Schizophrenia Bulletin vol. 48 no. 4 pp. 826–838, 2022
https://doi.org/10.1093/schbul/sbac029
Advance Access publication May 26, 2022

Increased Belief Instability in Psychotic Disorders Predicts Treatment Response to Metacognitive Training

D. J. Hauke*[,1,2,3], V. Roth[2], P. Karvelis[3], R. A. Adams[4,5], S. Moritz[6], S. Borgwardt[7,8], A. O. Diaconescu[3,9,10], and C. Andreou[7,8,10]

[1]Department of Psychiatry (UPK), University of Basel, Basel, Switzerland; [2]Department of Mathematics and Computer Science, University of Basel, Basel, Switzerland; [3]Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health (CAMH), Toronto, Canada; [4]Centre for Medical Image Computing, Department of Computer Science, University College London, London, UK; [5]Max Planck Centre for Computational Psychiatry and Ageing Research, University College London, London, United Kingdom; [6]Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf (UKE), Hamburg, Germany; [7]Department of Psychiatry and Psychotherapy, Translational Psychiatry Unit, University of Lübeck, Lübeck, Germany; [8]Center of Brain, Behaviour and Metabolism, University of Lübeck, Lübeck, Germany; [9]Department of Psychiatry, University of Toronto, Toronto, Canada; [10]These authors contributed equally to this work.

*To whom correspondence should be addressed; 250 College St., 12th Floor, Toronto, ON M5T 1R8, Canada; tel: +1 (416) 535-8501 ext. 30585, fax: +1 416-583-1207, e-mail: daniel.hauke@unibas.ch

Background and Hypothesis: **In a complex world, gathering information and adjusting our beliefs about the world is of paramount importance. The literature suggests that patients with psychotic disorders display a tendency to draw early conclusions based on limited evidence, referred to as the *jumping-to-conclusions bias*, but few studies have examined the computational mechanisms underlying this and related belief-updating biases. Here, we employ a computational approach to understand the relationship between jumping-to-conclusions, psychotic disorders, and delusions.** ***Study Design:*** **We modeled probabilistic reasoning of 261 patients with psychotic disorders and 56 healthy controls during an information sampling task—the fish task—with the Hierarchical Gaussian Filter. Subsequently, we examined the clinical utility of this computational approach by testing whether computational parameters, obtained from fitting the model to each individual's behavior, could predict treatment response to Metacognitive Training using machine learning.** ***Study Results:*** **We observed differences in probabilistic reasoning between patients with psychotic disorders and healthy controls, participants with and without jumping-to-conclusions bias, but not between patients with low and high current delusions. The computational analysis suggested that *belief instability* was increased in patients with psychotic disorders. Jumping-to-conclusions was associated with both increased belief instability and greater prior uncertainty. Lastly, belief instability predicted treatment response to Metacognitive Training at the individual level.** ***Conclusions:*** **Our results point towards increased belief instability as a key computational mechanism underlying probabilistic reasoning in psychotic disorders. We provide a proof-of-concept that this computational approach may be useful to help identify suitable treatments for individual patients with psychotic disorders.**

Key words: psychosis/jumping-to-conclusions/belief updating/Hierarchical Gaussian Filter/treatment response prediction/probabilistic reasoning

Introduction

Delusions have been defined as unfounded beliefs that are held with absolute conviction, despite strong evidence to the contrary and are resistant to change.[1] They occur in various forms: Prominently featuring among others are persecutory delusions—the belief that others deliberately intend to cause harm[2]—and grandiose delusions, believing that one has superior power, knowledge, or a special identity.[3] While delusions are key symptoms of schizophrenia, they also occur in other disorders with psychotic symptoms, such as delusional disorder, and psycho-affective disorders, including bipolar disorder.[4] It is therefore important to assume a transdiagnostic perspective and understand the mechanisms underlying delusion formation and persistence across psychotic disorders.

A substantial body of work has examined the relationship between reasoning biases and delusions.[5–7] Patients with psychotic disorders change their beliefs more than healthy controls (HC) when faced with evidence that

contradicts their current beliefs (ie, *disconfirmatory evidence*).[8–11] Another extensively studied bias is the *jumping-to-conclusions bias* (JTC), the tendency to draw conclusions based on limited evidence. JTC was found to be more prevalent in patients with psychotic disorders, especially in those with delusions.[6,7] Traditionally, JTC was assessed with the beads task, a probabilistic learning task, in which participants are asked to decide from which of two urns an experimenter is drawing a sequence of colored beads.[12,13] In another version of the task, the *fish task*,[14–16] participants are shown fish that a fisherman caught from one of two lakes with different ratios of colored fish and are asked to determine from which lake the fisherman was fishing.

While relationships between reasoning biases such as JTC, delusions, and psychotic disorders have been found across different tasks,[5–7] as of yet, it is unclear, whether JTC is contributing to delusion formation by increasing premature acceptance of implausible ideas,[6,17] or, whether it is merely an epiphenomenon of psychotic disorders[18] (see Ref.[7] for other biases). A third possibility is that JTC—and increased updating to disconfirmatory evidence—both reflect a noisy and unstable cognitive system that is more vulnerable to affective or habitual biases, thus enabling delusions without directly causing them.[19] Although answering this question may ultimately require longitudinal data, computational modeling allows us to study the computational mechanisms underlying behavioral differences across individuals. Computational models describe how changes in information processing give rise to observable differences in behavior. This approach is useful because there is often a many-to-many mapping between computational parameters and behavioral effects. For example, JTC could be caused by greater initial uncertainty, faster belief-updating, or noisier responding. Modeling allows the investigator to distinguish between these possibilities in each individual. This is potentially important, because specific computational mechanisms may relate to specific treatment effects (eg, blocking dopamine D2 receptors might reduce noisy responding, but not affect belief-updating).

Here, we employed a computational modeling approach to understand the relationship between JTC, psychotic disorders, and delusions. The research objective of this study was to dissect this relationship based on the computational mechanisms underlying belief updating in the fish task. To this end, we formulated three research questions (RQ):

RQ1: *What are the computational mechanisms underlying differences in probabilistic reasoning between HC and patients with psychotic disorders?*

RQ2: *What are the computational mechanisms underlying differences in probabilistic reasoning between individuals with and without JTC?*

RQ3: *What are the computational mechanisms underlying differences in probabilistic reasoning between patients with low and high current delusions?*

While computational analyses may provide relevant theoretical insight, the ultimate goal of understanding computational mechanisms is to improve patients' well-being. To examine the clinical utility of this approach, we investigated whether computational parameters predicted treatment response to Metacognitive Training,[20] an intervention that specifically targets reasoning biases. Based on previous results,[21] we expected that belief instability and decision noise would predict treatment response. This hypothesis was also based on the observation that several modules of Metacognitive Training are designed to make cognition more robust (supplementary material). In which case, those with the greatest belief instability or decision noise may stand to benefit most from a cognition-focused intervention.

Methods

Participants

Our sample consisted of $N = 333$ participants of three different studies.[22–24] All studies were approved by the local ethics committee and conducted in accordance with the most recent version of the Declaration of Helsinki. Participants provided written informed consent and were reimbursed for clinical assessments. We excluded 13 participants for which raters had indicated miscomprehension of task instructions and three participants due to incomplete probability ratings. The final sample ($N = 317$) consisted of 56 HC and 261 patients with psychotic disorders who met the criteria for a schizophrenia spectrum disorder diagnosis and experienced delusions currently or in the past (see supplementary material and Refs.[22–24]).

Metacognitive Training Intervention

In treatment trials,[22,24] patients were randomly allocated to treatment (Metacognitive Training, $n = 106$) or control intervention (computerized remediation intervention; CogPack; http://www.markersoftware.com/) using a fixed, pseudo-randomization schedule. Additionally, all participants continued treatment as usual. Clinical assessments were conducted by raters that were blind to the treatment allocation. Analyses were restricted to the baseline and post-intervention assessment (after 4[24]or 6[22] weeks).

Task

To assess probabilistic reasoning at baseline, we employed a graded estimates version[8] of the fish task.[14–16] Participants were instructed that a fisherman was fishing from one—and only one—of two lakes with different ratios of colored fish (80:20 in lake A and reversed in lake B; figure 1A). They were also instructed that these

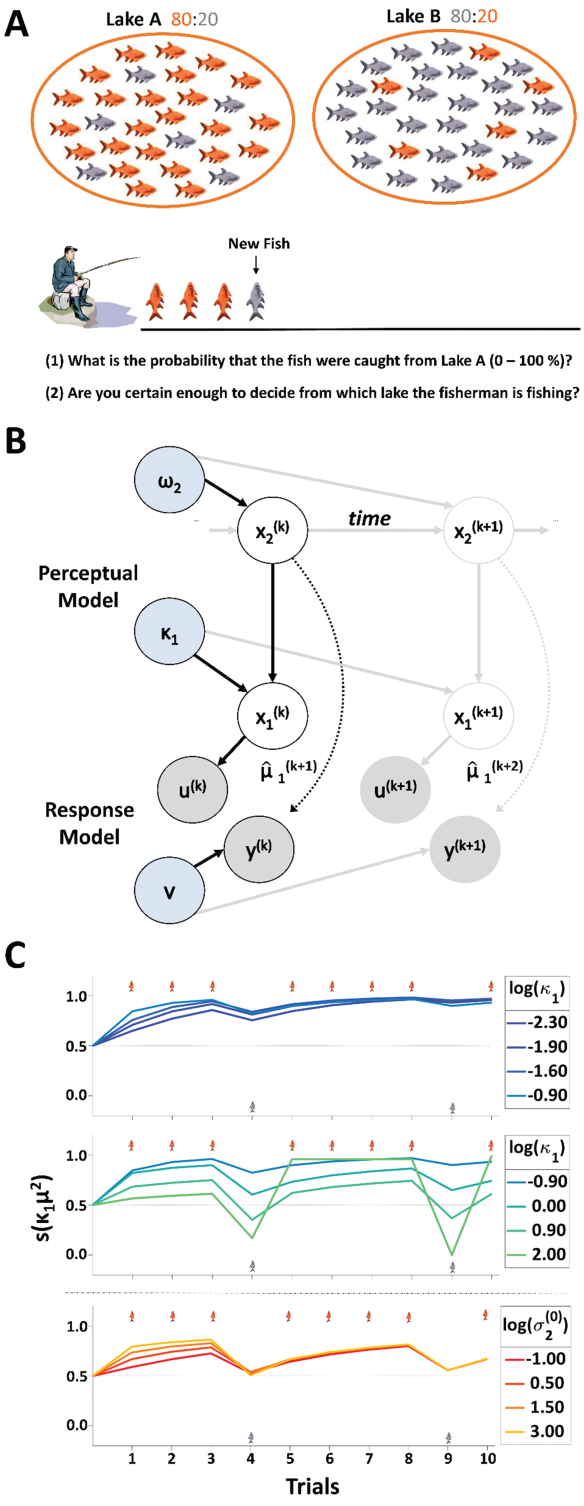

Fig. 1. Task and winning model. (A) Fish task. (B) Winning model. Graphical representation of the generative model adapted from Adams et al.[21] Observed quantities are denoted with gray circles. White circles represent hidden states and blue circles subject-specific parameters. Black lines indicate probabilistic network at trial k and gray lines at trial $k + 1$. Solid lines indicate generative model in the world, which participants infer on,[25,26] whereas dotted lines represent participants' inference on these states. (C) Simulation showing the impact of changing belief instability κ_1 and prior uncertainty $\sigma_2^{(0)}$. Displayed is the inferred probability that the fisherman is fishing from lake A $s\left(\kappa_1 \mu_2^{(k)}\right)$ for very low (upper panel) or low to high levels of belief instability (middle panel), and changing prior uncertainty (lower panel). All other parameters were fixed to the posterior medians. Increasing $log(\kappa_1)$ above approximately −0.9 leads to higher belief instability, as participants are changing their beliefs more rapidly when faced with disconfirmatory evidence. Increasing κ_1 in the very low range leads to larger belief updates early in the experiment. Note, however, that the exact value of κ_1 at which the model's behaviour undergoes this qualitative change depends on the other parameter values. Increasing $\sigma_2^{(0)}$ consistently leads to larger belief updates early in the experiment.

ratios did not change as the fisherman always threw the fish back into the water (sampling with replacement). Participants were presented with a sequence of ten fish. After each fish in the sequence they were asked (1) to estimate the probability that the fish were drawn from lake A (0%–100%) and (2) if they were certain enough to decide from which lake the fisherman was fishing and if so, what their conclusion was (ie, lake A or B).

Computational Modeling

Hierarchical Gaussian Filter. We modeled behavior with the Hierarchical Gaussian Filter (HGF).[27,28] This model was employed previously to understand probabilistic reasoning during the beads task in schizophrenia patients[21] and other symptoms of schizophrenia (eg, hallucinations[29] or paranoid delusions[30–32]). The HGF assumes that learning is driven by precision-weighted prediction error updates and that learners integrate prior and new information in a Bayes-optimal manner given their individual learning parameters, which are estimated from participants' behavior. These parameters can be understood as encoding an individual's approximation to Bayesian inference[28] and provide a concise summary of individual learning profiles. Differences in model parameters or architectures across participants can then be leveraged to understand the computational mechanisms underlying different populations. We closely followed the approach of Adams and colleagues,[21] briefly summarize below.

As Adams et al[21] we employed a 2-level, nonvolatile HGF (figure 1B; supplementary material) because participants experienced a stable environment (the fisherman was always fishing from the same lake) and the third level of the HGF captures learning about environmental volatility. This also helped reduce the number of free model parameters, which was important given the small number of available trials per participant.

The generative model assumes that a fish $u^{(k)}$ is drawn from the probability $x_1^{(k)}$ that the fisherman is fishing in one of the lakes (eg, lake A) on trial k (figure 1B). The state at the level above is the unbounded tendency $(-\infty, +\infty)$ of the fisherman fishing from lake A $x_2{}^{(k)}$, which is transformed to the probability $x_1{}^{(k)}$ using the

sigmoid transformation $s(x_2^{(k)})$, where $s(x_2^{(k)})$ is defined as: $s(z) = 1/(1 + e^{-z})$.

However, $x_2^{(k)}$ is not known to the participant and needs to be inferred when observing a sequence of fish. The participant's posterior estimate of $x_2^{(k)}$ on trial k is denoted $\mu_2^{(k)}$. Again, this unbounded tendency can be transformed into the participant's posterior estimate of the probability of the fish being fished from that lake with the range $[0, 1]$ using a sigmoid transformation as before $\hat{\mu}_1^{(k)} = s\left(\kappa_1 \mu_2^{(k)}\right)$, which is equivalent to the participant's prediction for the next trial (denoted with $\hat{}$). However, here, κ_1 represents a subject-specific parameter capturing the degree of *belief instability*.

Bayesian Model Selection. We formulated two competing hypotheses describing different learning mechanisms: Model 1: *Standard Bayesian belief updating* and Model 2: *Bayesian belief updating subject to belief instability* (controlled by κ_1). Note, that $\hat{\mu}_1^{(k)} = s\left(\kappa_1 \mu_2^{(k)}\right)$ reduces to a simple sigmoid transformation of $\mu_2^{(k)}$, if $\kappa_1 = 1$. In this case, Model 2 is reduced to Model 1. However, if $\kappa_1 > 1$, simulated participants will show increased belief instability leading them to quickly change their mind when confronted with disconfirmatory evidence (figure 1C), but also resulting in smaller updates when presented with consistent evidence (eg, fishes 5–8). Only Model 2 estimates κ_1 from participants' behavior (supplementary material) and thus tests the hypothesis that participants' learning can be better described by Bayesian belief updating subject to *belief instability*. In both models, we additionally estimated (1) $\sigma_2^{(0)}$, expressing the *prior uncertainty* at the beginning of the experiment, (2) ω_2, the constant component of the learning rate or the *evolution rate*, and (3) ν, capturing response stochasticity or *decision noise* (lower values indicate higher noise). All other parameters were fixed.

To arbitrate between hypotheses, we compared models with random-effects Bayesian model selection[33,34] and report protected exceedance probabilities φ and relative model frequencies f. Model and parameter recovery was assessed through simulations (supplementary material).

Statistical Analyses

We tested the three research questions with linear mixed-effects models including probability estimates as dependent variable and education, medication, sex, and study as covariates. JTC was defined as reaching a decision after ≤ 2 fish[35] and low and high current delusions were defined based on a median split of the Psychotic Symptom Rating Scales (PSYRATS)[36]: Delusion subscale. Additionally, we assessed trial-by-trial behavior with nonparametric Kruskal–Wallis tests Bonferroni-corrected for the number of trials ($n = 10$). Model parameters were compared using Kruskal–Wallis tests Bonferroni-corrected for the number of parameters ($n = 4$). We repeated analyses on an IQ-matched subsample of patients to exclude IQ as a confounding factor (supplementary material).

Treatment Response Prediction

Based on recent meta-analyses,[18,37] treatment response to Metacognitive Training was defined as a reduction in the Positive and Negative Syndrome Scale (PANSS)[38] positive symptom factor[39] by at least 20% at 4[24] or 6[22] week follow-up using percentage change scores.[40] Random forest classifiers[41] were trained to predict treatment response from either (1) the model-derived computational fingerprint of participants, ie, the four model parameters (κ_1, $\sigma_2^{(0)}$, ω_2, and ν), (2) participants' raw behavioral data (probability estimates and decisions) and a binary JTC indicator, or (3) clinical baseline data (PANSS items). Preprocessing included covariate correction[42] for sex, medication, and education and was embedded in k-fold cross-validation (supplementary material).

Results

Clinical and demographic characteristics are reported in table 1. Since, there was no conclusive evidence for increased JTC in patients with psychotic disorders ($\chi^2 = 3.435$, $p_{uncorr} = 0.064$), we analyzed HC and patients together in all subsequent analyses investigating JTC (for JTC analysis only in patients see supplementary material).

Behavioral Results

RQ1: Group Differences between HC and Patients with Psychotic Disorders. We found a significant group-by-trial interaction when comparing HC and patients with psychotic disorders ($F = 4.420$, $P < .001$; figure 2A), which held in IQ-matched subsamples (supplementary material). This effect was driven by a stronger decrease in probability for the more likely lake A in patients with psychotic disorders in trial 9 with one of the two rare (ie, disconfirmatory) fish ($\eta^2 = 0.028$, $P = .031$). We also observed trend-effects in trial 4 and 8, which did not survive Bonferroni correction, however ($\eta^2 = 0.015$, $p_{uncorr} = 0.028$, $P = .281$, and $\eta^2 = 0.020$, $p_{uncorr} = 0.012$, $P = .123$, respectively). None of the covariates was significant.

RQ2: Group differences Between Individuals with and without JTC. Comparing individuals with and without JTC, we found a significant JTC-by-trial interaction ($F = 11.598$, $P < .001$; figure 2B), which held when comparing IQ-matched subsamples (supplementary material). This effect was driven by increased probability estimates for lake A in individuals with JTC within the

Table 1. Sociodemographic and clinical characteristics

	Research Question 1 Psychosis			Research Question 2 Jumping-to-conclusions			Research Question 3 Delusions			Treatment Response to Metacognitive Training		
	HC $n = 56$	PSY $n = 261$	Statistic	JTC− $n = 174$	JTC+ $n = 143$	Statistic	D− $n = 131$	D+ $n = 126$	Statistic	R− $n = 63$	R+ $n = 43$	Statistic
Age_n	29_{56}	33_{261}	$\eta^2 = 0.003$	32_{174}	34_{143}	$\eta^2 = 0.001$	33_{131}	33_{126}	$\eta^2 = 0.002$	35_{63}	36_{43}	$\eta^2 = 0.001$
median [25th, 75th]	[27, 38]	[26, 45]	$P = .306$	[27, 43]	[26, 44]	$P = .500$	[26, 43]	[27, 45]	$P = .427$	[30, 45]	[30, 43]	$P = .750$
Sex	26/30	104/157	$\chi^2 = 0.826$	75/99	55/88	$\chi^2 = 0.699$	49/82	52/74	$\chi^2 = 0.402$	31/32	19/24	$\chi^2 = 0.258$
f/m			$P = .364$			$P = .403$			$P = .526$			$P = .611$
$Education_n$	13_{56}	12_{258}	$\eta^2 = 0.005$	13_{174}	12_{140}	$\eta^2 = 0.015$	13_{129}	12_{125}	$\eta^2 = 0.010$	13_{62}	12_{43}	$\eta^2 = 0.013$
median [25th, 75th]	[10, 13]	[10, 13]	$P = .200$	[10, 13]	[10, 13]	$\mathbf{P = .047}$	[10, 13]	[10, 13]	$P = .120$	[10, 13]	[10, 13]	$P = .252$
IQ_n	107_{56}	104_{181}	$\eta^2 = 0.033$	107_{133}	101_{104}	$\eta^2 = 0.022$	104_{83}	101_{96}	$\eta^2 = 0.001$	107_{45}	103_{27}	$\eta^2 = 0.016$
median [25th, 75th]	[101, 118]	[94, 112]	$\mathbf{P = .005}$	[100, 118]	[94, 112]	$\mathbf{P = .022}$	[95, 114]	[93, 112]	$P = .628$	[101, 118]	[96, 112]	$P = .280$
Medication	56/0	24/219	$\chi^2 = 188.630$	54/110	16/109	$\chi^2 = 7.058$	6/118	17/98	$\chi^2 = 6.783$	5/54	1/41	$\chi^2 = 1.630$
n/y			$\mathbf{P < .001}$			$\mathbf{P = .008}$			$\mathbf{P = .009}$			$P = .202$
$Draws\text{-}to\text{-}decision_n$	3_{56}	3_{181}	$\eta^2 = 0.007$	5_{174}	1_{143}	$\eta^2 = 0.772$	2_{131}	3_{126}	$\eta^2 = 0.010$	3_{63}	2_{43}	$\eta^2 = 0.013$
median [25th, 75th]	[2, 5]	[1, 5]	$P = .138$	[3, 6]	[1, 2]	$\mathbf{P < .001}$	[1, 5]	[2, 5]	$P = .109$	[2, 5]	[1, 4]	$P = .249$
$PANSS\ P_n$		14_{260}		14_{136}	14_{124}	$\eta^2 < 0.001$	10_{131}	19_{125}	$\eta^2 = 0.507$	11_{63}	15_{43}	$\eta^2 = 0.104$
median [25th, 75th]		[9, 19]		[10, 19]	[9, 19]	$P = .802$	[7, 13]	[15, 22]	$\mathbf{P < .001}$	[8, 16]	[13, 19]	$\mathbf{P < .001}$
$PANSS\ N_n$		12_{261}		11_{137}	13_{124}	$\eta^2 = 0.010$	10_{131}	13_{126}	$\eta^2 = 0.053$	11_{63}	14_{43}	$\eta^2 = 0.061$
median [25th, 75th]		[9, 16]		[9, 15]	[9, 17]	$P = .111$	[8, 16]	[10, 18]	$\mathbf{P < .001}$	[8, 17]	[11, 19]	$\mathbf{P = .012}$
$PANSS\ E_n$		11_{261}		10_{137}	11_{124}	$\eta^2 = 0.013$	10_{131}	12_{126}	$\eta^2 = 0.092$	10_{63}	12_{43}	$\eta^2 = 0.044$
median [25th, 75th]		[9, 13]		[9, 13]	[9, 15]	$P = .067$	[8, 12]	[10, 15]	$\mathbf{P < .001}$	[9, 11]	[10, 15]	$\mathbf{P = .031}$
$PANSS\ Ds_n$		16_{261}		15_{137}	16_{124}	$\eta^2 = 0.005$	12_{131}	18_{126}	$\eta^2 = 0.227$	14_{63}	18_{43}	$\eta^2 = 0.113$
median [25th, 75th]		[12, 20]		[12, 19]	[12, 21]	$P = .250$	[10, 16]	[15, 22]	$\mathbf{P < .001}$	[11, 17]	[13, 23]	$\mathbf{P < .001}$
$PANSS\ Do_n$		14_{261}		14_{137}	14_{124}	$\eta^2 = 0.009$	12_{131}	16_{126}	$\eta^2 = 0.200$	13_{63}	14_{43}	$\eta^2 = 0.015$
median [25th, 75th]		[12, 18]		[11, 17]	[12, 19]	$P = .135$	[11, 15]	[14, 20]	$\mathbf{P < .001}$	[11, 17]	[12, 18]	$P = .212$
$PSYRATS\ D_n$		8_{257}		9_{136}	5_{121}	$\eta^2 = 0.005$	0_{131}	14_{126}	$\eta^2 = 0.787$	2_{61}	0_{43}	$\eta^2 = 0.019$
median [25th, 75th]		[0, 14]		[0, 14]	[0, 14]	$P = .270$	[0, 1]	[11, 17]	$\mathbf{P < .001}$	[0, 11]	[0, 12]	$P = .164$
$PSYRATS\ H_n$		0_{257}		0_{136}	0_{121}	$\eta^2 < 0.001$	0_{128}	0_{125}	$\eta^2 = 0.799$	0_{61}	0_{43}	$\eta^2 = 0.039$
median [25th, 75th]		[0, 4]		[0, 6]	[0, 4]	$P = .739$	[0, 0]	[0, 21]	$\mathbf{P < .001}$	[0, 0]	[0, 12]	$\mathbf{P = .045}$

Note: Reported are uncorrected *P*-values and test statistics of either χ^2-tests for categorical variables or Kruskal–Wallis tests for all other variables for healthy controls (HC) or patients with psychotic disorders (PSY), participants without (JTC−) or with (JTC+) jumping-to-conclusion bias (decision after ≤ 2 fish), patients with low (D−) or high (D+) current delusions (split half based on median of Psychotic Symptom Rating Scales (PSYRATS)[36]: Delusion subscale), and patients without (R−) or with (R+) treatment response to Metacognitive Training[17] defined as at least 20% decrease in the Positive and Negative Syndrome Scale (PANSS)[38] positive factor[39] compared to baseline. P: Positive symptoms. N: Negative symptoms. G: General symptoms. E: Excitement. Ds: Distress. Do: Disorganization. D: Delusions. H: Hallucinations. Bold values indicate *p*-values significant at $p<0.05$.

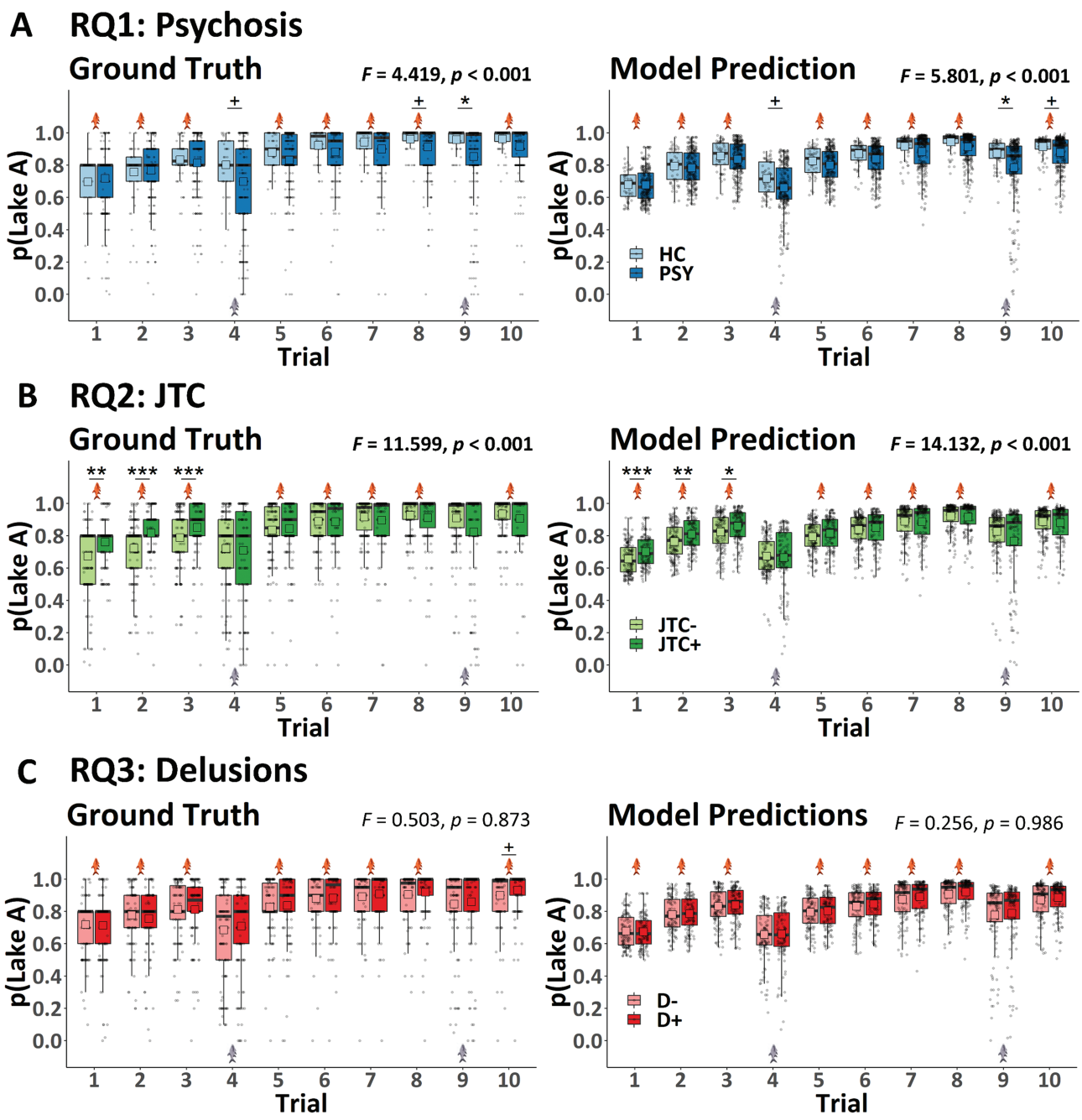

Fig. 2. Behavioral effects versus model predictions. (A) Behavioral effects for Research Question 1 (RQ1): Comparing behavior in the fish-task between healthy controls (HC) and patients with psychotic disorder (PSY). (B) Behavioral effects for Research Question 2 (RQ2): Comparing behavior between individuals without (JTC−) and with (JTC+) jumping-to-conclusion bias (decision after ≤ 2 fish). C Behavioral effects for Research Question 3 (RQ3): Comparing behavior between patients with low (D−) and high (D+) current delusions (split half based on median of Psychotic Symptom Rating Scales (PSYRATS)[36]: Delusion subscale). *F*- and *P*-values indicate results of ANCOVAs corrected for education, medication, sex, and study. Y-axis: Participants' estimates of the probability that the fisherman was fishing from lake A (see question (1) in Figure 1A). Left panels: Behavioral effects. Right panels: Model prediction of the winning model. Horizontal lines and squares in boxplots represent median and mean, respectively. Boxes span the 25th to 75th quartiles and whiskers extend from hinges to the largest and smallest value that lies within 1.5 × interquartile range. Asterisks indicate significance of nonparametric Kruskal–Wallis tests at: ***$P < 0.001$, **$P < 0.01$, and *$P < 0.05$, using Bonferroni correction, or at $^{+}P < .05$ uncorrected. Note, that Bonferroni correction is likely to be too conservative as responses were correlated across trials.

first three trials of the fish sequence (trial 1: $\eta^2 = 0.043$, $P = .002$, trial 2: $\eta^2 = 0.077$, $P < .001$, trial 3: $\eta^2 = 0.056$, $P < .001$). Furthermore, we observed a significant main effect of medication as medicated individuals estimated lower probabilities overall ($F = 7.138$, P=.008), but none of the other covariates.

RQ3: Group Differences between Patients with Low and High Current Delusions. There was no evidence for differences in probabilistic reasoning between patients with low or high current delusions (delusion-by-trial interaction: $F = 0.503$, $P = .873$; figure 2C). We observed a trend-effect of delusions in trial 10 that did not survive Bonferroni correction ($\eta^2 = 0.017$, $p_{uncorr} = 0.019$, $P = .194$). Among the covariates, we only found a significant main effect of education suggesting that longer education was associated with higher probability estimates overall ($F = 4.016$, $P = .046$).

Modeling Results

Bayesian Model Selection and Model Recovery. Model selection strongly suggested that Model 2: *Bayesian belief updating subject to belief instability* was the most likely mechanism explaining behavior across all groups ($\varphi = 100.00\%$, $f = 98.94\%$; figure 3; see supplementary material for original model space by Adams et al[21]). The model recovery analysis indicated that these models were distinguishable.

Posterior Predictive Checks and Parameter Recovery. To confirm that the winning model captured the behavioral effects of interest, we conducted posterior predictive checks. Repeating the behavioral analysis on the winning model's predictions confirmed that this model recapitulated the interaction effects observed in patients with psychotic disorders (RQ1) and individuals with JTC (RQ2), as well as the absence of a delusion effect (RQ3) in accordance with the behavioral analysis (figure 2). Our parameter recovery analysis indicated good recovery for

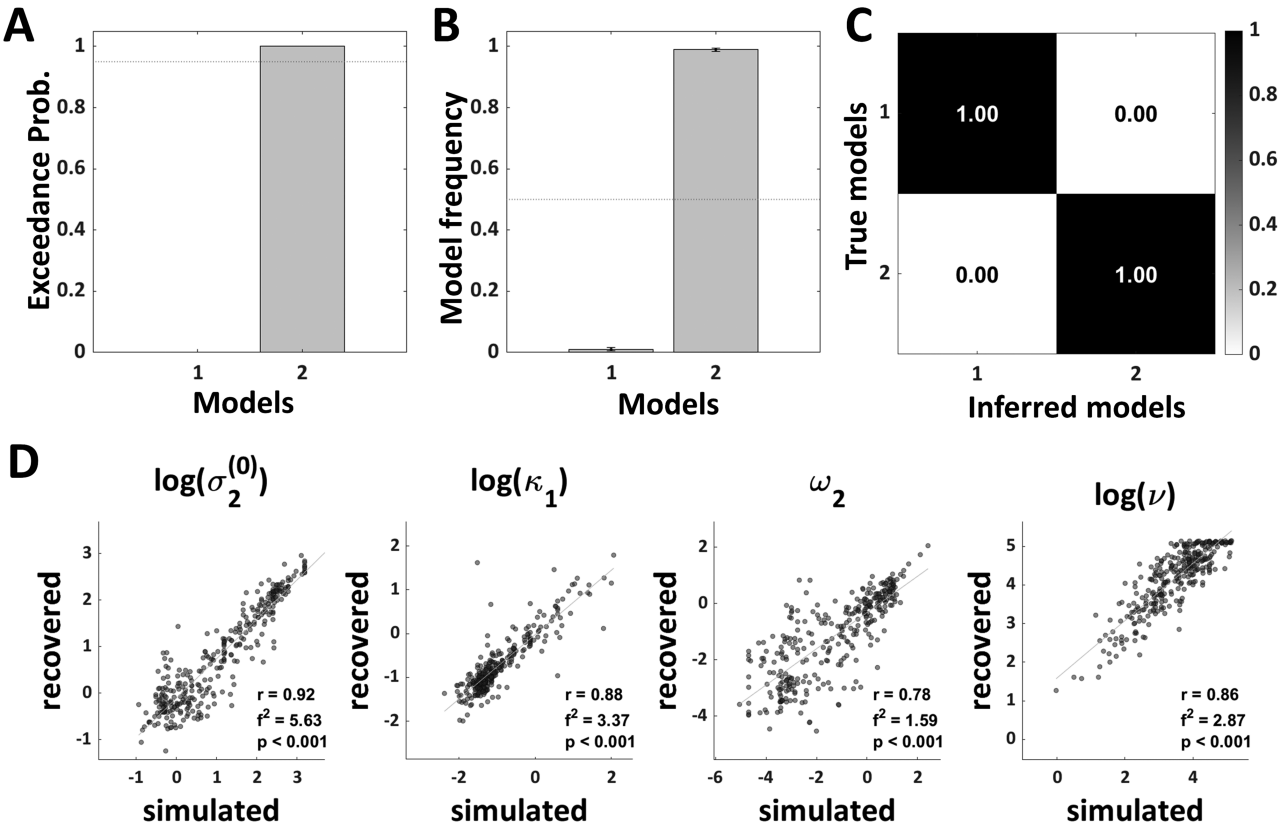

Fig. 3. Bayesian model selection and recovery analyses. Results of random-effects Bayesian model selection.[33,34] **A**: Protected exceedance probabilities. The dashed line indicates 95% exceedance probability. **B**: Expected model frequencies as a measure of effect size. The dashed line indicates chance model frequencies (ie, $1/\#\text{models} = 50$ % with two models). (C) Model recovery analysis. We simulated 20 synthetic datasets based on the empirical parameter estimates. The sample size of each synthetic dataset was chosen to be equivalent to the empirical sample size ($N = 317$) and the noise level was set based on the empirically estimated decision noise ν_{est}. Each simulation was initialized using different random seeds ($n = 20$ seeds) to account for the stochasticity of the simulation. We subsequently re-inverted both models—the generating and the competing—on the simulated data to determine, whether we could recover the true model that generated the data. The gray scale indicates protected exceedance probability averaged across all random seeds. Please, see also supplementary material for further details on the simulation analysis. **(D)** Parameter recovery analysis for the winning model. Based on the simulation analysis described in C, we assessed how accurately the parameters generating the data ("simulated") corresponded to the parameters that were estimated when re-inverting the same model on that data ("recovered") by computing Pearson correlations and effect sizes using Cohen's f^2, where an $f^2 \geq 0.35$ can be considered a large effect size.[43] We thus interpret $f^2 \geq 0.35$ as evidence for a good parameter recovery. This figure displays the recovery results for one of the random seeds, but all other seeds were comparable. In all simulations recovery was good for all parameters (ie, Cohen's $f^2 \geq 0.35$).

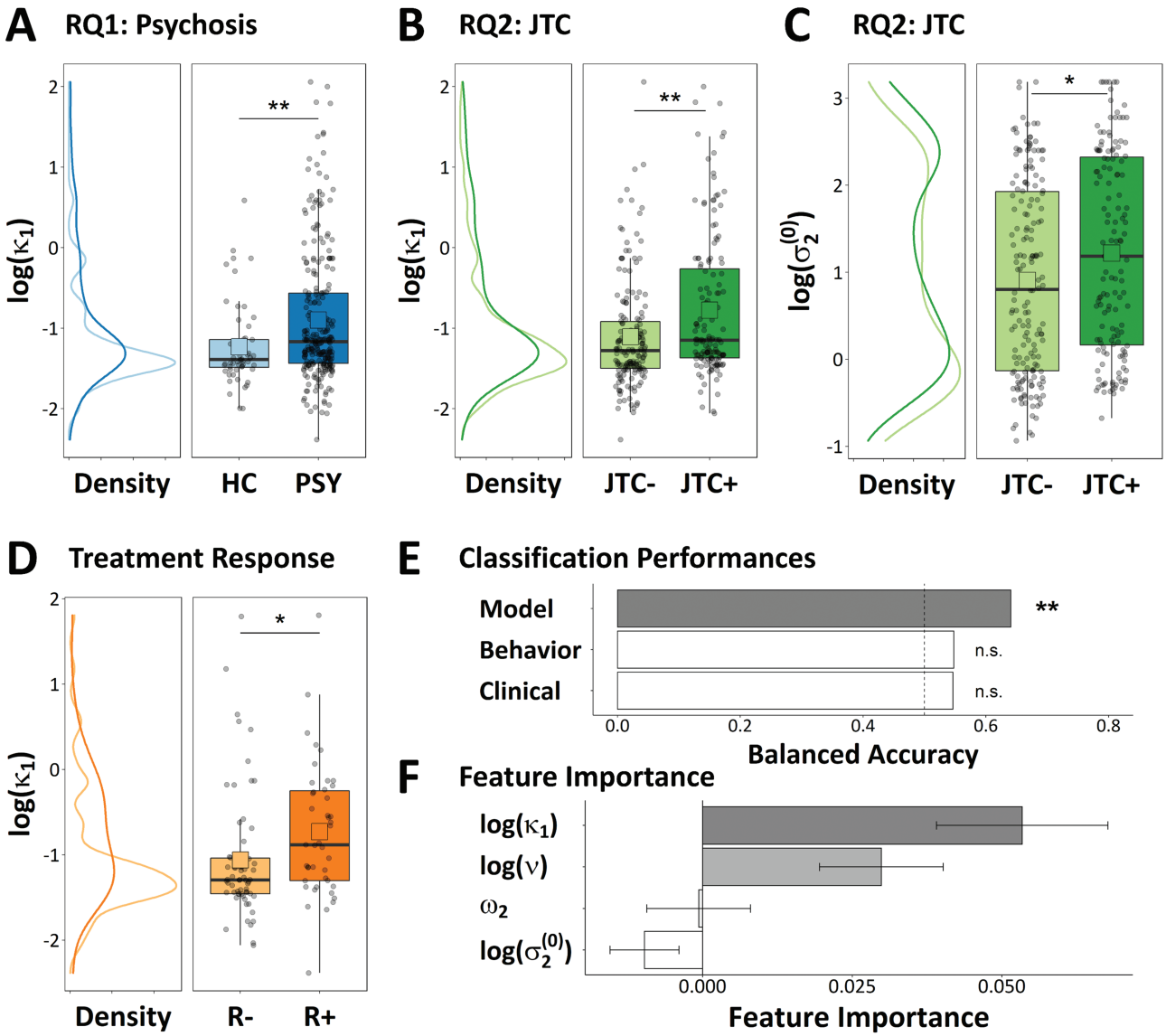

Fig. 4. Parameter group effects and treatment response prediction. (A) Belief instability κ_1 across healthy controls (HC) and patients with psychotic disorders (PSY). (B) Belief instability κ_1 and (C) prior uncertainty $\sigma_2^{(0)}$ across individuals without (JTC−) and with (JTC+) jumping-to-conclusions bias (decision after ≤ 2 fish). (D) Belief instability κ_1 across patients, who showed either no response (R−) or a response (R+) to Metacognitive Training defined as 20% decrease compared to baseline in the Positive and Negative Syndrome Scale (PANSS)[38] positive factor according to factor solution by van der Gaag.[39] RQ: Research question. Horizontal lines and squares in boxplots represent median and mean, respectively. Boxes span the 25th to 75th quartiles and whiskers extend from hinges to the largest and smallest value that lies within 1.5 × interquartile range. Asterisks indicate significance of nonparametric Kruskal–Wallis tests at: $^{**}P < .01$, and $^{*}P < .05$, using Bonferroni correction. (E) Classification performance of random forest trained on either the winning models' parameters (Model), raw behavioral data (probability estimates and choices) and a jumping-to-conclusion bias indicator (Behavior), or on PANSS baseline items (Clinical) to predict treatment response. Asterisks indicate significant permutation test with 1000 label permutations at: $^{**}P < .01$, n.s.: not significant. (F) Feature importance for the random forest trained on winning models' parameters. Bar size corresponds to mean and error bars to standard deviation across cross-validation folds.

all parameters in all simulations (ie, Cohen's $f^2 > 0.35$; figure 3D; supplementary material). Next, we tested for group differences in model parameters.

Parameter Group Effects

RQ1: Parameter Effects between HC and Patients with Psychotic Disorders. First, we found that patients were characterized by significantly larger belief instability κ_1 compared to HC ($\eta^2 = 0.033$, $P = .005$; figure 4A), which was reproduced in IQ-matched subsamples (supplementary material). Increased belief instability κ_1 likely explained the increased updating in response to disconfirmatory evidence that was observed behaviorally (figure 1C). None of the other parameters showed a significant effect.

RQ2: Parameter Effects Between Individuals with and Without JTC. Second, individuals with JTC displayed significantly larger belief instability κ_1 ($\eta^2 = 0.038$, $P = .002$; figure 4B), but also increased prior uncertainty $\sigma_2^{(0)}$ ($\eta^2 = 0.0208$, $P < .050$ ($P = .0499$); figure 4C), which likely accounted for the initial increase in belief updating found in individuals with JTC (figure 1C). Both effects remained significant, when comparing IQ-matched subsamples (supplementary material). None of the other parameters significantly differed across JTC groups.

RQ3: Parameter Effects Between Patients with Low and High Current Delusions. Lastly, we found no significant effect of current delusions on any model parameters. Based on an alternative definition of the delusion groups (supplementary material), we identified a trend-effect of increased decision noise ν in patients with any current clinically-relevant delusions (PANSS P1: Delusions ≥ 3, $\eta^2 = 0.0126$, $p_{uncorr} = 0.046$, $P = .186$). To assess the relationship with other symptoms, we computed Kendall rank correlations between all four model parameters and the five PANSS factors,[39] or the PSYRATS[36] delusion and hallucination subscales. We found only a trend-effect suggesting that increased decision noise ν was associated with higher PSYRATS hallucination scores ($\tau = -0.114$, $p_{uncorr} = 0.016$, $P = .130$).

Treatment Response Prediction

Increased belief instability κ_1 was significantly associated with better treatment response at the group level ($\eta^2 = 0.074$, $P = .021$, figure 4D). Subsequently, we also investigated, whether treatment response could be predicted at the individual level.

The classifier trained on model parameters predicted treatment response with 64% balanced accuracy (BAC), which was significantly greater than chance, indicated by a permutation test ($P = .001$, figure 4E; area under the curve (AUC): 0.67, sensitivity (SE): 0.53, specificity (SP): 0.76, positive predictive value (PPV): 0.60, negative predictive value (NPV): 0.71). This model's performance was mainly driven by belief instability κ_1, followed by decision noise ν (figure 4F).

To evaluate whether the modeling step was necessary for this performance, we also trained a classifier directly on the raw behavioural data. This model could not predict treatment response above chance (BAC: 0.55, $P = .127$, figure 4E; AUC: 0.63, SE: 0.38, SP: 0.71, PPV: 0.49, NPV: 0.63).

Lastly, to investigate whether treatment response could be equally well or even better predicted using clinical measures that are more readily available in clinical practice, we trained the third model on clinical baseline information. Despite differences in symptom expression at baseline, this model did not predict treatment response above chance (BAC: 0.55, $P = .139$, figure 4E; AUC: 0.58, SE: 0.41, SP: 0.68, PPV: 0.47, NPV: 0.63) suggesting that the model-based analysis indeed uncovered additional clinically-relevant information.

Discussion

We employed a computational modeling approach to understand belief updating dynamics during the fish task and their relationship with psychotic disorder diagnosis (RQ1), JTC (RQ2), and current delusions (RQ3). Comparing two competing mechanisms, we found that *belief updating subject to belief instability* best explained participants' behaviour in our study. This model was well-recoverable and could reproduce differences in probabilistic reasoning associated with psychotic disorders and a propensity to jump to conclusions. Analyzing parameters of the winning model, we obtained two major results: First, we found that probabilistic reasoning in patients with psychotic disorders was explained by the model through increased belief instability. Second, our results suggest that belief instability differentiated patients who responded from those who did not respond to a Metacognitive Training intervention, both at the group level and the individual level.

Learning Mechanisms Underlying Psychotic Disorders and Jumping-to-conclusions

Despite analyzing a different task in a more heterogeneous patient population, we replicated previous findings by Adams et al,[21] which suggested that abnormal belief updating in patients with schizophrenia performing the beads task may be explained by increased belief instability κ_1. Our results also offer a possible explanation for JTC as a general cognitive trait across HC and patients as we found an increase in prior uncertainty associated with JTC that explained this effect. Importantly, both associations held in a subsample, which was matched for IQ (supplementary material) and were not accounted for by differences in education, or medication. Additionally, we found a significant increase in belief instability in participants with JTC, which remains challenging to interpret. Based on simulations (figure 1C), the most likely explanation is that this increase in belief instability explained differences in belief updating when participants were faced with disconfirmatory evidence (fish 9) that the behavioural analysis did not identify due to a lack of power. However, we cannot rule out that κ_1 also partially explained increased initial updating for those participants, where the parameter assumed very low values.

Related Modeling Work

Although we replicated Adams and colleagues' findings[21] of a relative increase in belief instability in patients with psychotic disorders, we note that absolute belief instability in our sample was smaller. Furthermore, unlike

others,[21,44] we only found trend-effects linking increased decision noise with symptom severity, although feature importance measures indicated that decision noise was relevant for treatment response prediction. This divergence may be explained by differences between clinical populations (schizophrenia vs psychotic disorders) or different tasks that were used (beads vs fish task), and possibly ensuing differences in task comprehension.

In contrast to other results,[5–7] Baker et al[45] found that delusion severity correlated with more conservative behaviour, primarily, in a condition with high uncertainty (60:40 beads ratio), which their model explained through increased reliance on priors. Intuitively, this agrees with belief rigidity—by definition a hallmark of delusions. The authors used a performance-contingent monetized beads task with endowment. The impact of performance-contingent versus flat payments (as in our case) is not entirely clear. Some authors argued that the payment mode may affect cognitive strategies employed, for example by setting new goals or spending cognitive resources on strategy development.[46,47] Furthermore, it is possible that endowments led to more loss-averse (conservative) instead of risk-seeking (liberal) behaviour. Due to differences in environmental uncertainty and payment structure, a direct comparison is difficult. However, our model appears to capture a mechanism underlying psychotic disorders in general and not specifically related to current delusions.

Other computational approaches were employed to characterize belief updating in schizophrenia.[48,49] Using a task related to ours, but without any sequential updating, Jardri et al[49] suggested that schizophrenia is likely characterized by an overcounting of sensory information. Increased prior uncertainty has a comparable effect in early trials, because it increases the magnitude of belief updates, leading to stronger weighing of sensory information early on (figure 1C). However, we found that belief instability, rather than prior uncertainty, differentiated patients with psychotic disorders from HC. Increasing belief instability primarily results in exaggerated belief updates, when faced with disconfirmatory evidence specifically, not an overcounting of any evidence.

Can Belief Instability be Leveraged to Predict Treatment Response?

At the group level, we found that belief instability significantly differed in patients who responded to an intervention targeting cognitive biases. Intriguingly, greater belief instability (ie, more extreme pathology) related to better treatment response. One speculative explanation for this is that increased belief instability may indicate a vulnerable cognitive system, which places individuals at higher risk of being susceptible to delusional ideas,[19] but also more amenable to a therapy designed to make cognition more robust.

Subsequent analyses suggested that model parameters also predicted *individual* treatment response with 64% accuracy. Bearing in mind that treatment response prediction constitutes one of the most challenging problems in psychiatry and that Metacognitive Training was merely an add-on treatment in patients already treated with antipsychotics, we believe this to be an encouraging result. Given previous evidence,[50] it is interesting to note that neither JTC nor clinical baseline measures predicted individual treatment response above chance. This finding may suggest that the model-derived computational fingerprint contains additional clinically-relevant information about inference mechanisms. This prognostic model may be a valuable screening instrument for clinical trials, or help reduce the therapy load on patients with motivational deficits. However, the accuracy based on model parameters alone is likely not sufficient to justify clinical implementation. Nonetheless, this model can provide a valuable component of a sequential prognostic test battery, together with other clinical or neurophysiological predictors, as proposed previously for transition-to-psychosis[51] or negative symptom prediction.[52]

To summarize, two notable benefits of this approach are (1) the interpretability of the predictors and (2) the simplicity of the assessment, since the model relies on very little data per participant. Task and model fitting can be performed fast rendering it attractive for clinical applications, but the results still need to be replicated in different research sites.

A striking aspect of our results is that despite evident relationships between psychotic disorders and both behavioral and computational measures—and the potential for computational parameters to predict treatment outcome—we did not find any relationship between these measures and current delusions, even though these tasks were designed to assess reasoning biases thought to contribute to delusions themselves. Our findings add to a growing literature including meta-analyses,[53] large case-control,[54] and population-based studies[55] that find weak or absent correlations between delusion and beads task measures.

Limitations

Certain limitations merit attention: First, we only modeled ten trials per subject. While this increases clinical applicability, obtaining precise parameter estimates from such sparse data is challenging. Surprisingly, we could still recover parameters and were able to pinpoint computational mechanisms. Second, although we carefully controlled several confounders (education, medication, premorbid IQ), other confounders cannot be ruled out (eg, socioeconomic status). More fine-grained measures of socioeconomic status should be included in future studies.[45] Thirdly, participants were not incentivized to respond quickly. Fast decisions could reflect patients' desire

to end the experiment soon. However, participants were required to complete all trials rendering this unlikely. Furthermore, it is unclear, how monetization affects the cognitive processes involved. Fourth, even though we defined treatment response as change scores and despite our finding that baseline symptoms did not predict treatment response above chance, we cannot exclude influence of regression-to-the-mean effects on the treatment response prediction analysis presently. Lastly, without a clinical control group, we could not assess the specificity of increased belief instability, which is an important avenue for future research.

Future Directions

Future studies are required to examine the physiological basis of belief instability. A candidate mechanism is *N*-methyl-D-aspartate receptor (NMDAR) hypofunction[56,57] as a recent pharmacological study suggests that NMDAR functioning is linked to probabilistic reasoning during the beads task.[58] If this relationship can be confirmed, treatment response prediction to pharmacological interventions targeting glutamate metabolism (eg, d-serine or glycine), may be a promising avenue of research. Furthermore, future research is required to assess, whether model parameters allow stratifying patients for clinical trials using Metacognitive Training or similar interventions. Lastly, this model-based approach can also inform the design of new interventions that target belief instability specifically to assess whether such interventions can improve patients' well-being.

Conclusions

In conclusion, our results suggest that increased belief instability may be a key computational mechanism underlying probabilistic reasoning in patients with psychotic disorders. Furthermore, we provide a proof-of-concept that this computational parameter can potentially be leveraged to predict clinically-relevant outcomes.

Supplementary Material

Supplementary material is available at https://academic.oup.com/schizophreniabulletin/.

Funding

This work was supported by the Brain & Behavior Research Foundation (2012 NARSAD Young Investigator Grant, 18749 to CA); the German Research Foundation (DFG) (AN 970/4-1 and MU 2705/2-1 to CA); Federal Ministry of Education and Research (BMBF) (Mo 969/6-1); the Swiss National Science Foundation (Doc.Mobility, 200054 to DJH, Ambizione, PZ00P3_167952 to AOD); the Krembil Foundation (to AOD); the Novartis Foundation for medical-biological Research (19A011 to AOD); the Medical Research Council (Skills Development Fellowship, MR/S007806/1 to RAA); the National Institute for Health Research (NIHR) University College London Hospitals (UCLH) Biomedical Research Centre and the Centre for Medical Image Computing (CMIC) (Platform Grant, EP/M020533/1 to RAA).

Conflict of Interest

The authors have declared that there are no conflicts of interest in relation to the subject of this study.

References

1. Jaspers K. *Allgemeine Psychopathologie*. Berlin: J. Springer; 1913.
2. Freeman D, Garety PA. Comments on the content of persecutory delusions: does the definition need clarification? *Br J Clin Psychol*. 2000;39(4):407–414. doi:10.1348/0144665001634000.
3. Knowles R, McCarthy-Jones S, Rowse G. Grandiose delusions: a review and theoretical integration of cognitive and affective perspectives. *Clin Psychol Rev*. 2011;31(4):684–696. doi:10.1016/j.cpr.2011.02.009.
4. Appelbaum PS, Robbins PC, Roth LH. Dimensional approach to delusions: comparison across types and diagnoses. *Am J Psychiatry*. 1999;156(12):1938–1943. doi:10.1176/ajp.156.12.1938.
5. Ross RM, McKay R, Coltheart M, Langdon R. Jumping to conclusions about the beads task? A meta-analysis of delusional ideation and data-gathering. *Schizophr Bull*. 2015;41(5):1183–1191. doi:10.1093/schbul/sbu187.
6. Dudley R, Taylor P, Wickham S, Hutton P. Psychosis, delusions and the "Jumping to Conclusions" reasoning bias: a systematic review and meta-analysis. *Schizophr Bull*. 2015;42(3):652–665. doi:10.1093/schbul/sbv150.
7. McLean BF, Mattiske JK, Balzan RP. Association of the jumping to conclusions and evidence integration biases with delusions in psychosis: a detailed meta-analysis. *Schizophr Bull*. 2017;43(2):344–354. doi:10.1093/schbul/sbw056.
8. Young HF, Bentall RP. Probabilistic reasoning in deluded, depressed and normal subjects: effects of task difficulty and meaningful versus non-meaningful material. *Psychol Med*. 1997;27(2):455–465. doi:10.1017/S0033291796004540.
9. Garety P. Reasoning and delusions. *Br J Psychiatry*. 1991;159(S14):14–18. doi:10.1192/S0007125000296426.
10. Fear CF, Healy D. Probabilistic reasoning in obsessive–compulsive and delusional disorders. *Psychol Med*. 1997;27(1):199–208. doi:10.1017/S0033291796004175.
11. Peters E, Garety P. Cognitive functioning in delusions: a longitudinal analysis. *Behav Res Ther*. 2006;44(4):481–514. doi:10.1016/j.brat.2005.03.008.
12. Phillips LD, Edwards W. Conservatism in a simple probability inference task. *Exp Psychol*. 1966;72(3):346–354. doi:10.1037/h0023653.
13. Huq SF, Garety PA, Hemsley DR. Probabilistic judgements in deluded and non-deluded subjects. *Q J Exp Psychol Sect A*. 1988;40(4):801–812. doi:10.1080/14640748808402300.

14. Woodward TS, Munz M, LeClerc C, Lecomte T. Change in delusions is associated with change in "jumping to conclusions." *Psychiatry Res*. 2009;170(2):124–127. doi:10.1016/j.psychres.2008.10.020
15. Moritz S, Veckenstedt R, Hottenrott B, Woodward TS, Randjbar S, Lincoln TM. Different sides of the same coin? Intercorrelations of cognitive biases in schizophrenia. *Cogn Neuropsychiatry*. 2010;15(4):406–421. doi:10.1080/13546800903399993.
16. Speechley WJ, Whitman JC, Woodward TS. The contribution of hypersalience to the "jumping to conclusions" bias associated with delusions in schizophrenia. *J Psychiatry Neurosci*. 2010;35(1):7–17. doi:10.1503/jpn.090025.
17. Freeman D, Garety P. Advances in understanding and treating persecutory delusions: a review. *Soc Psychiatry Psychiatr Epidemiol*. 2014;49(8):1179–1189. doi:10.1007/s00127-014-0928-7.
18. Van Oosterhout B, Smit F, Krabbendam L, Castelein S, Staring ABP, Van Der Gaag M. Metacognitive training for schizophrenia spectrum patients: a meta-analysis on outcome studies. *Psychol Med*. 2016;46(1):47–57. doi:10.1017/S0033291715001105.
19. Adams RA, Vincent P, Benrimoh D, Friston KJ, Parr T. Everything is connected: inference and attractors in delusions. *Schizophr Res*. Published online 2021. doi:10.1016/j.schres.2021.07.032.
20. Moritz S, Veckenstedt R, Randjbar S, Vitzthum F, Woodward TS. Antipsychotic treatment beyond antipsychotics: metacognitive intervention for schizophrenia patients improves delusional symptoms. *Psychol Med*. 2011;41(9):1823–1832. doi:10.1017/S0033291710002618.
21. Adams RA, Napier G, Roiser JP, Mathys C, Gilleen J. Attractor-like dynamics in belief updating in schizophrenia. *J Neurosci*. 2018;38(44):9471–9485. doi:10.1523/JNEUROSCI.3163-17.2018.
22. Andreou C, Wittekind CE, Fieker M, et al. Individualized metacognitive therapy for delusions: a randomized controlled rater-blind study. *J Behav Ther Exp Psychiatry*. 2017;56:144–151. doi:10.1016/j.jbtep.2016.11.013.
23. Andreou C, Steinmann S, Leicht G, Kolbeck K, Mulert C. fMRI correlates of jumping-to-conclusions in patients with delusions: connectivity patterns and effects of metacognitive training. *NeuroImage Clin*. 2018;20:119–127. doi:10.1016/j.nicl.2018.07.004.
24. Moritz S, Veckenstedt R, Bohn F, et al. Complementary group Metacognitive Training (MCT) reduces delusional ideation in schizophrenia. *Schizophr Res*. 2013;151:61–69. doi:10.1016/j.schres.2013.10.007.
25. Daunizeau J, den Ouden HEM, Pessiglione M, Kiebel SJ, Stephan KE, Friston KJ. Observing the Observer (I): Meta-Bayesian Models of Learning and Decision-Making. *PLoS One*. 2010;5(12):e15554. doi:10.1371/journal.pone.0015554.
26. Daunizeau J, den Ouden HEM, Pessiglione M, Kiebel SJ, Friston KJ, Stephan KE. Observing the observer (II): deciding when to decide. *PLoS One*. 2010;5(12):e15555. doi:10.1371/journal.pone.0015555.
27. Mathys CD, Daunizeau J, Friston KJ, Stephan KE. A Bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci*. 2011;5(May):1–20. doi:10.3389/fnhum.2011.00039.
28. Mathys CD, Lomakina EI, Daunizeau J, et al. Uncertainty in perception and the Hierarchical Gaussian Filter. *Front Hum Neurosci*. 2014;8:1–24. doi:10.3389/fnhum.2014.00825.
29. Powers AR, Mathys C, Corlett PR. Pavlovian conditioning – induced hallucinations result from overweighting of perceptual priors. *Science (80-)*. 2017;357(6351):596–600. doi:10.1126/science.aan3458.
30. Diaconescu AO, Hauke DJ, Borgwardt S. Models of persecutory delusions: a mechanistic insight into the early stages of psychosis. *Mol Psychiatry*. 2019;24(9):1258–1267. doi:10.1038/s41380-019-0427-z.
31. Reed EJ, Uddenberg S, Suthaharan P, et al. Paranoia as a deficit in non-social belief updating. *Elife*. 2020;9:1–55. doi:10.7554/eLife.56345.
32. Suthaharan P, Reed EJ, Leptourgos P, et al. Paranoia and belief updating during the COVID-19 crisis. *Nat Hum Behav*. 2021;5:1190–1202. doi:10.1038/s41562-021-01176-8.
33. Rigoux L, Stephan KE, Friston KJ, Daunizeau J. Bayesian model selection for group studies - revisited. *Neuroimage*. 2014;84:971–985. doi:10.1016/j.neuroimage.2013.08.065.
34. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. Bayesian model selection for group studies. *Neuroimage*. 2009;46(4):1004–1017. doi:10.1016/j.neuroimage.2009.03.025.
35. Catalan A, Tognin S, Kempton MJ, et al. Relationship between jumping to conclusions and clinical outcomes in people at clinical high-risk for psychosis. *Psychol Med*. Published online 2020:1–9. doi:10.1017/S0033291720003396.
36. Haddock G, McCarron J, Tarrier N, Faragher EB. Scales to measure dimensions of hallucinations and delusions: the psychotic symptom rating scales (PSYRATS). *Psychol Med*. 1999;29(4):879–889. doi:10.1017/S0033291799008661.
37. Eichner C, Berna F. Acceptance and efficacy of metacognitive training (MCT) on positive symptoms and delusions in patients with schizophrenia: a meta-analysis taking into account important moderators. *Schizophr Bull*. 2016;42(4):952–962. doi:10.1093/schbul/sbv225.
38. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13(2):261–276. doi:10.1093/schbul/13.2.261.
39. van der Gaag M, Hoffman T, Remijsen M, et al. The five-factor model of the Positive and Negative Syndrome Scale II: a ten-fold cross-validation of a revised model. *Schizophr Res*. 2006;85(1):280–287. doi:10.1016/j.schres.2006.03.021.
40. Lachar D, Bailley SE, Rhoades HM, Varner R V. Use of BPRS-A percent change scores to identify significant clinical improvement: accuracy of treatment response classification in acute psychiatric inpatients. *Psychiatry Res*. 1999;89(3):259–268. doi:10.1016/S0165-1781(99)00114-6.
41. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32. doi:10.1023/A:1010933404324.
42. Koutsouleris N, Borgwardt S, Meisenzahl EM, Bottlender R, Möller HJ, Riecher-Rössler A. Disease prediction in the at-risk mental state for psychosis using neuroanatomical biomarkers: results from the FePsy study. *Schizophr Bull*. 2012;38(6):1234–1246. doi:10.1093/schbul/sbr145.
43. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. New York: Routledge; 1988.
44. Moutoussis M, Bentall RP, El-Deredy W, Dayan P. Bayesian modelling of Jumping-to-Conclusions bias in delusional patients. *Cogn Neuropsychiatry*. 2011;16(5):422–447. doi:10.1080/13546805.2010.548678.
45. Baker SC, Konova AB, Daw ND, Horga G. A distinct inferential mechanism for delusions in schizophrenia. *Brain*. 2019;142(6):1797–1812. doi:10.1093/brain/awz051.

46. Bonner SE, Sprinkle GB. The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. *Account Organ Soc.* 2002;27(4):303–345. doi:10.1016/S0361-3682(01)00052-6.
47. Locke EA, Shaw KN, Saari LM, Latham GP. Goal setting and task performance: 1969–1980. *Psychol Bull*. 1981;90(1):125–152. doi:10.1037/0033-2909.90.1.125.
48. Jardri R, Denève S. Circular inferences in schizophrenia. *Brain*. 2013;136(11):3227–3241. doi:10.1093/brain/awt257.
49. Jardri R, Duverne S, Litvinova AS, Denève S. Experimental evidence for circular inference in schizophrenia. *Nat Commun*. 2017;8. doi:10.1038/ncomms14218.
50. Leanza L, Studerus E, Bozikas VP, Moritz S, Andreou C. Moderators of treatment efficacy in individualized metacognitive training for psychosis (MCT+). *J Behav Ther Exp Psychiatry*. 2020;68:101547. doi:10.1016/j.jbtep.2020.101547.
51. Clark SR, Schubert KO, Baune BT. Towards indicated prevention of psychosis: using probabilistic assessments of transition risk in psychosis prodrome. *J Neural Transm*. 2015;122(1):155–169. doi:10.1007/s00702-014-1325-9.
52. Hauke DJ, Schmidt A, Studerus E, et al. Multimodal prognosis of negative symptom severity in individuals at increased risk of developing psychosis. *Transl Psychiatry*. 2021;11(1). doi:10.1038/s41398-021-01409-4.
53. Ashinoff BK, Singletary NM, Baker SC, Horga G. Rethinking delusions: a selective review of delusion research through a computational lens. *Schizophr Res.* 2021. doi:10.1016/j.schres.2021.01.023.
54. Tripoli G, Quattrone D, Ferraro L, et al. Jumping to conclusions, general intelligence, and psychosis liability: findings from the multi-centre EU-GEI case-control study. *Psychol Med*. 2021;51(4):623–633. doi:10.1017/S003329171900357X.
55. Croft J, Teufel C, Heron J, et al. A computational analysis of abnormal belief-updating processes and their association with psychotic experiences and childhood trauma in a UK birth cohort. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2021. doi:10.1016/j.bpsc.2021.12.007.
56. Moghaddam B, Javitt D. From revolution to evolution: the glutamate hypothesis of schizophrenia and its implication for treatment. *Neuropsychopharmacology*. Published online 2012:4–15. doi:10.1038/npp.2011.181.
57. Javitt DC, Zukin SR, Heresco-Levy U, Umbricht D. Has an angel shown the way? Etiological and therapeutic implications of the PCP/NMDA model of schizophrenia. *Schizophr Bull*. 2012;38(5):958–966. doi:10.1093/schbul/sbs069.
58. Strube W, Marshall L, Quattrocchi G, et al. Glutamatergic contribution to probabilistic reasoning and jumping to conclusions in schizophrenia: a double-blind, randomized experimental trial. *Biol Psychiatry*. 2020;88(9):687–697. doi:10.1016/j.biopsych.2020.03.018.