Psychological Medicine

cambridge.org/psm

Disrupted state transition learning as a computational marker of compulsivity

Paul B. Sharp[1,2,3], Raymond J. Dolan[1,2] and Eran Eldar[3]

[1]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, London, UK; [2]Wellcome Centre for Human Neuroimaging, University College London, London, UK and [3]The Hebrew University of Jerusalem, Jerusalem, IL, USA

Original Article

Cite this article: Sharp PB, Dolan RJ, Eldar E (2023). Disrupted state transition learning as a computational marker of compulsivity. *Psychological Medicine* **53**, 2095–2105. https://doi.org/10.1017/S0033291721003846

Received: 17 December 2020
Revised: 28 August 2021
Accepted: 2 September 2021
First published online: 24 September 2021

Key words:
Anxiety; compulsivity; computational modelling; model-based learning; transdiagnostic

Author for correspondence:
Paul B. Sharp,
E-mail: paul.sharp@mail.huji.ac.il

Abstract

Background. Disorders involving compulsivity, fear, and anxiety are linked to beliefs that the world is less predictable. We lack a mechanistic explanation for how such beliefs arise. Here, we test a hypothesis that in people with compulsivity, fear, and anxiety, learning a probabilistic mapping between actions and environmental states is compromised.

Methods. In Study 1 ($n = 174$), we designed a novel online task that isolated state transition learning from other facets of learning and planning. To determine whether this impairment is due to learning that is too fast or too slow, we estimated state transition learning rates by fitting computational models to two independent datasets, which tested learning in environments in which state transitions were either stable (Study 2: $n = 1413$) or changing (Study 3: $n = 192$).

Results. Study 1 established that individuals with higher levels of compulsivity are more likely to demonstrate an impairment in state transition learning. Preliminary evidence here linked this impairment to a common factor comprising compulsivity and fear. Studies 2 and 3 showed that compulsivity is associated with learning that is too fast when it should be slow (i.e. when state transition are stable) and too slow when it should be fast (i.e. when state transitions change).

Conclusions. Together, these findings indicate that compulsivity is associated with a dysregulation of state transition learning, wherein the rate of learning is not well adapted to the task environment. Thus, dysregulated state transition learning might provide a key target for therapeutic intervention in compulsivity.

Introduction

Although it is known that compulsivity, fear, and anxiety are associated with beliefs that external threats and internal worries are less controllable (Cartwright-Hatton & Wells, 1997; Grupe & Nitschke, 2013; Mathews, 1990), we lack a useful model to explain how such beliefs might develop. Psychological theorists propose that such beliefs stem from a lack of self-efficacy in controlling one's environment (Bandura, 1978), but this does not explain the nature of the dysfunction that produces a lack of self-efficacy. Here, we test a novel hypothesis that a deficit in state transition learning – specifically, the ability to internalize probabilistic maps of how actions lead to new states in the environment – is associated with fear, anxiety, and compulsivity symptoms (Zorowitz, Momennejad, & Daw, 2020).

An effort to relate these symptoms to state transition learning extends nascent work on how psychopathology is marked by difficulty in utilizing internal maps of one's environment (so-called 'cognitive maps') for decision-making (Behrens et al., 2018; Tolman & Honzik, 1930). Early computational psychiatry work has shown that 'model-based' behavioral control, which uses a cognitive map of state transitions to plan one's actions, is disrupted in individuals who show high compulsivity (Gillan et al., 2020a, 2020b, 2016; Rouhani et al., 2019). This work, however, has not considered that individuals may differ in how they learn state transitions. Consequently, model-based deficits were attributed to a failure to *use* state transition knowledge, the learning of which was assumed to be intact (Konovalov & Krajbich, 2020). Investigating state transition learning can thus determine whether compulsivity-related model-based deficits are due to a more fundamental impairment in learning how to transition between states. Doing so can additionally test for potential anxiety- or fear-related deficits in transition learning, which may be independent from the degree to which state transition knowledge is used for decision-making (e.g. Gillan et al., 2020b).

To address these issues, Study 1 tested whether difficulty in learning state transitions is associated with the symptoms of compulsivity, fear, and anxiety. For this purpose, we designed a task environment where it is difficult to learn state transitions, yet once learned it is trivial to use them for decision-making (Fig. 1). Moreover, unlike previous investigations, we specifically avoided administering rewards during the learning phase. This design enabled us to isolate state transition learning from other facets of learning and planning.

To investigate the relationship between transition learning and psychopathology, we sampled participants from the general population, and assessed them for a range of transdiagnostic

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

CAMBRIDGE UNIVERSITY PRESS

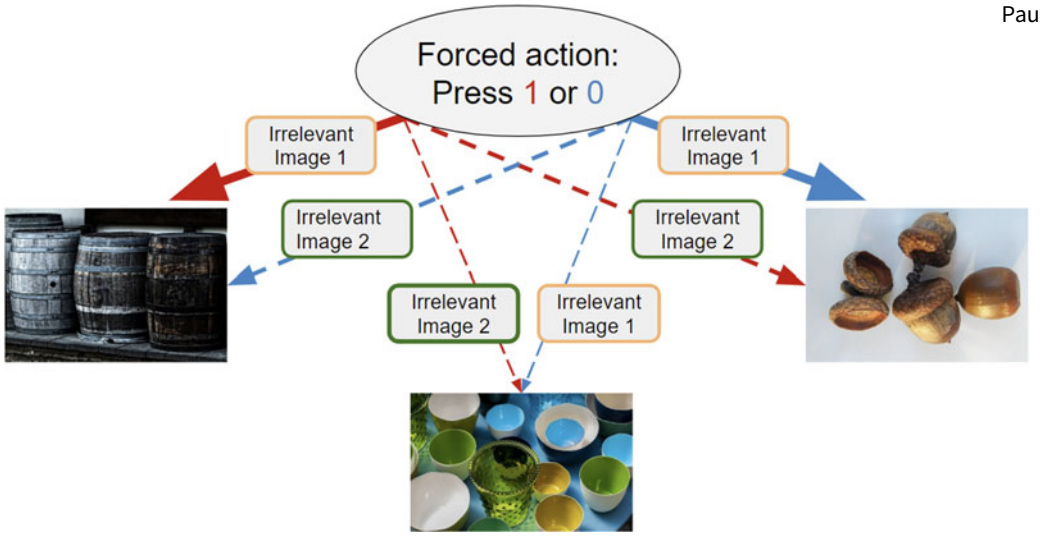


Fig. 1. Study 1: one-step revaluation task learning phase. Participants played a one-step revaluation task where two actions could each lead to three possible states. Depicted here is the learning phase wherein participants were instructed to take an action (pressing either 1 or 0), and observe to which of three possible states (each denoted by an emotionally-neutral image) their action led. Each action led to each of the three images with different probabilities. The red arrows above symbolize transitions from action 1, and blue arrows from action 0. The thick arrow indicates the common transition, occurring 5/10 times (per action), the middle-thickness dashed arrow indicates the uncommon transition, occurring 3/10 times, and the thinnest dashed arrows the rarest transitions, occurring 2/10 times. Participants chose each action 10 times, and saw an 'irrelevant' image of either neutral or emotional valence presented before the outcome state. The instructions clarified these additional images were irrelevant and should be ignored. Irrelevant images were displayed for 500 msec. Analyses revealed a lack of a role of these irrelevant images on the relationship between psychopathology and state transition learning.

symptoms, including those associated with depression, mania, worry, sympathetic hyperarousal, and compulsivity. Using this approach, we show that individuals high in compulsivity demonstrate impaired transition learning. In this regard, preliminary evidence implicated a transdiagnostic factor spanning compulsivity and fear.

Our novel one-step revaluation task is capable of isolating state transition learning from planning because it separates the learning and decision phases. However, for the same reason, it is not suited for modeling the computations underlying state transition learning. Thus, to more precisely characterize the computations involved in state transition learning, we complemented Study 1 with two independent studies comprising tasks which measured decision-making as state transition learning was ongoing. In Study 2, we re-analyzed publicly available two-step task data (Gillan, Kosinski, Whelan, Phelps, & Daw, 2016) to measure the rate of learning with regards to stable state transitions. In Study 3, we used a novel multi-goal pursuit task (Sharp, Russek, Huys, Dolan, & Eldar, 2021) to measure the rate of learning with regards to continuously changing state transitions. In both studies, the best-fitting model separately quantified the rate of state transition learning and the use of transition knowledge for making choices. Fitting the model to each individual participant revealed that compulsivity is associated with excessively fast transition learning with regards to stable transitions (which demand slower learning) and excessively slow transition learning with regards to changing state transitions (which demand faster learning).

Thus, across three independent studies, we demonstrate that compulsivity is associated with dysregulated state transition learning that fails to adapt to the statistics governing state transitions.

Methods and materials

Study 1: isolating state transition learning disruptions

Participants

We used the Gorilla Experiment Builder (www.gorilla.sc) to develop and host our experiment (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020). Participants were recruited online through Prolific recruiting service (https://www.prolific.co/). The only inclusion criteria for the study were that participants spoke fluent English and were over 18 years old. The study (final sample = 174) comprised 119 male participants (68%). We recruited 219 participants based on a power analysis that provided 80% power to detect a small-medium effect size ($r \leqslant 0.2$), given similar effect sizes across a relevant prior literature examining links between psychopathology and assays of model-based learning (Alvares, Balleine, & Guastella, 2014; Gillan et al., 2016; Voon et al., 2015). Forty-five participants (20%) did not meet the criterion of accurately identifying the common image (to which an action led 80% of the time) in a practice round, and were thus excluded from further analysis. Participants gave written informed consent before taking part in the study, which was approved by the university's ethics review board.

Psychopathology assay

Before participants completed the online task, they filled out five questionnaires covering a range of transdiagnostic types of psychopathology. Worry was measured via the 16-item Penn State Worry Questionnaire (PSWQ; Meyer, Miller, Metzger, & Borkovec, 1990). Anxious arousal was measured with the 10-questions mini version of the Mood and Anxiety Symptom Questionnaire – Anxious Arousal subscale (MASQ; Casillas & Clark, 2000). Obsessive compulsiveness was measured with the Obsessive Compulsive Inventory – Revised (OCI-R; Foa et al., 2002). Depressive symptoms were measured with the Becks Depression Inventory – II (BDI; Beck, Steer, & Brown, 1996). Mania symptoms were measured with the Altman Self-Rating Mania Scale (ASRM; Altman, Hedeker, Peterson, & Davis, 1997; see online Supplementary Fig. S1 for full distributions of questionnaires).

Although IQ, age, and gender are known to covary with compulsivity and related forms of psychopathology, we chose not to measure them in the data we collected for two reasons. First, because we investigate a learning impairment, controlling for

https://doi.org/10.1017/S0033291721003846 Published online by Cambridge University Press

IQ, age, or gender might remove meaningful variance and diminish our ability to detect relations with psychopathology. Second, and most importantly, we do not question whether compulsivity is associated with known deficits in model-based control, as this association has already been demonstrated to exceed the effects of IQ, age, and gender (Gillan et al., 2016). Our main goal was to determine how state transition learning is related to compulsivity and whether it explains compulsivity-related deficits in the utilization of state transition knowledge for model-based control.

One-step revaluation task

Participants played a one-step revaluation task where two actions could each lead to three possible states. In a learning phase (Fig. 1), subjects were forced via instruction to take an action (pressing either 1 or 0), and observe to which of three possible states their action led. Each action led to each of the three images with different probabilities. Participants were told that the better they learned the states that commonly followed each action, the more money they would win, though unbeknownst to subjects the payment schedule for each round was fixed at £7. Participants chose each action 10 times, with the two actions randomly interleaved. Additionally, before the presentation of the outcome state, an irrelevant image of either neutral or emotional valence was presented on screen.

After participants completed the learning phase, participants faced a test phase in which they were presented with each of the relevant outcome-state images from the learning phase. Upon seeing the image, participants had to choose the action that delivered them to that image most often in the learning phase (e.g. choosing action 1 when shown the image of the barrels). No feedback was presented after a choice was made so that performance would solely reflect what was learned about transitions in the learning phase. Note that both actions always led to rare states with the same, low probability (20%), and thus there was no optimal action to take in the test phase for such queries (Fig. 1, transitions to cups image). We included the 'rare state' because we learned in piloting that this was necessary to remove a ceiling effect.

The experiment started with an easy practice round, where each action led to its 'common' image 16 out of 20 times (e.g. action 1 leading to the image of the barrels in Fig. 1), and twice to each of the other two images. Participants then played five blocks (henceforth, 'experimental conditions') each of which comprised a learning phase (Fig. 1; 20 forced choice trials) and a test phase (three queries). Four blocks differed in whether the irrelevant emotional stimuli were presented during learning along common or uncommon transitions and were of either positive or negative valence. The remaining block included only neutral stimuli.

Image stimuli were taken from the OASIS image dataset (Kurdi, Lozano, & Banaji, 2017). All task-relevant images were selected to have neutral emotional content (i.e. valence between 3.5 and 4.5 on a seven-point scale, and arousal was below 3.5). Irrelevant emotional images were selected with high arousal (>5) and valence (<3 for negative, >5 for positive).

Of note, the original intention of the task was to test whether the interaction between task-irrelevant affective images and pre-existing worry impacted state transition learning. Specifically, we hypothesized that chronic worry would produce an attentional bias to negative distracting images, and this would facilitate greater memory for such transitions. As such, we predicted those with chronic worry would perform better at test when negative distractors were paired with common transitions, and worse when negative images were paired rare transitions. However, contrary to this prediction, no effects were found for the pairing of negative images with common (mode = −0.39, CI −0.96 to 0.27) or rare transitions (mode = 11, CI −0.51 to 0.70; see online Supplementary Fig. S2).

Modeling whether state transition learning was impaired in the novel revaluation task

To quantify learning, we analyzed participants' choices from test queries for which there was an optimal action (acorns and barrels in Fig. 1). A Bayesian logistic regression model was used to predict whether the participant correctly learned the state transition matrix that characterized each block, as evidenced by their choices in the test phase. Specifically, the logistic regression predicts whether each participant choice was correct as a function of: a grand-mean intercept, a random-effect participant baseline, a random-effect experimental condition, and three fixed-effects for each transdiagnostic psychopathology factor. The intercept was drawn from a normal prior where the prior was centered on the grand mean (mean = 0.6, S.D. = 1). Each random-effect was drawn from a group-level prior whose mean was 0. We estimated hierarchically the group-level variances on each random effect. We kept all priors on variances across fixed- and random-effects to an upper bound of 3. The hyperpriors for the variance of the participant baseline were Uniform(0,3) whereas each fixed-effect for each transdiagnostic psychopathology factor was drawn from Normal(0,3). The participant-specific baseline and experimental condition effects were instantiated in STAN as unordered categorical variables.

To derive the transdiagnostic factors, we used the Factor Analyzer Python package (https://pypi.org/project/factor-analyzer/) with a promax rotation, which allows factors to be partially correlated, along with default parameters.

For hypothesis testing, we defined a Region Of Practical Equivalence (ROPE; Kruschke, 2018) representing the range of insignificant effect sizes, and compared it to the 95% most credible parameter values (i.e. the 95% Highest Density Intervals; HDI). The ROPE was defined as 10% of the standard deviation in the estimate of the baseline effect, signifying a small effect (0 ± 0.01).

Posterior probability distributions of effects were estimated using a Hamiltonian Monte Carlo (HMC) sampling procedure using a python implementation of STAN statistical software (Stan Development Team, 2020). We changed the adapt_delta parameter to 0.9 and max tree depth to 18, to reduce divergences and improve sampling efficiency.

After fitting each model, we re-scaled all categorical parameter estimates by enforcing a sum-to-zero constraint (i.e. the sum of coefficients for the different levels of the categorical variable equals 0). This procedure improves the interpretability of categorical variable coefficients and regression intercept by ensuring each coefficient represents the deviation for the corresponding categorical level from baseline, which is captured by the intercept (Kruschke & Liddell, 2018). In all models, all estimated parameters had good indicators of reliable sampling from the posterior, including r-hat below 1.1 and effective sample size above 1000 for all parameters.

Study 2: quantifying disruptions in state transition learning when transitions are stable

To investigate how disrupted state transition learning in compulsivity manifests in model-based control when state transitions are

https://doi.org/10.1017/S0033291721003846 Published online by Cambridge University Press

stable, we fit computational models of state transition learning to a publicly available two-step task data ($n = 1413$; Gillan et al., 2016). Briefly, the two-step task is a trial-and-error learning task that requires participants first choose an action that leads them to a second state, from which they can choose another action in order to attain reward. Participants are instructed in training about the probability of common and rare transitions, which do not change over the course of the task, but participants are not told which of the two transitions for a given action is most likely and thus they must learn that through experience. We empirically tested here whether such learning is less effective in some individuals than in others, such that accounting for these individual differences would improve our ability to model participants' choices.

Previous models of this task learn state transitions by a simple counting heuristic (Otto, Raio, Chiang, Phelps, & Daw, 2013). We refer to this as the 'Typical model'. Specifically, each trial, one of two possible transition matrices, is inferred as the true transition matrix that determines the probability first-stage actions transition second-stage states (e.g. either action 1 leads to state 1 70% of time, or to state 2 70% of the time). Here the columns are the possible two actions and the rows are possible states each action can transition to:

$$T1_{true} = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \text{ or } T2_{true} = \begin{bmatrix} 0.3 & 0.7 \\ 0.7 & 0.3 \end{bmatrix} \quad (1)$$

The transition matrix is inferred by using a running counter of how many times each action transitioned to each second stage state. For instance, if a subject's initial choice was action 1, and they transitioned to state 1, it would be encoded in the following 'counting' matrix:

$$T_{counting} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad (2)$$

At any point, $T1_{true}$is inferred if $T_{counting}(1,1) + T_{counting}(2,2) > T_{counting}(1,2) + T_{counting}(2,1)$, where the indices in the parenthesis reflect (row,column) in $T_{counting}$. If the inequality is the converse, $T2_{true}$is inferred. If both sides of the inequality are equal, the average of $T1_{true}$ and $T2_{true}$ is inferred.

Here, we tested an alternative model that incrementally learns state transitions as a consequence of state transition prediction errors (Gläscher, Daw, Dayan, & O'Doherty, 2010). We refer to this model as the incremental state transition learning (ISTL) model (online Supplementary Fig. S3). This model quantifies each individual's rate of state transition learning, which when excessively low or high can result in inaccurate estimates of true state transition probabilities. For instance, if one transitions from action 1 to state 2 at time t, the estimated probability of that transition, $P(s = 2|a = 1)_t$, is updated by the state prediction error $[1–P(s = 2|a = 1)_t]$ according to a learning rate, γ, which defines how much weight to give to the most recent prediction error:

$$P(s = 2|a = 1)_{t+1} = P(s = 2|a = 1)_t + \gamma(1 - P(s = 2|a = 1)_t) \quad (3)$$

At the same time, the complementary transition probability, from action 1 to state 1, is also updated so that the probabilities sum to 1:

$$P(s = 1|a = 1)_{t+1} = 1 - P(s = 2|a = 1)_{t+1} \quad (4)$$

By contrast, estimated state transitions for the action that was not taken decay toward the initial prior of 0.5:

$$P(s = 2|a = \text{not taken})_{t+1} = P(s = 2|a = \text{not taken})_t + \gamma(0.5 - P(s = 2|a = \text{not taken})_t) \quad (5)$$

and similarly for $s = 1$.

Action values for chosen actions were updated exactly as in Gillan et al. (2016), with a decay rate on action values that equals the complement of the learning rate, 1-α. Action values for unchosen actions were also updated as in Gillan et al. (2016) except that they had their own decay rate, D.

We tested participants' choices against several additional models as detailed in online Supplementary Fig. S3. None of these models was able to explain choices as well as the ISTL model, and thus we focus here on a comparison between the Typical and ISTL models.

Models were fit hierarchically via iterative importance sampling (Bishop, 2006), which has been shown to ensure high parameter and model recoverability (Eldar & Niv, 2015; Eldar, Roth, Dayan, & Dolan, 2018). The priors for this model-fitting procedure largely do not affect the results, because the procedure iteratively updates priors via likelihood-weighted resampling in order to converge on the distributions of parameters that maximize model evidence. As such, all parameters had 'naïve' priors. To ensure models and parameters were appropriately designed, we simulated data using the ISTL and the Typical models, and successfully recovered both the true parameter values (all correlations between true and fitted values above 0.5; online Supplementary Fig. S4) and the model that generated each simulated dataset (online Supplementary Note 1).

A model-agnostic behavioral signature of model-based behavior

To quantify individual differences in model-based control in empirical and simulated data without relying on a specific model, we used a behavioral signature established in prior work (Daw, Gershman, Seymour, Dayan, & Dolan, 2011). This model-agnostic signature comprises the proportion of engaging the switch or stay behavior that is predicted by model-based control following rare state transitions. Model-based control predicts one should switch more often from one's prior action if a state to which a rate transition led was rewarded, because one is more likely to reach the rewarded state via the action one did not take last trial. For the same reason, model-based control predicts one should stay with one's prior action more often if a rare-transitioned state was not rewarded. Model-basedness is thus computed via the following equation, where '#' denotes a tally:

$$\text{Model basedness} = \frac{\#\text{ Stay after rare unrewarded} + \#\text{ Switch after rare rewarded}}{\text{Total }\#\text{ Rare transitions}} \quad (6)$$

Relating compulsivity to state transition learning

We characterized state transition learning in each of the 1413 participants from Gillan et al. (2016) by estimating the state transition learning rate (γ) in the ISTL model that best fitted

https://doi.org/10.1017/S0033291721003846 Published online by Cambridge University Press

participants' choices. We then regressed compulsivity scores on the state transition learning rate as well as covariates of age, sex, and IQ in line with Gillan et al. (2016). We then tested whether the use of transition knowledge for planning (β_{MB}) mediated the aforementioned relation between γ and compulsivity. Due to a nonlinear relation between β_{MB} and γ, we log-transformed each to more fully account for shared variance (online Supplementary Fig. S5).

Study 3: quantifying disruptions in state transition learning when transitions change

To determine whether the finding of a compulsivity-related impairment in state transition learning was specific to environments with stable state transitions, we analyzed behavioral data from an additional learning task wherein state transitions changed throughout the task. This task was a component of a larger study ($n = 192$; Sharp et al., 2021) designed primarily to measure how individuals switch between punishment avoidance and reward pursuit goals when learning from reinforcement. Importantly, we assayed participants in this study for compulsivity using the same measure as utilized from Study 1, the OCI-R (Foa et al., 2002). Here, we describe relevant details required to understand how the task probed state transition learning. Full details of Study 3's task and hypotheses can be found in Sharp et al. (2021; Fig. 1).

To win points, participants had to learn from experience the probabilities that each of two actions (pressing '*j*' or '*g*') would transition them to each of two 'states' (presented as colored circles). Importantly, the four transition probabilities slowly drifted across the task according to independent random walks. After choosing an action, participants could reach any combination of states – none, one, or both. Thus, it was possible to reach two states simultaneously, or reach neither state. Participants were presented with an instructed goal each trial: either to seek the 'reward' state (i.e. the gold circle) as getting to this state would reward the participant with 1 point, or to avoid the 'punishment' state (i.e. black circle) as reaching this state would cost the participant 1 point. Reaching a state not referred to in the instructions had no consequence in terms of winning or losing points. However, observing whether the chosen action led to a presently irrelevant state could, via state transition learning, guide choices in subsequent trials wherein instructions made this state relevant. This aspect of the task allows an evaluation of state transition learning that is separate from value learning.

Given that state transitions could not be inferred from explicit instruction in this task, but rather had to be learned via experience, in line with prior work (Dayan, 1993; Russek et al., 2017) we modeled state transition learning as an associative, incremental learning process. Note, this learning process is identical to how state transition learning was modeled in Study 2:

$$P(\text{state} = \text{reward}|\text{press} = g) = P(\text{state} = \text{reward}|\text{press} = j) + \gamma(1 - P(\text{state} = \text{reward}|\text{press} = g)) \quad (7)$$

In this example, the model updates its estimate of the probability that pressing '*j*' leads to the 'reward' state after having just experienced that state transition. All state transitions probabilities are updated in this way every time a state transition is observed.

To choose an action, an agent computes the expected value of each action by multiplying the state transition probability with the values of the possible outcome states, as defined by the trial-specific, instructed goal. Here, the agent is facing an avoid punishment trial, for which reaching a black punishment state results in a loss of 1 point (i.e. a value of −1) and reaching a reward gold state delivers no points and is thus irrelevant. Actions values for pressing '*g*' are thus:

$$\begin{matrix}\text{Model-Based}\\ \text{Action-Values}\end{matrix} \begin{bmatrix} \widehat{Q_{\text{MBreward}}}(\text{press} = g) \\ Q_{\text{MBpunish}}(\text{press} = g) \end{bmatrix} = \begin{matrix}\text{State Transition Estimates}\end{matrix} \begin{bmatrix} P(\text{state} = \widehat{\text{reward}}|\text{press} = g) \\ P(\text{state} = \text{punish}|\text{press} = g) \end{bmatrix} \cdot \begin{matrix}\text{Instructed}\\ \text{Goal}\end{matrix} \widehat{\begin{bmatrix} 0 \\ -1 \end{bmatrix}} \quad (8)$$

Note the model-based action-value for reward (Q_{MBreward}) equals 0 above, as determined by the 0 in the 'Instructed Goal' vector.

In the best-fitting model for Study 3, three strategies influence decision-making in addition to a model-based strategy. A goal-perseveration strategy uses model-based state transition estimates to always both avoid a punishment state (Q_{GPpunish}) and seek the reward state (Q_{GPreward}). Thus, goal-perseveration values are computed by setting the goal vector in Eq. 9 to $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$. A model-free strategy computes action-values (Q_{MF}) solely based on points gained or lost previously in response to each action, thus neglecting state transitions and the presently instructed goal. Finally, an action perseveration strategy consists of a bias (Q_{AP}) to stay with the same action taken at the last trial irrespective of any task feature.

The action-values computed by each strategy are integrated via a weighted sum. Thus, each action value is multiplied by weights (e.g. β_{MBpunish} and β_{MBreward}) that reflect the degree to which the model utilizes the corresponding strategy. For the MB and GP strategies, these utilization weights can differ in strength for reward seeking and punishment avoidance. For the action 'press = *g*' (which we omit from the right side of the equation for concision) the weighted sum is computed as:

$$\begin{aligned} Q_{\text{Integrated}}(\text{press} = g) = {} & \beta_{\text{MBreward}} Q_{\text{MBpunish}} \\ & + \beta_{\text{MBpunish}} Q_{\text{MBpunish}} \\ & + \beta_{\text{GPreward}} Q_{\text{GPreward}} \\ & + \beta_{\text{GPpunish}} Q_{\text{GPpunish}} + \beta_{\text{MF}} Q_{\text{MF}} \\ & + \beta_{\text{AP}} Q_{\text{AP}} \end{aligned} \quad (9)$$

The integrated action value ($Q_{\text{Integrated}}(\text{press} = g)$) is then inputted into a softmax function to generate the policy:

$$P(\text{press} = g) = \frac{e^{Q_{\text{Integrated}}(\text{press} = g)}}{e^{Q_{\text{Integrated}}(\text{press} = g)} + e^{Q_{\text{Integrated}}(\text{press} = j)}} \quad (10)$$

$$P(\text{press} = \text{j}) = 1 - P(\text{press} = \text{g}) \quad (11)$$

https://doi.org/10.1017/S0033291721003846 Published online by Cambridge University Press

Since model-based utilization in this task differed across reward pursuit and punishment avoidance goals, we included both $\beta_{MBreward}$ and $\beta_{MBpunish}$ in subsequent regression analyses to determine how each were related to compulsivity. Importantly, an extensive model comparison showed that the model delineated above outcompeted alternative models that used only a subset of these decision strategies as well as several variations within them (Sharp et al., 2021 for full details).

Results

In the one-step revaluation task, a subject was deemed to have successfully learned state transitions for a given block of learning if they chose both actions that had the highest probability of transitioning them to the two states presented at test. Examination of the data showed participants learned the correct state transitions in the majority of cases. On average, participants performed correctly on 3.22 out of 5 conditions ($N = 174$, S.E.M. = 0.11), significantly greater than chance level (chance = 1.25, $t_{173} = 17.9$, $p < 0.001$).

Compulsivity is associated with poor state transition learning

Rank correlations between psychopathology dimensions and state transition learning showed that OCD symptoms, as measured by the OCI-R, were significantly related to worse transition learning ($r = -0.154$, $p = 0.04$; Fig. 2a). Given that this analysis could not account for co-morbidities between symptoms, we next sought to employ a transdiagnostic analysis that could reveal the contribution of additional symptoms to weakened state transition learning.

To account for psychopathological co-morbidities (Sharp, Miller, & Heller, 2015), and thus potentially improve our ability to capture underlying neurocomputational dysfunctions (Norbom et al., 2019), we conducted an exploratory factor analysis (Fig. 2b) on the set of questionnaire items we administered. A three-factor structure was implied by visual inspection of the scree plot. Parallel analysis supported a strategy of *not* reducing this number of factors, which aligns with prior work using similar questionnaires (Gillan et al., 2016). The three components of interest can largely be interpreted as (1) negative distress, (2) worry, and (3) obsessive-compulsive and fear (OC-Fear; see Fig. 2b). This result agrees with recent work that also found a factor comprising OC and fear/panic symptoms (Wise & Dolan, 2020), as well as with a wealth of psychometric literature that used exploratory factor analytic approaches and demonstrated an overlap between OCD and fear/panic symptomatology (Stasik, 2014; Stasik, Naragon-Gainey, Chmielewski, & Watson, 2012).

Out of the three transdiagnostic factors, only the OC-Fear factor significantly correlated with worse state transition learning ($r = -0.17$, $p = 0.03$; Fig. 2a). We additionally examined how all three factors relate to state transition learning within a single hierarchical Bayesian logistic regression model. The results again showed that only the OC-Fear factor (95% HDI −0.58 to −0.05, mode = −0.33) significantly related to poor state transition learning (Fig. 2c).

The impact of transition learning on model-based decision making

To test more specifically how state transition learning is disrupted in the context of model-based control, we modeled choice data ($n = 1413$; Gillan et al., 2016) from the two-step task using the ISTL model. We first used this model to determine if past signatures of reduced model-based control could be accounted for by a disruption in incremental state-transition learning. For this purpose, we simulated data by having the model 'play' the task with different transition learning rates, and fixing other parameters to fitted group means. In so doing, we show that having a suboptimal high state transition learning rate reproduces the empirical effect reported in Daw et al. (2011), which was deemed to reflect decreased model-based planning (Fig. 3a).

To determine whether participants learned state transitions incrementally or via counting, and whether they differed from one another in their transition learning rates, we conducted an extensive comparison between different computational models in terms of how well each model explained participants' choices (online Supplementary Fig. S3). We found that participants' choices were best explained by the ISTL model, which features an individually fit incremental learning rate for state transitions (log Bayes factor = 274.78 compared to the Typical model used to analyze two-step task data). Model recovery tests demonstrated our analysis successfully recovered the true model from simulated data, validating our model selection procedure (online Supplementary Note 1).

To determine whether differences among participants in model-based control reflected differences in state transition learning rate (γ) and in the use of transition knowledge for making choices (β_{MB}, 'model-based beta'), we fitted the values of these two parameters of the computational model to each participants' choices, and we regressed onto them a model-agnostic behavioral signature of model-based control. This signature ('model-basedness') tallies optimal switch behavior that should occur after participants experience rare transitions. We found that the model-based β, β_{MB} ($\beta = 0.01$, S.E. = 0.003, $p = 001$) and the state transition learning rate, γ ($\beta = -0.09$, S.E. = 0.006, $p < 0.001$) each predicted unique variance in model-based behavior.

Given that the recoverability of γ was expectedly low when β_{MB} was itself very small, we computed the same regression on a subsample of participants ($n = 409$) with a relatively high β_{MB} ($\beta_{MB} > 2.5$). We determined this threshold for β_{MB} by using increments of 0.5 and selecting the threshold that maximized parameter recoverability of state transition learning rate while also ensuring there was no significant difference in the mean level of compulsivity across the reduced and full samples (mean difference = −0.06, S.D. = 0.93, $t_{408} = 1.80$, $p = 0.07$). Enhanced recoverability when excluding low-β_{MB} participants manifested in a substantial increase in the correlation between estimated and true γ in simulated data ($r = 0.61$ compared to $r = 0.22$; difference between correlations: $Z_{408} = 6.46$, $p \leq 0.001$). The regression results for this subset of participants agreed with the results for the full sample (β_{MB}: $\beta = 0.043$, S.E. = 0.009, $p < 0.001$ and γ: $\beta = -0.054$, S.E. = 0.008, $p < 0.001$; Fig. 3b). Finally, we further confirmed the ability of the model to explain model-basedness in the empirical data using a posterior predictive check wherein we replicated the empirical effects noted above in synthetic data ($n = 400$) generated from the ISTL model and individual-level best-fitting parameters (β_{MB}: $\beta = 0.34$, S.E. = 0.006, $p < 0.001$ and γ: $\beta = -0.068$, S.E. = 0.007, $p \leq 0.001$; Fig. 3c).

In a stable environment, high-compulsivity participants update state-transition estimates too fast

To determine whether compulsivity is related to state transition learning (γ), we regressed the former onto the latter, controlling

https://doi.org/10.1017/S0033291721003846 Published online by Cambridge University Press

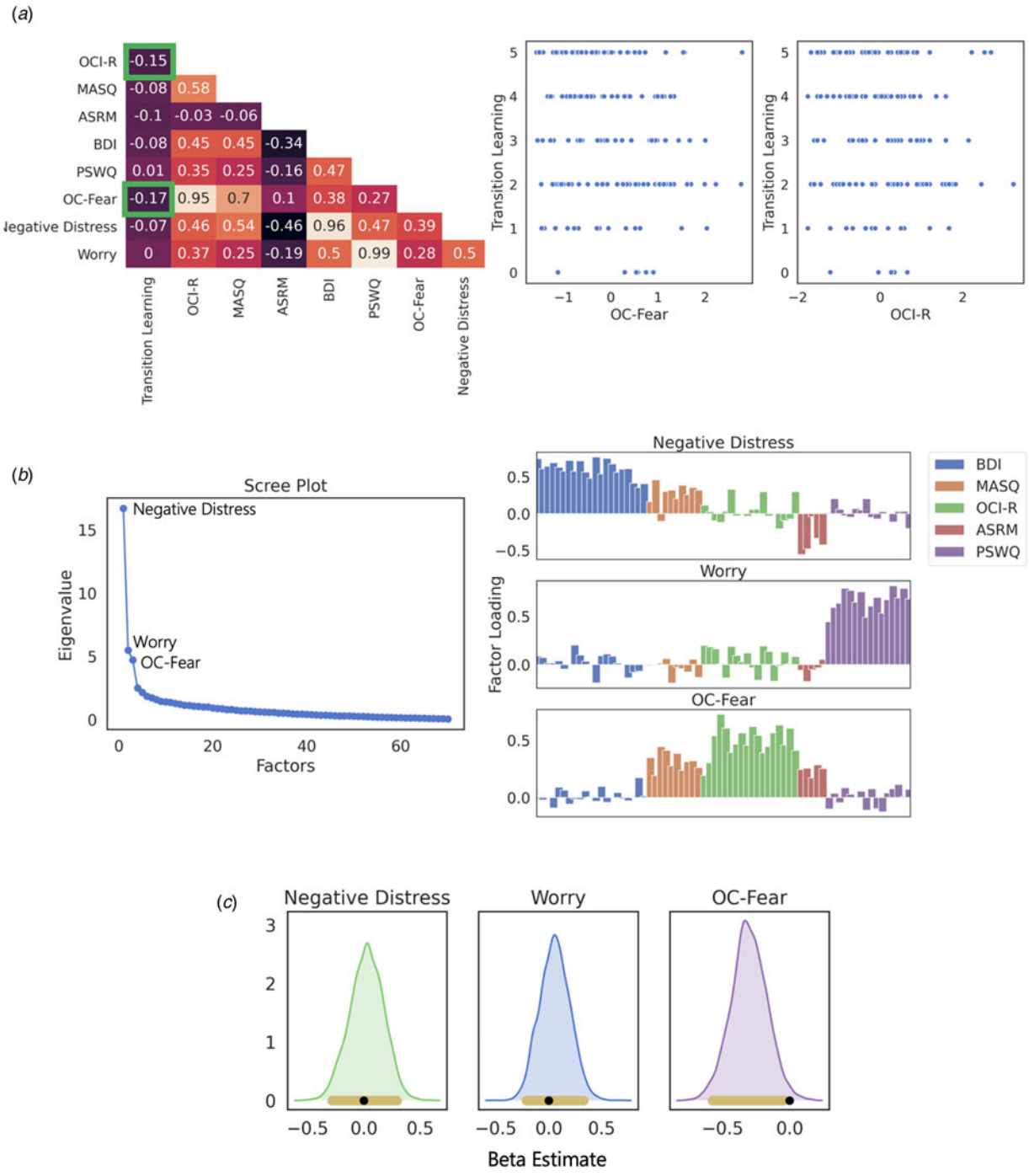


Fig. 2. Study 1: transition learning and psychopathology. (a) Correlations between psychopathology dimensions, transdiagnostic factors, and choice behavior. The left plot comprises correlations between psychopathology dimensions (denoted by their associated questionnaire's abbreviation), transdiagnostic factors derived from exploratory factor analysis, and how well participants learned state transitions. We outline in green the two significant associations involving transition learning in the task, and plot the raw data comprising these correlations in the middle and right plots ($p < 0.05$). (b) Exploratory factor analysis of psychopathology dimensions. Left is the scree plot showing that variance explained plateaus after the first three factors. The eigenvalues for the three components were 16.70, 5.50, and 4.70 for the Negative Distress, Worry, and OC-Fear factors, respectively. The bottom bar plots show the composition of each of these factors in terms of the factor loading of each individual question from all five questionnaires used. Each questionnaire is denoted by a different color. Note, the dataset met both the Bartlett's test of sphericity ($p \leq 0.001$) and Kaiser-Meyer-Olkin test of sampling adequacy (KMO = 0.83). (c) Posterior density plots estimating the effects of the three latent psychopathology factors on transition learning. The black bar denotes the ROPE and the yellow bar the 95% HDI. Only the OC-Fear components has the ROPE entirely outside its HDI. The width of the ROPE was defined as 10% of the standard deviation of the posterior distribution (i.e.±0.01).

for age, gender, and IQ. This showed that state transition learning and compulsivity were positively related ($\beta = 0.15$, S.E. = 0.05, $p = 0.003$). Next, replicating the core finding in Gillan et al. (2016), we showed that utilization of transition knowledge (β_{MB}) was inversely related to compulsivity ($\beta = -0.12$, S.E. = 0.03, $p < 0.001$). Whereas β_{MB} decreased with compulsivity, γ increased with compulsivity. Thus, that both were related to compulsivity cannot be explained by an identifiability tradeoff between them,

https://doi.org/10.1017/S0033291721003846 Published online by Cambridge University Press

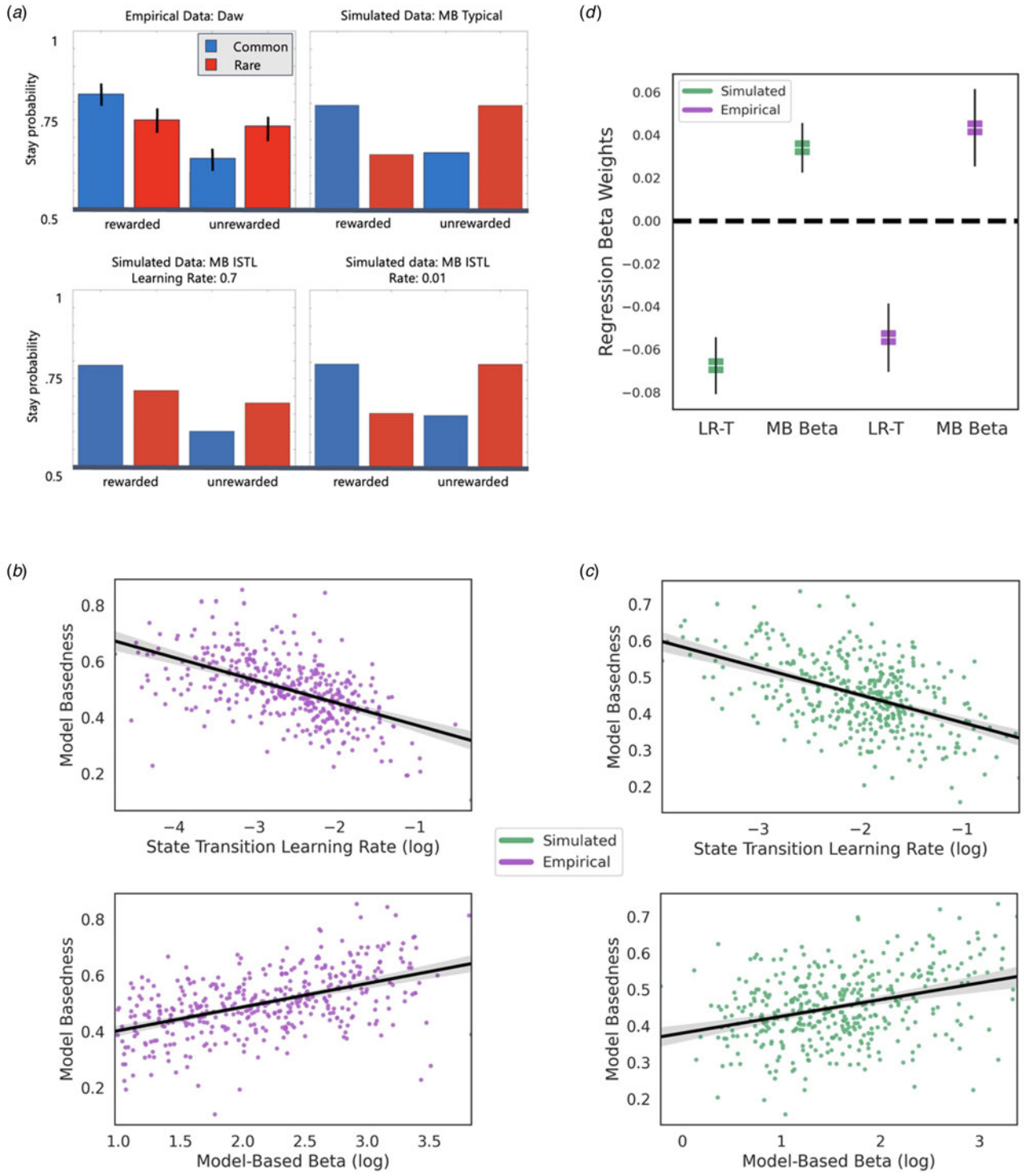


Fig. 3. Study 2: state transition learning rate and model-based behavior in the two-step task. (a) A high state transition learning rate can produce behavior resembling compromised model-based control. The plots show the proportion of times a participant (real or simulated) engaged the same action they deployed in the preceding trial ('Stay probability') conditioned on whether or not one's prior action was followed by a common or rare state transition and whether reward was administered. The behavioral signature of model-basedness is shown in the top right plot (Simulated Data: MB Typical). The deviation of Daw et al.'s real data (top left panel) from this signature was deemed to reflect reduced utilization of intact transition matrix knowledge when making decisions, relative to a competing model-free system. The bottom row depicts simulations of the ISTL algorithm which gradually learns the transition matrix from experienced state prediction errors. The bottom left plot shows that the same qualitative pattern found in Daw et al.'s empirical data (top left) can emerge due to a fast transition learning rate, even in the absence of a putative model-free system. (b) Model-basedness as a function of transition learning rate and model-based β in empirical data. Model-basedness quantifies the degree to which participants complied with the behavioral signature of model-based choice as shown in panel A (e.g. switching after a rare transition was rewarded; see Methods). The top subplot reflects how model-basedness inversely covaries with state transition learning rate whereas the bottom subplot shows the positive relationship between model-basedness and model-based β. Both were generated using a subsample of participants with high (>2.5) model-based control. (c) Model-basedness as a function of transition learning rate and model-based β in simulated data. As a post-predictive check, we simulated the data using the winning ISTL model and best-fitting parameters, which generated the same effects as the empirical data in (c). (d) Regression weights of computational parameters in explaining model-basedness.

https://doi.org/10.1017/S0033291721003846 Published online by Cambridge University Press

since this (anticorrelated) tradeoff would have made both parameters share the same relationship with compulsivity.

In a given task, a state transition learning rate that is too high or too low may result in misestimations of the true state transition probabilities. Thus, we next asked whether the higher state transition learning rates associated with compulsivity were suboptimal or well-suited to the experimental task. To test this, we simulated agents with varying state transition learning rates, while setting β_{MB} to a high value (within the empirical range) and all other parameters to group-fitted means. Each agent played the task 1 000 000 times and the average reward for each state transition learning rate was computed. The results showed that the higher learning rates associated with compulsivity were suboptimal, in the sense that they made the simulated agent win less reward (Fig. 4a).

Subsequently, we sought to determine whether the two parameters, state transition learning rate (γ) and model-based beta (β_{MB}), independently contribute to explaining compulsivity, or whether, as proposed by Seow et al. (2021), variations in γ can account for known compulsivity-related deficits in β_{MB}. Consistent with the latter possibility, we found that γ anticorrelated with β_{MB} ($\beta = -0.98$, S.E. = 0.04, $p < 0.001$). Thus, as all three variables were correlated with one another, we tested for mediation by regressing compulsivity on both β_{MB} and γ in the same model. This demonstrated that β_{MB} ($\beta = -0.10$, S.E. = 0.03, $p = 0.001$) fully mediated the relationship between γ and compulsivity ($\beta = 0.05$, S.E. = 0.06, $p = 0.40$). To further validate this result, we additionally performed the mediation analysis on a subsample of participants ($n = 409$) that included only participants with a β_{MB} estimate of 2.5 or greater, for which the true learning rate is more accurately recovered. We again found that γ was associated with compulsivity ($\beta = 0.14$, S.E. = 0.06, $p = 0.02$) and this relationship was fully mediated by β_{MB} (β_{MB}: $\beta = -0.23$, S.E. = 0.09, $p = 0.006$; γ: $\beta = 0.0004$, S.E. = 0.08, $p = 0.99$).

In a changing environment, high-compulsivity participants update state-transition estimates too slowly

To determine whether a compulsivity-related disruption in state transition learning is due to learning too quickly or due to a failure in adapting the rate of state transition learning to the pace at which state transitions change, we next examined learning about continuously changing state transitions. To do so, we re-analyzed data from a recently completed study in our lab (Sharp et al., 2021) in which participants were required to learn changing state transitions in order to pursue trial-specific goals [see Methods, Study 3 for brief task and model descriptions; Sharp et al. (2021) for in-depth details of task and model comparison]. Unlike in Study 2, here we found that compulsivity was associated with lower state-transition learning rates (i.e. γ; $\beta = -1.69$, S.E. = 0.8, $p = 0.035$).

To determine whether such slow state-transition learning was detrimental to task performance, as in Study 2, we simulated task performance and computed the average reward that could be earned with various state transition learning rates. This showed that the lower state transition learning rates associated with compulsivity were suboptimal (Fig. 4b). Thus, in a changing environment, which warrants faster learning than a stable environment, individuals with compulsivity update their estimates of state transitions at too slow a rate.

In this study, neither β_{MB} parameter was significantly associated with compulsivity. However, it is important to stress that

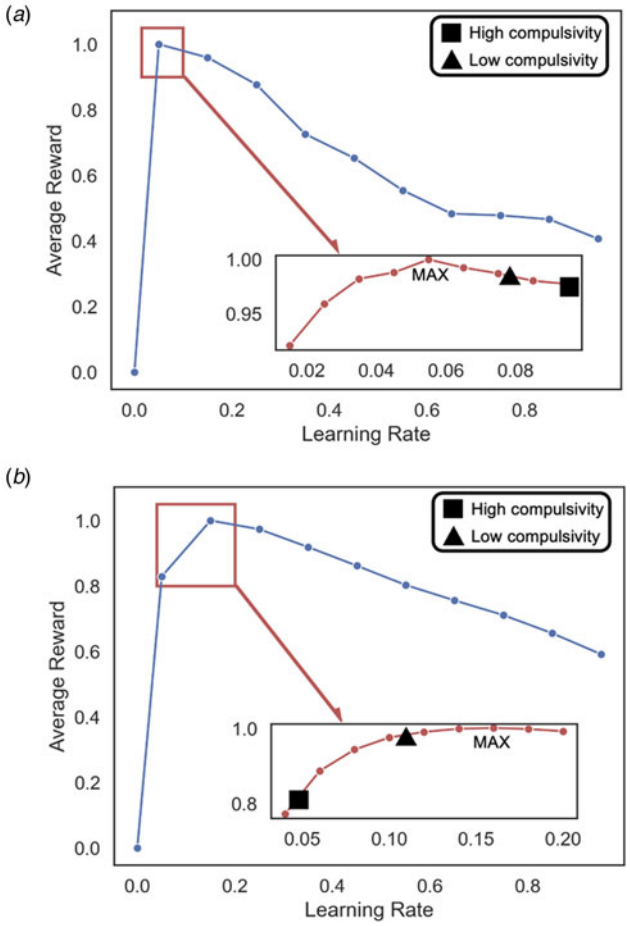


Fig. 4. Compulsivity-associated state transition learning rates are suboptimal. (a) Compulsivity is associated with sub-optimally fast state transition learning in a stable environment (Study 2). We simulated agents that played the exact same task as described in Gillan et al. (2016) and plot the min-max normalized average reward. We instantiated agents with a model-based β weight that maximized reward earned but was still within the tail of the empirical range, using a selection procedure for extreme values in skewed distributions (Rousseeuw & Hubert, 2011). We set all other parameters to their group-fitted medians for distributions that were highly skewed (*z*-score >4, which was the lowest among statistically significant *z*-scores), and group-fitted means otherwise. Agents played the game 100 000 times with different state transition learning rates [sampling from (0,1) in increments of 0.1; main plot], and 1 000 000 times within a region of interest around the optimal and empirical learning rates (denoted by red box) for increased precision. The low- and high-compulsivity groups included participants scoring <–1 and >1 on the standardized scale of compulsivity factor derived in Gillan et al. (2016). Medians from each group (due to their skew) are plotted on the inset plots. (b) Compulsivity is associated with sub-optimally slow state transition learning in a changing environment (Study 3). Plots were generated using the procedure described in panel A, here applied to the Study 3 model.

Study 3's task involved explicit instruction on a trial-by-trial basis to prompt model-based flexibility, which has been shown to significantly increase model-based control, and reduce between-subject variance in this regard (da Silva & Hare, 2020). Thus, β_{MB} values found here may not be comparable with those found in prior studies.

Discussion

We showed in Study 1 that compulsivity was associated with impaired learning of state transitions in a novel revaluation task that was designed to isolate the effects of transition learning.

https://doi.org/10.1017/S0033291721003846 Published online by Cambridge University Press

Additionally, Study 1 found preliminary evidence that state transition learning deficits are associated with a transdiagnostic factor which is influenced in part by symptoms of fear. We next demonstrated across two independent studies that compulsivity is related to state transition learning impairments, in both one- and two-step decision tasks, and in the presence of stable or changing state transitions. In Study 2, we showed that when state transitions are stable, compulsivity is associated with a sub-optimally high state transition learning rate, whereas Study 3 demonstrated that when state transitions change, compulsivity is associated with a sub-optimally low state transition learning rate.

These findings extend recent evidence that compulsivity is associated with greater uncertainty regarding state transitions (Fradkin, Adams, Parr, Roiser, & Huppert, 2020a, 2020b; Seow et al., 2021). For example, previous work implicated disruptions in uncertainty-related learning of shifting cue-outcome contingency (Fradkin, Ludwig, Eldar, & Huppert, 2020b). Here, our finding that compulsivity-related disruptions in the rate of state-transition learning depended on whether state transitions changed over the course of the task suggests that individuals with compulsivity have trouble estimating higher-order statistics, such as state transition volatility (e.g. Behrens, Woolrich, Walton, & Rushworth, 2007). Future work could seek to manipulate the rate of change of state transitions, within task, in order to determine if the aforementioned hypothesis can parsimoniously explain the present results as well as related findings in prior work (e.g. Fradkin et al., 2020a).

One caveat to Study 1's results is that they do not account for multiple comparisons across the different measures provided by the one-step revaluation task. However, the possibility of a false-positive is mitigated by the results of Study 2 and Study 3, which offer two 'conceptual replications' of Study 1's core finding – that compulsivity is associated with disrupted state transition learning. A conceptual replication tests the same hypothesis using different task designs to determine whether the core hypothesis is robust under task variation (Watts, Duncan, & Quan, 2018). It has been argued that a conceptual replication is one of the most effective ways to rule out false-positive findings (Crüwell et al., 2019).

By contrast, findings related to fear symptoms as captured by an anxious arousal measure were only implicated by Study 1 and thus should be considered preliminary. Nevertheless, we note that our finding of an OC-Fear factor aligns with recent psychometric literature and online computational psychiatry studies in large samples (Levin-Aspenson, Watson, Ellickson-Larew, Stanton, & Stasik-O'Brien, 2021; Stasik, 2014; Wise & Dolan, 2020) showing, across several factor analytic approaches, that fear and OCD symptoms converge. This is unsurprising given that, in many instances, fear of an aversive outcome (e.g. contamination) may partly be the cause of compulsions. That said, future work should determine if fear-related impairments in transition learning are entirely driven by compulsivity, especially given that compulsivity scores loaded more strongly onto our OC-Fear factor. In any case, this finding does not contradict a previously observed lack of association between anxiety and model-based control (Gillan et al., 2020b; Heller, Ezie, Otto, & Timpano, 2018), given the former studies did not specifically assay fear symptoms as distinct from classical anxiety symptoms (e.g. worry). Indeed, worry in Study 1 was not associated with state transition learning deficits.

The factor analysis employed in Study 1 was limited by the sample size afforded by study resources, which entailed a low item:participant ratio. Additionally, we could not mitigate this problem by performing the factor analysis on subscales (Gillan et al., 2020a, 2020b), as only one of five questionnaires was comprised of subscales. Thus, validation of the specific factor structure found and its relation to state transition learning would require direct replication on a larger sample size.

Finally, it is important to ask why the mediational hypothesis (Seow et al., 2021), that reduced utilization of model-based control mediates a relation between poor transition learning and compulsivity, was only found in Study 2. First, Study 3 included a one-step task wherein utilization of model-based control is potentially less effortful to leverage than in the two-step task. Second, and more importantly, Study 3's task included trial-by-trial instruction that repeatedly signaled to participants the need for model-based control. Prior work has shown that such instruction greatly improves model-based control (da Silva & Hare, 2020).

Future work should determine how other aspects of the utilization and consolidation of cognitive maps, such as episodic retrieval (Talmi, Lohnas, & Daw, 2019) and offline replay of trajectories (Eldar, Lièvre, Dayan, & Dolan, 2020), are affected in compulsivity. Finally, developmental work is crucial to understand whether a state transition learning deficit emerges before compulsive symptomatology, or whether it is an effect of this pathology.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291721003846

Acknowledgements. P.B.S. is supported by a Fulbright postdoctoral fellowship. E.E. is supported by NIH grants R01MH124092 and R01MH125564, ISF grant 1094/20 and US-Israel BSF grant 2019801. R.J.D. holds a Wellcome Trust Investigator award (098362/Z/12/Z). The Max Planck UCL Centre for Computational Psychiatry and Ageing Research is a joint initiative supported by the Max Planck Society and University College London. We thank Toby Wise, Evan Russek, Jessica McFadyen, and Isaac Fradkin for comments on previous drafts.

Conflict of interest. None.

References

Altman, E. G., Hedeker, D., Peterson, J. L., & Davis, J. M. (1997). The Altman self-rating mania scale. *Biological Psychiatry*, *42*(10), 948–955.

Alvares, G. A., Balleine, B. W., & Guastella, A. J. (2014). Impairments in goal-directed actions predict treatment response to cognitive-behavioral therapy in social anxiety disorder. *PLoS ONE*, *9*(4), e94778.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407.

Bandura, A. (1978). Reflections on self-efficacy. *Advances in Behavior Research and Therapy*, *1*(4), 237–269.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory-II. *San Antonio*, *78*(2), 490–498.

Behrens, T. E., Muller, T. H., Whittington, J. C., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron*, *100*(2), 490–509.

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Cartwright-Hatton, S., & Wells, A. (1997). Beliefs about worry and intrusions: The Meta-Cognitions Questionnaire and its correlates. *Journal of Anxiety Disorders*, *11*(3), 279–296.

Casillas, A., & Clark, L.A (2000) The Mini mood and anxiety symptom questionnaire (Mini-MASQ). (2000, May). Paper presented at the 72nd annual Meeting of the Midwestern Psychological Association, Chicago, IL.

Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J. C., … Schulte-Mecklenbeck, M. (2019). Seven easy steps to open science. *Zeitschrift für Psychologie*, *227*(4), 237–248.

da Silva, C. F., & Hare, T. A. (2020). Humans primarily use model-based inference in the two-stage task. *Nature Human Behavior*, *4*(10), 1–14.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215.

Dayan, P. (1993). Improving generalization for temporal difference learning: The successor representation. *Neural Computation*, *5*(4), 613–624.

Eldar, E., Lièvre, G., Dayan, P., & Dolan, R. J. (2020). The roles of online and offline replay in planning. *Elife*, *9*, e56911.

Eldar, E., & Niv, Y. (2015). Interaction between emotional state and learning underlies mood instability. *Nature Communications*, *6*(1), 1–10.

Eldar, E., Roth, C., Dayan, P., & Dolan, R. J. (2018). Decodability of reward learning signals predicts mood fluctuations. *Current Biology*, *28*(9), 1433–1439.

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The Obsessive-Compulsive Inventory: Development and validation of a short version. *Psychological Assessment*, *14*(4), 485.

Fradkin, I., Adams, R. A., Parr, T., Roiser, J. P., & Huppert, J. D. (2020a). Searching for an anchor in an unpredictable world: A computational model of obsessive compulsive disorder. *Psychological Review*, *127*(5), 672–699.

Fradkin, I., Ludwig, C., Eldar, E., & Huppert, J. D. (2020b). Doubting what you already know: Uncertainty regarding state transitions is associated with obsessive compulsive symptoms. *PLoS Computational Biology*, *16*(2), e1007634.

Gillan, C. M., Kalanthroff, E., Evans, M., Weingarden, H. M., Jacoby, R. J., Gershkovich, M., … Sokol, Y. (2020a). Comparison of the association between goal-directed planning and self-reported compulsivity vs obsessive-compulsive disorder diagnosis. *JAMA Psychiatry*, *77*(1), 77–85.

Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., & Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *Elife*, *5*, e11305.

Gillan, C. M., Vaghi, M. M., Hezemans, F. H., van Ghesel Grothe, S., Dafflon, J., Brühl, A. B., … Robbins, T. W. (2020b). Experimentally induced and real-world anxiety have no demonstrable effect on goal-directed behavior. *Psychological Medicine*, *51*(9), 1467–1478.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*(4), 585–595.

Grupe, D. W., & Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: An integrated neurobiological and psychological perspective. *Nature Reviews Neuroscience*, *14*(7), 488–501.

Heller, A. S., Ezie, C. C., Otto, A. R., & Timpano, K. R. (2018). Model-based learning and individual differences in depression: The moderating role of stress. *Behavior Research and Therapy*, *111*, 19–26.

Konovalov, A., & Krajbich, I. (2020). Mouse tracking reveals structure knowledge in the absence of model-based choice. *Nature Communications*, *11*(1), 1–9.

Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science*, *1*(2), 270–280.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206.

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the open affective standardized image set (OASIS). *Behavior Research Methods*, *49*(2), 457–470.

Levin-Aspenson, H. F., Watson, D., Ellickson-Larew, S., Stanton, K., & Stasik-O'Brien, S. M. (2021). Beyond distress and fear: Differential psychopathology correlates of PTSD symptom clusters. *Journal of Affective Disorders*, *284*, 9–17.

Mathews, A. (1990). Why worry? The cognitive function of anxiety. *Behaviour Research and Therapy*, *28*(6), 455–468.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn state worry questionnaire. *Behavior Research and Therapy*, *28*(6), 487–495.

Norbom, L. B., Doan, N. T., Alnæs, D., Kaufmann, T., Moberget, T., Rokicki, J., & Tamnes, C. K. (2019). Probing brain developmental patterns of myelination and associations with psychopathology in youths using gray/white matter contrast. *Biological Psychiatry*, *85*(5), 389–398.

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences*, *110*(52), 20941–20946.

Rouhani, N., Wimmer, G. E., Schneier, F. R., Fyer, A. J., Shohamy, D., & Simpson, H. B. (2019). Impaired generalization of reward but not loss in obsessive–compulsive disorder. *Depression and Anxiety*, *36*(2), 121–129.

Rousseeuw, P. J., & Hubert, M. (2011). Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(1), 73–79.

Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., Daw, N. D., & Daunizeau, J. (2017). Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLOS Computational Biology*, *13*(9), e1005768. doi: 10.1371/journal.pcbi.1005768.

Seow, T. X., Benoit, E., Dempsey, C., Jennings, M., Maxwell, A., O'Connell, R., … Gillan, C. M. (2021). Model-based planning deficits in compulsivity are linked to faulty neural representations of task structure. *Journal of Neuroscience*, *41*(30), 6539–6550.

Sharp, P. B., Miller, G. A., & Heller, W. (2015). Transdiagnostic dimensions of anxiety: Neural mechanisms, executive functions, and new directions. *International Journal of Psychophysiology*, *98*(2), 365–377.

Sharp, P.B., Russek, E.M., Huys, QJM, Dolan, RJ, & Eldar, E. (2021). Disengaging punishment avoidance is difficult for humans. *PsyArxiv*. https://psyarxiv.com/63fq8/.

Stan Development Team. (2020). Stan Modeling Language Users Guide and Reference Manual. https://mc-stan.org.

Stasik, S. M. (2014). Nosological placement of obsessive-compulsive and related disorders: A comprehensive structural analysis. *University of Notre Dame*.

Stasik, S. M., Naragon-Gainey, K., Chmielewski, M., & Watson, D. (2012). Core OCD symptoms: Exploration of specificity and relations with psychopathology. *Journal of Anxiety Disorders*, *26*(8), 859–870.

Talmi, D., Lohnas, L. J., & Daw, N. D. (2019). A retrieved context model of the emotional modulation of memory. *Psychological Review*, *126*(4), 455.

Tolman, E. C., & Honzik, C. H. (1930). Introduction and removal of reward, and maze performance in rats. *University of California Publications in Psychology*, *4*, 257–275.

Voon, V., Derbyshire, K., Rück, C., Irvine, M. A., Worbe, Y., Enander, J., … Bullmore, E. T. (2015). Disorders of compulsivity: A common bias towards learning habits. *Molecular Psychiatry*, *20*(3), 345–352.

Watts, T. W., Duncan, G. J., & Quan, H. (2018). Revisiting the marshmallow test: A conceptual replication investigating links between early delay of gratification and later outcomes. *Psychological Science*, *29*(7), 1159–1177.

Wise, T., & Dolan, R. J. (2020). Associations between aversive learning processes and transdiagnostic psychiatric symptoms in a general population sample. *Nature Communications*, *11*(1), 1–13.

Zorowitz, S., Momennejad, I., & Daw, N. D. (2020). Anxiety, avoidance, and sequential evaluation. *Computational Psychiatry*, *4*, 1–17.

https://doi.org/10.1017/S0033291721003846 Published online by Cambridge University Press