Overestimating environmental volatility increases switching behavior and is linked to activation of dorsolateral prefrontal cortex in schizophrenia

Short title: Volatility, choice switching and PFC in schizophrenia

L. Deserno[1,2,3*], R. Boehme[1,4*], C. Mathys[5,6,7,8], T. Katthagen[1,9], J. Kaminski[1,10], K. E. Stephan[6,8], A. Heinz[1,10], F. Schlagenhauf[1,2]

[1]Department of Psychiatry and Psychotherapy, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health
[2]Max-Planck-Institute for Human Cognitive and Brain Sciences, Leipzig;
[3]Department of Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, University of Leipzig, Leipzig;
[4]Center for Social and Affective Neuroscience, Linköping University, Linköping
[5]Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy
[6]Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London;
[7]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, London;
[8]Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich & ETH Zurich, Switzerland;
[9] Berlin School of Mind and Brain,Humboldt-Universität zu Berlin, Berlin;
[10]Berlin Institute of Health, Berlin, Germany
[11]Cluster of Excellence NeuroCure, Charite Universitätsmedizin, Berlin;
*equal contribution

Word count: Abstract 244, Manuscript 4332
3 Tables, 4 Figures, Supplemental Material

Correspondence:
Dr. Lorenz Deserno
Max Planck Institute for Human Cognitive and Brain Sciences
Department of Neurology
Stephanstraße 1A
04103 Leipzig
deserno@cbs.mpg.de

Abstract (244 / 250 words)

Background. Reward-based decision-making is impaired in schizophrenia, as reflected by increased switching between choices. The underlying cognitive mechanisms and associated neural signatures remain unknown. Reinforcement learning (RL) and hierarchical Bayesian learning account for this behavior in different ways. We hypothesized that enhanced switching during flexible reward-based decision-making in schizophrenia relates to higher-order beliefs about environmental volatility and examined the associated neural signatures.

Methods. 46 medicated schizophrenia patients and 43 controls underwent a reward-based decision-making task requiring flexible behavior to changing action-outcome contingencies during functional Magnetic Resonance Imaging (fMRI). Computational modeling of behavior was performed, including RL and the Hierarchical Gaussian Filter (HGF). The estimated learning trajectories informed the analysis of fMRI data.

Results. A three-level HGF accounted best for the observed choice data and revealed a heightened prior belief about environmental volatility and a stronger influence of volatility on lower-level learning of action-outcome contingencies in schizophrenia. This finding was replicated in an independent sample of unmedicated patients. Beliefs about environmental volatility were reflected by higher activity in dorsolateral prefrontal cortex (dlPFC) of patients compared to controls.

Conclusions. This study suggests a mechanistic explanation for instable behavior in schizophrenia: patients inferred the environment as being too volatile and thus overestimated environmental changes, leading to maladaptive choice switching. Our data suggest enhanced dlPFC activity related to beliefs about environmental volatility as a neural learning signature of instable behavior. Such detailed 'computational phenotyping' may provide useful information to dissect clinical heterogeneity and could improve prediction of outcome.

Introduction

Cognitive and motivational deficits represent important characteristics of schizophrenia that are associated with clinical and social outcome (1-5). Flexible reward-based learning and decision-making require the integration of cognition and motivation and are impaired in schizophrenia (6, 7). These impairments are present at the onset of the disorder, are independent of lower general IQ, remain stable over time (8, 9) and were proposed as neurocognitive markers with potential clinical utility (10). However, the underlying cognitive mechanisms and associated neural signatures remain to be understood.

Flexible reward-based learning and decision-making can be probed via variants of reversal learning. In such tasks, schizophrenia is characterized by increased switching between choice options (8, 11-16). The underlying mechanisms of this instable behavior remain unknown but can be targeted by computational modeling (17). For example, in 'reinforcement learning' (RL; 18), choices are selected based on their expected value, which are learned trial-by-trial by weighing reward prediction errors (RPEs) with the learning rate. RPEs are closely aligned with phasic dopamine transients (19, 20). Considering enhanced presynaptic dopamine synthesis capacity in schizophrenia (21, 22), this could translate into enhanced phasic dopamine transients in schizophrenia, which in turn might result in increased learning rates (23). Although this could theoretically account well for instable behavior observed in patients, increased learning rates were not found in schizophrenia (e.g. 16, for review see 17, 23, 24).

Theories of predictive coding (25) and hierarchical Bayesian inference have been used to postulate mechanisms underlying the symptoms of schizophrenia (26-29). Symptoms are understood as false inference about the world due to altered precision attributed to beliefs on different hierarchical levels. Dysfunction at higher levels, which are thought to extract and represent the general and stable features of the environment, might lead to experiencing the world as being more volatile, i.e. less stable and more surprising. This was applied with a strong focus on positive symptoms (30) and has received empirical

evidence (e.g. 31). Applying this framework to reward-based learning, beliefs about the probability of rewards are learned at lower levels but are also determined by learning about the volatility of reward probabilities at higher levels (32). This higher-level environmental volatility directly influences trial-by-trial learning from lower-level RPEs by scaling the belief update (via a ratio of precisions from both levels). Thus, a higher-order belief about environmental volatility could induce rapid updates of lower-level beliefs about reward probabilities and promote enhanced switching behavior in patients.

In schizophrenia, striatal and prefrontal activation is reduced during reward anticipation and receipt (33-35) but this can be ameliorated by antipsychotics (36). Reduced striatal RPE activity was observed in unmedicated (16) but not in medicated patients (14, 37). Neural correlates of hierarchical Bayesian learning were demonstrated in functional Magnetic Resonance Imaging (fMRI) studies in healthy individuals (38, 39) linking the concepts of volatility und uncertainty to activation in fronto-striatal circuits, in particular prefrontal sub-regions (32, 40). While neural correlates of hierarchical Bayesian learning were successfully used to distinguish between individuals with and without hallucinations and patients with and without psychosis (31), this has not yet facilitated an understanding of the cognitive and motivational processes underlying impairments in flexible decision-making.

Here, we used a modified reward-based reversal-learning task during fMRI in schizophrenia patients and healthy controls. Detailed computational modeling was applied to the behavioral data by comparing RL and a hierarchical Bayesian learning model, the Hierarchical Gaussian Filter (HGF, 41, 42). Learning trajectories from computational modeling informed the analysis of fMRI data. We hypothesized that enhanced switching during decision-making in schizophrenia relate to higher-order beliefs about the volatility of the environment and examine the associated neural signatures by informing the analysis of fMRI data by computational modeling.

Materials and Methods

Participants and instruments. 46 medicated schizophrenia patients and 43 healthy controls (HC) were included in the study (see Table 1). Handedness was assessed with the Edinburgh Handedness Inventory. Cognitive functioning was examined with six subscales selected from the Wechsler Adult Intelligence Scale (43). Symptoms were assessed with the Positive and Negative Syndrome Scale (PANSS) and with delusion and anhedonia items from the Scale for the Assessment of Positive Symptoms and Negative Symptoms (SAPS/SANS). Dosage of antipsychotics was standardized by chlorpromazine equivalents (44). Written informed consent was obtained from all participants. The study was performed in accordance with the Declaration of Helsinki and was approved by the local ethics committee of Charité Universitätsmedizin.

Task. Participants performed a two-choice decision-making task requiring flexible behavioral adaptation (Figure 1A; 45, 46, 47). In 160 trials, individuals decided between two cards each showing a different visual stimulus. The right or left location of stimuli was randomized over trials. After left or right button press (max. response time 1.5s), the selected card was highlighted and a monetary win (10 Eurocent coin) or a monetary loss (crossed 10 Eurocent coin) was shown for .5s. During the inter-trial interval, a fixation cross was presented (distributed jitter, range 1s-12.5s). If no response occurred in time, the message “too slow” appeared. One of two cards was initially assigned with a reward probability of 80% and a punishment probability of 20% and vice versa for the other card. Thus, the task had a simple higher-order structure (Figure 1B): a perfect anti-correlation between the reward probabilities associated with the two choices; whenever one card was associated with a high reward probability of 80%, the other card would be associated with a low reward probability of 20%. Reward contingencies were stable for the first 55 trials (‘pre-reversal’) and for the last 35 trials (‘post-reversal’). During the ‘reversal’ phase, reward contingencies changed four times, after 15 or 20 trials, respectively. In an instruction and training session before MRI scanning, participants

were informed that one of two cards had a superior chance of winning money. They were told that depending on their choice they could either win 0.10€ or lose 0.10€ per trial and to win as much as possible as the total gain was paid out. 20 training trials were performed with a different set of cards and without reversal. After training, participants were instructed that reward probabilities could change over time and to track such changes to win as much as possible. No other information on reversals or the anti-correlated task structure was given.

Analysis of choice behavior. Behavioral performance was quantified by counting 'correct' choices of the stimuli with high (80%) reward probability and was analyzed using repeated-measures ANOVA including the between-subject factor 'group' and the within-subject factor 'phase' (pre-reversal, reversal, post-reversal). Another repeated-measures ANOVA was used to test the effect of previous feedback on subsequent choices, i.e. repeating choices after reward ('win-stay') and punishment ('lose-stay').

Computational models of learning. Computational modeling methods in this paper comprised a model-free RL approach and a Bayesian learning model, the HGF. The HGF is grounded in the "Bayesian Brain hypothesis" and reflects the principles of predictive coding (25). For all models, RPE δ is used to update expectations for the chosen stimulus:

$$(1)\ \delta_{Q_C}^{(k)} = R^{(k)} - Q_c^{(k)}$$

Using the notation of RL, $Q_c^{(k)}$ represents the expectation for receiving reward or punishment by choosing the card *c* in trial k. R denotes the received outcome. RPE δ reflects the difference between received rewards and expectations. The update of $Q_c^{(k)}$is equal to the RPE $\delta_{Q_C}^{(k)}$ weighted by a parameter α, the learning rate:

$$(2)\ Q_c^{(k+1)} - Q_c^{(k)} = \alpha\delta_{Q_C}^{(k)}$$

The learning rate α reflects a constant rate of change in values throughout the series of trials k. Thus, with learning rates close to 1, PEs strongly affect expectations, while learning rates close to 0 lead to little influence of PEs on expectations.

The HGF describes learning as a process of inductive inference under uncertainty. This model deploys hierarchically organized states, in which learning at a higher-level state determines learning at a lower-level state by dynamically adjusting the lower level's learning rate. In our case, the top level represents a trial-by-trial computation of environmental volatility, an estimate of how likely a change in action-outcome contingencies is to occur. This top-level estimate dynamically modulates the learning rate of lower-level RPEs trial-by-trial (see Figure 2A for a graphical illustration). More formally, the evolution of each state $\mathrm{x_i}$ is defined as a Gaussian random walk, where the change in each x_i can be inferred, using a variational Bayesian scheme, giving the posterior mean μ_i^k and variance σ_i^k at each level (with i being an index of the level). In our implementation, action-outcome contingencies $x_2^{(k)}$ evolved as a Gaussian random walk at the logit scale; the probability for a choice being rewarded (i.e., $x_1 = 1$) at a given trial k is $p(x_1 = 1) = s\left(x_2^{(k)}\right)$, where $s(\cdot)$ is the logistic sigmoid function. Crucially, the step size of the Gaussian random walk of x_2 depends on the next higher level x_3:

$$(3)\ \mathrm{p}\left(\mathrm{x}_2^{(\mathrm{k})} \middle| \mathrm{x}_2^{(\mathrm{k}-1)}, \mathrm{x}_3^{\mathrm{k}}\right) = \mathcal{N}\left(x_2^{(k)}; x_2^{(k-1)}, \exp\left(\kappa x_3^{(k)} + \omega\right)\right)$$

The width of the Gaussian is defined by the parameters ω and κ. The term $\kappa x_3^{(k)}$ expresses the influence of the third-level environmental volatility $x_3^{(k)}$ on the second level and κ captures inter-individual differences in coupling of the levels. $x_3^{(k)}$ evolves in the same manner, except that the variance of the Gaussian is a constant ϑ (because there is no higher level):

$$(4)\ p\left(x_3^k \middle| x_3^{k-1}, \vartheta\right) = \mathcal{N}\left(x_3^{(k)};\ x_3^{(k-1)}, \vartheta\right)$$

In our implementation, ω was fixed at the second level because we were particularly interested in learning about environmental volatility at the third level ϑ and its influence

on the lower level κ. Variational inversion of the HGF shows that trial-by-trial updates of posterior means at each level i are proportional to the PE from the level below weighted by a precision ratio (compare equation 2 for the RL equivalent), where precision $\pi_i^{(k)}$ is defined as inverse variance $\pi_i^{(k)} = 1/\sigma_i^{(k)}$.

$$(6)\ \mu_i^k \propto \frac{\hat{\pi}_{i-1}^{(k)}}{\pi_i^{(k)}} \delta_{i-1}^{(k)}$$

For an exact derivation of the precision weights and the precision-weighted PEs, we refer to previous methodological papers (41, 42). In addition to the three-level HGF (HGF3), we also include a two-level variant (HGF2) to test the hypothesized superiority of HGF3.

HGF and RL provide different ways to learn expectations about rewards. As described above, both types of learning update expectations of the chosen card only ("single-update", SU) and the expectation about the unchosen card remains unchanged. Based on the anti-correlated task structure, we implemented a variant of each of the learning models updating the unchosen card simultaneously, i.e. an increase of the chosen card implies a decrease of the unchosen card ("double-update", DU). Please see Supplemental Material for the implementation of this in RL and HGF. Both variants (SU and DU) of each learning model (RL, HGF2, HGF3) were fit to the choice data resulting in six models so far (RL-SU, RL-DU, HGF2-SU, HGF2-DU, HGF3-SU, HGF3-DU).

Decision models. For each trial k, $\mathrm{Q}^{(\mathrm{k})}$ (RL) or $\hat{\mu}_{i=1}^{(k)}$ (HGF) were transformed to choice probabilities by using a logistic function, the softmax:

$$(11)\ \mathrm{p(t,a)} = \frac{\exp(\beta \mathrm{Q(t,a)} + \rho * \mathrm{rep}(\mathrm{r_{t-1}}))}{\sum \exp(\beta \mathrm{Q(t,a')} + \rho * \mathrm{rep}(\mathrm{r_{t-1}}))}$$

In binary choice tasks with correlated reward-probabilities, there is strong autocorrelation of choices (perseveration), which is captured by estimating the parameter ρ. 'rep' indicates whether the outcome of the previous action was a reward or punishment. Thus, ρ_{win} and ρ_{pun} reflect differences in choice perseveration after rewards and punishment

separately. The parameter β was fixed to 1. This parameter β captures how tightly choice probabilities follow learned reward expectations, i.e. whether individuals tend to exploit these expectations. Other decision models, i.e. estimating β as a parameter, were tested but had lower evidence (see Supplemental Material). We also tested the possibility that the parameter β is directly influenced by trial-by-trial environmental volatility as computed on the third-level of HGF3. This additional model tests the hypothesis that dynamic beliefs about environmental volatility influence how closely individuals follow learned expectations by setting $\beta = \exp(-\mu_3^{(k)})$ and allows fitting the prior belief at the third level $\mu_3^{(0)}$ as an additional parameter of HGF3. Since this cannot be estimated for RL or HGF2, it only led to the addition of two more models to the model space (RL-SU, RL-DU, HGF2-SU, HGF2-DU, HGF3-SU, HGF3-DU, HGF3-SU-V, HGF3-DU-V).

Model fitting. Models were fitted using the HGF toolbox 3.15 (http://www.translationalneuromodeling.org/tapas/; 41, 42) applying a quasi-Newton algorithm for optimization. For prior means and variances of parameters, see S-Table 1.

Model selection. The negative variational free energy (an approximation to the log model evidence) was used to perform random-effects Bayesian Model Selection (BMS, 48) (49). In addition to relative BMS, inferred parameters could be confounded by poor absolute model fit, i.e. a model performs poorly because the inferred parameters do not explain the observed data better than chance (16, 50). A subject was classified as 'not fit better than chance' in case the log-likelihood of the choice data relative to the number of trials did not exceed .55 (see Supplemental Material).

Model parameters. Parameters of the best-fitting model were compared between groups using t-tests or the non-parametric Mann-Whitney-U-Test if assumptions of normality were violated. Bonferroni-correction was applied according to the number of parameters

(5, $p<0.01$). Simulations of the task were run based on the inferred parameters to reproduce the observed choice behavior.

Statistical analysis of fMRI data. For acquisition and preprocessing of fMRI data, see Supplemental Material. An event-related analysis was applied to the images on two levels using the general linear model approach in SPM8. Individual (random-effects) model parameters were used to generate regressors provided by the best fitting HGF-learning model. On the first level, one regressor spanned the entire trial from cue to feedback as in a previous study (39). The following five modeling-based trajectories were added as parametric modulators to this regressor to best capture the different aspects of the hierarchical inference process: precision-weighted PEs from the second and third level (ε_2 and ε_3), precision weights from the second and third level (ψ_2 and ψ_3) and the third level volatility (μ_3). Parametric modulators were not serially orthogonalized. Invalid trials (no choice within response window) were modeled separately. Regressors were convolved with the canonical hemodynamic response function in SPM8 and its temporal derivative. The six movement parameters were included in the model as regressors of no interest as well as the first derivative of translational movement with respect to time. An additional regressor was included censoring scan-to-scan movement >1 mm.

For between-group second-level analysis, a random-effects model including t-contrasts of all five modeling-based trajectories (precision-weighted PEs ε_2 and ε_3, precision weights ψ_2 and ψ_3 and the third-level volatility μ_3 was estimated.

Results

Sample. Both groups did not differ in age, gender and handedness (Table 1).

Behavioral data. Repeated-measures ANOVA on choices of the stimulus with 80% reward probability revealed that participants' performance differed between phases (dropping in the reversal phase, main effect of phase F=23.74, p<.001). Patients chose the better card less frequently irrespectively of task phases (Figure 1C, main effect of group F=14.52, p<.001, phase x group interaction F=1.87, p=.16). Thus, the factor phase was dropped from further analyses.

Repeated-measures ANOVA on the probability of choice repetition showed that participants were overall more likely to stay with the previous action after rewards compared to losses (main effect of feedback F=369.80, p<.001). Patients were less likely to stick to the previous action but this did not depend on the feedback received in the previous trial (Figure 1D, main effect of group F=27.77, p<.001, feedback x group interaction F=.02, p=.89). A potential mechanism underlying this instable behavior in schizophrenia remains unknown and the following computational modeling aims to bridge this gap.

Computational modeling. BMS revealed a three-level HGF with double-updating together with a decision model capturing choice repetition separately after rewards and punishments and third-level environmental volatility linked to decision noise (Figure 2A) as the most plausible model (HGF3-DU-V, posterior probability (PP)=50.1%, followed by a single-update RL model PP=14.5% with protected exceedance probabilities of 99.8%, and 0.3%, Table 2). While this model (HGF3-DU-V) was clearly superior in HC alone (PP=64.4%, PXP=100%) and remained first-ranking in patients with a PP of 22.6%, the differences between models were marginal in the patient group (PXP=12.6%).

15 patients and 1 HC were identified as not fit better than chance by any of the models (S-Figure 1A). This led to inclusion of 31 patients and 42 HC for all further modeling-

based analyses and altered BMS only marginally (Table 2). Repeating the analysis of win-stay and lose-stay behavior as reported before with three groups (HC, patients-fit, patients-nofit), revealed a group x feedback interaction (F=20.68, p<.001) showing that reduced lose-stay behavior characterized both patients-fit and patients-nofit while a pronounced reduction of win-stay behavior was present in patients-nofit (see Supplemental Material, S-Figure 1B-D).

Comparison of parameters (Table 3, Figure 2B and 2C) revealed that $\mu_3^{(0)}$ was higher in patients (z=3.21, p<0.01). This prior belief about environmental volatility also had a stronger influence on lower-level updating as κ was enhanced in patients (z=2.51, p<0.01). Learning about environmental volatility ϑ did not differ between groups (z=.73, p=.47). Regarding the decision model, choice repetition after rewards did not differ between groups (z=.88, p=.38) while staying after losses was significantly lower in patients (z=2.27, p=.03) but did not survive Bonferroni-correction.
Notably, between-group findings on behavioral data were reproduced when running the task based on the inferred parameters, which represents an important validation of the model's ability to capture the observed behavioral data (see supplemental Material).

Replication in unmedicated patients. In an independent sample of unmedicated patients (n=24) and HC (n=24), who performed a comparable reversal learning task (16), we replicated between-group findings on learning parameters when fitting the same variant of HGF (HGF-DU-V). For statistics, see Table 3. This replication remained significant when excluding participants not fit better than chance (23 HC, 13 patients).

Relationship with symptom dimensions. We explored the relationship of the two learning parameters that differed between groups with measures of cognition (6 measures) and psychopathology (7 measures) within patients (Table 1). Regarding cognition, we found negative correlations between the initial estimate of the third-level environmental volatility

$\mu_3^{(0)}$ and measures of cognitive flexibility surviving Bonferroni-correction (p<.0083, TMT B, r=-.56, p=<.001; DSST, r=-.56, p<.001, S-Table 2, S-Figure 3). No such correlations reached significance between κ and cognition nor between the two computational parameters and any of the seven symptom scales. The strongest, albeit non-significant, correlations were observed between κ and severity of delusions (r=.30, p =.10) and between $\mu_3^{(0)}$ and anhedonia (r=.32, p =.09). For an overview of all explorative correlations, see Supplemental Material (S-Table 2).

FMRI – task effects (pooled across groups). Activity related to ε_2 (a precision-weighted RPE) peaked in bilateral ventral striatum and ventromedial PFC among other regions (p-$FWE_{wholebrain}$<.05, Figure 3A, S-Table 3) and also in midbrain (p-$FWE_{midbrain\text{-}voi}$<.05, S-Table 9), a well-known network associated with RPEs. In a conjunction analysis, there was large overlap between ε_2 and the RPE from RL (S-Table 8; S-Figure 4). In contrast, third-level precision-weighted RPE (ε_3) was associated with activation in prefrontal and parietal regions as well as in the left insula (Figure 3A, S-Table 4).

BOLD response related to second-level precision (ψ_2) was found in bilateral insula, dorsolateral prefrontal cortex (dlPFC), cingulate cortex, parietal cortex, caudate and thalamus (Figure 3B, S-Table 5). Interestingly, BOLD signal associated with the third-level precision (ψ_3) estimate was largely overlapping with that of the second-level precision ψ_2 (Figure 3B, S-Table 6).

Environmental volatility (μ_3) co-varied with BOLD signal in similar areas to precision estimates ψ_2 and ψ_3, i.e. bilateral insula, cingulate cortex, parietal cortex and thalamus (S-Figure 5). However, there was additional activation associated with μ_3 located in prefrontal regions in the superior, middle and inferior frontal gyrus and in middle temporal gyrus and globus pallidus (Figure 4A, S-Table 7, S-Figure 5).

FMRI – between-group effects. For environmental volatility μ_3, a group difference between HC and patients was found in the right dlPFC (F-contrast, corrected for main effect of μ_3 over all participants, [x=34, y=44, z=24], F=19.89, z=4.24, p_{FWE}=0.038, Figure 4B). Post-hoc analysis revealed stronger activity related to volatility in dlPFC of patients compared to HC (t=4.46, z=4.4, p_{FWE}=0.019, Figure 4C). There was no other significant difference between the groups for any other regressors.

Discussion

To the best of our knowledge, this is the first study to apply hierarchical Bayesian learning to choice and fMRI data of schizophrenia patients during reward-based decision-making. We present two core findings: First, medicated patients both overestimated their prior higher-level belief about the volatility of the environment and exhibited an increased influence of volatility estimates on lower-level learning of action-outcome contingencies. This provides a computational explanation of the increased switching behavior seen in patients with schizophrenia, both in this and in previous studies (8, 11-16). We replicate this finding in an independent cohort of unmedicated patients. Second, medicated patients displayed higher dlPFC activity related to beliefs about environmental volatility. This points towards a prominent role of this region in promoting instable behavior in schizophrenia.

Replicating previous studies (8, 11-16), patients were impaired in flexible decision-making driven by enhanced switching between choice options. Using computational modeling of hierarchical Bayesian learning, we demonstrate a possible underlying learning mechanism: an enhanced prior belief about the environment being volatile combined and a stronger influence of volatility on lower-level learning of action-outcome contingencies. In consequence, patients' beliefs about which choice was currently being rewarded fluctuated more strongly and led to increased choice switching. This is compatible with a deficient representation of the expected value of choices in schizophrenia (6). Importantly, the inferred computational parameters reproduced the observed data well. Furthermore, we replicated these findings in an independent sample of unmedicated patients suggesting elevated beliefs about environmental volatility as core mechanism of impaired flexible decision-making in schizophrenia independent of medication status. Consistently, administration of ketamine, as a model of psychosis in HC, related to an inability to stabilize behavior according to an internal model of action-outcome contingencies, although the modeling approach was different (51). Enhanced

estimation of changing context probabilities was also found in schizophrenia in the non-reward domain (52). Our presented mechanism destabilizes behavior in the face of probabilistic (negative) outcomes and patients who performed the task according to the model indeed showed increased switching behavior primarily after punishments. Thus, patients were more likely to infer negative events as being indicative of changes in action-outcome contingencies, despite the possibility of probabilistically misleading negative feedback. In a task with negative feedback being deterministically informative about changes in environmental contingencies, our model predicts improved performance in patients.

On the neural level, we found a stronger association of trial-by-trial beliefs about environmental volatility with BOLD response in dlPFC of patients. This result is in line with the suggestion of instable prefrontal representations that along with (hyper-) flexible updating of information impede the detection of regularities in probabilistic environments and therefore render the system (hyper-) sensitive for any new input (53, 54). In the same vein, the decision model function of our winning model implemented a direct link between the beliefs about environmental volatility and choice stochasticity. This interpretation is also supported by a negative correlation of this prior belief about environmental volatility with cognitive measures requiring cognitive flexibility and speed in terms of accuracy under time pressure. When performing well-known cognitive tasks during fMRI, meta-analyses showed reduced prefrontal activation compared to controls when contrasting ('relevant') task conditions to ('irrelevant') control conditions (55, 56) likely to reflect the same mechanism.

No differences were observed for other learning signatures such as lower- and higher-level precision weights and precision-weighted PEs from the HGF as well as RPE from RL. While precision-weighted RPEs have not been investigated in schizophrenia so far, the results regarding RPE activity in schizophrenia are mixed. In medicated patients no significantly different RPE BOLD signal was observed in two recent studies (13, 14) in

line with our current finding. On the other hand, in unmedicated patients striatal RPE activity was reduced (16), which suggests an effect of antipsychotic medication similar as for reward anticipation (36, 57, 58).

As our model revealed consistent results in medicated und unmedicated patients, one may be tempted to assume similar neurochemical processes in terms of reduced NMDA receptor functioning (27, 59) leading to dysfunctional cortico-cortical information processing (54). In unmedicated patients, this was proposed to result in increased striatal dopamine turnover (27, 59). Enhanced (and spontaneous) phasic dopamine transients could lead to exaggerated lower-level belief updates and switching behavior (23), while enhanced tonic stimulation of dopamine D2 receptors could support the assignment of 'high' precision to a higher-order belief by blocking inputs to the PFC (60, 61). The latter reasoning is in line with the clinical persistence of positive symptoms such as delusional ideation. While antipsychotic medication is helpful in treating these symptoms via blockade of primarily striatal D2 receptors, it also disinhibit inputs to the prefrontal cortex after negative outcomes leading to instable or somewhat (hyper-) flexible behavior. While we emphasize the speculative nature of these thoughts, one may presume a common higher-level dysfunction associated with experiencing the world as instable but different lower-level processes leading to distinct aberrant experiences and behaviors. This also suggests (dys-)connectivity between prefrontal and other cortical and subcortical regions, in particular midbrain and striatum, to modulate learning signals (27). In-vivo measures of dopamine and glutamate (62) are promising candidates to be closely linked to these computations.

An important limitation of the present study is the lack of any longitudinal clinical data. Future research should focus on trans-diagnostic computational studies or meta-analyses to test whether the presented mechanism of switching behavior represents a potential computational phenotype specific for schizophrenia and the psychosis

spectrum or whether this characterizes any group of individuals exhibiting instable behaviors. In this regard, one may also refine the direct link between environmental volatility and decision noise applied in our model.

In summary, we present a computational mechanism putatively underlying instable behavior in schizophrenia during flexible reward-based decision-making: an increased higher-order belief about the environment being volatile combined and a stronger influence of volatility on lower-level learning, which was present in medicated und unmedicated patients. In medicated patients, this was accompanied by enhanced activity related to environmental volatility in dlPFC, suggesting this region in mediating the instable behavior seen in patients. Such a detailed computational understanding of cognitive and motivational functioning in schizophrenia could in the long run provide clinically useful information to enable differential diagnostics, e.g. the identification of subgroups of patients with different outcome (63), but could also enable differential indications, e.g. in terms of treatment responses to antipsychotic drugs. These further steps can be accomplished within the Computational Psychiatry (60, 64-67) framework applied in the present study.

Author contributions. LD, CM, KES & FS designed the study. LD, RB, TK & JK performed research. LD & RB analyzed the data, CM & FS supervised data analysis. LD & RB wrote the initial version of the manuscript. LD, RB, TK, JK, CM, KES, AH, FS read and corrected versions of the manuscript.

Acknowledgements. This study was supported by the Max Planck Society and grants from the German Research Foundation awarded to FS (DFG SCHL1969/1-2, DFG SCHL 1969/2-2, SCHL1969/4-1). KES acknowledges support by the René and Susanne Braginsky Foundation and the University of Zurich.

Financial Disclosures. All authors report no biomedical financial interests or potential conflicts of interest.

References

1. Bowie CR, Leung WW, Reichenberg A, McClure MM, Patterson TL, Heaton RK, et al. (2008): Predicting schizophrenia patients' real-world behavior with specific neuropsychological and functional capacity measures. *Biol Psychiatry*. 63:505-511.
2. Green MF (1996): What are the functional consequences of neurocognitive deficits in schizophrenia? *Am J Psychiatry*. 153:321-330.
3. Nuechterlein KH, Subotnik KL, Green MF, Ventura J, Asarnow RF, Gitlin MJ, et al. (2011): Neurocognitive predictors of work outcome in recent-onset schizophrenia. *Schizophr Bull*. 37 Suppl 2:S33-40.
4. Gonzalez-Ortega I, de Los Mozos V, Echeburua E, Mezo M, Besga A, Ruiz de Azua S, et al. (2013): Working memory as a predictor of negative symptoms and functional outcome in first episode psychosis. *Psychiatry Res*. 206:8-16.
5. Dominguez Mde G, Viechtbauer W, Simons CJ, van Os J, Krabbendam L (2009): Are psychotic psychopathology and neurocognition orthogonal? A systematic review of their associations. *Psychological bulletin*. 135:157-171.
6. Gold JM, Waltz JA, Prentice KJ, Morris SE, Heerey EA (2008): Reward processing in schizophrenia: a deficit in the representation of value. *Schizophr Bull*. 34:835-847.
7. Barch DM, Dowd EC (2010): Goal representations and motivational drive in schizophrenia: the role of prefrontal-striatal interactions. *Schizophr Bull*. 36:919-934.
8. Murray GK, Cheng F, Clark L, Barnett JH, Blackwell AD, Fletcher PC, et al. (2008): Reinforcement and reversal learning in first-episode psychosis. *Schizophr Bull*. 34:848-855.
9. Leeson VC, Robbins TW, Matheson E, Hutton SB, Ron MA, Barnes TR, et al. (2009): Discrimination learning, reversal, and set-shifting in first-episode schizophrenia: stability over six years and specific associations with medication type and disorganization syndrome. *Biol Psychiatry*. 66:586-593.
10. Ragland JD, Cohen NJ, Cools R, Frank MJ, Hannula DE, Ranganath C (2012): CNTRICS imaging biomarkers final task selection: Long-term memory and reinforcement learning. *Schizophr Bull*. 38:62-72.
11. Waltz JA, Gold JM (2007): Probabilistic reversal learning impairments in schizophrenia: further evidence of orbitofrontal dysfunction. *Schizophr Res*. 93:296-303.
12. Waltz JA, Kasanova Z, Ross TJ, Salmeron BJ, McMahon RP, Gold JM, et al. (2013): The roles of reward, default, and executive control networks in set-shifting impairments in schizophrenia. *PLoS One*. 8:e57257.
13. Culbreth AJ, Gold JM, Cools R, Barch DM (2016): Impaired Activation in Cognitive Control Regions Predicts Reversal Learning in Schizophrenia. *Schizophr Bull*. 42:484-493.
14. Culbreth AJ, Westbrook A, Xu Z, Barch DM, Waltz JA (2016): Intact Ventral Striatal Prediction Error Signaling in Medicated Schizophrenia Patients. *Biological psychiatry : cognitive neuroscience and neuroimaging*. 1:474-483.
15. Reddy LF, Waltz JA, Green MF, Wynn JK, Horan WP (2016): Probabilistic Reversal Learning in Schizophrenia: Stability of Deficits and Potential Causal Mechanisms. *Schizophr Bull*. 42:942-951.
16. Schlagenhauf F, Huys QJ, Deserno L, Rapp MA, Beck A, Heinze HJ, et al. (2014): Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. *Neuroimage*. 89:171-180.
17. Deserno L, Boehme R, Heinz A, Schlagenhauf F (2013): Reinforcement learning and dopamine in schizophrenia: dimensions of symptoms or specific features of a disease group? *Front Psychiatry*. 4:172.
18. Sutton RS, Barto AG (1998): *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.
19. Schultz W, Dayan P, Montague PR (1997): A neural substrate of prediction and reward. *Science*. 275:1593-1599.
20. Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH (2013): A causal link between prediction errors, dopamine neurons and learning. *Nat Neurosci*. 16:966-973.
21. Howes OD, Kambeitz J, Kim E, Stahl D, Slifstein M, Abi-Dargham A, et al. (2012): The nature of dopamine dysfunction in schizophrenia and what this means for treatment. *Arch Gen Psychiatry*. 69:776-786.
22. Heinz A (2002): Dopaminergic dysfunction in alcoholism and schizophrenia--psychopathological and behavioral correlates. *Eur Psychiatry*. 17:9-16.
23. Maia TV, Frank MJ (2017): An Integrative Perspective on the Role of Dopamine in Schizophrenia. *Biol Psychiatry*. 81:52-66.

24. Deserno L, Heinz A, Schlagenhauf F (2017): Computational approaches to schizophrenia: A perspective on negative symptoms. *Schizophr Res*. 186:46-54.
25. Rao RP, Ballard DH (1999): Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci*. 2:79-87.
26. Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ (2013): The computational anatomy of psychosis. *Front Psychiatry*. 4:47.
27. Stephan KE, Friston KJ, Frith CD (2009): Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. *Schizophr Bull*. 35:509-527.
28. Corlett PR, Frith CD, Fletcher PC (2009): From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology (Berl)*. 206:515-530.
29. Sterzer P, Mishara AL, Voss M, Heinz A (2016): Thought Insertion as a Self-Disturbance: An Integration of Predictive Coding and Phenomenological Approaches. *Front Hum Neurosci*. 10:502.
30. Fletcher PC, Frith CD (2009): Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat Rev Neurosci*. 10:48-58.
31. Powers AR, Mathys C, Corlett PR (2017): Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science*. 357:596-600.
32. Rushworth MF, Behrens TE (2008): Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat Neurosci*. 11:389-397.
33. Juckel G, Schlagenhauf F, Koslowski M, Wustenberg T, Villringer A, Knutson B, et al. (2006): Dysfunction of ventral striatal reward prediction in schizophrenia. *Neuroimage*. 29:409-416.
34. Schlagenhauf F, Sterzer P, Schmack K, Ballmaier M, Rapp M, Wrase J, et al. (2009): Reward feedback alterations in unmedicated schizophrenia patients: relevance for delusions. *Biol Psychiatry*. 65:1032-1039.
35. Radua J, Schmidt A, Borgwardt S, et al. (2015): Ventral striatal activation during reward processing in psychosis: A neurofunctional meta-analysis. *JAMA psychiatry*.1243-1251.
36. Nielsen MO, Rostrup E, Wulff S, Bak N, Broberg BV, Lublin H, et al. (2012): Improvement of brain reward abnormalities by antipsychotic monotherapy in schizophrenia. *Arch Gen Psychiatry*. 69:1195-1204.
37. Dowd EC, Frank MJ, Collins A, Gold JM, Barch DM (2016): Probabilistic Reinforcement Learning in Patients With Schizophrenia: Relationships to Anhedonia and Avolition. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 1:460-473.
38. Behrens TE, Woolrich MW, Walton ME, Rushworth MF (2007): Learning the value of information in an uncertain world. *Nat Neurosci*. 10:1214-1221.
39. Iglesias S, Mathys C, Brodersen KH, Kasper L, Piccirelli M, den Ouden HE, et al. (2013): Hierarchical Prediction Errors in Midbrain and Basal Forebrain during Sensory Learning. *Neuron*. 80:519-530.
40. Rushworth MF, Noonan MP, Boorman ED, Walton ME, Behrens TE (2011): Frontal cortex and reward-guided learning and decision-making. *Neuron*. 70:1054-1069.
41. Mathys C, Daunizeau J, Friston KJ, Stephan KE (2011): A bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci*. 5:39.
42. Mathys CD, Lomakina EI, Daunizeau J, Iglesias S, Brodersen KH, Friston KJ, et al. (2014): Uncertainty in perception and the Hierarchical Gaussian Filter. *Front Hum Neurosci*. 8:825.
43. Wechsler D (1955): *Wechsler Adult Intelligence Scale Manual*. New York: Psychological Corporation.
44. Gardner DM, Murphy AL, O'Donnell H, Centorrino F, Baldessarini RJ (2010): International consensus study of antipsychotic dosing. *Am J Psychiatry*. 167:686-693.
45. Boehme R, Deserno L, Gleich T, Katthagen T, Pankow A, Behr J, et al. (2015): Aberrant Salience Is Related to Reduced Reinforcement Learning Signals and Elevated Dopamine Synthesis Capacity in Healthy Adults. *J Neurosci*. 35:10103-10111.
46. Reiter AM, Heinze HJ, Schlagenhauf F, Deserno L (2017): Impaired Flexible Reward-Based Decision-Making in Binge Eating Disorder: Evidence from Computational Modeling and Functional Neuroimaging. *Neuropsychopharmacology*.
47. Reiter AMF, Deserno L, Kallert T, Heinze HJ, Heinz A, Schlagenhauf F (2016): Behavioral and neural signatures of reduced updating of alternative options in alcohol-dependent patients during flexible decision-making. *J Neurosci*.
48. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009): Bayesian model selection for group studies. *Neuroimage*. 46:1004-1017.

49. Rigoux L, Stephan KE, Friston KJ, Daunizeau J (2014): Bayesian model selection for group studies - revisited. *Neuroimage*. 84:971-985.
50. Stephan KE, Schlagenhauf F, Huys QJ, Raman S, Aponte EA, Brodersen KH, et al. (2017): Computational neuroimaging strategies for single patient predictions. *Neuroimage*. 145:180-199.
51. Vinckier F, Gaillard R, Palminteri S, Rigoux L, Salvador A, Fornito A, et al. (2015): Confidence and psychosis: a neuro-computational account of contingency learning disruption by NMDA blockade. *Molecular psychiatry*.
52. Kaplan CM, Saha D, Molina JL, Hockeimer WD, Postell EM, Apud JA, et al. (2016): Estimating changing contexts in schizophrenia. *Brain*. 139:2082-2095.
53. Durstewitz D, Seamans JK (2008): The dual-state theory of prefrontal cortex dopamine function with relevance to catechol-o-methyltransferase genotypes and schizophrenia. *Biol Psychiatry*. 64:739-749.
54. Lewis DA, Gonzalez-Burgos G (2006): Pathophysiologically based treatment interventions in schizophrenia. *Nat Med*. 12:1016-1022.
55. McTeague LM, Huemer J, Carreon DM, Jiang Y, Eickhoff SB, Etkin A (2017): Identification of Common Neural Circuit Disruptions in Cognitive Control Across Psychiatric Disorders. *Am J Psychiatry*.appiajp201716040400.
56. Glahn DC, Ragland JD, Abramoff A, Barrett J, Laird AR, Bearden CE, et al. (2005): Beyond hypofrontality: a quantitative meta-analysis of functional neuroimaging studies of working memory in schizophrenia. *Hum Brain Mapp*. 25:60-69.
57. Radua J, Schmidt A, Borgwardt S, Heinz A, Schlagenhauf F, McGuire P, et al. (2015): Ventral Striatal Activation During Reward Processing in Psychosis: A Neurofunctional Meta-Analysis. *JAMA Psychiatry*.1243-1251.
58. Schlagenhauf F, Juckel G, Koslowski M, Kahnt T, Knutson B, Dembler T, et al. (2008): Reward system activation in schizophrenic patients switched from typical neuroleptics to olanzapine. *Psychopharmacology (Berl)*. 196:673-684.
59. Krystal JH, Perry EB, Jr., Gueorguieva R, Belger A, Madonick SH, Abi-Dargham A, et al. (2005): Comparative and interactive human psychopharmacologic effects of ketamine and amphetamine: implications for glutamatergic and dopaminergic model psychoses and cognitive function. *Arch Gen Psychiatry*. 62:985-994.
60. Adams RA, Huys QJ, Roiser JP (2016): Computational Psychiatry: towards a mathematically informed understanding of mental illness. *J Neurol Neurosurg Psychiatry*. 87:53-63.
61. Frank MJ, O'Reilly RC (2006): A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol. *Behav Neurosci*. 120:497-517.
62. Gleich T, Deserno L, Lorenz RC, Boehme R, Pankow A, Buchert R, et al. (2015): Prefrontal and Striatal Glutamate Differently Relate to Striatal Dopamine: Potential Regulatory Mechanisms of Striatal Presynaptic Dopamine Function? *J Neurosci*. 35:9615-9621.
63. Brodersen KH, Deserno L, Schlagenhauf F, Lin Z, Penny WD, Buhmann JM, et al. (2014): Dissecting psychiatric spectrum disorders by generative embedding. *NeuroImage Clinical*. 4:98-111.
64. Huys QJ, Maia TV, Frank MJ (2016): Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat Neurosci*. 19:404-413.
65. Stephan KE, Mathys C (2014): Computational approaches to psychiatry. *Curr Opin Neurobiol*. 25:85-92.
66. Wang XJ, Krystal JH (2014): Computational Psychiatry. *Neuron*. 84:638-654.
67. Heinz A (2017): *A New Understanding of Mental Disorders - Computational Models for Dimensional Psychiatry*. MIT Press

Table 1 Sample Characteristics. Group means with standard deviations and range in brackets are reported; for group comparisons two-sample t-test or x^2 tests were used. *=Bonferoni-corrected

	Schizophrenia patients (N=46)	*Healthy controls (N=43)*	Sig.
Age (years) (46/43)	35.07 ± 7.596 (22-52)	34.40 ± 8.24 (22-51)	.69
Gender (46/43)	14 female / 32 male	13 female / 30 male	.98
Handedness (44/40)	38r 4l 2b	32r 2l 6b	.22
EHI (44/40)	70.71 ± 53.13 (-100-100)	71.68 ± 47.55 (-100-100)	.93
Smoking (44/39)	35 smoking, 9 non-smoking	15 smoking, 24 non-smoking	<.01
Cognitive measures			
Verbal Intelligence (45/42)	98.22 ± 11.83 (70-115)	103.81 ± 8.03 (80-115)	.01
Word list (46/42)	8.72 ± 1.60 (4-10)	9.26 ± 1.38 (4-10)	.09
Working memory (46/42)	6.46 ± 2.14 (4-13)	7.48 ± 2.36 (3-14)	.04
TMT A (45/41)	36.67 ± 14.69 (16-80)	26.12 ± 8.98 (10-50)	<.01*
TMT B (45/41)	87.50 ± 50.49 (37-304)	59.44 ± 21.51 (29-122)	<.01*
Processing speed (46/42)	63.70 ± 19.10 (29-111)	76.69 ± 11.32 (49-99)	<.01*
Clinical measures			
PANSS Positive (46)	20.83 ± 6.83 (7-39)	/	/
PANSS Negative (46)	22.17 ± 7.64 (8-38)	/	/
PANSS GP (46)	40.50 ± 11.41 (16-62)	/	/
PANSS Total (46)	83.57 ± 22.31 (33-134)	/	/
SAPS Delusion (45)	21.18 ± 10.70(0-48)	/	/
SANS Anhedonia (45)	12.11 ± 4.97 (4-21)	/	/
CPZ (44) in mg	380.27 ± 194.11 (33.25-1020)	/	/

Table 2 Bayesian Model Selection. Exceedance probabilities (PX) describe the probability of a model to exceed all other models in the comparison set, the probability that expected posterior probabilities (PP) differ. Protected exceedance probabilities (PXP) protect against the risk that differences in model evidences occur by chance; HC = healthy controls, SZ = schizophrenia patients, RL = Reinforcement Learning, HGFX = Hierarchical Gaussian Filter with 2 or 3 levels, SU = single update, DU = double update, HGF3-**-V = three-level HGF with environmental volatility linked to decision noise with either SU or DU.

		RL-SU	RL-DU	HGF2-SU	HGF2-DU	HGF3-SU	HGF3-DU	HGF3-SU-V	HGF3-DU-V
HC+SZ *all n=89*	*PP*	14.5	6.1	3.5	7.8	3.8	8.8	5.4	50.1
	XP	0	0	0	0	0	0	0	100
	PXP	0	0	0	0	0	0	0	**99.8**
HC *all n=43*	*PP*	5.7	5.9	3.0	5.9	3.0	6.1	6.1	64.4
	XP	0	0	0	0	0	0	0	100
	PXP	0	0	0	0	0	0	0	**100**
SZ *all n=46*	*PP*	17.23	8.7	10.0	11.9	11.7	12.9	5.0	22.6
	XP	21.0	.9	1.8	4.0	3.7	5.9	.1	62.6
	PXP	12.5	12.5	12.5	12.5	12.5	12.5	12.5	**12.6**
HC+SZ *fit n=73*	*PP*	14.5	6.1	3.5	7.8	3.8	8.8	5.4	50.1
	XP	0	0	0	0	0	0	0	100
	PXP	0	0	0	0	0	0	0	**99.8**
HC *fit n=42*	*PP*	5.9	5.7	3.1	6.1	3.0	6.4	6.2	63.7
	XP	0	0	0	0	0	0	0	100
	PXP	0	0	0	0	0	0	0	**100**
SZ *fit n=31*	*PP*	16.7	10.2	9.6	11.4	11.1	12.3	5.1	23.6
	XP	16.9	1.8	1.3	3.0	2.6	4.2	.1	70.1
	PXP	12.5	12.5	12.5	12.5	12.5	12.5	12.5	**12.6**

Table 3. Between-group comparisons of model parameters using t-tests or the non-parametric Mann-Whitney-U-Test if assumptions of normality were violated. Bonferroni-correction was applied according to the number of parameters (5, p<0.01). HC = healthy controls, SZ = schizophrenia patients.

	HC (n=42)	*SZ (n=31)*	*Test statistic*
Learning model			
μ_3	-.84 ± .49	-.47 ± .48	z=3.21, p<.01
κ	.87± .60	1.35 ± 1.12	z=2.51, p<.01
υ	-6.00 ± .02	-5.99 ± .05	z=.73, p=.47
Decision model			
ρ_{rew}	.97 ± .57	1.08 ± .49	z=.88, p=.38
ρ_{pun}	.08 ± 0.31	-.12 ± .44	z=2.27, p=.03
Replication sample: Schlagenhauf et al. 2014			
	HC (n=24)	unmedicated SZ (n=24)	
Learning model			
μ_3	-1.17 ± .61	-.43 ± .70	z=3.88, p<.01
κ	.61 ± .58	1.56 ± 1.21	z=3.46, p<.01
υ	-6.09 ± .02	-5.99 ± .06	z=1.35, p=.18
Decision model			
ρ_{rew}	.70 ± .74	.43 ± .66	z=1.35, p=.18
ρ_{pun}	.-07 ± 0.59	-.28 ± .77	z=1.05, p=.30

Figure 1: A. Trial sequence from the decision-making task. B. Reward probabilities of both choice options were perfectly anti-correlated and were stable for the first 55 trials ('pre-reversal'), changed four times, after 15 or 20 trials, in the 'reversal' phase and remained stable for the last 35 trials ('post-reversal'). C. Percent choices of the stimulus with 80% reward probability was significantly lower in patients (main effect of group F=14.52, p<.001). D. Patients were less likely to repeat the previous action independent of feedback received in the previous trial (main effect of group F=27.77, p<.001, feedback x group interaction F=.02, p=.89).

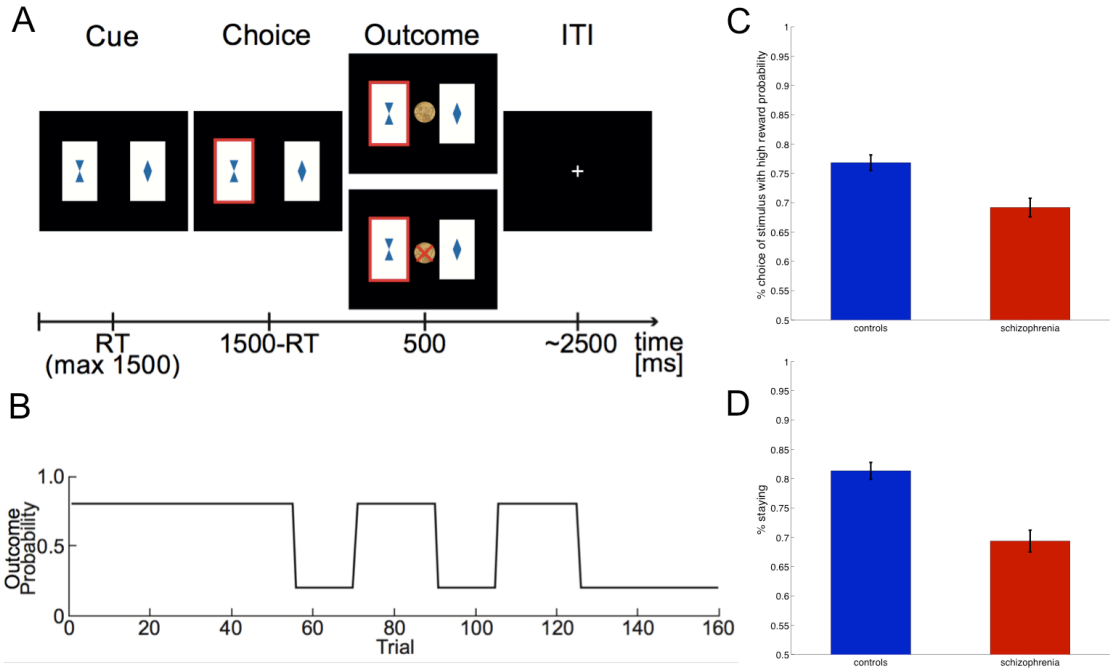


Figure 2. A) Graphical illustration. The HGF deploys hierarchically organized states x_i, in which learning about environmental volatility at a higher-level state x_3 determines lower-level learning about reward probabilities x_2. The constant step size ϑ reflects learning about environmental volatility. The constant ω was fixed across all individuals and κ reflects inter-individual differences in the coupling between the levels. The best-fitting model was a three-level implementation (HGF3-DU-V) with double-updating (not illustrated) together with a decision model capturing choice repetition separately after rewards and punishments (ρ), third-level environmental volatility linked to decision noise $\beta = \exp(-\mu_3^{(k)})$ and the prior belief about environmental volatility $\mu_3^{(0)}$ as an additional parameter inferred from the data. B) The initial estimate of environmental volatility is higher in patients. C) The coupling between the third level (environmental volatility) and the second level is stronger in schizophrenia.

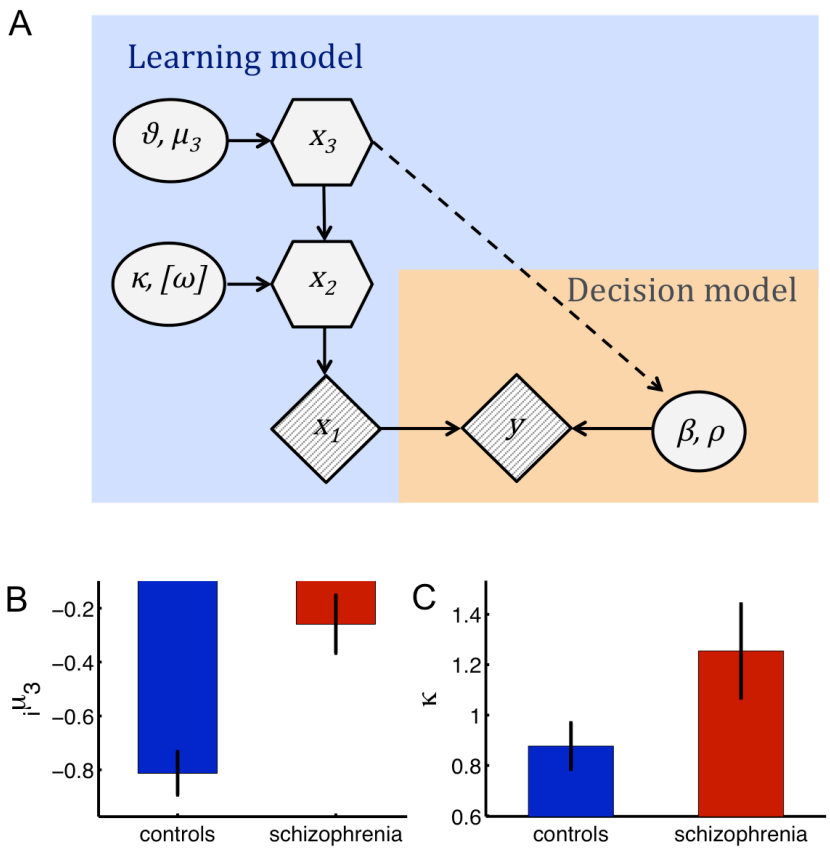


Figure 3: BOLD signal across all participants related to A) precision-weighted prediction errors from the second (red) and third (blue) level and (B) to precision weight from the second (red) and third (green) level, overlap in yellow. Both at p-FWE$_{wholebrain}$<0.05, k=10.

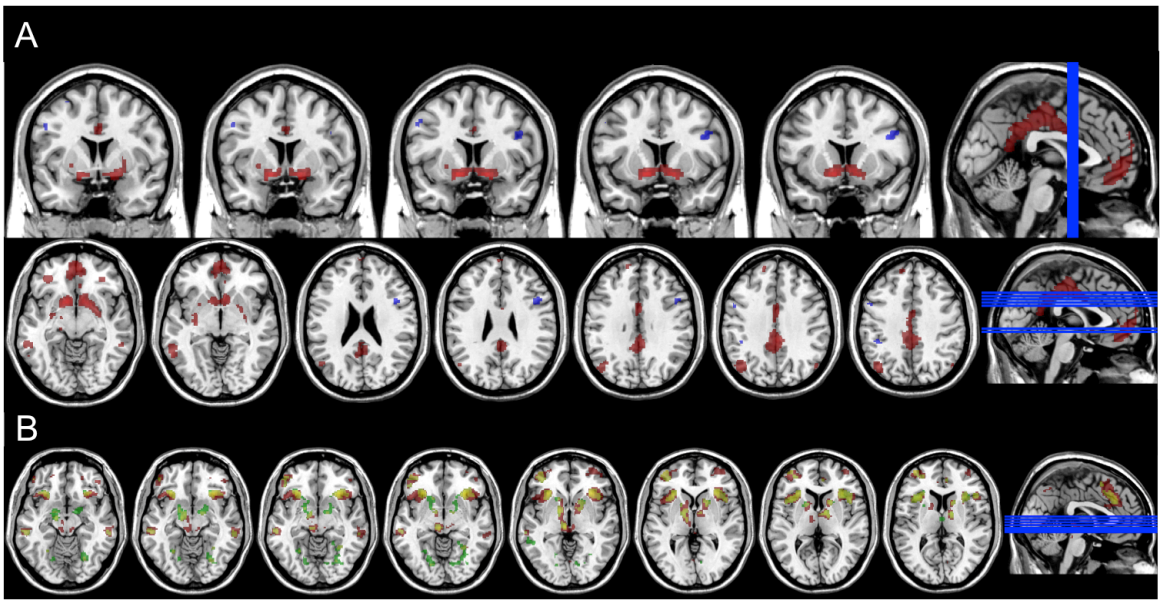
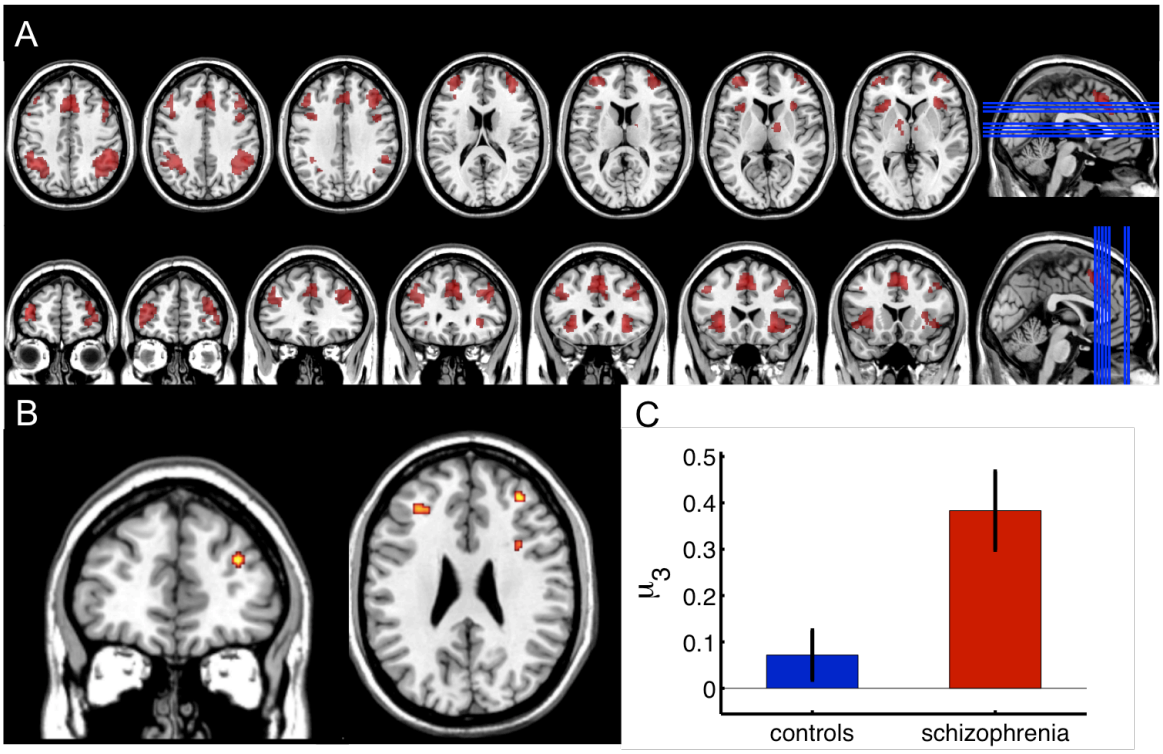


Figure 4: A) Across all participants, BOLD signal related to volatility estimate μ_3 from the third level at p-FWE<0.05 for the whole brain, k=10. B) μ_3-related BOLD signal in the dorsolateral prefrontal cortex (dlPFC) differs between in patients and controls (F-contrast displayed at p<.001 uncorrected; corrected for main effect of μ_3 over all participants, [x=34, y=44, z=24], F=19.89, z=4.24, p=0.038). C) A post-hoc analysis of parameter estimates at the peak of the group difference [x=34, y=44, z=24] showed that this was driven by heightened activity related to environmental volatility in dlPFC of patients compared to controls (t=4.46, z=4.4, p=0.019).



Overestimating environmental volatility increases switching behavior and is linked to activation of dorsolateral prefrontal cortex in schizophrenia

Short title: Volatility, choice switching and PFC in schizophrenia

Supplemental Material

L. Deserno[1,2,3*], R. Boehme[1,4*], C. Mathys[5,6,7,8], T. Katthagen[1], J. Kaminski[1], K. E. Stephan[6,8], A. Heinz[1,10], F. Schlagenhauf[1,2]

[1]Department of Psychiatry and Psychotherapy, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health
[2]Max-Planck-Institute for Human Cognitive and Brain Sciences, Leipzig;
[3]Department of Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, University of Leipzig, Leipzig;
[4]Center for Social and Affective Neuroscience, Linköping University, Linköping
[5]Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy
[6]Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London;
[7]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, London;
[8]Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich & ETH Zurich, Switzerland;
[9] Berlin School of Mind and Brain,Humboldt-Universität zu Berlin, Berlin;
[10]Cluster of Excellence NeuroCure, Charite Universitätsmedizin, Berlin;

Correspondence:
Dr. Lorenz Deserno
Max Planck Institute for Human Cognitive and Brain Sciences
Department of Neurology
Stephanstraße 1A
04103 Leipzig
deserno@cbs.mpg.de

Supplemental Methods

Computational models of learning. In the case of the applied task, both HGF and RL provide different ways to learn expectations about rewards. In the main text, we describe how both algorithms update expectations of the chosen card c only ("single-update", SU), which implies that there is no update of the expectation about the unchosen card:

$$(1)\ Q_{uc}^{(k+1)} = Q_{uc}^{(k)}$$

Correspondingly, for the HGF:

$$(2)\ \mu_{i=2,uc}^{(k)} = \mu_{i=2,uc}^{(k-1)}$$

Based on the anti-correlated task structure, one can implement a variant of each of the learning models updating the unchosen card simultaneously, i.e. an increase of the chosen card implies a decrease of the unchosen card uc ("double-update", DU), which can be written as:

$$(3)\ Q_{uc}^{(k)} = 1 - Q_{c}^{(k)}$$

$$(4)\ \mu_{i=2,uc}^{(k)} = 1 - \mu_{i=2,c}^{(k)}$$

Both SU and DU variants of each learning model were fit to the choice data.

Decision models. In binary choice tasks with correlated reward-probabilities, there is strong autocorrelation of choices (perseveration). In our decision model (equation 5, main manuscript), this is captured by estimating a parameter ρ, which changes the inflection point of the sigmoid function, thus, biasing towards an overall tendency to stay or switch irrespectively of the learned expectations, which we split up to ρ_{win} and ρ_{pun} to reflect differences in choice perseveration after rewards and punishment separately. This is in line with previous observations that choice perseveration differs strongly after rewards and punishment. Note that this implementation of capturing choice perseveration is equivalent to fitting parameters for rewards and punishments as part of the learning model (so-called reinforcement sensitivities) as in previous work (1, 2), which changes the asymptote of expectations in the learning model instead of the

inflection point of the sigmoid decision model. We name these decision models 'RP', in each of which the parameter β was fixed to 1 to avoid over-fitting. In addition, we also estimate all learning models with a decision model with no such repetition parameters but β as a parameter and with no repetition parameters ('BETA'). The parameter β captures the steepness of the sigmoid and thereby reflects (inverse) decision noise by determining how tightly choice probabilities follow learned reward expectations. The eight learning models with β capturing inverse decision noise provided overall inferior fit when performing a family comparison (RP XP_{all}=.9669, BETA XP_{all}=.0331; RP XP_{HC}=.8667, BETA XP_{HC} = .1333; RP XP_{SZ}=.9884, BETA XP_{SZ} = .0154).

Classification of subjects not fit better than chance. In choice tasks like ours, a subject can be classified as fit better than chance when the negative log-likelihood (-LL) of the model significantly exceeds .55 given by exp(-LL/ntrials) corresponding to $p<.05$ when performing a binomial test. This procedure was applied to all individuals to avoid the possibility that between-group differences in model parameters are confounded by differences in model fit (1, 3-6).

Functional Magnetic Resonance Imaging. Functional Magnetic Resonance Imaging (fMRI) imaging was performed using a 3 Tesla Siemens Trio scanner to acquire gradient echo T2*-weighted echo-planar images with blood oxygenation level dependent contrast. Covering the whole brain, 36 slices were acquired in oblique orientation at 20° to AC-PC line in ascending order with 2.5-mm thickness, 3x3mm² in-plane voxel resolution, 0.5-mm gap between slices, TR=2s, TE=22ms and a flip angle α=90°. Prior to functional scanning, a field map was collected to account for individual homogeneity differences of the magnetic field. T1-weighted structural images were also acquired (TR=1300ms, TE=3.46ms, flip=10°, matrix=240×256, voxel size: 1×1×1mm, slices=170).

Preprocessing of fMRI data. FMRI data were analyzed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). For preprocessing, images were

corrected for delay of slice time acquisition. Voxel-displacement maps were estimated based on field maps. All images were realigned to correct for motion and were also corrected for distortion and the interaction of distortion and motion. The images were spatially normalized into the Montreal Neurological Institute (MNI) space using the normalization parameters generated during the segmentation of each subject's anatomical T1 scan; spatial smoothing was applied with an isotropic Gaussian kernel of 6mm full width at half maximum. Prior to first-level statistical analysis, data were high-pass filtered with a cutoff of 128s.

Supplemental Results

Behavioral data in HC, SZ-fit and SZ-nofit. Repeating the analysis of win-stay and lose-stay behavior with three groups (HC-fit, SZ-fit, SZ-nofit) revealed that all participants were more likely to repeat the previous action after rewards versus losses (main effect of feedback F=371.91, p= 8.65e-33) and patients were overall less likely to switch independent of the feedback received in the previous trial (significant main effect of group F=32.99, p= 2.48e-11). However, as mentioned in the main manuscript, there was also a significant group x feedback interaction (F=20.68, p=4.81e-08) showing that reduced lose-stay behavior characterized both SZ-fit and SZ-nofit (S-Figure 1). A reduction of win-stay behavior was pronounced in SZ-nofit only (S-Figure 1). Performing the same analysis with two groups (HC-fit, SZ-fit; main effect of feedback F=636.30, p=3.60e-37; main effect of group F=10.22, p=.0021) confirmed a significant interaction between group x feedback interaction (F=6.79, p=.0111). Thus, in these patients, the tendency to switch was particularly driven by enhanced switching after negative feedback. This is in line and was even corroborated by computational modeling showing that these patients showed a reduced tendency to stay after punishments but not after rewards that remained significant beyond the differences in learning (see Results section in main manuscript). In addition, the latter differences in learning provide a mechanistic explanation for the switching behavior observed in this and previous studies. For discussion, see main manuscript.

Clinical distinction between subgroups. When further exploring the subgroups, SF-nofit patients did not differ from SF-fit patients in any of the seven measures of positive and negative symptoms (Table 1, all p's>.3910). Interestingly, SZ-nofit differed from SZ-fit patients in terms of certain domains of cognitive functioning such as verbal memory (word list p=.0329), working memory (digit span p=.0271) but not verbal IQ (p=.1024), attention (TMTA p=.5158) or cognitive speed (DSST p=.3091) but trend-wise with respect to cognitive flexibility (TMTB p=.0697). However, this analysis should be

considered exploratory as none of the effects reached correction for multiple comparisons for the 6 cognitive tests ($p<.0083$).

Simulation of choice data by inferred parameters. In addition to relative model selection and classification of subjects' fit better than chance, a further important procedure to assure that the inferred parameters capture the data in a meaningful manner is to simulate choice data based on the inferred parameters. Because modeling-derived parameters can only be meaningfully interpreted when they account for the observed data better than chance, this analysis was reduced to 42 HC-fit and 31 SZ-fit (10 simulations per subject). The SZ-fit patients overall switched more than HC (*stay:* HC_{obs} .8184 ± .0892, SZ_{obs} .7594 ± .0895, t=2.7850, p=.0069) and this was well reproduced by the inferred parameters (*stay:* HC_{sim} .8014 ± .0753, SZ_{sim} .7540 ± .0756, t=2.6559, p=.0098). When analyzing staying after rewards and punishments separately, as reported above, choices in SZ-fit were driven by a pronounced tendency to switch after punishment relative reward as compared to HC-fit as indicated by group x feedback interaction (F=6.79, p=.0111; *win-stay:* HC_{obs} .9419 ± .0539, SZ_{obs} .9128 ± .0565, t=2.2327, p=.0287; *lose-stay:* HC_{obs} .5800 ± .1489, SZ_{obs} .4675 ± .1598, t=3.0928, p=.0028) next to significant main effects of feedback (F=636.30, p=3.60e-37) and group (F=10.22, p=.0021). These results were fully reproduced by the inferred parameters (group x feedback interaction F=6.14, p=.0156; *win-stay:* HC_{sim} .9150 ± .0556, SZ_{sim} .8869 ± .0599, t=2.0659, p=.0425; *lose-stay:* HC_{obs} .6122 ± .1200, SZ_{obs} .5197 ± .1331, t=3.1052, p=.0027) next to a significant main effects of feedback (F=665.83, p=8.4011e-38) and group (F=9.8987, p=.0024).

Neural coding of reward prediction errors based on Reinforcement Learning. As in previous studies (1), we also explored coding of prediction errors using a regressor of reward prediction errors from the RL model. As previously, we found that BOLD signal covaried with RPEs in the ventral and dorsal striatum, posterior cingulate cortex,

dorsolateral and lateral PFC, parietal cortex, hippocampus, amygdala and the cerebellum (S-Table 8, S-Figure 3A). There was no significant difference between the groups. However, this signal largely overlapped with the precision-weighted PE at the second level of the HGF (Figure 3A, S-Table 3, S-Figure 3B), which represents, in the case of our tasks, a precision-weighted RPE.

References

1. Schlagenhauf F, Huys QJ, Deserno L, Rapp MA, Beck A, Heinze HJ, et al. (2014): Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. *Neuroimage*. 89:171-180.
2. Huys QJ, Pizzagalli DA, Bogdan R, Dayan P (2013): Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. *Biology of mood & anxiety disorders*. 3:12.
3. Reiter AM, Heinze HJ, Schlagenhauf F, Deserno L (2017): Impaired Flexible Reward-Based Decision-Making in Binge Eating Disorder: Evidence from Computational Modeling and Functional Neuroimaging. *Neuropsychopharmacology*.
4. Reiter AMF, Deserno L, Kallert T, Heinze HJ, Heinz A, Schlagenhauf F (2016): Behavioral and neural signatures of reduced updating of alternative options in alcohol-dependent patients during flexible decision-making. *J Neurosci*.
5. Huys QJ, Cools R, Golzer M, Friedel E, Heinz A, Dolan RJ, et al. (2011): Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS Comput Biol*. 7:e1002028.
6. Huys QJ, Eshel N, O'Nions E, Sheridan L, Dayan P, Roiser JP (2012): Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput Biol*. 8:e1002410.
7. Iglesias S, Mathys C, Brodersen KH, Kasper L, Piccirelli M, den Ouden HE, et al. (2013): Hierarchical Prediction Errors in Midbrain and Basal Forebrain during Sensory Learning. *Neuron*. 80:519-530.
8. Bunzeck N, Duzel E (2006): Absolute coding of stimulus novelty in the human substantia nigra/VTA. *Neuron*. 51:369-379.

S-Table 1. Prior means and variances of parameters in computational models**.** Repetition parameters of the decision model ρ_{win} and ρ_{pun} were estimated native space. Learning rate α in RL was estimated in logit-space. Parameters of HGF are all in log space.

	Prior Mean	Prior Variance
Decision Model for all learning models		
Softmax		
ρ_{rew}	1	1
ρ_{pun}	0	1
Learning models		
RL		
α	.5	1
HGF		
ω (HGF2+3)	-1.5	0
κ (HGF3)	1	1
ϑ (HGF3)	-6	1
$\mu_3^{(0)}$ (HGF3-V)	0	1

S-Table 2. Explorative correlations (Spearman's rank correlation coefficient) between model parameters that differed between groups and measures of cognitive functioning as well as symptom domains within patients fit better than chance by the model (n=31).
* indicates significance surviving Bonferoni-correction.

	$\mu_3^{(0)}$		κ	
	r	p	r	p
Cognitive measures				
Verbal Intelligence (31)	-.35	.06	-.26	.17
Word list (31)	-.18	.33	-.15	.43
Working memory (31)	-.22	.23	-.25	.18
TMT A (30)	-.46	.01	-.33	.08
TMT B (30)	-.56*	<.001*	-.37	.05
Processing speed (31)	.-56*	<.001*	-.22	.23
Clinical measures				
PANSS Positive (31)	.25	.17	.26	.16
PANSS Negative (31)	.24	.19	.06	.75
PANSS GP (31)	.28	.12	.30	.10
PANSS Total (31)	.31	.09	.21	.25
SAPS Delusion (30)	.16	.39	.30	.22
SANS Anhedonia (30)	.32	.09	.07	.72
CPZ in mg (30)	.11	.57	.31	.09

S-Table 3. ε2 related activation in all participants, p-FWE$_{wholebrain}$<0.05, k=10

Region	Cluster		Coordinates			T	p (FWE)
Medial Frontal Gyrus	1539	L	-2	52	-4	7.33	<0.001
			-16	64	8	6.54	
			0	66	18	6.30	
			0	-12	52	6.17	
			0	-18	54	6.03	
			0	-24	52	5.86	0.001
			-4	40	-18	5.54	0.003
			-2	62	28	5.44	0.005
Cingulate Gyrus		R	2	-36	36	8.22	<0.001
			2	2	34	6.97	
		L	-4	-18	44	6.78	
			-6	-46	36	6.64	
			-2	-8	42	5.84	0.001
			-10	-32	40	5.47	0.004
Posterior Cingulate		L	-6	-52	14	6.55	<0.001
			-2	-52	22	6.07	
		R	4	-52	22	6.11	
			6	-52	14	5.87	0.001
Paracentral Lobule		R	6	-34	50	6.07	<0.001
		L	-2	-30	48	6.16	
Precuneus		L	-10	-38	44	5.06	0.024
Nucleus accumbens	1441	L	-12	10	-10	10.94	<0.001
		R	14	8	-12	10.38	
			30	-12	6	7.11	
Putamen		R	26	-2	-8	6.33	
			24	4	4	5.35	0.007
		L	-30	-18	6	6.82	<0.001
			-30	-14	-4	6.35	
			-28	-6	10	6.20	
			-30	-8	-4	6.19	
Amygdala		L	-26	-6	-12	6.59	
		R	20	-8	-14	6.16	
		L	-16	-8	-12	5.88	
Hippocampus		L	-24	-18	-16	6.69	
			-32	-20	-18	5.96	
			-30	-10	-20	5.50	0.003
			-28	-28	-12	5.06	0.024
Anterior Cingulate	1138	R	6	30	-10	5.49	0.004
			6	34	-4	4.99	0.033
		L	-4	42	4	5.37	0.006
Medial Frontal Gyrus		R	4	50	-10	8.00	<0.001
			4	38	-16	7.86	

Middle Temporal Gyrus	549	L	-58	-54	-2	7.25	<0.001
			-62	-42	-10	6.05	
			-56	-44	-4	6.04	
			-50	-66	26	5.61	
			-46	-64	20	5.41	0.005
Superior Temporal Gyrus		L	-54	-62	12	5.05	0.026
Precuneus			-38	-74	42	6.49	<0.001
Inferior Parietal Lobule			-40	-70	38	7.04	
Middle Frontal Gyrus	214	L	-34	36	-10	6.87	<0.001
			-38	38	-10	6.60	
			-22	32	-16	6.34	
Inferior Frontal Gyrus		L	-46	34	-10	5.74	0.001
Middle Frontal Gyrus	158	L	-18	30	46	5.03	0.027
Superior Frontal Gyrus		L	-16	34	54	6.03	<0.001
			-14	42	46	5.93	
			-12	50	40	5.72	0.001
			-14	52	36	5.72	0.001
Middle Frontal Gyrus	101	R	34	36	-12	6.45	<0.001
Middle Temporal Gyrus	72	R	56	-2	-18	6.20	<0.001
Precentral Gyrus	63	R	16	-26	62	5.38	0.006
			24	-28	64	5.31	0.008
			28	-24	56	5.30	
			24	-24	58	5.19	0.014
			36	-28	62	4.97	0.036
Hippocampus	41	R	26	-16	-18	6.02	<0.001
Inferior Frontal Gyrus	35	L	-44	38	6	5.53	0.003
			-50	32	6	5.01	0.03
Middle Temporal Gyrus	31	R	56	-42	-10	6.09	<0.001
Inferior Temporal Gyrus		R	64	-12	-20	5.59	0.002
Inferior Parietal Lobule	21	R	50	-68	38	5.97	<0.001
Putamen	17	L	-26	6	-2	5.58	<0.001
Claustrum	16	L	-38	-2	2	5.97	<0.001
Inferior Temporal Gyrus	13	L	-56	-34	-16	5.58	0.002

S-Table 4. ε3 related activation in all participants, p-FWE$_{wholebrain}$<0.05, k=10

Region	Cluster		Coordinates			T	p (FWE)
Inferior Frontal Gyrus	84	R	44	12	26	6.44	<0.001
Inferior Parietal Lobule	28	L	-42	-38	40	5.69	0.001
Middle Frontal Gyrus	26	L	-30	2	60	5.48	0.004
Precuneus	20	L	-32	-50	50	5.17	0.015
Inferior Frontal Gyrus	19	L	-50	4	36	5.30	0.009
Precuneus	7	R	10	-72	44	5.30	0.009
Insula	5	L	-40	8	24	5.46	0.004
Superior Frontal Gyrus	4	L	-4	10	54	5.03	0.027
Precuneus	3	L	-30	-74	28	5.22	0.012
Putamen	2	L	-14	10	-12	4.96	0.036
Inferior Parietal Lobule	2	L	-40	-48	44	4.92	0.044

S-Table 5. Ψ2 related activation in all participants, p-FWE$_{wholebrain}$<0.05, k=10

Region	Cluster		Coordinates			T	p (FWE)
Cingulate Gyrus	564		6	28	38	7.77	<0.001
			10	18	42	7.32	
		L	-6	24	38	6.27	
Anterior Cingulate		R	4	38	28	5.34	0.007
Inferior Frontal Gyrus	348	L	-28	24	-6	7.87	<0.001
Claustrum			-30	20	8	6.61	
			-32	18	2	6.46	
Insula			-40	16	8	6.09	
Inferior Frontal Gyrus	332	R	30	24	-8	8.45	<0.001
Middle Frontal Gyrus	180	R	36	40	24	6.01	<0.001
			42	30	30	5.91	0.001
			36	24	34	5.57	0.002
			36	50	18	5.12	0.019
Inferior Parietal Lobule	125	R	52	-46	38	6.26	<0.001
			44	-36	40	5.51	0.003
Inferior Parietal Lobule	118	L	-42	-58	52	5.51	0.003
			-42	-52	44	5.45	0.004
			-46	-52	54	5.33	0.008
			-48	-50	50	5.07	0.024
Middle Frontal Gyrus	111	L	-38	56	12	6.65	<0.001
			-40	54	-2	5.85	0.001
			-46	48	-8	5.15	0.016
Cerebellum	109	L	-32	-58	-30	7.36	<0.001
Cerebellum	92	R	36	-62	-28	7.24	<0.001
			46	-66	-26	5.44	0.005
Middle Frontal Gyrus	86	L	-36	28	30	5.77	0.001
Inferior Frontal Gyrus	53	L	-36	4	32	6.01	<0.001
Middle Frontal Gyrus			-42	10	36	5.49	0.004
Red Nucleus	52	L	-6	-24	-4	7.47	<0.001
Globus Pallidus	41	L	-14	2	-2	6.03	<0.001

Thalamus			-12	-10	0	5.29	0.009
Inferior Parietal Lobule	40	L	-46	-38	42	6.32	<0.001
Thalamus	29	R	12	-6	8	5.68	0.001
Precentral Gyrus	24	R	44	2	38	5.21	0.013
Superior Frontal Gyrus	27	L	-6	12	54	5.78	0.001
Middle Frontal Gyrus			-34	-2	48	5.67	0.002
Cingulate Gyrus			-4	14	46	5.01	0.030
Caudate	20	R	14	12	8	5.80	0.001
Inferior Parietal Gyrus	14	L	-32	-48	38	5.2	0.013
Inferior Frontal Gyrus	11	R	48	22	10	5.53	0.003
Middle Temporal Gyrus	11	L	-56	-30	-10	5.25	0.010
Insula	10	R	44	16	4	5.54	0.003

S-Table 6. Ψ3 related activation in all participants, p-FWE$_{wholebrain}$<0.05, k=10

Region	Cluster		Coordinates			T	p (FWE
Medial Frontal Gyrus	126	R	6	26	40	6.01	<0.001
Cingulate Gyrus		R	10	18	42	6.15	
			8	12	46	5.89	0.001
Insula		L	-32	20	10	5.57	0.002
			-36	18	10	5.52	0.003
		R	34	22	6	5.56	
Claustrum		L	-28	24	4	5.49	0.004
Inferior Frontal Gyrus	104	R	30	24	-8	6.34	<0.001
Inferior Frontal Gyrus	97	L	-28	24	-6	5.90	0.001
Claustrum	86	R	36	-62	-28	6.53	<0.001
			32	-68	-22	5.65	0.002
	44	L	-32	-56	-30	6.28	<0.001
Red nucleus	12	L	-6	-24	-4	5.91	<0.001
Inferior Frontal Gyrus	11	L	-36	4	32	5.35	0.007
Anterior Cingulate	11	R	8	28	30	5.39	0.006
		L	-6	30	30	5.36	
Cingulate Gyrus			-8	22	38	5.34	0.007

S-Table 7. μ3 related activation in all participants, p-FWE$_{wholebrain}$<0.05, k=10

Region	Cluster		Coordinates			T	p (FWE)
Middle Frontal Gyrus	1829	R	38	40	26	7.90	<0.001
			42	30	30	7.44	
			36	58	14	7.07	
			38	24	34	7.03	
			40	50	20	6.99	
			40	6	58	6.91	
			34	4	54	6.75	
			46	26	42	6.46	
			40	58	6	6.16	0.001
			26	4	46	5.99	<0.001
			46	52	0	5.75	0.001
Inferior Frontal Gyrus			38	8	30	5.56	0.003
			44	10	24	5.36	0.007
Superior Frontal Gyrus			32	58	-2	5.96	<0.001
			36	54	16	7.05	
Precentral Gyrus			42	6	36	6.55	
Inferior Parietal Lobule	1463	R	50	-46	40	8.37	<0.001
			50	-46	46	7.83	
			52	-42	52	7.59	
			50	-38	42	7.50	
			38	-46	42	7.18	
			38	-58	40	7.03	
			46	-50	54	6.64	
			36	-64	50	6.30	
			42	-34	38	6.29	
Superior Parietal Lobule			42	-64	50	6.62	
Middle Frontal Gyrus	1164	L	-38	56	12	8.75	<0.001
			-46	26	34	6.83	
			-28	50	12	6.70	
			-46	50	-6	6.52	
			-34	30	24	6.46	
			-40	10	34	6.27	
			-42	32	40	6.17	
			-44	34	24	6.08	
			-38	26	30	5.91	0.001
			-40	40	20	5.28	0.009
Inferior Frontal Gyrus			-46	50	2	6.93	<0.001
			-34	6	34	6.23	
			-40	2	28	5.01	0.030
Superior Frontal Gyrus			-38	48	28	5.49	0.004
Medial Frontal Gyrus	1007	R	6	26	40	8.19	<0.001
			4	24	44	7.94	

Superior Frontal Gyrus		L	-4	12	54	7.20	<0.001
Cingulate Gyrus		R	8	16	46	8.64	<0.001
		L	-2	22	46	7.69	
Anterior Cingulate			-6	30	30	5.83	0.001
Inferior Parietal Lobule	926	L	-34	-50	40	7.64	<0.001
			-44	-50	42	7.57	
			-48	-40	44	7.09	
			-46	-52	54	6.27	
			-42	-58	52	6.05	
			-32	-64	42	5.61	0.002
Superior Parietal Lobule			-34	-66	50	5.43	0.005
Cerebellum	615	L	-34	-58	-30	8.50	<0.001
			-14	-80	-22	6.98	
			-8	-80	-30	6.10	
			-26	-74	-24	6.00	
Claustrum	519	L	-30	24	-2	8.55	
			-30	20	8	7.22	
Insula			-36	18	4	6.99	
			-40	16	2	6.89	
Inferior Frontal Gyrus			-48	18	-4	5.50	0.003
Insula	449	R	32	20	8	6.54	<0.001
Inferior Frontal Gyrus			30	24	-6	8.36	
			44	22	-10	6.18	
			48	20	-6	5.94	
			40	20	-4	5.88	0.001
Cerebellum	176	R	34	-64	-28	9.15	<0.001
Thalamus	105	R	12	-4	8	6.58	<0.001
Middle Temporal Gyrus	102	R	62	-30	-10	6.22	<0.001
			54	-28	-12	5.21	0.013
Globus Pallidus	94	L	-14	0	0	6.69	<0.001
Thalamus			-10	-12	2	5.71	0.001
Brainstem	65	L	-6	-24	-4	7.92	<0.001
Middle Frontal Gyrus	31	L	-26	2	48	5.47	0.004
Globus Pallidus	29	R	16	2	-2	6.13	<0.001
Cerebellum	25	R	16	-82	-20	5.35	0.007
Middle Frontal Gyrus	22	R	24	46	-12	5.83	0.001
	21	L	-48	12	46	5.70	0.001
	17	R	46	44	-8	5.5	0.003

S-Table 8. Reward prediction error related activation (RPE δ from reinforcement learning model), over all participants, FWE corrected $p < 0.05$, k = cluster size (>=10).

Region	Cluster		Coordinates			T	p (FWE)
Superior Frontal Gyrus	3908	L	-16	36	52	9.07	<0.001
			-14	54	30	7.84	
		R	14	54	36	7.32	
			18	36	54	7.31	
Medial Frontal Gyrus		L	-2	64	8	8.61	
			-4	58	4	8.34	
			-8	56	-12	7.36	
		R	2	66	20	8.54	
			6	56	-4	8.40	
			2	64	24	8.01	
			2	42	-14	7.96	
			4	36	-12	7.74	
			2	62	28	7.64	
Middle Frontal Gyrus			30	34	-14	7.33	
Anterior Cingulate		L	-6	40	-8	8.40	
		R	4	40	-8	7.72	
Cingulate		L	-2	-48	32	7.81	
			-6	-46	34	7.74	
			-10	-48	36	7.25	
			-10	-32	40	6.10	0.002
			0	-2	36	6.05	
			-4	-8	44	5.92	0.004
		R	2	-34	36	9.29	<0.001
			2	-10	38	6.67	
			4	-14	36	6.35	0.001
Posterior Cingulate			-8	-56	16	8.66	<0.001
	1845	L	-2	-52	22	9.50	<0.001
Paracentral Lobule		R	6	-34	48	5.39	0.025
Putamen	1218	L	-14	8	-10	9.56	<0.001
			-28	-6	4	9.14	
			-30	-6	-6	8.19	
			-26	8	2	6.71	
Nucleus accumbens		L	-8	12	-10	9.25	
Amygdala			-26	-8	-18	9.47	
Hippocampus			-28	-20	-18	8.42	
Claustrum			-34	-6	-2	8.64	
			-38	0	4	6.85	
Globus Pallidus			-14	-8	-10	6.16	0.001
Middle Frontal Gyrus	1039	L	-42	38	2	8.42	<0.001
			-36	38	-10	7.46	
			-40	34	-14	7.33	
			-42	50	-4	6.98	

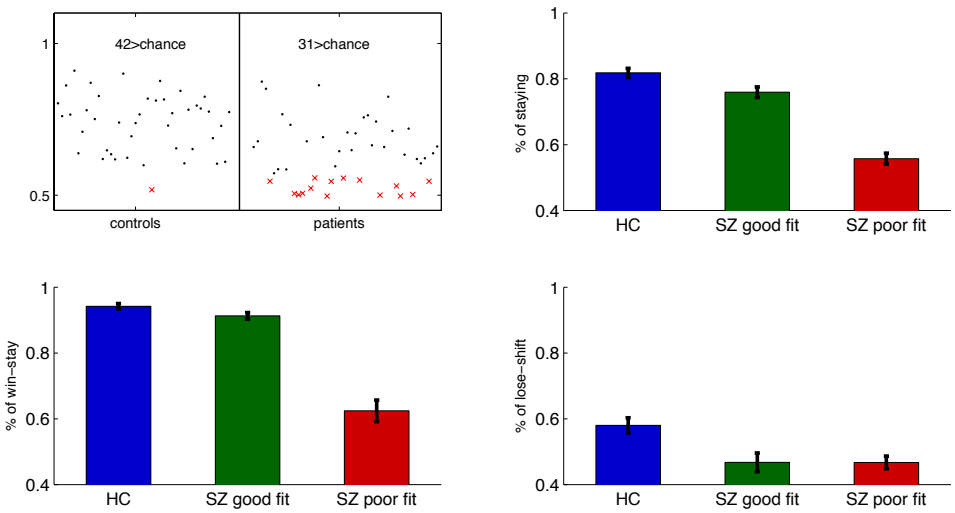
			-46	44	-4	6.84	
			-22	34	-14	6.64	
Inferior Frontal Gyrus			-50	38	8	7.54	
			-46	38	-10	7.20	
			-40	26	-16	6.94	
			-26	30	-16	6.71	
			-50	32	-8	6.41	0.001
			-54	32	0	6.40	
			-52	30	8	6.37	
			-46	48	8	5.93	0.003
Putamen	1013	R	30	-14	2	8.78	<0.001
			26	2	4	8.09	
			28	-4	0	7.82	
			30	-2	-10	6.03	0.002
Nucleus accumbens			12	10	-10	9.89	<0.001
Amygdala			24	-6	-18	6.67	
			26	-6	-26	6.21	0.001
Hippocampus			30	-12	-20	7.40	<0.001
			28	-20	-16	7.00	
Claustrum			34	2	2	6.43	0.001
Middle Temporal Gyrus	769	L	-50	-66	26	8.31	<0.001
Angular Gyrus			-52	-68	30	8.26	
Precuneus			-40	-68	36	7.77	
Superior Parietal Lobule			-42	-64	50	6.20	0.001
			-36	-74	48	6.09	0.002
Cerebellum	502	L	-40	-76	-32	9.35	<0.001
			-36	-80	-28	7.99	
			-46	-70	-32	7.68	
			-14	-86	-28	7.60	
			-26	-76	-30	7.08	
			-24	-80	-28	7.01	
			-28	-82	-24	6.62	
			-6	-84	-28	6.62	
			-32	-70	-34	6.41	0.001
			-18	-72	-32	6.03	0.002
			-34	-64	-34	5.31	0.033
Inferior Temporal Gyrus	387	L	-56	-16	-18	7.21	<0.001
			-62	-10	-20	6.97	
			-62	-16	-20	6.84	
			-60	-54	-8	6.02	0.002
Middle Temporal Gyrus	387	L	-64	-48	-4	7.05	<0.001
			-60	-34	-10	7.02	
			-54	-36	-14	5.91	0.004
			-58	-50	-12	5.74	0.007
Inferior Parietal Lobule	318	R	54	-60	42	6.61	<0.001

			52	-66	36	6.50	
			46	-66	40	6.15	0.001
			44	-72	42	6.09	0.002
			44	-68	46	6.03	
Superior Parietal Lobule			42	-64	50	5.82	0.005
			50	-58	26	5.36	0.028
			46	-60	28	5.30	0.034
			54	-60	24	5.29	0.036
Cerebellum	160	R	20	-42	-24	7.02	<0.001
			16	-46	-22	6.91	
			22	-54	-24	6.57	
			28	-48	-26	6.05	0.002
Middle Temporal Gyrus	144	R	62	-38	-14	7.07	<0.001
			64	-30	-14	6.75	
			64	-42	-4	6.32	0.001
			58	-42	-12	6.13	0.002
Cerebellum	113	R	26	-82	-22	6.53	<0.001
			20	-74	-22	5.81	0.005
			22	-86	-28	5.70	0.008
			10	-84	-26	5.38	0.026
Thalamus	89	R	14	-6	14	6.44	0.001
Caudate			18	-2	22	6.25	
	69	L	-14	-10	18	6.36	0.001
Thalamus			-8	-4	16	5.97	0.003
Middle Frontal Gyrus	57		34	30	52	6.51	<0.001
Inferior Temporal Gyrus	44	R	50	-8	-22	5.73	0.007
			60	-14	-22	5.71	0.008
Middle Temporal Gyrus			62	-8	-14	6.01	0.003
			56	-20	-20	5.39	0.025
Medial Frontal Gyrus	39	L	0	-12	52	6.10	0.002
Cerebellum	30	R	44	-74	-28	6.98	<0.001
	22	L	-22	-56	-20	6.14	0.002
	17	L	-26	-42	-30	6.2	0.001
Superior Parietal Lobule	14		-26	-40	56	6.03	0.002
Culmen	13	L	-18	-38	-24	5.57	0.013
Insula	11	R	36	2	22	6.29	0.001
Inferior Parietal Lobule	10	L	-64	-36	24	5.65	0.010

S-Table 9. Small volume corrections for region of interests (ROI) in midbrain and ventral striatum. Bilateral ROIs were used for midbrain as in Iglesias et al. 2016 based on a study by Bunzeck and Dutzel (2006) and for ventral striatum as in Schlagenhauf et al. (2014) (1, 7, 8).

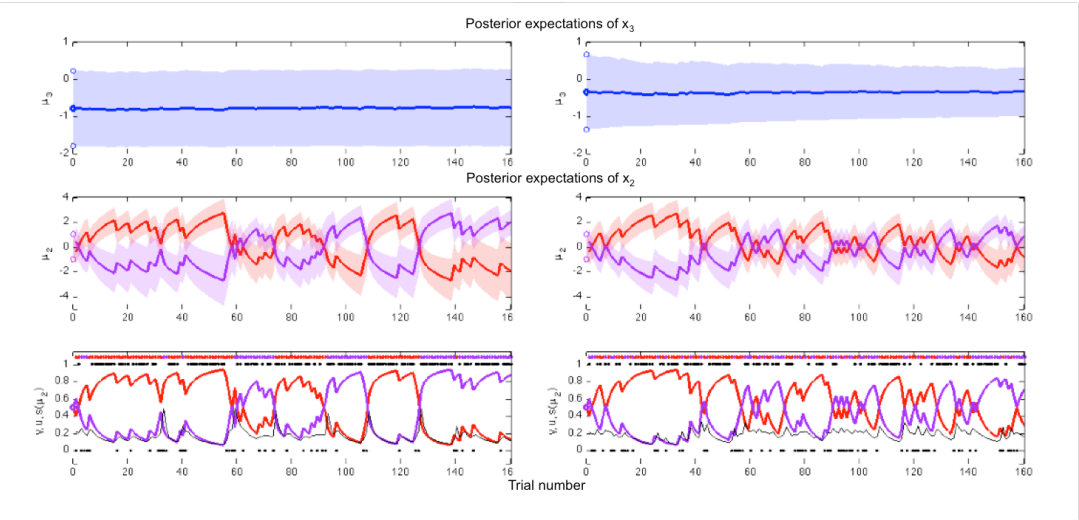
Region			**Coordinates**			**T/F**	**p (FWE)**
Second level precision-weighted prediction error ε2							
Midbrain ROI	All HC and SZ, T	L	-8	-18	-10	3.94	.01
	HC vs SZ, F	R	12	-18	-14	12.97	.08
Ventral striatum ROI	All HC and SZ, T	$FWE_{wholebrain}$<.05, see S-Table 3					
	HC vs SZ, F	R	16	2	-6	6.23	.35
Reward prediction error from reinforcement learning model δ							
Midbrian ROI	All HC and SZ, T	L	-6	-14	-10	4.03	.02
	HC vs SZ, F	R	14	-18	-12	11.60	.17
Ventral striatum ROI	All HC and SZ, T	$FWE_{wholebrain}$<.05, see S-Table 8					
	HC vs SZ, F	L	-14	4	-4	9.49	.12

S-Figure 1. A) Classification above (black dots) and below (red crosses) chance and its influence on B) overall stay behavior and C) win-stay and D) lose-stay behavior.



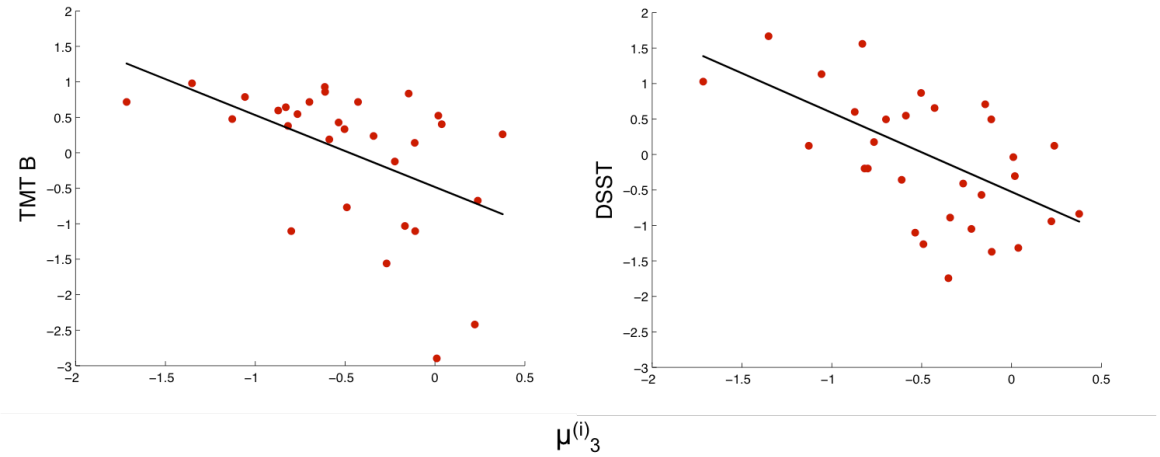
S-Figure 2. Exemplary trajectories of one healthy controls and one patient with imu3 and kappa around the group mean illustrating the differences in learning on the second and the third level of the HGF.

HC31 SZ65

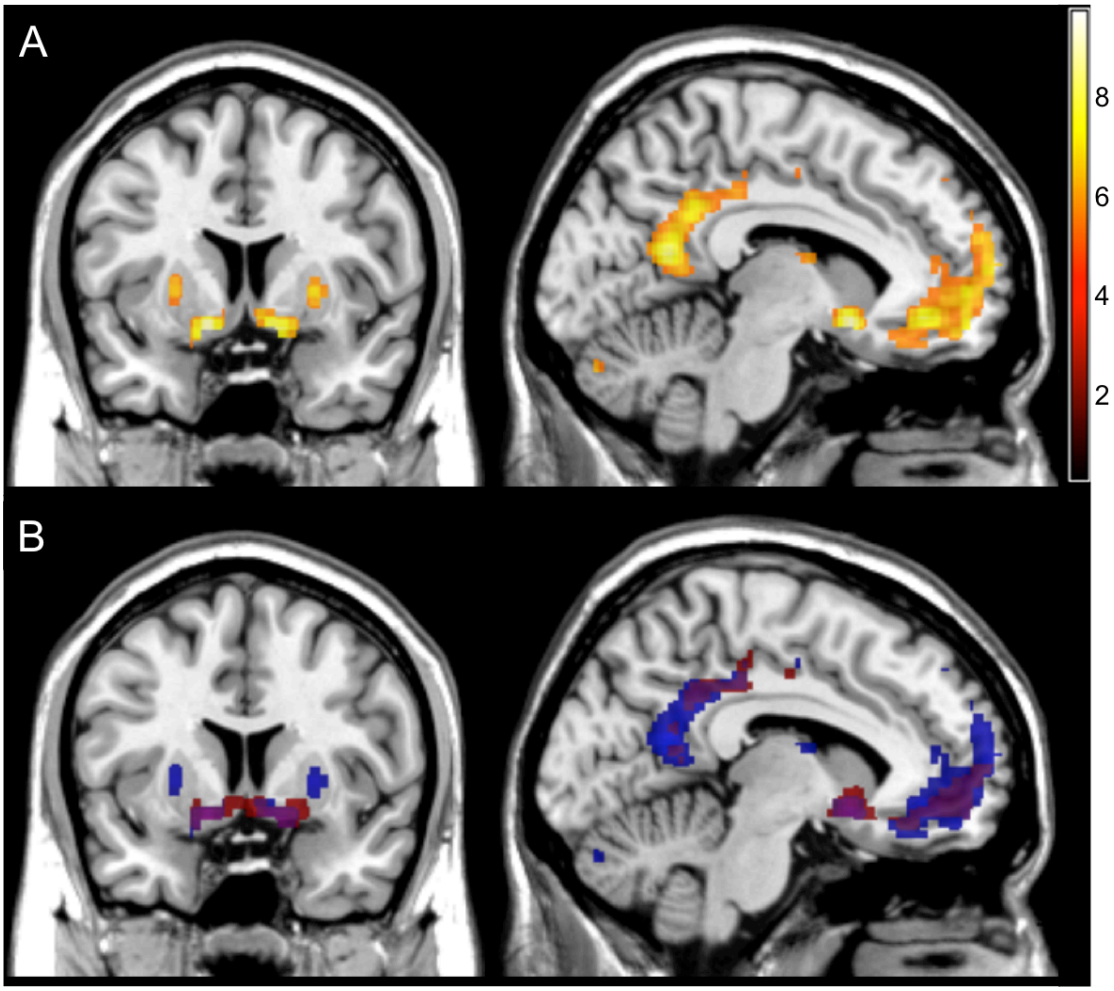
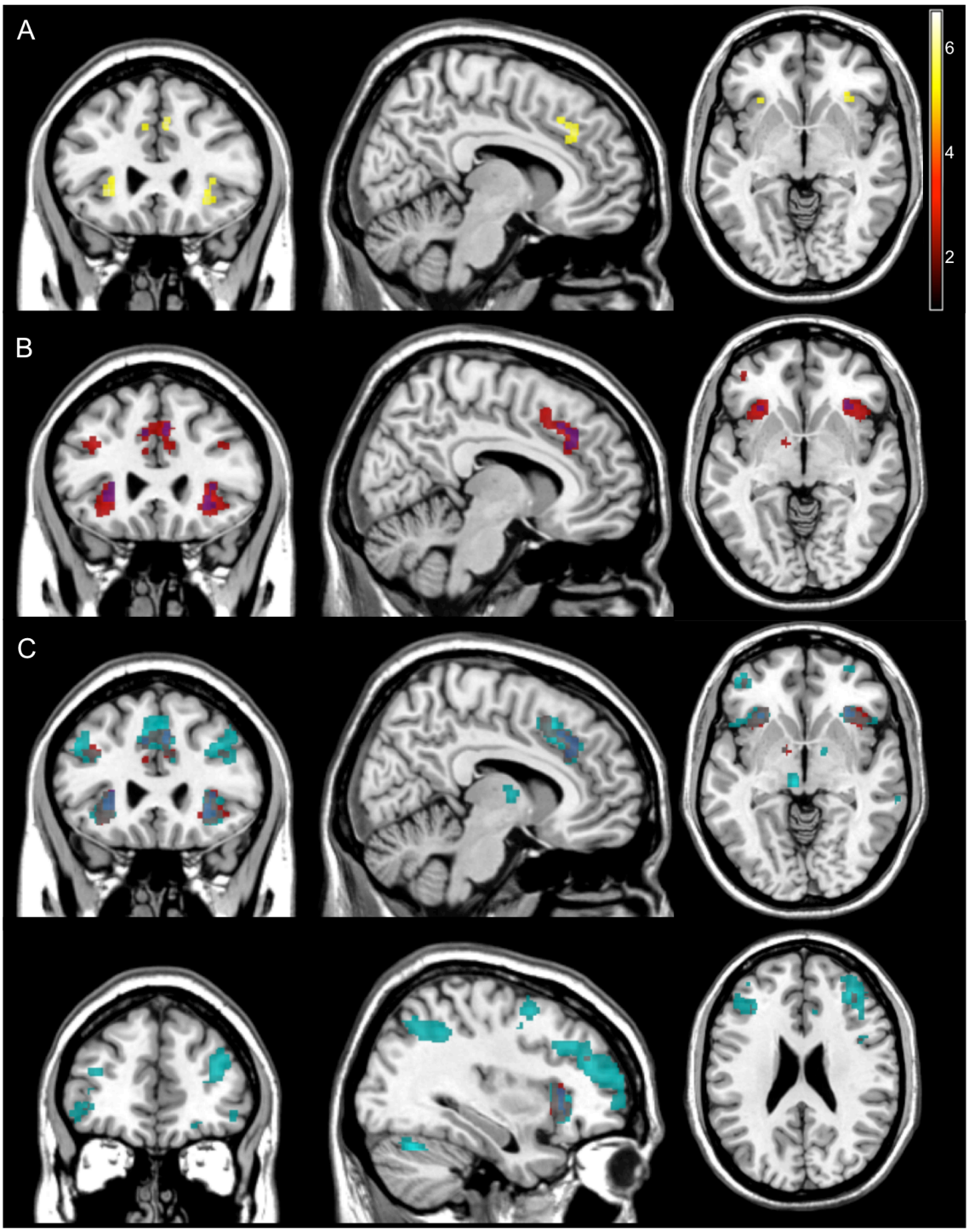


$\mu^{(i)}_3$=-.79; κ=.53 $\mu^{(i)}_3$=-.35; κ=1.77

S-Figure 3. Bonferoni-corrected significant correlations between $\mu_3^{(0)}$ and z-scored measures of cognitive flexibility (TMTB) and cognitive speed (DSST) in patients fit better than chance by the model (n=31), Spearman rank correlation coefficient -.56, p<.001 for TMB and -.56, p<.001 for DSST.



S-Figure 4. A) PE-related signals based on RL model, p-FWE$_{wholebrain}$<0.05, k=10, [7 -8]; B) Overlay of RRE from RL model in blue and second-level precision weighted RPE from HGF model in red, p-FWE$_{wholebrain}$<0.05, k=10, [7 -8];



S-Figure 5. A) BOLD signal related to third-level precision, p-FWE$_{wholebrain}$<0.05, k=10, [8 26 -4]; B) overlay of second-level (red) and third-level precision (blue), p-FWE$_{wholebrain}$<0.05, k=10; C) overlay of second-level (red) and third-level precision (blue), environmental volatility (turqoise) at [8 26 -4] and [34 44 24], p-FWE$_{wholebrain}$<0.05, k=10;