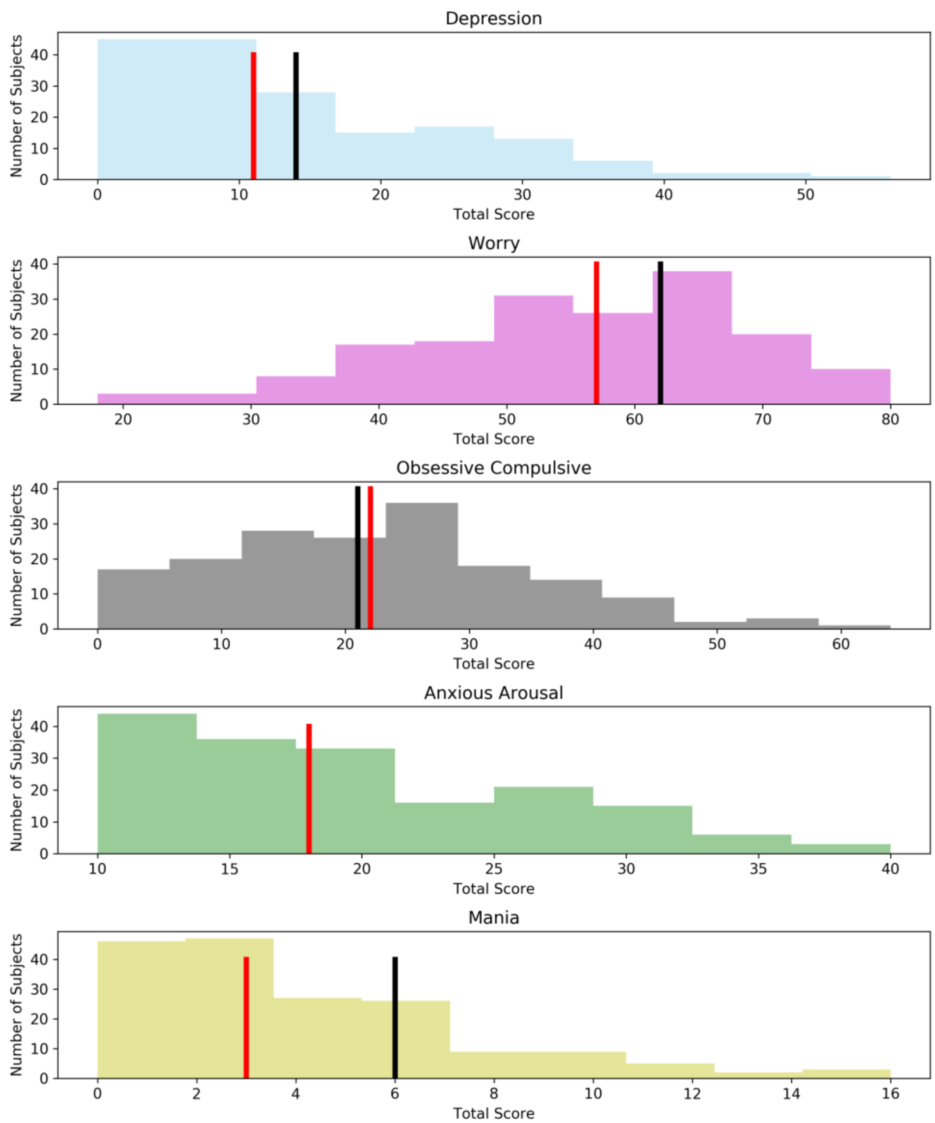

Figure S1. Self-report distributions. Frequency plots for each self-report measure used to assay various pathologies. In each plot, the black bar denotes cut-offs of mild but clinically-significant symptoms, and the red bar denotes the median. Note that for most measures, the median is either slightly below or above the clinical cut-off; only mania has a large gap between the two. Cut-offs for pathological worry (cutoff=62) was determined in Curtiss and Klemanski (2015), for depression (cutoff-14 for at least mild depression) in Beck et al. (1996), for obsessive-compulsive symptoms (cutoff=21) in Foa et al. (2002) and mania (cutoff=6) in Altman et al. (1997). We used the median to define the MASQ-AA short form clinical cutoff (cutoff=18) which is close to a very similar questionnaire, the MASQ-D30 Anxious Arousal subscale (cutoff=17).

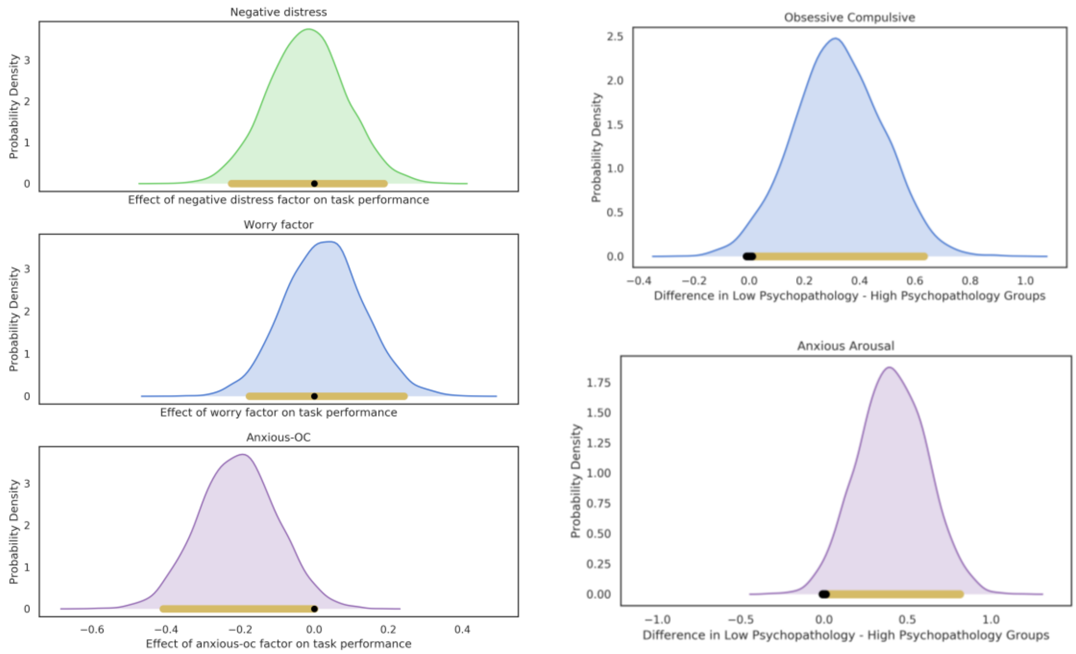

Figure S2. Effects of psychopathology on performance in the one-step revaluation task using non-binarized data. The left three plots come from the model parametrically relating orthogonal psychopathology factors derived from the exploratory factor analysis to task performance. Negative Distress (95% HDI: [-.17, .15], mode=0) and Worry factors (95% HDI:[-.14, .18], mode=0.01) were unrelated to task performance, whereas the Anxious-OC factor (95% HDI:[-.33,-.02], mode= -.19) had a negative relationship with task performance. The right two graphs comprise the group analyses wherein having lower anxious arousal (95% HDI:[.02,.81], mode=.40] or lower obsessive-compulsiveness (95% HDI:[.006, .63], mode=.31) trended towards being significantly related better task performance.

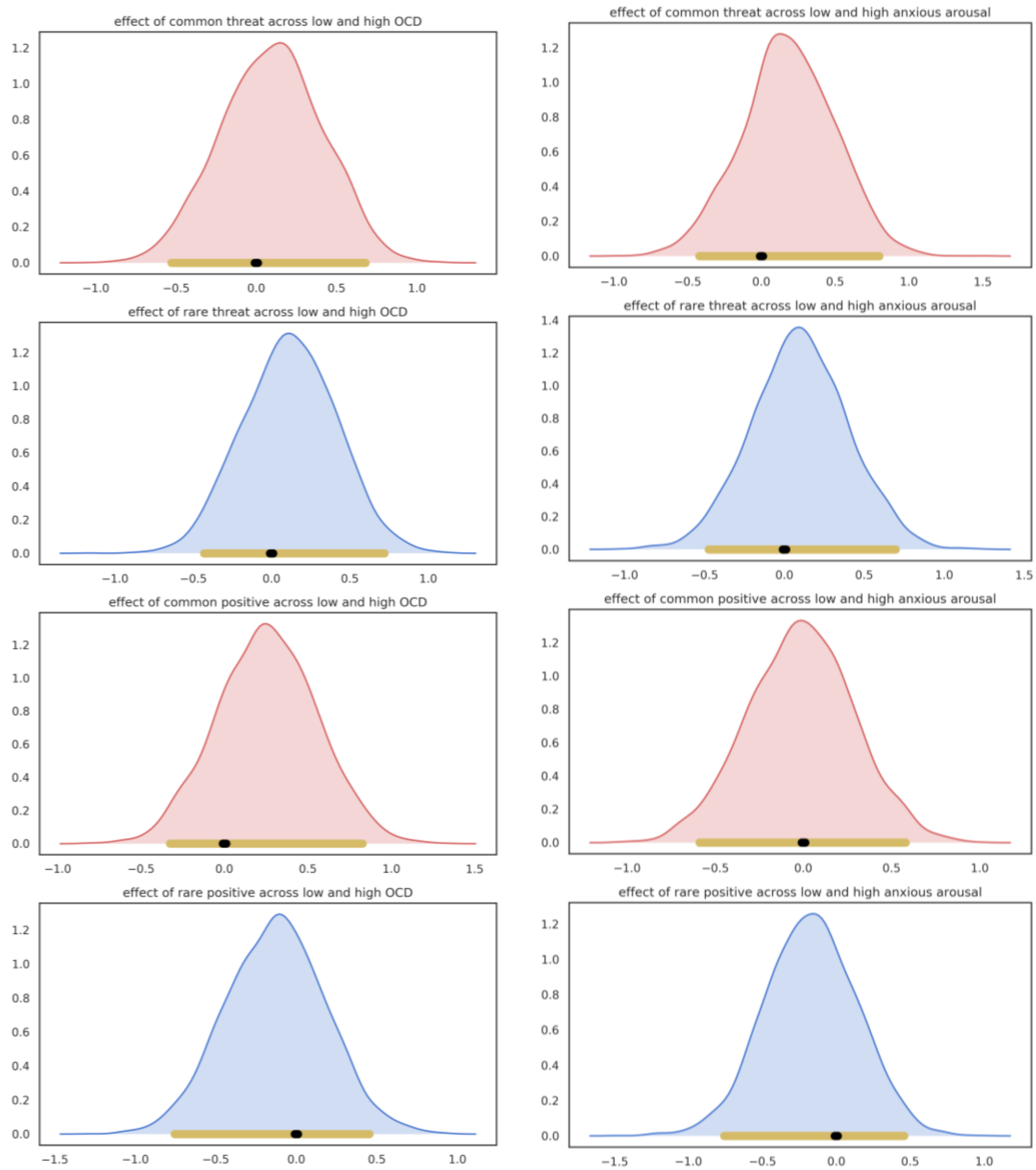

Figure S3. Interaction between condition effect and psychopathology. In each plot, the terms "positive" or "threat" denote whether or not a positive or negative image was interposed between state transitions, and the terms "common" or "rare" denote the probability of the state transition (common = 50%, rare=30%; the remaining state transition was of non-interest because both actions led to the final state 20% of the time) No interaction approached a significant effect, as noted by the ROPE being near the mode of each interaction effect. We additionally tested whether differences between condition (e.g., common positive – rare positive) depended on psychopathology group, all of which were similarly non-significant.

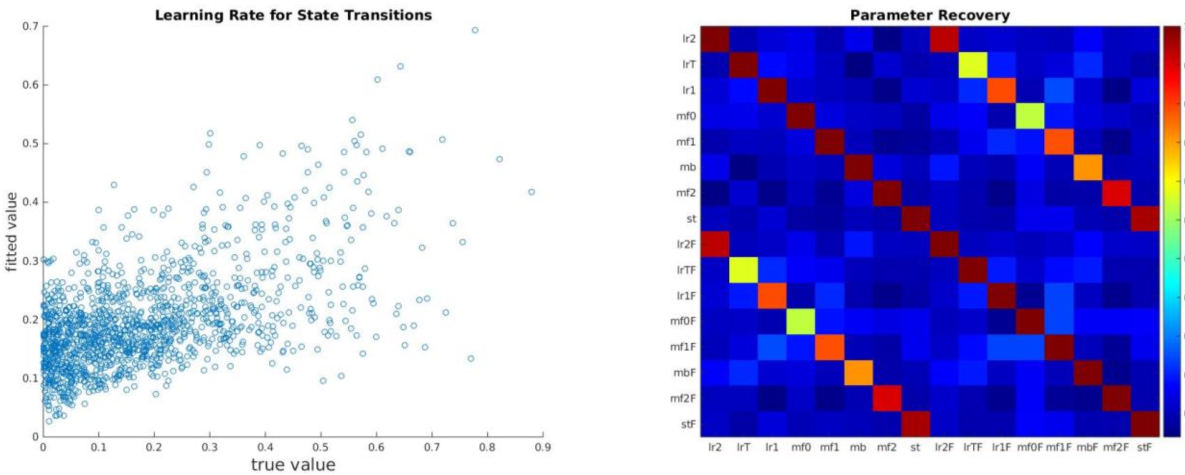

Figure S4. Both diagrams comprise data simulated from the winning model (2AV + LR) and best-fit group hyperparameters. The simulation was contained 400 generated agents over 200 trials of the two-step task. The group hyperparameters are as follows: **Learning rates:** Learning Rate 1st stage ~ beta(0.47,0.48)**,** Learning Rate 2nd stage ~ beta(0.99,0.68)**,** Learning Rate state transitions ~ beta(0.79,3.74). **Softmax beta weights 1st action,** Model based beta ~ gamma(0.50,4.09) **,** Model free beta TD(0) ~ gamma(0.56,0.65)**,** Model free beta TD(1) ~ gamma (1.78,1.28)**,** Stickiness beta ~ gamma(1.47,0.80). **Softmax beta weight 2nd action**: MF2 beta ~ gamma (3.68,0.69). The left is a scatter plot of the true state transition learning rates used in a simulation (x-axis) and the fitted state transition learning rates. The right heat map comprises the full set of correlations between fitted and true parameters in the simulation and subsequent model fitting. Simulating the the winning model and the second-best fitting model ('2AV+Count'), we recovered the generative model 10 out of 10 times. The mean difference in iBIC when generating data with the winning model, '2AV+LR' was 332.99, and was 196.41 when generating the data with '2AV+Count'.

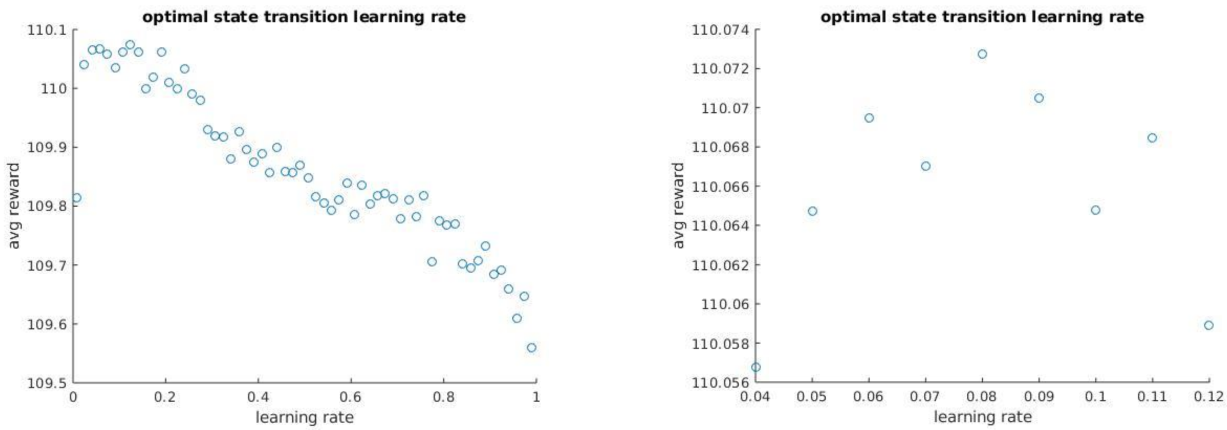

Figure S5. Determining the state transition learning rate that maximizes reward in the two-step task as configured in Gillan et al. (2016) We simulated agents that played the exact same task as described in Gillan et al. (2016) with parameters set to the fitted group average from the winning model (‘2AV+LR’, see Model Descriptions). We removed model-free behaviour (setting softmax beta weights on model-free action values to 0) from affecting first-stage actions to determine the best state-transition learning rate under the assumption that agents emply the normative model-based strategy. We had each agent play the game 10,000 times, and sampled five different model-based beta parameters [5,10,20,100,1000] to determine what model-based beta maximized reward. After demonstrating that a model-based beta value of 100 yielded the maximum reward across the full range [0,1] of possible state transition learning rates (left image), we zoomed in on the region that yielded the greatest rewards [0.04,0.13]. For each transition learning rate (in increments of .01) in this range, we simulated agents 10,000,000 times to reduce random noise in estimates of average reward earned. Doing so revealed that the optimal state transition learning rate was 0.08.