phenotypes are computational parameters that provide trait-like, mechanistic descriptions of psychological, behavioral, or neural processes and have the potential to guide psychiatric treatment (1,2). One area where computational approaches have proven especially promising is in describing schizotypal traits across the severity spectrum as aberrant inference or learning (3–6).

Under normative accounts of statistical inference, the degree to which new observations reflect meaningful events in our environment should depend on our beliefs about several distinct sources of uncertainty, such as our beliefs about outcome uncertainty (how reliable or noisy we feel our observation was) and environmental uncertainty (how likely we believe such an event to be) (7,8). Overestimating either the precision of sensory sensations or the probability of unlikely events can both lead to attributing too much saliency to internal sensations and random external events, leading to a psychological state where the world seems imbued with significant experiences that provoke bizarre explanations (4,9,10).

The challenge of separating outcomes arising from noise in observation from outcomes that reflect latent environmental changes has been operationalized in probabilistic reversal learning tasks. Empirical research using various variations of learning tasks has shown that patients with schizophrenia tend to update their beliefs more rapidly (11,12), overestimate the likelihood of a contextual change (13), and are more likely to switch between choices in two-choice tasks (14,15). Computational work implies that these behavioral patterns result from an overestimation of and uncertainty about the task volatility (16–18); however, several important issues remain to be resolved.

First, all computational work in this domain only investigated either beliefs about volatility or beliefs about observation noise and did not account for both simultaneously and within one common framework. This is an important limitation because the two mechanisms are computationally interdependent, and abnormalities in one might be attributed to abnormalities in the other (19). They might also represent two different factors of the illness and recruit different brain circuits (20,21). Second, findings of increased updating in patients with schizophrenia are hard to reconcile with a range of studies showing that patients demonstrate reduced belief updating and impaired flexibility (22–25), and display overconfidence across a range of social, cognitive, and learning tasks (26–30). Third, a major setback in defining outputs of computational models as clinically relevant phenotypes has been that their test-retest reliability is either unknown (31), or modest at best (32–34). Despite recent progress demonstrating fair to good reliability of parameters in reinforcement

learning tasks (35–40), test-retest scores of parameters representing beliefs about more abstract task features, such as task volatility or outcome variance, are simply not known. Finally, it is not clear to what degree the proposed belief updating patterns are specific to positive symptoms, rather than to the syndrome of schizophrenia more generally, and whether they represent transdiagnostic patterns on a whole spectrum of clinical severity.

The present study aimed to address these issues. We first introduced a model that describes how individuals dynamically update their beliefs about environmental volatility as well as observational noise when both can dynamically change throughout the task. We then determine the minimal task conditions under which satisfactory reliability of computational parameters is feasible. We then empirically investigated whether the parameters of the computational model are reliable and clinically meaningful measures. To this end, we first conducted a test-retest study in healthy volunteers and looked at which of the three computational patterns predicts subclinical delusional ideation (Experiment 1). We then replicated and extended the findings of Experiment 1 in a larger data set of healthy participants (Experiment 2), and a data set of patients with schizophrenia (Experiment 3).

Methods and Materials

Computational Modelling

How an observer updates their beliefs about a latent cause in an environment where the (sensory) observations are noisy and the underlying cause changes has been operationalized within the Predictive Inference Task (Figure 1A). In one commonly used cover story, the participants need to guess the position of a hidden helicopter by simply observing where items dropped from that helicopter appear on the ground (23). Observational noise is introduced by the dropped items randomly deviating from the actual position of the helicopter and the latent change is operationalized by the helicopter changing its position. The only way to notice the change is to infer it from the dropped items. Participants are asked to predict the location of the next drop and (optionally) how confident they are in their prediction. In this task, a prediction error is simply the difference between the predicted and actual outcome, and the (model agnostic) learning rate is the change in the response from one trial to the next, normalized by the prediction-error. Behavioral measures include average learning rates, the adaptation of the learning rate based on outcome variance and the effect of the prediction error on confidence ratings.

To model how participants form beliefs about the underlying state and the associated observational noise, we use a meta-Bayesian framework(41,42). We define the observers's assumed generative

model of how observations are generated and invert this model, which gives us a set of update equations that amount to a generalization of the hierarchical Gaussian filter (8,43,44). In a well-established two-level version of the HGF the mean latent variable x and its volatility x_v are modeled as states, and the observational noise is a fixed parameter α (8). Here, we modify this model to instead model observational noise as a state, x_α , with its own volatility. This allows us to infer not only the participants' mean beliefs and uncertainties about the overall mean μ_x , environmental volatility μ_v , but also their mean beliefs about observational noise μ_α . Importantly, the evolutions of the three described states are governed by three agent-specific parameters: variance volatility parameter ω_α (the rate of change of the variance state), environmental volatility parameter ϑ (the rate of change of the volatility of the mean), and tonic volatility parameter ω (the rate of change of the mean x , after accounting for the environmental volatility, see Supplementary Table 1 for update equations and Supplementary Figure 1).

The likelihood function (or the decision model) probabilistically maps beliefs about the mean μ_x to behavior (y) with a Gaussian distribution:

$$y^{(t)} = N(\mu^{(t)}, \exp(\eta)),$$

where η stands for decision noise, or measurement error (on a log scale), which represents the participant-level error term in the parameter fitting, and therefore captures behavior of subjects not accounted for by the model (such as lack of attention, motivation, or other random behavior).

Parameter estimation was done both without pooling and hierarchically (Refer to Supplementary Methods for details on parameter estimation as well as ICC calculation).

Experiment 1

Participants and procedures

Fifty-five ($N = 55$) healthy adults took part in four in-person testing sessions separated by at least 1 week, out of which $N = 45$ finished all four sessions and were included in the final analysis. On each session participants performed a behavioral battery that included the Predictive inference task, as well as an exploration-exploitation task and a losing of control task (45) that will be published elsewhere. In the third session, participants filled out the Peters et al. Delusion Inventory (PDI)(46) and in the fourth session they played the Two-step task (47,48). The participants earned a base fee of € 40 and could earn up to € 80 through bonus payments from the tasks. All participants gave written informed consent, and the experiment was approved by the ethics committee of the University of Vienna (# 1918/2015).

The predictive inference task

In the adapted Predictive inference task (49) task, participants are required to predict the outcome of each trial. The outcome is a random number between 0 and 100, drawn from a normal Gaussian distribution with a certain mean and standard deviation, with a rebounding boundary. In contrast to the task by Nassar and colleagues, both the mean and the standard deviation change several times throughout the task. The probability of the mean changing is 0.1 except for the first 3 trials of each block and the probability of the standard deviation to change when the mean changed is 0.4, and could be either high (15) or low (5). Participants could express their confidence with a longer button press and performance was incentivized based on prediction errors (the smaller the prediction error, the higher the reward). Participants played 240 trials in four sessions with four different trajectories that did not change across participants. The task was embedded in a cover story, where an alien mines for gold. For a detailed description of the task see Supplementary Methods.

The two-step task

The two-step task is often used as a behavioral measure for the capacity for model-based decision making (47,48), whereby keeping a “model” of the mapping between step one and step two can improve points earned in the task (see Supplementary Methods for details). For 5 participants, the data of the task was not saved, therefore, a total of 40 participants were included in the analysis of these data.

Behavioral analysis

For frequentist mixed linear models, the lme4 package in R was used, and for Bayesian mixed linear models, the brms package was used. For frequentist models we report Confidence Intervals (CI)s and p values (P). for Bayesian models we report Credibility intervals (CrI) and the probability mass of the posterior distribution that lies above (or below) zero (P). For details on behavioral analysis see Supplementary Methods.

Experiment 2

In the second experiment, we reanalyzed data from a previously published online study(50). Here, a sample of 437 MTurk participants (Table 1) played a version of the Predictive inference task that deviated from that in Experiment 1 in several ways. First, the outcome was generated with constant noise, but under varying volatility; second, the cover story required of participants to catch particles on a circle; third, confidence was expressed on a visual analogue scale after each prediction; and fourth, participants could lose points if their prediction was poor enough (see Supplementary Methods and (50) for details).The participants filled out a wide range of questionnaires assessing general psychopathology including the Short Scales for Measuring Schizotypy that consists of four

subscales (51). In our main analysis we used the Unusual Experience (UE) subscale and the Introvertive Anhedonia (IA) subscale that are phenomenologically related to positive and negative symptoms of psychosis, respectively. In an exploratory fashion, we also looked at how our model parameters relate to the anxious-depression factor of the factor analysis reported in (50).

Experiment 3

In the third experiment, we reanalyzed data from a previously published study with patients with schizophrenia (23). Data was used from 102 participants (Table 1) with a diagnosis of schizophrenia or schizoaffective disorder collected at the Maryland Psychiatric Research Center, University of Maryland School of Medicine. Participants were stable outpatients, most treated with antipsychotic medications. Inclusion criteria in patients were a presence of a schizophrenia spectrum disorder in patients. Exclusion criteria were a presence of a neurological disorder, or a cognitively impairing medical disorder. The severity of positive symptoms was assessed with the Brief Psychiatric Rating Scale (BPRS (52)), averaging the items grandiosity, suspiciousness, unusual thought content, and hallucinations. Negative symptoms were estimated with the average rating across all items of the Scale for the Assessment of Negative Symptoms (SANS (53)). For 2 participants, symptom data were missing, and participants were excluded. In the Predictive inference task performance was incentivized in two different ways across two different task sessions, each lasting for 100 trials. In the reward seeking condition, participants could earn points and in the loss avoiding condition, participants would have points taken away from their initial endowment in a performance contingent way. In contrast to Experiment 1, no confidence data was collected, and the cover story was built around a helicopter obscured by clouds dropping items (see Supplementary Methods for details).

Results

Computational Model

We used simulations to determine under which conditions the recovery and test-retest reliability of computational parameters in an experiment would be sufficient. We found that parameter recovery is generally excellent when trial numbers are adequate (Supplementary Figure 2). We then simulated behavior on two different trajectories and calculated ICCs (type ICC3k, see Supplementary Methods). Simulated ICCs for all three parameters (even when comparing task trajectories with various trial lengths) were excellent with sufficient trial numbers conditional on decision noise (Figure 2A, Supplementary Figure 3), in contrast to binary task structures and associated models (Figure 2B).

Experiment 1

Model and behavior

Participants (N = 45, see Table 1 for demographics) in our task tend to adapt their learning rate ($b = -0.167$, SE = 0.011, $d = 0.168$, $t(139.34) = -15.761$, $P < 0.001$, Figure 3A), beliefs about variance ($b = 0.108$, SE = 0.015, $d = 0.190$, $t(55.245) = 7.199$, $P < 0.001$, Figure 3B) and volatility ($b = 0.14519$, SE = 0.01588, $d = 0.265$, $t(52.504) = 9.141$, $P < 0.001$, Figure 3C) across the two variance block. Model comparison revealed that the model with all three parameters outperformed the two other models with fixed ϑ (Supplementary Table 2 and 3).

Hierarchically derived ICC scores were the following: for ω the mean was 0.787 (95% CrI [0.681, 0.874]), for ϑ the mean was 0.557 (95% CrI [0.261, 0.819]) and for ω_α the mean was 0.491 (95% CrI [0.264, 0.711], Figure 3D), indicating at least moderate ICCs. However, when looking at the ICC scores (ICC3k, absolute raters) of averaged trial-by-trial trajectories calculated with a separate linear model (Figure 3E), we found excellent test-retest scores for the precision-weighted learning rates ψ_1 (ICC = 0.93, 95% CI [0.88, 0.96]), for mean beliefs about volatility μ_v (ICC = 0.92, 95% CI [0.87, 0.95]), as well as noise μ_α (ICC = 0.92, 95% CrI [0.88, 0.96]). Note that the ICC scores are lower when parameter estimation was done without pooling (Supplementary Figures 4, 5 and 6), or when considering single rater ICC scores.

We defined three participant-level behavioral patterns that reflect implicit learning rates (B_U), learning rate adaptation to variance (B_L), and confidence adaptation to prediction error (B_C). All three had excellent ICC (all mean ICC > 0.9, Figure 3F, Supplementary Figure 7) and were differentially related to the model parameters (Figure 3G-I, Supplementary Table 4-6). Furthermore, we found that ϑ correlated negatively to the performance in the Two-step task ($r = -0.371$, $P = 0.048$, Figure 3J).

Delusional ideation is associated with the environmental volatility parameter

We found that higher PDI was related to increased environmental volatility parameter ϑ (effect size $d = 0.687$, 95% CrI [0.027, 1.373], $P(d < 0) = 0.02$, Figure 4A, B), and no evidence for a decrease in the tonic volatility parameter ω ($d = 0.176$, 95% CrI [-0.407, 0.771], $P(d < 0) = 0.286$ and no conclusive evidence of an effect on the noise volatility parameter ω_α ($d = -0.548$, 95% CrI [-1.317, 0.198], $P(d > 0) = 0.072$, Supplementary Table 7). We found no robust evidence of PDI's effect on the behavioral pattern associated with ϑ , confidence behavior ($b_{PE:PDI} = 0.009$, 95% CrI [-0.011, 0.029], $P(b < 0) = 0.186$, Figure 4c).

Experiment 2

We analyzed the data from (50) with the same computational model as in Experiment 1 and investigated how the model parameters relate to schizotypal traits. Associating the scores of the positive schizotypy (UE subscale) with the model parameters, controlling for age, gender, and IQ we found positive correlations with the environmental volatility parameter ϑ ($d = 0.092$, 95% CrI [-0.001, 0.186], $P(d < 0) = 0.026$, Figure 4D), with no conclusive evidence for an effect on ω_α (Supplementary Table 8). For the negative schizotypy (IA subscale) (Figure 4E), we found weak evidence of a negative correlation with ω_α ($d = -0.081$, 95% CrI [-0.179, 0.017], $P(d > 0) = 0.053$) and ω ($d = -0.071$, 95% CrI [-0.168, 0.023], $P(d > 0) = 0.071$), but no evidence for an association with ϑ (Supplementary Table 9). We also found that UE subscale modulated the effect of prediction error on confidence ratings ($d = 0.035$, 95% CrI [0.015, 0.055], $P(d < 0) < 0.001$, Figure 4F, Supplementary Table 10), but IA subscale did not ($d = 0.01$, 95% CrI [-0.009, 0.029], $P(d < 0) = 0.154$).

As an exploratory measure, we investigated how our model parameters correlate with the “anxious-depression” transdiagnostic factor of the factor analysis, and found that the factor was negatively related to ω ($d = -0.125$, 95% CrI [-0.222, -0.03], $P(d > 0) = 0.006$, Supplementary Table 12), as well as ω_α ($d = -0.111$, 95% CrI [-0.21, -0.013], $P(d > 0) = 0.013$, see Supplementary Materials for models with other subscales and factors of the factor analysis).

Experiment 3

The behavior in this experiment was analyzed with the same model as in Experiments 1 and 2, whereby we estimated the parameters for the two incentivization conditions (reward seeking and loss avoiding) separately, giving six parameters in the belief model. We first looked at the association between the six parameters and the positive symptoms and found that the environmental volatility parameter ϑ estimated in the reward seeking condition was a significant predictor of positive ($d = 0.192$, 95% CrI [-0.013, 0.397], $P(d < 0) = 0.034$, Figure 5A, C, Supplementary Table 14), while we found no credible support for ω ($d = -0.058$, 95% CrI [-0.274, 0.149], $P(d > 0) = 0.287$) nor ω_α ($d = -0.147$, 95% CrI [-0.358, 0.067], $P(d > 0) = 0.088$). We also found no evidence for an association between the parameters estimated in the loss avoidance condition on positive symptoms (all $P > 0.368$). To ensure the robustness of the main effect of ϑ (seek) we ran several other models and found the effect remains stable when including decision noise in both conditions, age, and gender, and finally IQ. We also repeat the analysis by excluding patients with poorest performance (for a details and results of these effects see Supplementary Table 16).

When looking at negative symptoms, we found that in the loss avoiding condition, there was a negative association between ω_α ($d = -0.282$, 95% CrI [-0.498, -0.057], $P(d > 0) = 0.007$, Figure 5B, D,

Supplementary Table 15) and ω ($d = -0.208$, 95% CrI [-0.428, 0.027], $P(b>0) = 0.039$), but not ϑ ($d = 0.071$, 95% CrI [-0.149, 0.282], $P(d<0) = 0.274$). We found no evidence for an association of negative symptoms with the model parameters in the reward seeking condition (all $P > 0.222$).

Finally, we investigated whether the association between ϑ (in the seek condition) and symptoms was specific to the positive symptoms. By comparing the effect sizes using Fisher's z-transform method, we found no convincing differences between the two correlations ($z = 1.110$, $P = 0.133$). To get a more detailed view of how the combinations of model parameters relate to the two symptom variables we conducted a penalized canonical correlation analysis (54) (Figure 5E, for details see Supplementary Methods). Using two latent canonical variables we found that positive symptoms and ϑ in the seek condition load on to correlated canonical variables ($r = 0.254$, $p = 0.038$). In contrast ϑ in the seek condition only has a very weak correlations with the canonical variable that negative symptoms correlate with.

Discussion

We introduce a computational model that describes how individuals with schizophrenia and schizotypal traits learn in environments where both observational noise and the volatility of environmental changes can vary. We demonstrate that positive schizotypal traits and positive symptoms of schizophrenia are related to higher uncertainty about the task volatility, even when accounting for how participants update their beliefs about observational uncertainty.

Previous work suggests that schizotypal traits correlate with volatility estimates across both clinical and non-clinical samples (16,18), however, not accounting for participants' perception of outcome reliability limits the conclusions that can be drawn from these studies (19,55). Importantly, even in stable task environments (without latent changes), patients with schizophrenia do not down-weight their updating in response to higher outcome variance (56) and fail to attenuate sensory signals (6,20,57,58). In our data, positive symptoms in chronic patients and schizotypal traits were associated with rigid beliefs about outcome variance, however the effects did not reach the significance thresholds. Instead, our findings suggest that positive schizotypy and positive symptoms are associated with an inability or unwillingness to extract and utilize higher-order statistical features of the environment.

Proposed biological implementations of hierarchical Bayesian inference suggest that these processes are encoded at higher levels of the neural hierarchy (20,59,60). Neuroimaging studies report that volatility estimates are encoded in the dorsal prefrontal cortex(61,62) and this encoding of volatility might be different in patients with schizophrenia compared to healthy controls (16). It has been

proposed that positive symptoms result from a disruption of prefrontal and hippocampal circuits through hypofunction of glutamatergic receptors (63). This abnormality could reduce the precision of top-down predictions and make the world seem more uncertain and surprising (6). In this context, symptoms may emerge as a compensatory mechanism to explain a seemingly chaotic world (64). This notion is supported by modelling work depicting how having weak priors over possible explanations can serve as a “protective belt” (65), allowing unpredictable events to be explained away with a new hypothesis (e.g., a new change has occurred), while at the same time being consistent with the former belief being held with high confidence and rigidity (66). This is also in-line with our data where higher uncertainty about volatility was associated with higher confidence and reduced model-based behavior.

Using penalized canonical correlation analysis, we find some evidence for the claim that this pattern is related to positive but not negative symptom domains, although a direct comparison of effects using simpler methods failed to show this specificity. We thus regard these findings as tentative and preliminary, and a larger dataset will likely be required for more conclusive evidence (or lack thereof). Interestingly, negative symptoms in patients were related to reduced beliefs about observational noise when avoiding losses. A suggestive pattern in the same direction was observed for negative schizotypy and, more reliably, in the anxious-depression transdiagnostic psychiatric dimension. Previous work in humans and rodents has shown, that estimates of outcome uncertainty and learning rate adaptations are encoded in the dopaminergic midbrain (67–70), and amygdala (71) and were shown to be reduced in anxiety (72). Our results support the notion that negative symptoms, similarly to traits related to anxiety and depression, may be associated with such misestimations of outcome uncertainty, particularly in aversive contexts.

Despite the consistency of our main findings across the three studies, several pertinent issues must be considered. First, there were some important differences between the three experiments. The two measures of schizotypal traits in Experiments 1 and 2 were not the same. While convergent validity between them is, to our knowledge, yet to be established, our results support the notion that the two scales capture similar aspects of schizotypal psychopathology. The task versions across the three experiments also had small but important differences, such as the manner of responding, the cover story, and incentivization schemes. Second, the effects sizes of main effects were small to moderate. Meta-analytic evidence indicates that this is usually the case when considering specific associations of belief updating patterns with positive symptoms (73,74), and such effect sizes can still have a clinically meaningful implications (75), especially given the possibility of repeated measurement, which current standardized questionnaires are not suited for. Third, the significance threshold we used was not very conservative. To mitigate overfitting, we include predictors in a

single model and use weakly informative shrinkage priors for regularization. Taken together, these limitations imply that more direct replications of the results of each (or either) experiment are valuable, as is a more systematic investigation of the effects of various task modifications. Finally, we showed that ICCs of computational parameters can sometimes be improved by using continuous inputs and hierarchical estimation, as shown previously (39,40,76), however, test-retest studies in clinical populations are needed to investigate to what degree these results generalize to patients.

In conclusion, we show that mechanistic descriptions of how individuals with psychotic symptoms and psychotic-like traits process information when faced with various sources of uncertainty have the potential to become reliable, transdiagnostic and clinically meaningful computational phenotypes.

Journal Pre-proof

Data availability

Data and code for Experiment 1 is available online (<https://github.com/nacemikus/jget-schizotypy.git>).

Acknowledgements

This research work was funded by the Vienna Science and Technology Fund (VRG13-007) awarded to Claus Lamm and by the Aarhus University Research Foundation Grant (AUFF-E-2019-7-10) awarded to Christoph Mathys. A previous version of this manuscript was published on medRxiv (<https://medrxiv.org/content/10.1101/2023.11.27.23299069v1>). We want to express our gratitude to Matthew Nassar and colleagues, as well as Tricia Seow and Claire Gillan for freely providing data used in Experiments 2 and 3. We thank Angus MacDonald, Jessica Arend, Graham Murray and two anonymous reviewers for constructive comments on the manuscript.

Competing Interest Statement

NM is a cofounder of TiliaHealth, a precision psychiatry company, that did not fund this work. CM and CL report no biomedical financial interests or potential conflicts of interest.

References

1. Paulus MP, Huys QJM, Maia T V. (2016): A Roadmap for the Development of Applied Computational Psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 1. pp 386–392.
2. Stephan KE, Mathys CD (2014): Computational approaches to psychiatry. *Curr Opin Neurobiol* 25: 85–92.
3. Corlett PR, Frith CD, Fletcher PC (2009): From drugs to deprivation: A Bayesian framework for understanding models of psychosis. *Psychopharmacology (Berl)* 206: 515–530.
4. Fletcher PC, Frith CD (2009): Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat Rev Neurosci* 10: 48–58.
5. Stephan KE, Baldeweg T, Friston KJ (2006): Synaptic Plasticity and Dysconnection in Schizophrenia. *Biological Psychiatry*, vol. 59. pp 929–939.
6. Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ (2013): The Computational Anatomy of Psychosis. *Front Psychiatry* 4: 1–26.
7. Friston KJ (2009): The free-energy principle: a rough guide to the brain? *Trends Cogn Sci* 13: 293–301.
8. Mathys CD, Daunizeau J, Friston KJ, Stephan KE (2011): A bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci* 5: 39.
9. Jaspers K (1997): *General Psychopathology*. JHU Press.
10. Kapur S (2003): Psychosis as a state of aberrant salience: A framework linking biology, phenomenology, and pharmacology in schizophrenia. *Am J Psychiatry* 160: 13–23.
11. Garety PA, Hemsley DR, Wessely S (1991): Reasoning in deluded schizophrenic and paranoid patients biases in performance on a probabilistic inference task. *J Nerv Ment Dis* 179: 194–201.
12. Morawetz C, Bode S, Baudewig J, Jacobs AM, Heekeren HR (2016): Neural representation of emotion regulation goals. *Hum Brain Mapp* 37: 600–620.
13. Kaplan CM, Saha D, Molina JL, Hockeimer WD, Postell EM, Apud JA, *et al.* (2016): Estimating changing contexts in schizophrenia. *Brain* 139: 2082–2095.
14. Schlagenhauf F, Huys QJM, Deserno L, Rapp MA, Beck A, Heinze HJ, *et al.* (2014): Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. *Neuroimage* 89:

171–180.

15. Fromm S, Katthagen T, Deserno L, Heinz A, Kaminski J, Schlagenhauf F (2023): Belief Updating in Subclinical and Clinical Delusions. *Schizophr Bull Open* 4: 1–11.
16. Deserno L, Boehme R, Mathys CD, Katthagen T, Stephan KE, Heinz A, Schlagenhauf F (2019): Volatility estimates increase choice switching and relate to prefrontal activity in schizophrenia. *Biol Psychiatry Cogn Neurosci Neuroimaging*. <https://doi.org/10.1016/j.bpsc.2019.10.007>
17. Henco L, Diaconescu AO, Lahnakoski JM, Brandi ML, Hörmann S, Hennings J, *et al.* (2020): Aberrant computational mechanisms of social learning and decision-making in schizophrenia and borderline personality disorder. *PLoS Comput Biol* 16: 1–22.
18. Reed EJ, Uddenberg S, Suthaharan P, Mathys CD, Taylor JR, Groman SM, Corlett PR (2020): Paranoia as a deficit in non-social belief updating. *Elife* 9: 1–55.
19. Piray P, Daw ND (2021): A model for learning based on the joint estimation of stochasticity and volatility. *Nat Commun* 12. <https://doi.org/10.1038/s41467-021-26731-9>
20. Sterzer P, Adams RA, Fletcher PC, Frith CD, Lawrie SM, Muckli L, *et al.* (2018): The Predictive Coding Account of Psychosis. *Biol Psychiatry* 84: 634–643.
21. Iglesias S, Tomiello S, Schneebeli M, Stephan KE (2017): Models of neuromodulation for computational psychiatry. *Wiley Interdiscip Rev Cogn Sci* 8: 1–22.
22. Baker SC, Konova AB, Daw ND, Horga G (2019): A distinct inferential mechanism for delusions in schizophrenia. *Brain* 142: 1797–1812.
23. Nassar MR, Waltz JA, Albrecht MA, Gold JM, Frank MJ (2021): All or nothing belief updating in patients with schizophrenia reduces precision and flexibility of beliefs. *Brain* 1–17.
24. Doll BB, Waltz JA, Cockburn J, Brown JK, Frank MJ, Gold JM (2014): Reduced susceptibility to confirmation bias in schizophrenia. *Cogn Affect Behav Neurosci* 14: 715–28.
25. Waltz JA, Gold JM (2007): Probabilistic reversal learning impairments in schizophrenia: Further evidence of orbitofrontal dysfunction. *Schizophr Res* 93: 296–303.
26. Moritz S, Ramdani N, Klass H, Andreou C, Jungclaussen D, Eifler S, *et al.* (2014): Overconfidence in incorrect perceptual judgments in patients with schizophrenia. *Schizophr Res Cogn* 1: 165–170.
27. Köther U, Veckenstedt R, Vitzthum F, Roesch-Ely D, Pfueller U, Scheu F, Moritz S (2012): “Don’t

- give me that look” - Overconfidence in false mental state perception in schizophrenia.
Psychiatry Res 196: 1–8.
28. Hahn S, Moritz S, Elmers J, Scheunemann J (2021): Do you like cliff-hangers? Objective versus subjective need for closure in the schizophrenia spectrum. *Schizophr Res* 238: 20–26.
29. Hoven M, Lebreton M, Engelmann JB, Denys D, Luigjes J, van Holst RJ (2019): Abnormalities of confidence in psychiatry: an overview and future perspectives. *Transl Psychiatry* 9.
<https://doi.org/10.1038/s41398-019-0602-7>
30. Rossi-Goldthorpe RA, Leong YC, Leptourgos P, Corlett PR (2021): Paranoia, self-deception and overconfidence. *PLOS Comput Biol* 17: e1009453.
31. Browning M, Carter CS, Chatham C, Den Ouden H, Gillan CM, Baker JT, *et al.* (2020): Realizing the Clinical Potential of Computational Psychiatry: Report From the Banbury Center Meeting, February 2019. *Biol Psychiatry* 88: e5–e10.
32. Enkavi AZ, Eisenberg IW, Bissett PG, Mazza GL, Mackinnon DP, Marsch LA, Poldrack RA (2019): Large-scale analysis of test – retest reliabilities of self-regulation measures. *Proc Natl Acad Sci U S A* 1–6.
33. Schaaf J V., Weidinger L, Molleman L, van den Bos W (2023): Test–retest reliability of reinforcement learning parameters. *Behav Res Methods*. <https://doi.org/10.3758/s13428-023-02203-4>
34. Neuser MP, Kühnel A, Kräutlein F, Teckentrup V, Svaldi J, Kroemer NB (2023): Reliability of gamified reinforcement learning in densely sampled longitudinal assessments. *PLOS Digit Heal* 2: e0000330.
35. Loosen AM, Seow TXF, Hauser TU (2022): Consistency within change: Evaluating the psychometric properties of a widely-used predictive-inference task. *PsyArXiv* 62.
36. Waltmann M, Schlagenhaut F, Deserno L (2022): Sufficient reliability of the behavioral and computational readouts of a probabilistic reversal learning task. *Behav Res Methods* 54: 2993–3014.
37. Mkrtchian A, Valton V, Roiser JP (2023): Reliability of Decision-Making and Reinforcement Learning Computational Parameters. *Comput Psychiatry* 7: 30–46.
38. Toyama A, Katahira K, Kunisato Y (2023): Examinations of Biases by Model Misspecification and Parameter Reliability of Reinforcement Learning Models. *Comput Brain Behav* 6: 651–670.

39. Katahira K (2016): How hierarchical models improve point estimates of model parameters at the individual level. *J Math Psychol* 73. <https://doi.org/10.1016/j.jmp.2016.03.007>
40. Katahira K, Oba T, Toyama A (2024): Does the reliability of computational models truly improve with hierarchical modeling? Some recommendations and considerations for the assessment of model parameter reliability: Reliability of computational model parameters. *Psychon Bull Rev.* <https://doi.org/10.3758/s13423-024-02490-8>
41. Daunizeau J, Den Ouden H, Pessiglione M, Kiebel SJ, Stephan KE, Friston KJ (2010): Observing the observer (I): Meta-bayesian models of learning and decision-making. *PLoS One* 5. <https://doi.org/10.1371/journal.pone.0015554>
42. Daunizeau J, Den Ouden H, Pessiglione M, Kiebel SJ, Friston KJ, Stephan KE (2010): Observing the observer (II): Deciding when to decide. *PLoS One* 5. <https://doi.org/10.1371/journal.pone.0015555>
43. Mathys CD, Lomakina EI, Daunizeau J, Iglesias S, Brodersen KH, Friston KJ, Stephan KE (2014): Uncertainty in perception and the Hierarchical Gaussian Filter. *Front Hum Neurosci* 8. <https://doi.org/10.3389/fnhum.2014.00825>
44. Weber LA, Waade PT, Legrand N, Møller AH, Stephan KE, Mathys CD (2023): The generalized Hierarchical Gaussian Filter. 1–38.
45. Mikus N, Mancinelli F, Lamm C, Mathys CD (n.d.): A novel task for measuring sensitivity to loss of control.
46. Peters E, Joseph S, Day S, Garety PA (2004): Measuring delusional ideation: The 21-item Peters et al. Delusions Inventory (PDI). *Schizophr Bull* 30: 1005–1022.
47. Daw ND, Niv Y, Dayan P (2005): Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci* 8: 1704–11.
48. Kool W, Cushman FA, Gershman SJ (2016): When Does Model-Based Control Pay Off? ((J. X. O'Reilly, editor)). *PLoS Comput Biol* 12: e1005090.
49. Nassar MR, Wilson RC, Heasley B, Gold JI (2010): An Approximately Bayesian Delta-Rule Model Explains the Dynamics of Belief Updating in a Changing Environment. *J Neurosci* 30: 12366–12378.
50. Seow TXF, Gillan CM (2020): Transdiagnostic Phenotyping Reveals a Host of Metacognitive Deficits Implicated in Compulsivity. *Sci Rep* 10: 1–11.

51. Mason O, Linney Y, Claridge G (2005): Short scales for measuring schizotypy. *78*: 293–296.
52. Overall JE, Gorham DR (1962): The brief psychiatric rating scale. *Psychol Rep* 10: 799–812.
53. Andreasen NC (1989): The Scale for the Assessment of Negative Symptoms (SANS): conceptual and theoretical foundations. *Br J psychiatry* 155: 49–52.
54. Witten DM, Tibshirani R, Hastie T (2009): A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* 10: 515–534.
55. Pulcu E, Browning M (2019): The Misestimation of Uncertainty in Affective Disorders. *Trends Cogn Sci* 23: 865–875.
56. Haarsma J, Fletcher PC, Griffin JD, Taverne HJ, Ziauddeen H, Spencer TJ, *et al.* (2020): Precision weighting of cortical unsigned prediction error signals benefits learning, is mediated by dopamine, and is impaired in psychosis. *Mol Psychiatry*. <https://doi.org/10.1038/s41380-020-0803-8>
57. Weilhhammer V, Röd L, Eckert AL, Stuke H, Heinz A, Sterzer P (2020): Psychotic experiences in schizophrenia and sensitivity to sensory evidence. *Schizophr Bull* 46: 927–936.
58. Schmack K, Schnack A, Priller J, Sterzer P (2015): Perceptual instability in schizophrenia: Probing predictive coding accounts of delusions with ambiguous stimuli. *Schizophr Res Cogn* 2: 72–77.
59. Heinz A, Murray GK, Schlagenhaut F, Sterzer P, Grace AA, Waltz JA (2019): Towards a Unifying Cognitive, Neurophysiological, and Computational Neuroscience Account of Schizophrenia. *Schizophr Bull* 45: 1092–1100.
60. Lisman JE, Grace AA (2005): The hippocampal-VTA loop: Controlling the entry of information into long-term memory. *Neuron*. <https://doi.org/10.1016/j.neuron.2005.05.002>
61. Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007): Learning the value of information in an uncertain world. *Nat Neurosci* 10: 1214–21.
62. McGuire JT, Nassar MR, Gold JI, Kable JW (2014): Functionally Dissociable Influences on Learning Rate in a Dynamic Environment. *Neuron* 84: 870–881.
63. Howes OD, McCutcheon R, Stone J (2015): Glutamate and dopamine in schizophrenia: An update for the 21st century. *J Psychopharmacol* 29: 97–115.
64. Corlett PR, Taylor JR, Wang XJ, Fletcher PC, Krystal JH (2010): Toward a neurobiology of delusions. *Prog Neurobiol* 92: 345–369.

65. Gershman SJ (2018): How to never be wrong. *Psychon Bull Rev* 1.
66. Erdmann T, Mathys CD (2022): A generative framework for the study of delusions. *Schizophr Res* 245: 42–49.
67. Diederer K MJ, Ziauddeen H, Vestergaard MD, Spencer T, Schultz W, Fletcher PC (2017): Dopamine modulates adaptive prediction error coding in the human midbrain and striatum. *J Neurosci* 37: 1708–1720.
68. Diederer K MJ, Spencer T, Vestergaard MD, Fletcher PC, Schultz W (2016): Adaptive Prediction Error Coding in the Human Midbrain and Striatum Facilitates Behavioral Adaptation and Learning Efficiency Adaptive Prediction Error Coding in the Human Midbrain and Striatum Facilitates Behavioral Adaptation and Learning Efficiency. *Neuron* 90: 1127–1138.
69. Tobler PN, Fiorillo CD, Schultz W (2005): Adaptive coding of reward value by dopamine neurons. *Science (80-)* 307: 1642–1645.
70. Fiorillo CD, Tobler PN, Schultz W (2003): Discrete Coding of Reward Dopamine Neurons. *Science (80-)* 299: 1898–1902.
71. Ledoux J, Daw ND (2018): Surviving threats: Neural circuit and computational implications of a new taxonomy of defensive behaviour. *Nat Rev Neurosci* 19: 269–282.
72. Browning M, Behrens TEJ, Jochem G, O'Reilly JX, Bishop SJ (2015): Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nat Neurosci* 18: 590–596.
73. McLean BF, Mattiske JK, Balzan RP (2017): Association of the jumping to conclusions and evidence integration biases with delusions in psychosis: A detailed meta-analysis. *Schizophr Bull* 43: 344–354.
74. Dudley R, Taylor P, Wickham S, Hutton P (2016): Psychosis, delusions and the “Jumping to Conclusions” reasoning bias: A systematic review and meta-analysis. *Schizophr Bull* 42: 652–665.
75. Funder DC, Ozer DJ (2019): Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Adv Methods Pract Psychol Sci* 2: 156–168.
76. Zech H, Waltmann M, Lee Y, Reichert M, Bedder RL, Rutledge RB, *et al.* (2022): Measuring self-regulation in everyday life: Reliability and validity of smartphone-based experiments in alcohol use disorder. *Behav Res Methods*. <https://doi.org/10.3758/s13428-022-02019-8>

Journal Pre-proof

Tables

Table 1: Demographic data

	N	Gender	Age Mean (SD)	Race	Parental Education Mean (SD)	Data collection site
Experiment 1	45	68,9% F (N = 31), 31,1% M (N = 14)	25,31 (4,14)	NA	NA	In person, Vienna, Austria
Experiment 2	437	44,9 % F (N = 196), 55,1% M (N = 241)	37,54 (10,39)	NA	NA	Online, USA
Experiment 3	100	33,3% F (N = 33) 66,6% M (N = 66)	37,08 (10,05)	49 C, 40 AA, 4 AS, 6 M/O	14,56 (2,93)*	In person, Baltimore, USA

AA = African-American; AS = Asian; C = Caucasian, F = female; M = male; M/O = mixed/other; NA = data not available

* Data for parental education is missing for 5 participants in Experiment 3.

Figure Legends

Figure 1. Computational model. **A**, Observations are generated from an unobservable dynamically changing Gaussian variable with varying variance. Observers make predictions based on their beliefs and use the prediction errors to update their beliefs about the underlying states. Belief updates are regulated by the precision weight ψ_1 : ratio of the belief about outcome precision (π_u) and the belief about the expected mean precision (π_x), which is in turn governed by both prior belief precision or uncertainty σ_x and belief about volatility μ_v . **B**, The generative model is defined as two parallel cascades of Gaussian random walks, where the first governs the evolution of the underlying mean (x), and its (log) volatility (x_v), and the second governs the evolution of the outcome variance (x_α), also on a log scale. **C**, Model inversion through the hierarchical Gaussian filter provides both the mean and the precision of beliefs about the mean, variance, and volatility of the latent variable.

Figure 2. Simulated intraclass correlations. **A**, ICC3k (absolute raters) scores in the simulation of behavior in the predictive inference task with three levels of decision noise η (1, 16, or 48). **B**, Binary reversal one-armed bandit task, where the trajectories (correct/incorrect) are generated with the same probability of reversal as in (A) and the probability of correct in each block being either 0.2, 0.8, or 0.5. The decision noise is parametrized through the softmax inverse temperature set as 48, 8, or 1 (from low to high noise). Note, that in tasks with binary outcome variables, the probability of outcome (p) determines the outcome variance ($p(1-p)$) therefore making it indistinguishable from prior beliefs.

Figure 3. Model performance in Experiment 1. **A – C**, Learning rate on a log scale (A), beliefs about variance (B) and volatility (C) across the variance blocks. Parameter estimation was performed

hierarchically. All dependent variables were scaled within participant, then averaged within participant and variance block and then across participants. Bar plots and error bars represent means and standard errors across participants (with $N = 45$). **D**, Empirical ICC scores derived from the hierarchical model (means and 95% CrI). **E**, Single and average fixed rater ICCs (means and 95% CI) of model derived trial-by-trial precision-weighted learning rates (ψ_1), beliefs about outcome variance (μ_α) and volatility (μ_v) meaned within participant. **F**, ICCs (means and 95% CI) plotted for the participant-level random slope of prediction error effect on confidence on the next trial (B_C), signed prediction error on signed update (B_U), and the effects of variance on learning rate (B_L). **G – I**, Relationship of random slopes B_U , B_L , and B_C on model parameters. **J**, Correlation between the performance in the Two-step task and the environmental volatility parameter ϑ in session 1 ($N = 40$).

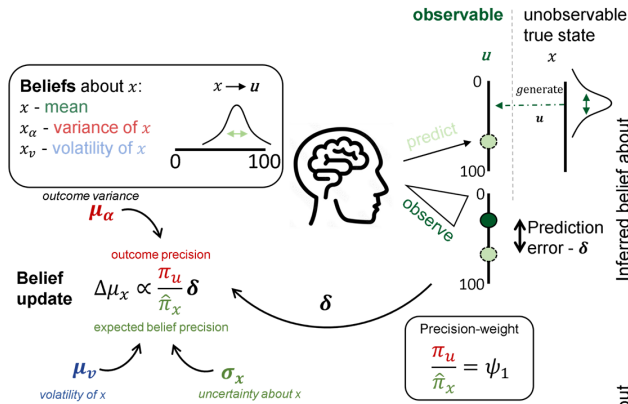
Figure 4. Model parameters and schizotypal traits in Experiment 1 (A-C) and 2 (D-F). **A**, Posterior distributions of effect sizes of the effect of each model parameter (averaged per participant) regressed against PDI, plotted over means and CrIs. **B**, Correlation plot of ϑ and PDI. PDI was log transformed and scaled. **C**, Plotting confidence ratings after change points in Experiment 1. Points and error bars represent means and standard errors (across participants). Line and shade represent the posterior predictive mean and 95% predictive interval of a linear regression model predicting confidence from prediction errors. **D**, Plotting posterior distributions of effects of each model parameter, over means and CrIs of effect sizes when positive schizotypy (UE subscale) was regressed against the three parameters controlling for age, gender and IQ. **E**, Same as (**D**) but for negative schizotypy (IA subscale). **F**, Plotting confidence ratings after change points for the median split of the UE. Points and error bars represent means and standard errors (across participants). Line and shade represent the posterior predictive mean and 95% Predictive interval of a linear regression model predicting confidence from prediction errors. PDI - Peters et al. Delusional Inventory. UE - Unusual Experience subscale of Short Scales for Measuring Schizotypy, IA - Introvertive Anhedonia subscale of Short Scales for Measuring Schizotypy.

Figure 5. Modelling results from a sample of patients with chronic schizophrenia in Experiment 3. **A-B**, A total of six parameters (three parameters of the belief model for both reward seeking and loss avoiding condition) were used as predictor variables in two models predicting Positive symptoms measured by the BPRS (**A**) and negative symptoms measured by the SANS (**B**). Plotting posterior distributions of effects of each model parameter, over means and CrIs of effect sizes across two incentivization condition. **C-D**, Correlation plot for ϑ in the seek condition and positive symptoms (**C**) and negative symptoms (**D**). **E**, Results of a penalized CCA. We used CCA to find the maximal correlation between a linear combination of computational parameters from the belief model with the linear combination of the two symptoms domains. Bootstrapping with 5000

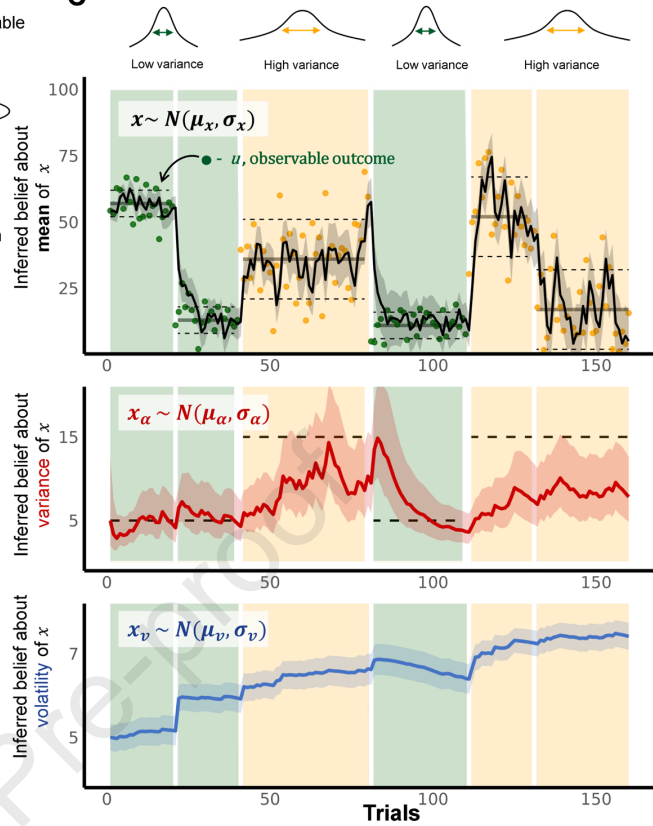
permutations was used to obtain p-values. BPRS - Brief Psychiatric Rating Scale, SANS - Scale for the Assessment of Negative Symptoms, CCA – Canonical correlation analysis

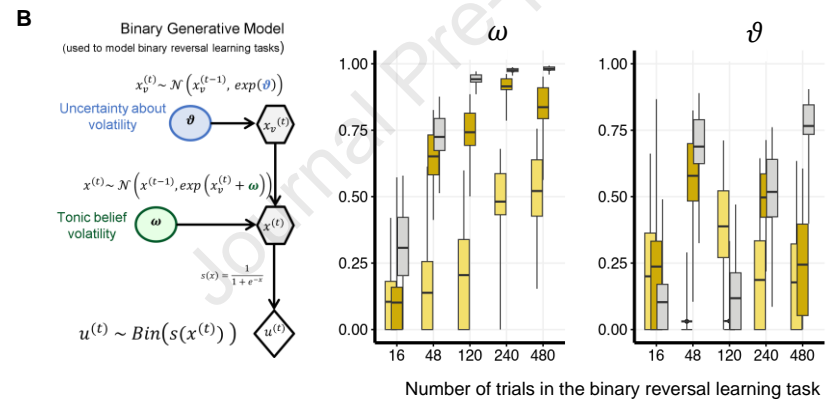
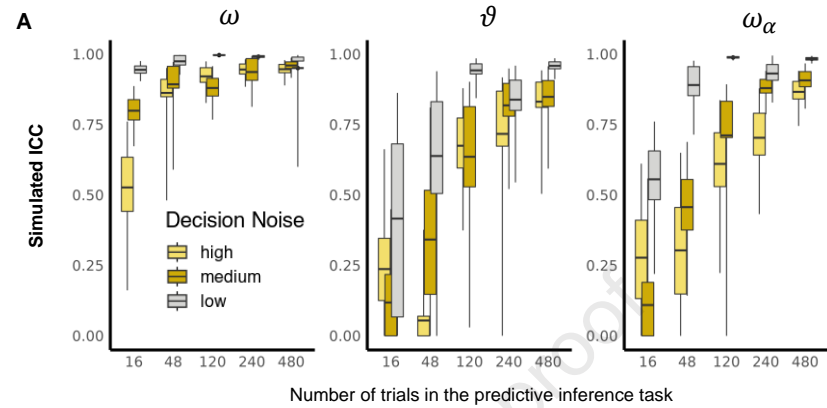
Journal Pre-proof

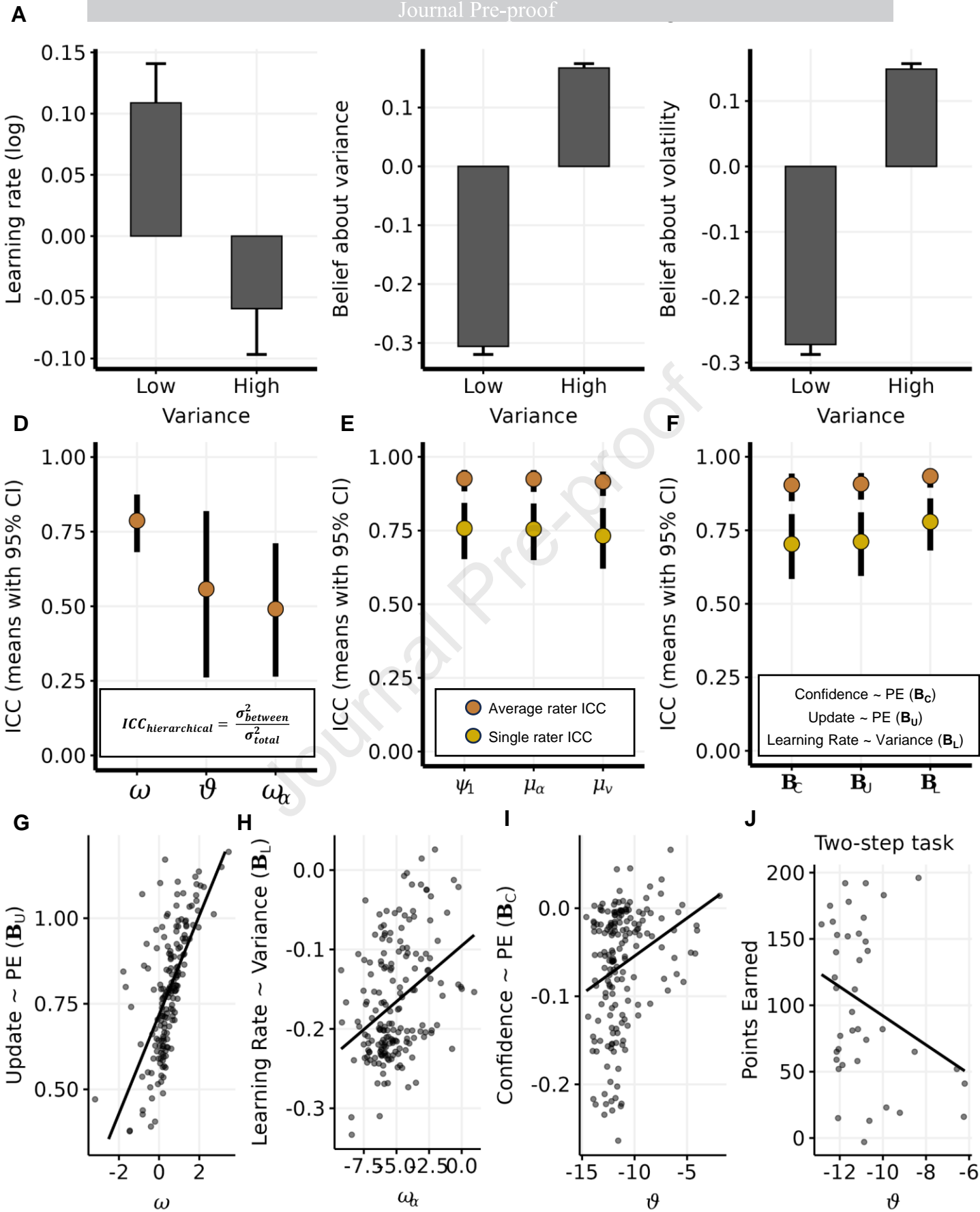
A



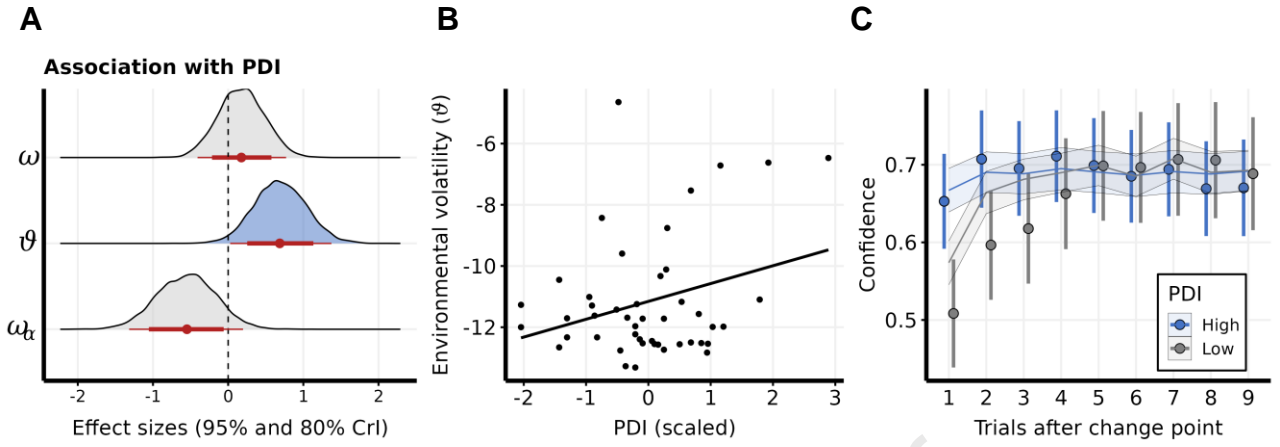
C



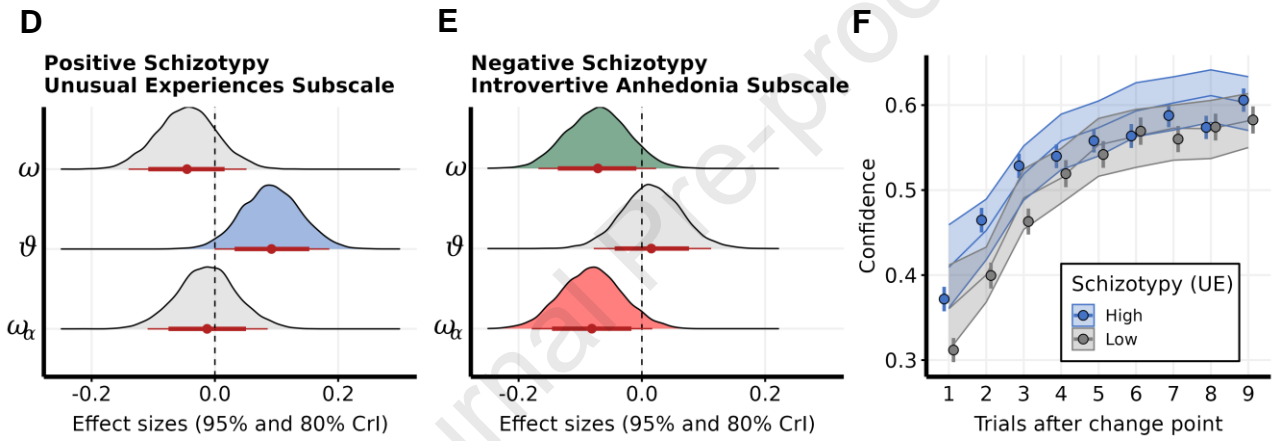




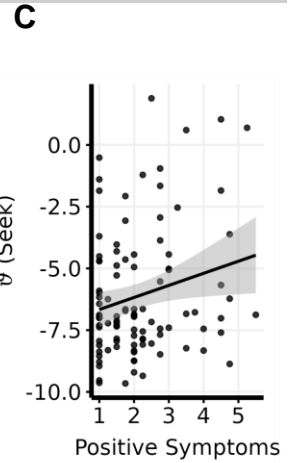
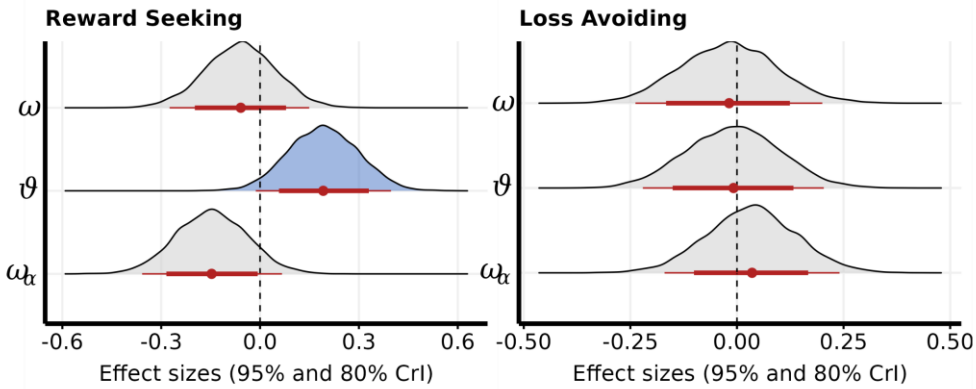
Experiment 1 (Test-retest study, N = 45)



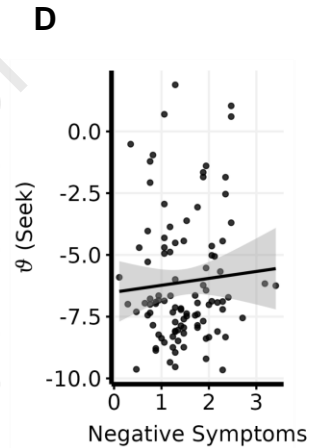
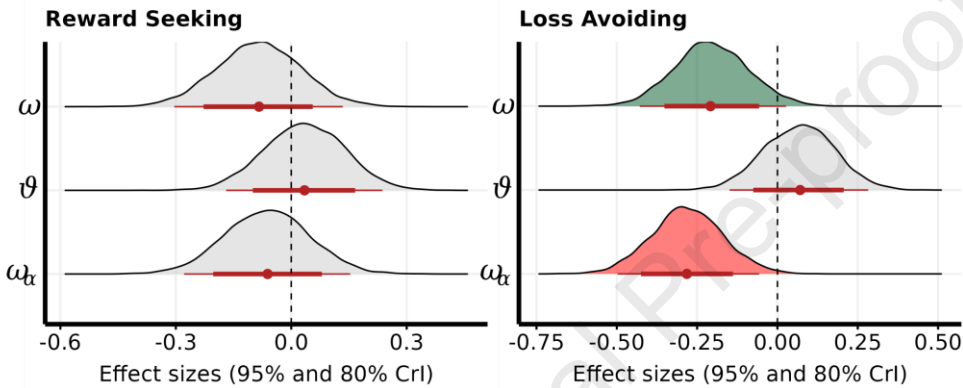
Experiment 2 (Online study, N = 437)



A Association with Positive symptoms (GLM 1)



B Association with Negative symptoms (GLM 2)



E Penalized Canonical Correlation Analysis

Correlation of Original Variables with Canonical Variables

