Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making

Lorenz Deserno[a,b,c,1], Quentin J. M. Huys[d,e], Rebecca Boehme[c], Ralph Buchert[f], Hans-Jochen Heinze[a,b,g], Anthony A. Grace[h,i,j], Raymond J. Dolan[k,l], Andreas Heinz[c,m], and Florian Schlagenhauf[a,c]

[a]Max Planck Fellow Group "Cognitive and Affective Control of Behavioral Adaptation", Max Planck Institute for Human Cognitive and Brain Sciences, 04130 Leipzig, Germany; [b]Department of Neurology, Otto von Guericke University, 39118 Magdeburg, Germany; [c]Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charité–Universitätsmedizin Berlin, 10115 Berlin, Germany; [d]Translational Neuromodeling Unit, Institute for Biomedical Engineering, University of Zurich and Swiss Federal Institute of Technology (ETH) Zurich, 8032 Zurich, Switzerland; [e]Department of Psychiatry, Psychotherapy and Psychosomatics, Hospital of Psychiatry, University of Zurich, 8032 Zurich, Switzerland; [f]Department of Nuclear Medicine, Charité–Universitätsmedizin Berlin, 10115 Berlin, Germany; [g]Leibniz Institute for Neurobiology, Otto von Guericke University, 39118 Magdeburg, Germany; Departments of [h]Neuroscience, [i]Psychiatry and [j]Psychology, University of Pittsburgh, Pittsburgh, PA 15260; [k]Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, United Kingdom; [l]Humboldt Universität zu Berlin, Berlin School of Mind and Brain, 10115 Berlin, Germany; and [m]Cluster of Excellence NeuroCure, Charité–Universitätsmedizin Berlin, 10115 Berlin, Germany

Edited by Ranulfo Romo, Universidad Nacional Autonóma de México, Mexico City, D.F., Mexico, and approved December 23, 2014 (received for review September 11, 2014)

Dual system theories suggest that behavioral control is parsed between a deliberative "model-based" and a more reflexive "model-free" system. A balance of control exerted by these systems is thought to be related to dopamine neurotransmission. However, in the absence of direct measures of human dopamine, it remains unknown whether this reflects a quantitative relation with dopamine either in the striatum or other brain areas. Using a sequential decision task performed during functional magnetic resonance imaging, combined with striatal measures of dopamine using [^{18}F]DOPA positron emission tomography, we show that higher presynaptic ventral striatal dopamine levels were associated with a behavioral bias toward more model-based control. Higher presynaptic dopamine in ventral striatum was associated with greater coding of model-based signatures in lateral prefrontal cortex and diminished coding of model-free prediction errors in ventral striatum. Thus, interindividual variability in ventral striatal presynaptic dopamine reflects a balance in the behavioral expression and the neural signatures of model-free and model-based control. Our data provide a novel perspective on how alterations in presynaptic dopamine levels might be accompanied by a disruption of behavioral control as observed in aging or neuropsychiatric diseases such as schizophrenia and addiction.

dopamine | decision making | reinforcement learning | PET | fMRI

Human choice behavior is influenced by both habitual and goal-directed systems (1). For example, having enjoyed a delicious dinner makes another subsequent visit to the same restaurant more likely. Upon returning at a later point, another visit could happen reflexively when walking past the restaurant, or alternatively be planned and involve reflection, for instance, by checking recent customer reviews to bolster against possible changes. These two decision modes differ fundamentally in terms of their control over actions and associated outcome consequences. Reflexive habitual preferences are retrospective and arise from a slow accumulation of rewards via iterative updating of expectations (2), for example by repeating dinner at the same place after having previously enjoyed tasty food there. In contrast, goal-directed behavior requires a prospective consideration of future outcomes associated with a set of actions (3). For example, knowledge that the chef has changed and subsequent reviews have been less good should reduce one's expectations. Thus, in the face of such change, a goal-directed system can adapt quickly, whereas a habitual system needs to experience an actual outcome before it can alter behavior in an adaptive manner (4). This dual-system theory has been formalized within computational models of learning that update expectations based on past rewards ("model-free") or map possible actions to their potential outcomes ("model-based") (5). There is evidence that model-based learning signals during the acquisition of task structure are encoded within prefrontal–parietal cortices, whereas model-free learning signals are encoded in ventral striatum (6). In the sequential decision task used here, a neural dissociation between the two systems has been less easy to define, with prefrontal cortex (PFC) and ventral striatum coding both model-free learning signals and additional model-based signatures (7).

An unresolved question centers on what factors relate to the degree to which an individual's choices reflect the dominance of either model-free or model-based systems of control. Among neuromodulators, dopamine has repeatedly been linked to this balance (1, 8–12), although it is important to acknowledge that other neuromodulatory agents are likely to also play a role (13). Traditionally, dopamine is associated with model-free learning, representing a teaching signal used to update expectations,

Significance

Whether humans make choices based on a deliberative "model-based" or a reflexive "model-free" system of behavioral control remains an ongoing topic of research. Dopamine is implicated in motivational drive as well as in planning future actions. Here, we demonstrate that higher presynaptic dopamine in human ventral striatum is associated with more pronounced model-based behavioral control, as well as an enhanced coding of model-based signatures in lateral prefrontal cortex and diminished coding of model-free learning signals in ventral striatum. Our study links ventral striatal presynaptic dopamine to a balance between two distinct modes of behavioral control in humans. The findings have implications for neuropsychiatric diseases associated with alterations of dopamine neurotransmission and a disrupted balance of behavioral control.

Author contributions: L.D., Q.J.M.H., and F.S. designed research; L.D. and R. Boehme performed research; L.D., Q.J.M.H., R. Boehme, R. Buchert, and F.S. analyzed data; and L.D., Q.J.M.H., R. Boehme, R. Buchert, H.-J.H., A.A.G., R.J.D., A.H., and F.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. Email: deserno@cbs.mpg.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1417219112/-/DCSupplemental.

for example via a temporal difference reward prediction error (14, 15). Potential correlates of this dopamine learning signal have been reported in functional magnetic resonance imaging (fMRI) studies in humans (e.g., ref. 16). On the other hand, individual variation of striatal presynaptic dopamine, quantified using neurochemical imaging, is known to positively relate to variability in "prefrontal" cognitive capacities (17, 18), which might also limit the capacity for model-based learning (19). Indeed, depletion of presynaptic dopamine precursors and Parkinson's disease both compromised goal-directed behavior in a devaluation experiment and a slips-of-action test, whereas habitual learning remained intact (20, 21). Furthermore, a pharmacological challenge with L-DOPA, a manipulation known to boost overall brain dopamine levels, has been shown to enhance model-based over model-free choices in a sequential decision-making task (12). These studies raise the possibility that a balance between model-free and model-based control is intimately related to variations in dopamine levels but they are agnostic as to the likely locus of this influence.

A radiolabeled variant of L-DOPA, [^{18}F]DOPA, allows quantification of individual levels of presynaptic dopamine in vivo by using positron emission tomography (PET) (22). Schlagenhauf et al. (23) used this methodology to show an inverse relationship between ventral striatal presynaptic dopamine levels and an fMRI signal that indexed ventral striatal model-free learning signals. Ventral striatal presynaptic dopamine levels are a candidate marker for a balance between model-free and model-based control in light of evidence that ventral striatal lesions impair model-based learning (24), whereas ventral striatal activation encodes a signature of both model-free and model-based learning (7). Furthermore, as mentioned above, presynaptic dopamine levels in ventral striatum were negatively correlated with ventral striatal model-free learning signals (23).

Here, we combine a two-step sequential decision task during fMRI with [^{18}F]DOPA PET to quantify interindividual differences in striatal presynaptic dopamine levels. Our hypothesis was that interindividual variation in presynaptic levels of striatal dopamine relate to behavioral and neural signatures of model-based and model-free control.

Results

Model-Free Versus Model-Based Control. A balance between model-free and model-based choice behavior was assessed using a two-step decision task in 29 healthy participants (Fig. 1 *A* and *B*). In this task, subjects make two sequential choices between stimulus pairs to receive a monetary reward. At the first stage, each choice option led commonly (70% probability) to one of two pairs of stimuli and rarely (30% probability) to the other pair. After entering the second stage, a second choice was followed by monetary reward or zero outcome, delivered according to slowly changing Gaussian random walks to facilitate continuous updating of action values. A purely model-based learner exploits probabilities in the transition structure from the first to the second stage, whereas a purely model-free learner neglects this task structure. It has been shown that behavior shows influences of both systems (7) (Fig. S1) and at an individual level a balance between model-free and model-based control can be quantified by a hybrid model. This hybrid model combines the decision values of two algorithms according to a weighting factor ω. One algorithm involves model-free temporal difference learning, whereas the other performs a model-based tree search by using explicitly instructed transition probabilities to prospectively update first-stage decision values (*SI Text*). A higher weighting parameter ω indicates a bias toward model-based choices and is our primary measure of interest. The models were implemented as in the original paper (7), and in line with previous studies (7, 12), a hybrid model again best explained choice behavior as shown in a Bayesian model selection procedure (exceedance probability = 0.98; Table S1; ref. 25).

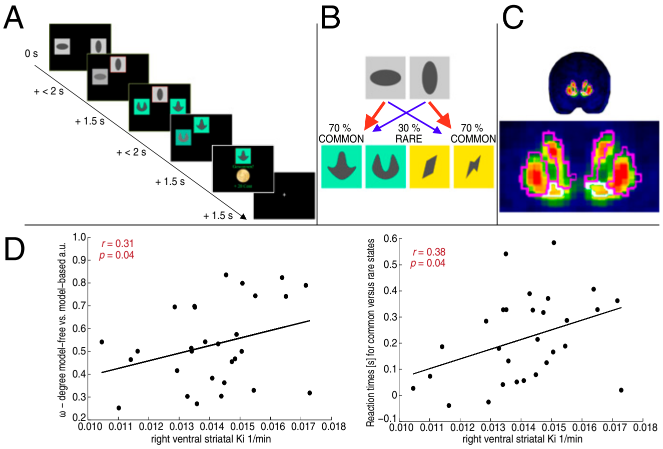

Fig. 1. Behavioral task and relation to presynaptic dopamine. (*A*) Exemplary trial sequence of the two-step decision task and timing. (*B*) Illustration of the state transition matrix. (*C*) Mean voxelwise K_i map of 29 participants and borders of striatal regions of interest. (*D*) Correlation between right ventral striatal K_i and the balance of model-free and model-based choices ω ($r = 0.31$; $P = 0.04$) and between right ventral striatal K_i and the reaction times for common versus rare states ($r = 0.38$; $P = 0.04$).

Striatal Dopamine and a Balance of Behavioral Control. To test whether striatal presynaptic dopamine levels relate to a balance between model-free and model-based choice behavior, we used the weighting parameter ω derived from computational modeling (Table S2) as dependent variable in a linear regression analysis with a quantitative metric of F-DOPA uptake (K_i) from right and left ventral and remaining striatum as independent variables (Fig. 1*C*). This revealed a significant positive relation between K_i in right ventral striatum and the parameter ω (ventral striatum—right: $\beta = 0.43$, $t = 2.16$, $P = 0.04$; left: $\beta = 0.10$, $t = 0.40$, $P = 0.70$; remaining striatum—right: $\beta = 0.10$, $t = 0.34$, $P = 0.73$; left: $\beta = -0.46$, $t = 1.48$, $P = 0.15$; Fig. 1*D*). We repeated this linear regression analysis with presynaptic dopamine from ventral striatum, caudate, and putamen for each hemisphere. As in the initial regression analysis, this revealed that right ventral striatal presynaptic dopamine alone related to the weighting parameter ω (ventral striatum—right: $\beta = 0.46$, $t = 2.22$, $P = 0.04$; left: $\beta = 0.07$, $t = 0.33$, $P = 0.74$; caudate—right: $\beta = -0.04$, $t = 0.14$, $P = 0.89$; left: $\beta = -0.03$, $t = 0.10$, $P = 0.92$; putamen—right: $\beta = 0.09$, $t = 0.33$, $P = 0.74$; left: $\beta = -0.46$, $t = 1.68$, $P = 0.11$). This positive relationship was also consistent with findings from an analysis of stay–switch behavior at the first stage as a function of right ventral striatal presynaptic dopamine (*SI Text*, Fig. S2). In line with our hypothesis, ventral striatal presynaptic dopamine levels were associated with a behavioral bias toward model-based choices.

Our finding of a positive relation between ventral striatal presynaptic dopamine and model-based control indicates that a model-based system is more engaged as a function of higher ventral striatal presynaptic dopamine. This relationship can also be probed via an analysis of second-stage reaction times. In our task, a model-based learner uses knowledge about state transitions and second-stage reaction time differences between common versus rare states should reflect the level of involvement in model-based control. When comparing common and rare states, we found that second-stage reaction times differed significantly (paired t test: mean difference, 218 ± 165 ms SD; $t = 7.10$; $P < 0.001$; Fig. S3). Note that model-free learning cannot account for this effect because it neglects the state transition matrix. Reaction times were significantly slower in rare compared with common states and individual variability in this reaction time difference (most likely slowing down in rare states; Fig. S3) positively related to the parameter ω ($r = 0.59$, $P = 0.001$; Fig. S4), where the latter was inferred independently of reaction times using computational modeling. Crucially, a positive relation between the second-stage reaction time difference for rare versus

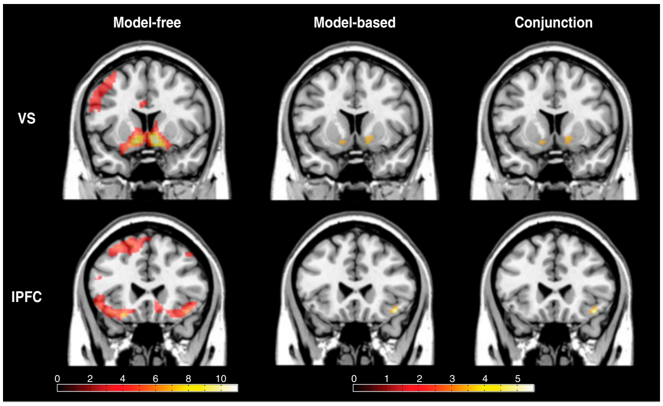

Fig. 2. fMRI results. Model-free prediction errors (*Left*), additional model-based signals (*Middle*), and the conjunction of both (*Right*) in ventral striatum (VS, *Upper*) and lateral prefrontal cortex (lPFC, *Lower*). For display purposes, all statistical maps are thresholded at a minimum T value of 3.24 (corresponding to $P < 0.001$, uncorrected) with a cluster extent $k = 20$. For details, see Table S3.

common states was linked to right ventral striatal presynaptic dopamine (linear regression analysis: ventral striatum—right: $\beta = 0.47$, $t = 2.33$, $P = 0.03$; left: $\beta = 0.03$, $t = 0.14$, $P = 0.89$; remaining striatum—right: $\beta = 0.07$, $t = 0.22$, $P = 0.83$; left: $\beta = -0.32$, $t = -1.02$, $P = 0.32$; Fig. 1*D*). The latter relationship was specific for the second-stage reaction time difference comparing common with rare states, whereas no relationship was evident between presynaptic dopamine levels in ventral striatum and overall reaction times at the second stage of the task (Fig. S4). This analysis further supports the idea that higher levels of ventral striatal presynaptic dopamine relate to more pronounced model-based control in rare task states, where the computational cost of model-based inference is expected to result in slower reaction times.

Neural Signatures of Model-Free and Model-Based Choices. We first replicated the results reported by Daw et al. (7), who showed that ventral striatal blood oxygen level-dependent (BOLD) signals reflect model-free as well as model-based components. Following the same analytic strategy, we first sought to identify brain regions where BOLD responses covaried with model-free prediction errors. We then asked whether these BOLD signals might also incrementally reflect model-based components, by including the difference between model-based and model-free prediction errors as an additional regressor (for details, see *Experimental Procedures*). Positive correlations with model-free prediction errors were observed in a prefrontal-striatal network, including sectors of lateral and medial PFC bilaterally as well as bilateral ventral striatum [$P < 0.05$, familywise error (FWE)-corrected at the peak level for the whole brain; Fig. 2 and Table S3]. The effect of additional model-based components reached significance in the same regions, namely bilateral ventral striatum, right lateral PFC, and medial PFC ($P < 0.05$, FWE-corrected at the peak level for the respective bilateral regions of interest; Fig. 2 and Table S3). The conjunction of model-free and model-based effects reached significance in right lateral PFC and bilateral ventral striatum ($P < 0.05$, FWE-corrected at the peak level for the respective bilateral regions of interest; Fig. 2).

Ventral Striatal Dopamine and Ventral Striatal Model-Free Learning Signals. In previous work (23), we presented evidence for a negative relationship between right ventral striatal presynaptic dopamine levels and model-free prediction errors in right ventral striatum. To replicate this finding, we extracted parameter estimates of model-free prediction errors in right ventral striatum at peak coordinates [$x = 16$, $y = 8$, $z = -8$] from the conjunction contrast within an 8-mm sphere. In an analysis restricted to right ventral striatum based on previous work (23), we again found a negative relationship between ventral striatal coding of model-free prediction errors and ventral striatal presynaptic dopamine levels ($r = -0.37$; $P < 0.05$; Fig. 3*A*). This correlation also remained significant when controlling for presynaptic dopamine levels from other striatal regions (*SI Text*) and when perfoming a voxelwise analysis (*SI Text*, Fig. S5).

Ventral Striatal Dopamine and Neural Model-Based Signatures. Here, we asked whether right ventral striatal presynaptic dopamine levels related to encoding of model-based information. We extracted parameter estimates of the model-based difference regressor for lateral PFC [$x = 42$, $y = 24$, $z = -14$] and ventral striatum [$x = 16$, $y = 8$, $z = -8$] at peak coordinates of the conjunction contrast (surrounded by 8-mm spheres), which were then subjected to an ANOVA with the factor "region" and right ventral striatal K_i as a covariate. We found a significant region by K_i interaction ($F = 5.10$; $P < 0.05$), driven by a significant positive relation between ventral striatal K_i with model-based signatures in lateral PFC ($r = 0.39$; $P < 0.05$; Fig. 3*B*) but not in ventral striatum ($r = -0.07$; $P > 0.7$). This correlation also remained significant when controlling for presynaptic dopamine levels from other striatal regions (*SI Text*) and when perfoming a voxelwise analysis (*SI Text*, Fig. S5). Note that the sensitivity of the PET technique does not allow accurate measures of cortical levels of presynaptic dopamine.

Discussion

Here, we demonstrate that ventral striatal presynaptic dopamine reflects a balance in the behavioral and neural signatures of model-free and model-based control in a two-stage sequential decision-making task. Higher levels of presynaptic dopamine in right ventral striatum were positively related to a greater disposition to make model-based choices. Crucially, higher levels of presynaptic dopamine in right ventral striatum were also associated with stronger model-based coding in lateral PFC and diminished coding of model-free prediction errors in ventral striatum.

Ventral Striatal Dopamine and a Model-Based System. It has been shown previously, using an identical task to the one used here, that administration of L-DOPA increases model-based over model-free choices (12). Using PET, we now demonstrate that interindividual differences in ventral striatal presynaptic dopamine levels are related to this bias toward model-based control. This accords with other studies that report enhanced cognitive capacities in subjects with higher levels of striatal F-DOPA uptake (17, 18). Cognitive capacity, particularly as it relates to working memory function, is also linked to the extent to which individuals exploit model-based control (19). Conceptually, this pattern of results can be explained in a framework of uncertainty-based competition between the two decision systems (5). Thus, participants with higher levels of presynaptic dopamine can be thought of as encoding model-based estimates with higher certainty. At a neural level, we demonstrate that ventral striatal presynaptic dopamine levels relate positively to coding of model-based signatures in lateral PFC and are accompanied by a bias toward more model-based choices. It is conceivable that higher levels of presynaptic dopamine enable lateral PFC to code cognitively demanding model-based information with greater

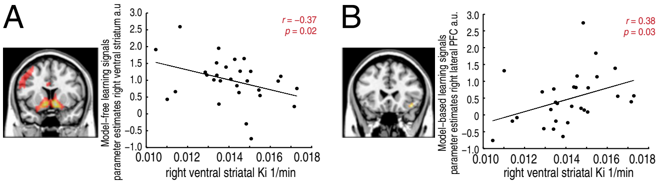

Fig. 3. Presynaptic dopamine and neural learning signatures. Correlation between right ventral striatal presynaptic dopamine K_i and (*A*) model-free learning signals in right ventral striatum ($r = -0.37$; $P = 0.02$) and (*B*) model-based signatures in right lateral prefrontal cortex ($r = 0.38$; $P = 0.03$).

precision, thereby increasing certainty in model-based estimates. As a consequence, a model-based system may exert a greater influence on behavioral control. In a similar vein, dopamine is implicated in a modulation of PFC maintenance processes via a gating of cortical gain, rendering coding of relevant environmental information more robust against noise (11, 26, 27). Indeed, the importance of lateral PFC for model-based inference is supported by findings that theta-burst transcranial magnetic stimulation compromises model-based control in humans (28).

Our analysis of second-stage reaction times, which were affected by the state transition matrix, showed that a response time difference for rare versus common states was positively related to a bias toward more model-based choices. Intriguingly, this reaction time difference for rare versus common states positively correlated with ventral striatal presynaptic dopamine. These results are consistent with an engagement of a slower, computationally more costly model-based system (1, 3). Engagement of a model-based system is more likely after rare transitions as these trials are associated with increased uncertainty in representing an anticipated sequence of actions and outcomes. Furthermore, ventral striatal tonic dopamine is implicated in signaling average reward rates (29), a theoretical proposal that has received recent empirical support (e.g., ref. 30). Nevertheless, in the context of the task used here, ventral striatal presynaptic dopamine levels were not related to invigoration per se as represented by overall reaction times. In participants who used a more model-based strategy, one possible explanation is that faster reaction times in common versus rare states reflect higher expectation of average reward rates, resulting in greater invigoration for a specific action–outcome sequence. However, the role of expected average reward rates, invigoration, and model-based learning requires experimental designs tailored to address this question.

Ventral Striatal Dopamine and a Model-Free System. High levels of ventral striatal presynaptic dopamine can also influence a model-free system as suggested by the inverse correlation with ventral striatal model-free prediction errors, a replication of previous findings (23). This indicates that participants with high levels of ventral striatal presynaptic dopamine show a bias toward a more pronounced model-based form of control and are also characterized by a diminished coding of ventral striatal model-free prediction errors. The hypothesis of uncertainty-based competition (5) might also account for this finding under a premise that higher presynaptic dopamine levels result in larger phasic prediction error dopamine transients. In the reinforcement learning account, this corresponds to an increase in a learning rate within a model-free system. With high model-free learning rates, model-free values change more quickly. Thus, over the course of learning, value changes are more pronounced for single events and a value estimate at a given point in time represents an average across fewer experiences. This could in turn result in greater uncertainty of model-free estimates. Such uncertainty would reduce the weight attached to predictions by a model-free system.

There is substantial evidence that high levels of presynaptic dopamine exert a detrimental effect on NoGo-learning from negative prediction errors and promote Go-learning from positive prediction errors (31). Interestingly, in a previous study (12) as well as in our data, an alternative model with separate learning rates for positive and negative updating provided an inferior fit to the observed choices during the sequential decision task (*SI Text*) and failed to account for the observed enhancing effect of L-DOPA on model-based behavior in the previous study (12). However, we had only Go-trials and future studies with paradigms designed to disentangle a potential role of Go- and NoGo-learning and learning from positive and negative prediction errors in model-free and model-based control are required.

Ventral Striatal Dopamine and a Balance of the Two Systems. Ventral striatal presynaptic dopamine may exert its influence on a balance between the two systems by directly affecting an arbitrator that chooses between the two. Here, it is important to note that model-based signals modulated by ventral striatal presynaptic dopamine levels were located to the inferior part of the lateral PFC. Activation at close coordinates has recently been reported to covary with the reliability of estimates arising from the two decision systems as inferred from a hierarchical computational model (32). The latter finding links the inferior section of lateral PFC to an arbitration process. We note that the study by Lee et al. (32) extends the idea of uncertainty-based competition by identifying two PFC regions, the inferior lateral PFC and the frontopolar cortex, involved in the arbitration of the two systems by weighting the reliability of the predictions from each system. With respect to the present study, this also underlines the importance of the association of model-based signatures in inferior lateral PFC with ventral striatal presynaptic dopamine levels hinting at the possibility that these dopamine levels may be directly involved in the arbitration process. State prediction errors for implicit transition learning were expressed in parietal and dorsolateral PFC (6, 32). Future studies should study locally distinct learning signals in lateral PFC (32) and their hierarchical organization as suggested by models of lateral PFC function (33, 34).

Mechanistic Considerations. With regard to mechanisms, it is important to take into account the intricacies of dopamine neurotransmission. In animal research, learning new reward contingencies is causally linked to time-locked, phasic activation of dopamine neurons (35). We acknowledge that neither fMRI learning signals nor F-DOPA update kinetics can match the dynamical properties of these directly recorded signals. However, phasic dopamine release in ventral striatum selectively facilitates context-dependent inputs to ventral striatal neurons via activation of D_1 receptors (36). This ventral striatal activation removes inhibition of midbrain dopamine neurons resulting in an increase in firing of dopamine neurons leading to an enhanced tonic dopamine influence on ventral striatum (36), potentially indexed by activity of dopa decarboxylase. Thus, larger phasic dopamine transients, which happen in response to unexpected events, may reduce the weight attached to a model-free system and allow model-based inputs to dominate. This could in turn be reflected in overall higher presynaptic dopaminergic activity. Such changes have been demonstrated in animal research (36), and it is conceivable that a long-term dominance of such activity might be reflected in higher presynaptic dopamine levels, as assessed here via F-DOPA PET. Although speculative, this notion is supported by evidence for reliability of F-DOPA uptake quantifications in healthy individuals over a period of 1 y (37). Thus, relatively higher presynaptic dopamine levels could preferentially facilitate signals, which are thought to carry important, context-dependent, model-based information (36). A possible neural architecture for these signals includes the hippocampus and prefrontal cortex (38). In the present study, we did not observe model-based signatures in the hippocampus, which may well be due to the applied analytic strategy and the task design (3), but show that interindividual variability in ventral striatal presynaptic dopamine levels coincide with a greater coding of model-based information in lateral PFC. This finding also resonates with the notion of disrupted presynaptic dopamine function in neurological and psychiatric illnesses (e.g., refs. 39 and 40).

Regarding the neural instantiation of both control systems, animal research has highlighted a dissociation between dorsolateral and dorsomedial striatum, with dorsolateral lesions disrupting habit formation, whereas dorsomedial lesions impact on goal-directed control (41, 42). In the present study, we did not observe a relationship between striatal presynaptic dopamine in either caudate nucleus (the homolog of dorsomedial striatum) or putamen (the homolog of dorsolateral striatum) and model-based fMRI effects (*SI Text*). This may be due to several factors including the choice of experimental task, the type of neural measurement, and also limited homologies between neuroanatomical structures in rodents and primates (43, 44). Furthermore, evidence indicates these structures may encode model-based and model-free value signals (45), quantities that were not assessed here. However, these issues and inconsistencies require clarification in future translational research.

Limitations

The correlative design we deploy precludes any conclusions about causality. This is important when considering factors that may determine individual variability in presynaptic dopamine levels in the healthy population. Here, the orchestration of dopamine and other neuromodulators at a system level should be taken into account. For example, serotonin interferes with aversive processing (46) and learning from negative prediction errors (47), whereas cholinergic influences are linked to an encoding of precision-weighted prediction errors (48). These processes undoubtedly contribute to behavioral control and underline a requirement for a more unified view (49). However, the association between a balance of behavioral control and ventral striatal presynaptic dopamine levels, as demonstrated in the present study, supports the idea that ventral striatum is an important nexus where several inputs converge (50). It remains an open question as to whether the association between ventral striatal presynaptic dopamine and a relative dominance of model-based control in our sequential decision task generalizes to other instances of goal-directed learning and cognitive control. Furthermore, the interpretation of lateralization with respect to right ventral presynaptic dopamine measures is challenging, although this lateralization effect replicates a previous fMRI-PET study (23). Lateralization effects have been reported in human PET studies of the dopamine system (e.g., refs. 51 and 52) and also with respect to the association of these dopamine measurements with reward and motivation (53, 54). However, results in the present study were derived from right-handed participants alone, and all reported correlations remained significant when controlling for dopamine measures from right and left striata.

Conclusion

In summary, we show that interindividual differences in human ventral striatal presynaptic dopamine levels reflect a balance in behavioral and neural signatures of model-free and model-based control. Extending pharmacological challenge findings (12), higher ventral striatal presynaptic dopamine levels were correlated with a bias toward more model-based control. Higher presynaptic dopamine levels were associated with stronger coding of model-based information in lateral PFC and diminished coding of model-free prediction errors in ventral striatum. The link between presynaptic dopamine levels and a balance between model-free and model-based behavioral control has implication for aging as well as psychiatric diseases such as schizophrenia or addiction.

Experimental Procedures

Participants. Twenty-nine right-handed participants (11 females) with a mean age of 28.35 ± 4.95 y (range, 20–39) were included. The research ethics committee of the Charité Universitätmedizin approved the study, and written informed consent was obtained from the participants.

Task. A two-step decision task was implemented as in previous studies (7, 12). The task consisted of a total of 201 trials with two choice stages within each trial. At each stage, participants had to give a forced choice (maximum decision time, 2 s) between two stimuli presented either on two gray boxes at the first stage or two pairs of differently colored boxes at the second stage (Fig. 1). All stimuli were randomly assigned to the left and right position on the screen. The chosen stimulus was surrounded with a red frame, moved to the top of the screen after completion of the 2-s decision phase and remained there for 1.5 s. Subsequently, participants entered the second stage, and a reward was delivered after a second-stage choice. Reward probabilities of second-stage stimuli were identical to those of Daw et al. (7). Each first-stage choice was associated with one pair of the second-stage stimuli via a fixed transition probability of 70%, which did not change during the experiment. Trials were separated by an exponentially distributed intertrial interval with a mean of 2 s. Before the experiment and similar to Daw et al. (7), participants were explicitly informed that the transition structure would stay constant throughout the task. Additionally, information was provided about the independence of reward probabilities and their dynamic change over the course of the experiment. Participants were instructed to maximize reward, which they received as monetary payout after completion of the task. Before entering the scanner, participants performed a shortened version of the task (55 trials) with different reward probabilities and stimuli.

Computational Modeling. As in previous studies, we fit a hybrid model to the observed behavioral data (7, 12). This model weights the relative influence of model-free and model-based choice values, which only differ with respect to first-stage values. This weighting, the relative influence of both systems on first-stage values, is expressed via the parameter ω. The special cases of this model refer to $\omega = 1$ or $\omega = 0$ reflecting purely model-based or purely model-free control over first-stage values, respectively. For details on the model itself, fitting, and model selection, see *SI Text*.

Magnetic Resonance Imaging. Functional imaging was performed using a 3-tesla Siemens Trio scanner to acquire gradient echo T2*-weighted echo-planar images with BOLD contrast. Covering the whole brain, 40 slices were acquired in oblique orientation at 20° to the anterior commissure–posterior commissure line and in interleaved order with 2.5-mm thickness, $3 \times 3\text{-mm}^2$ in-plane voxel resolution, 0.5-mm gap between slices, repetition time of 2.09 s, echo time of 22 ms, and a flip angle α of 90°. Before functional scanning, a field map was collected to account for individual homogeneity differences of the magnetic field. T1-weighted structural images were also acquired.

Analysis of fMRI Data. fMRI data were analyzed using SPM8 (www.fil.ion.ucl.ac.uk/spm/software/spm8/). For preprocessing of fMRI data, see *SI Text*. Before statistical analysis, data were high-pass filtered with a cutoff of 128 s. An event-related analysis was applied to the images on two levels using the general linear model approach as implemented in SPM8. As in the original paper by Daw et al. (7), the analysis comprised two time points within each trial when prediction errors arise: at onsets of the second stage and at reward delivery. Prediction errors at second-stage onsets compare values of first- and second-stage stimuli and can therefore be varied with respect to the weighting parameter ω of the hybrid algorithm. Both time points were entered into the first-level model as one regressor, which was parametrically modulated by (*i*) model-free prediction errors and (*ii*) by the difference between model-based and model-free prediction errors, which refers to the partial derivative of the value function with respect to ω and reflects the difference between model-based and model-free values. For details of the first-level model, see *SI Text*. Two contrasts of interest, model-free prediction errors and the difference regressor reflecting additional model-based predictions, were taken to a second-level random-effects model. For correction of multiple comparisons, FWE correction was applied using small-volume correction for bilateral volumes of interest of the ventral striatum (as obtained in the IBASPM atlas as part of the WFU Pick Atlas), lateral PFC (comprising the middle and inferior frontal gyrus as part of Automated Anatomic Labeling Atlas), and medial PFC (comprising the superior medial frontal and medial orbital gyrus as part of Automated Anatomic Labeling Atlas).

Positron Emission Tomography. Data were acquired using a Philips Gemini TF16 time-of-flight PET/CT scanner in 3D mode. After a low-dose transmission CT scan for attenuation correction, a dynamic 3D "list-mode" emission recording lasting 60 min was started simultaneously with i.v. injection of 200 MBq of F-DOPA as a slow bolus. The emission data were framed into 20 dynamic frames (3 × 20 s, 3 × 1 min, 3 × 2 min, 3 × 3 min, 7 × 5 min, 1 × 6 min) and reconstructed with an isotropic voxel size of 2 mm.

Analysis of PET Data. PET data were analyzed using SPM8. For preprocessing of PET data, see *SI Text*. A quantitative measure of dopamine synthesis capacity (K_i) was obtained voxel-by-voxel using the Gjedde–Patlak linear graphical analysis with the cerebellum as reference region (55). Frames recorded between 20 and 60 min of the emission recording were used for linear fit. The time activity curve of the cerebellum (excluding Vermis) was extracted using a mask from the WFU Pick Atlas. Mean K_i values were extracted from the voxelwise maps using the same mask of ventral striatum as for the fMRI analysis and a corresponding mask of remaining striatal parts taken from the same atlas (compare Fig. 1).

Combination of PET and Behavioral Data. Right and left K_i from ventral and remaining striatum were entered as independent variables into a linear regression analysis with modeling-derived balance of model-free and model-based choice behavior ω as dependent variable.

Combination of PET and fMRI Data. The main focus of the present study was to examine the relationship between presynaptic dopamine and additional model-based brain signals. Specifically, we aimed to answer the question whether presynaptic dopamine relates to model-based signatures in ventral striatum or PFC. Parameter estimates were extracted at peak coordinates (surrounded with 8-mm spheres) of the conjunction of model-free and model-based effects. First, and based on previous work (23), parameter estimates of right ventral striatal model-free prediction errors were correlated with K_i from right ventral striatum. Second, parameter estimates of additional model-based effects in right ventral striatum and right lateral PFC were entered into a repeated-measures ANOVA with the factor region. K_i from right ventral striatum was entered as a covariate. For multimodal imaging analysis, K_i from right ventral striatum was chosen because it explained individual differences in the weight of model-free and model-based decisions.

ACKNOWLEDGMENTS. We thank Anne Pankow, Teresa Katthagen, Yu Fukuda, and Tobias Gleich for assistance during fMRI data acquisition and Stephan Lücke for organization and assistance during FDOPA PET. This study was supported by grants from the German Research Foundation (to F.S. and A.H.) [Deutsche Forschungsgemeinschaft (DFG) SCHL 1969/1-1, DFG SCHL1969/2-1, and DFG HE2597/14-1 as part of DFG FOR 1617]. L.D. and F.S. were supported by the Max Planck Society. Q.J.M.H. (DFG RA1047/2-1) and R. Boehme (GRK 1123/2) received funding from the German Research Foundation. R.J.D. is supported by a Wellcome Trust Senior Investigator Award (098362/Z/12/Z). A.H. received funding from the German Federal Ministry of Education and Research (01GQ0411; 01QG87164; NGFN Plus 01 GS 08152 and 01 GS 08159).

1. Dolan RJ, Dayan P (2013) Goals and habits in the brain. *Neuron* 80(2):312–325.
2. Dickinson AD (1985) Action and habits: The development of behavioural autonomy. *Philos Trans R Soc Lond B Biol Sci* 308(1135):67–78.
3. Doll BB, Simon DA, Daw ND (2012) The ubiquity of model-based reinforcement learning. *Curr Opin Neurobiol* 22(6):1075–1081.
4. Balleine BW, Dickinson A (1998) Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37(4-5):407–419.
5. Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci* 8(12):1704–1711.
6. Gläscher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66(4):585–595.
7. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69(6):1204–1215.
8. Cools R (2011) Dopaminergic control of the striatum for high-level cognition. *Curr Opin Neurobiol* 21(3):402–407.
9. Hiroyuki N (2014) Multiplexing signals in reinforcement learning with internal models and dopamine. *Curr Opin Neurobiol* 25:123–129.
10. Schultz W (2013) Updating dopamine reward signals. *Curr Opin Neurobiol* 23(2):229–238.
11. Seamans JK, Yang CR (2004) The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Prog Neurobiol* 74(1):1–58.
12. Wunderlich K, Smittenaar P, Dolan RJ (2012) Dopamine enhances model-based over model-free choice behavior. *Neuron* 75(3):418–424.
13. Dayan P (2012) Twenty-five lessons from computational neuromodulation. *Neuron* 76(1):240–256.
14. Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* 16(5):1936–1947.
15. Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275(5306):1593–1599.
16. D'Ardenne K, McClure SM, Nystrom LE, Cohen JD (2008) BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science* 319(5867):1264–1267.
17. Cools R, Gibbs SE, Miyakawa A, Jagust W, D'Esposito M (2008) Working memory capacity predicts dopamine synthesis capacity in the human striatum. *J Neurosci* 28(5):1208–1212.
18. Vernaleken I, et al. (2007) "Prefrontal" cognitive performance of healthy subjects positively correlates with cerebral FDOPA influx: An exploratory [^{18}F]-fluoro-L-DOPA-PET investigation. *Hum Brain Mapp* 28(10):931–939.
19. Otto AR, Gershman SJ, Markman AB, Daw ND (2013) The curse of planning: Dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychol Sci* 24(5):751–761.
20. de Wit S, Barker RA, Dickinson AD, Cools R (2011) Habitual versus goal-directed action control in Parkinson disease. *J Cogn Neurosci* 23(5):1218–1229.
21. de Wit S, et al. (2012) Reliance on habits at the expense of goal-directed control following dopamine precursor depletion. *Psychopharmacology (Berl)* 219(2):621–631.
22. Kumakura Y, Cumming P (2009) PET studies of cerebral levodopa metabolism: A review of clinical findings and modeling approaches. *Neuroscientist* 15(6):635–650.
23. Schlagenhauf F, et al. (2013) Ventral striatal prediction error signaling is associated with dopamine synthesis capacity and fluid intelligence. *Hum Brain Mapp* 34(6):1490–1499.
24. McDannald MA, Lucantonio F, Burke KA, Niv Y, Schoenbaum G (2011) Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. *J Neurosci* 31(7):2700–2705.
25. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. *Neuroimage* 46(4):1004–1017.
26. Braver TS, Cohen JD (1999) Dopamine, cognitive control, and schizophrenia: The gating model. *Prog Brain Res* 121:327–349.
27. Moran RJ, Symmonds M, Stephan KE, Friston KJ, Dolan RJ (2011) An in vivo assay of synaptic function mediating human cognition. *Curr Biol* 21(15):1320–1325.
28. Smittenaar P, FitzGerald TH, Romei V, Wright ND, Dolan RJ (2013) Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. *Neuron* 80(4):914–919.
29. Niv Y, Daw ND, Joel D, Dayan P (2007) Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology (Berl)* 191(3):507–520.
30. Beierholm U, et al. (2013) Dopamine modulates reward-related vigor. *Neuropsychopharmacology* 38(8):1495–1503.
31. Frank MJ, Seeberger LC, O'reilly RC (2004) By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science* 306(5703):1940–1943.
32. Lee SW, Shimojo S, O'Doherty JP (2014) Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 81(3):687–699.
33. Badre D, Hoffman J, Cooney JW, D'Esposito M (2009) Hierarchical cognitive control deficits following damage to the human frontal lobe. *Nat Neurosci* 12(4):515–522.
34. Koechlin E, Ody C, Kouneiher F (2003) The architecture of cognitive control in the human prefrontal cortex. *Science* 302(5648):1181–1185.
35. Steinberg EE, et al. (2013) A causal link between prediction errors, dopamine neurons and learning. *Nat Neurosci* 16(7):966–973.
36. Goto Y, Grace AA (2005) Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nat Neurosci* 8(6):805–812.
37. Egerton A, Demjaha A, McGuire P, Mehta MA, Howes OD (2010) The test-retest reliability of ^{18}F-DOPA PET in assessing striatal and extrastriatal presynaptic dopaminergic function. *Neuroimage* 50(2):524–531.
38. Goto Y, Grace AA (2008) Dopamine modulation of hippocampal-prefrontal cortical interaction drives memory-guided behavior. *Cereb Cortex* 18(6):1407–1414.
39. Howes OD, et al. (2012) The nature of dopamine dysfunction in schizophrenia and what this means for treatment. *Arch Gen Psychiatry* 69(8):776–786.
40. Rakshi JS, et al. (1999) Frontal, midbrain and striatal dopaminergic function in early and advanced Parkinson's disease A 3D [^{18}F]dopa-PET study. *Brain* 122(Pt 9):1637–1650.
41. Yin HH, Knowlton BJ, Balleine BW (2004) Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur J Neurosci* 19(1):181–189.
42. Yin HH, Ostlund SB, Knowlton BJ, Balleine BW (2005) The role of the dorsomedial striatum in instrumental conditioning. *Eur J Neurosci* 22(2):513–523.
43. Knutson B, Gibbs SE (2007) Linking nucleus accumbens dopamine and blood oxygenation. *Psychopharmacology (Berl)* 191(3):813–822.
44. Balleine BW, O'Doherty JP (2010) Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* 35(1):48–69.
45. Wunderlich K, Dayan P, Dolan RJ (2012) Mapping value based planning and extensively trained choice in the human brain. *Nat Neurosci* 15(5):786–791.
46. Heinz AJ, Beck A, Meyer-Lindenberg A, Sterzer P, Heinz A (2011) Cognitive and neurobiological mechanisms of alcohol-related aggression. *Nat Rev Neurosci* 12(7):400–413.
47. den Ouden HE, et al. (2013) Dissociable effects of dopamine and serotonin on reversal learning. *Neuron* 80(4):1090–1100.
48. Moran RJ, et al. (2013) Free energy, precision and learning: The role of cholinergic neuromodulation. *J Neurosci* 33(19):8227–8236.
49. Cools R, Nakamura K, Daw ND (2011) Serotonin and dopamine: Unifying affective, activational, and decision functions. *Neuropsychopharmacology* 36(1):98–113.
50. Goto Y, Grace AA (2008) Limbic and cortical information processing in the nucleus accumbens. *Trends Neurosci* 31(11):552–558.
51. Hietala J, et al. (1999) Depressive symptoms and presynaptic dopamine function in neuroleptic-naive schizophrenia. *Schizophr Res* 35(1):41–50.
52. Vernaleken I, et al. (2007) Asymmetry in dopamine D(2/3) receptors of caudate nucleus is lost with age. *Neuroimage* 34(3):870–878.
53. Martin-Soelch C, et al. (2011) Lateralization and gender differences in the dopaminergic response to unpredictable reward in the human ventral striatum. *Eur J Neurosci* 33(9):1706–1715.
54. Tomer R, Goldstein RZ, Wang GJ, Wong C, Volkow ND (2008) Incentive motivation is associated with striatal dopamine asymmetry. *Biol Psychol* 77(1):98–101.
55. Patlak CS, Blasberg RG (1985) Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. Generalizations. *J Cereb Blood Flow Metab* 5(4):584–590.

Supporting Information

Deserno et al. 10.1073/pnas.1417219112

SI Text

SI Experimental Procedures

Participants. Eighteen male and 11 female participants resulting in a total of 29 right-handed participants were included. They had a mean age of 28.35 ± 4.95 y (range, 20–39). Data from five additional participants were excluded (>3 SDs of missed trials = two participants; >3 SDs of K_i values in left striatum = one participant; technical failure of MRI acquisition = one participant; nonattendance to fMRI session = one participant). All participants had no history of psychiatric or neurological disorder (according to SCID interview; ref. 1) and normal or corrected-to-normal vision. All participants underwent task-based fMRI on one day and F-DOPA PET took place separately before or after the fMRI session (mean, 20.85 ± 17.90 d; range, 1–56). This study was approved by the local ethics committee. Participants gave written informed consent.

Behavioral Data Analysis. Stay–switch behavior was analyzed as a function of reward (reward or no reward) and state (common or rare). Individual stay probabilities were subjected to a repeated-measures ANOVA with factors reward and state (Fig. S1). Second-stage reaction times were also analyzed as a function of reward and state and then subjected to a repeated-measures ANOVA with factors reward and state (Fig. S3).

Computational Model. The aim of model-free and model-based algorithms is to learn values for each of the stimuli, which appear in the task as three pairs (s_A, s_B, s_C). s_A refers to the first-stage stimuli, and s_B and s_C to the two pairs of second-stage stimuli. a refers to the chosen stimuli. The indices i and t denote the stage (i = 1 for S_A at the first stage and i = 2 for S_B or S_C at the second stage) and time in trials, respectively.

First, the model-free algorithm was SARSA(λ) (2):

$$Q_{MF}\left(s_{i,t+1},a_{i,t+1}\right)=Q_{MF}\left(s_{i,t},a_{i,t}\right)+\alpha_i\delta_{i,t}, \quad \textbf{[S1]}$$

$$\delta_{i,t}=r_{i,t}+Q_{MF}\left(s_{i+1,t},a_{i+1,t}\right)-Q_{MF}\left(s_{i,t},a_{i,t}\right). \quad \textbf{[S2]}$$

Notably, $r_{1,t}=0$ and $Q_{MF}(s_{3,t},a_{3,t})=0$. We allow different learning rates for each stage. Furthermore, we allow for an additional stage-skipping update of first-stage values by introducing another parameter λ, which connects the two stages and allows the reward prediction error at the second stage to influence first-stage values:

$$Q_{MF}\left(s_{1,t+1},a_{1,t+1}\right)=Q_{MF}\left(s_{1,t},a_{1,t}\right)+\alpha_1\lambda\delta_{2,t}. \quad \textbf{[S3]}$$

It is worth mentioning that λ additionally accounts for the main effect of reward as observed in the analysis of first-stage stay–switch behavior but not for an interaction of reward and state. Instead, the influence of learning values for the transition matrix accounts for this interaction.

Second, the model-based algorithm learns values in a forward-planning way and computes first-stage values by simply multiplying the better option at the second stage with the transition probabilities (3):

$$\begin{aligned}Q_{MB}\left(s_A,a_j\right)=&P\left(S_B|S_A,a_j\right)\max\, Q_{MF}\left(s_B,a\right)\\&+P\left(S_C|S_A,a_j\right)\max\, Q_{MF}\left(s_c,a\right).\end{aligned} \quad \textbf{[S4]}$$

Note that this approach simplifies transition learning because transition probabilities are not learned explicitly. This approach is in line with the task instructions, and Daw et al. (3) refer to a simulation, which verified that this approach outperforms incremental learning of the transition matrix (3); this was identically applied in another paper (4) but also compare work that implemented incremental learning of transition matrices (5, 6).

Third, the hybrid algorithm connects model-free and model-based decision values at the first stage:

$$Q\left(s_A,a_j\right)=\omega Q_{MB}\left(s_A,a_j\right)+(1-\omega)\,Q_{MF}\left(s_A,a_j\right). \quad \textbf{[S5]}$$

At the second stage, $Q=Q_{MB}=Q_{MF}$. Importantly, ω gives a weighting of the relative influence of model-free and model-based values and is therefore the model's parameter of most interest. The special cases of this model refer to ω = 1 or ω = 0 reflecting purely model-based or purely model-free control over first-stage values, respectively.

Finally, we transform values into action probabilities using a softmax for Q:

$$p\left(a_{i,t}=a|s_{i,t}\right)=\frac{\exp\left(\beta_i\left[Q\left(s_{i,t},a\right)+\rho*\mathrm{rep}(a)\right]\right)}{\sum_{a'}\exp\left(\beta_i\left[Q\left(s_{i,t},a'\right)+\rho*\mathrm{rep}(a')\right]\right)}. \quad \textbf{[S6]}$$

Here, β controls the stochasticity of the choices, and we assume this to be different between the two stages. The additional parameter ρ captures first-stage choice perseveration, and rep is an indicator function that equals 1 if the previous first-stage choice was the same. Models were implemented as in the original paper by Daw et al. (3).

Model Fitting. In summary, the algorithm has a total of seven parameters and can be reduced to its special cases ω = 1 and ω = 0. We fit bounded parameters by transforming them to a logistic (α, λ, ω) or exponential (β) distribution to render normally distributed parameter estimates. To infer the maximum-a-posteriori estimate of each parameter for each subject, we set the prior distribution to the maximum-likelihood given the data of all participants and then used expectation–maximization. For an in-depth description, please compare refs. 7 and 8.

Model Selection. First, we report the Bayesian information criterion (BIC) based on the negative log-likelihood (Table S1). Second, we compute the model evidence by integrating out the free parameters. This integral was approximated by sampling from the prior distribution, and we therefore add the subscript "int" to the BIC (Table S1; refs. 7 and 8). Third, we submit this integrated likelihood to the spm_BMS function contained in SPM8 for a random-effects model selection that computes so-called exceedance probabilities (9).

Analysis of fMRI Data. Preprocessing included correction for delay of slice time acquisition. Voxel-displacement maps were estimated based on acquired field maps. All images were realigned to correct for motion and 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 and resampled into 2-mm isotropic voxels; subsequently, all images were spatially smoothed with an isotropic Gaussian kernel of 8 mm full width at half maximum.

First-Level Statistics of fMRI Data. An event-related analysis was applied to the images on two levels using the general linear model approach as implemented in SPM8. Individual (random-effects) model parameters were used to generate regressors for the analysis of fMRI data. In line with Daw et al. (3), the analysis focused on prediction errors at two time points within each trial: onsets of second-stage stimuli and onsets of reward delivery but only prediction errors at second-stage onsets differ between model-free and model-based algorithms and therefore vary with respect to the weighting parameter ω of the hybrid algorithm. Both time points were entered to the first-level model as one regressor, which was parametrically modulated by (*i*) model-free prediction errors and (*ii*) by the difference of model-based and model-free prediction errors. This second parametric modulator refers to the partial derivative of the value function with respect to ω and reflects the difference between model-based and model-free values at the first stage. Note that this difference regressor equals zero at reward delivery, because both algorithms converge at this time point. To avoid any confound of the neural results due to activity differences between these two time points per se, the difference regressor was mean-centered within subject. Furthermore, the time point of reward delivery was additionally included as a separate regressor. Finally, the design also included first-stage onsets, which were not in the focus of the present analysis, also with two parametric modulators, the softmax probability for choosing one of the two first-stage probabilities as well as its partial derivative with respect to ω. Invalid trials (no choice within response window) were also modeled. All regressors were convolved with the canonical hemodynamic response function as provided by SPM8 and its temporal derivative. The six movement parameters from the realignment 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.

Analysis of PET Data. First, head motion was corrected. Realignment started from frame 7, because earlier frames did not provide sufficient anatomical information. The transformation matrix for frame 7 was applied to frames 1–6. Using the mean over all frames, frames were coregistered to the individual T1 image. As for the fMRI data, all frames were spatially normalized into the MNI space based on normalization parameters generated during the segmentation of each subject's anatomical T1 scan.

SI Results

Behavioral Raw Data and Ventral Striatal Presynaptic Dopamine. To prove consistency of the observed relationship between right ventral striatal presynaptic dopamine and the parameter ω from computational modeling with behavioral raw data, the sample was split up at the median of right ventral striatal dopamine measures, and in each group, two participants neighboring the median were left out. This was done because the study design was correlational and participants were not sampled from the extreme ends of the distribution of presynaptic dopamine measures. This resulted in two groups of 12 participants each with high and low right ventral striatal presynaptic dopamine, respectively. It is important to note that the positive direction of this effect was not at stake for this supplemental analysis given the positive pharmacological effect of L-DOPA on model-based choices reported in an independent sample by Wunderlich et al. (4). First-stage stay probabilities were then analyzed in a repeated-measures ANOVA as a function of reward, state, and dopamine group. First, and given the observed correlation with the parameter ω, the two groups differed significantly on the weighting parameter (mean ω for low ventral striatal dopamine, 0.48 ± 0.16, and for high ventral striatal dopamine, 0.62 ± 20; t = 1.76, P = 0.04, one-tailed). Notably, mean ω shows that high right ventral striatal dopamine is associated with a higher degree of model-based learning, whereas low ventral striatal dopamine is associated with maintained model-based control and no dominance of model-free control. In the same vein, analysis of stay–switch behavior at the first stage as a function of reward, state, and right ventral striatal dopamine revealed a trendwise significant interaction of reward by state by dopamine group (F = 1.82, P = 0.09, one-tailed; Fig. S2). It can be clearly depicted from the bar plots that this effect was mainly driven by a higher stay probability in rewarded-common trials in participants with high levels of ventral striatal dopamine (t = 1.81, P = 0.042, one-tailed), whereas staying in nonrewarded rare trials was only trendwise significant (t = 1.3, P = 0.10, one-tailed). However, and in line with the modeling analysis, it is clearly visible from Fig. S2 that both groups show aspects of model-based behavior and that relatively higher right ventral striatal presynaptic dopamine is associated with pronounced model-based behavior. In turn, there was no evidence for relatively lower ventral striatal dopamine being associated with predominantly model-free control.

Computational Modeling.

Inferring ω from simulated data. One of our reviewers asked for the stability of inferring ω in a complex space of parameters. We provide an analysis that shows that ω can indeed be inferred reliably in complex space by refitting 50 simulations of each participant's data (inferred from observed data: mean ω, 0.5294 ± 0.1798; inferred from simulated data: mean ω, 0.4915 ± 0.1971; correlation of r = 0.9389). Also, the results of the regression model regarding striatal measures of presynaptic dopamine remained significant when using ω inferred from simulated data (ventral striatum—right: β = 0.43, t = 2.17, P = 0.04; left: β = 0.050, t = 0.23, P = 0.82; remaining striatum—right: β = 0.10, t = 0.35, P = 0.73; left: β = –0.51, t = 1.65, P = 0.11).

Separate learning rates for wins and losses at the second stage. We were also asked to provide evidence whether the hybrid model favored by Bayesian model selection also outperforms a slightly modified version by adding separate learning rates for reward and punishment at the second stage. When comparing this variant of the hybrid against the more parsimonious version of the model, we found the "original" and more restrictive model without separate learning rates for reward and punishment to be superior in terms of random-effects Bayesian model selection (exceedance probability hybrid, 76.07; exceedance probability hybrid-rew-pun 23.93).

Correlation with Midbrain Presynaptic Dopamine. In line with a reviewer's suggestion, we also extracted presynaptic dopamine measures from the midbrain. However, we did not observe a significant correlation between midbrain K_i and a balance of behavioral control ω (r = –0.13; P = 0.50).

Neuroimaging.

fMRI-PET correlations: Control for presynaptic dopamine in other striatal regions. All fMRI -PET correlations between right ventral striatal dopamine and right ventral striatal model-free prediction errors or right lateral prefrontal model-based signals, respectively, remained significant when performing linear regression analyses with all dopamine measures as predictors and the right ventral striatal model-free signals as dependent variable (ventral striatum—right: β = –0.43, t = 2.09, P = 0.048; left: β = 0.12, t = 0.52, P = 0.61; remaining striatum—right: β = 0.16, t = 0.54, P = 0.60; left: β = –0.05, t = 0.14, P = 0.89) or the right lateral prefrontal model-based learning signals as dependent variable (ventral striatum—right: β = –0.44, t = 2.14, P = 0.04; left: β = 0.16, t = 0.70, P = 0.50; remaining striatum—right: β = –0.24, t = 0.81, P = 0.43; left: β = –0.09, t = 0.27, P = 0.79). In line, explorative correlations with presynaptic dopamine in other striatal regions showed no significant results (all values of r < 0.14, values of P > 0.49).

Voxelwise fMRI-PET correlations. Although the parameter estimates extracted for testing fMRI-PET correlations were unbiased for covariation with dopamine measures, we additionally report voxelwise group-level analyses. First, for the negative correlation between model-free prediction errors and right ventral striatal presynaptic dopamine, we found a significant effect in right ventral striatum ($x = 10, y = 20, z = -4, t = 3.18$, p-FWE = 0.040; and $x = 8, y = 8, z = 0, t = 3.15$, p-FWE = 0.042; Fig. S5) but not in lateral PFC ($x = 48, y = 28, z = -8, t = 1.70$, p-FWE = 0.315) and medial PFC ($x = 0, y = 64, z = 4, t = 0.85$, p-FWE = 0.749).

For the positive correlation between right ventral striatal presynaptic dopamine and model-based learning signals in the lPFC, we report a region by dopamine interaction in the main manuscript, thereby demonstrating that this positive correlation is present between right ventral striatal presynaptic dopamine and model-based signals in lPFC but not between right ventral striatal presynaptic dopamine and model-based signals in the right ventral striatum. It should be noted that such an interaction of region by dopamine cannot be tested in a voxelwise analysis. However, when testing for a voxelwise positive association between right ventral striatal presynaptic dopamine and model-based learning signals, we found a significant effect in right lPFC ($x = 40, y = 24, z = -16, t = 3.12$, p-FWE = 0.033; Fig. S5) but not in mPFC ($x = 12, y = 56, z = 24, t = 0.95$, p-FWE = 0.658) and ventral striatum ($x = 14, y = 8, z = -12, t = 0.73$, p-FWE = 0.572). This positive association between model-based signals and presynaptic dopamine was also present when averaging dopamine across right and left ventral striata ($x = 40, y = 24, z = -16, t = 2.95$, p-FWE = 0.046) but not for presynaptic dopamine from left ventral striatum alone ($x = 40, y = 24, z = -14, t = 1.64$, p-FWE = 0.357). For small-volume corrections, we used large anatomical masks applied for corrections of the task effects per se (lPFC, mPFC, striatum) and restricted them to those voxels that were indeed activated for model-free or model-based learning signals, respectively.

1. First MB, Spitzer RL, Gibbon M, Williams J (2001) *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition with Psychotic Screen (SCID-I/P W/ PSY SCREEN)* (New York State Psychiatric Institute, New York).
2. Sutton RS, Barto AG (1998) *Reinforcement Learning: An Introduction* (MIT, Cambridge, MA).
3. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69(6):1204–1215.
4. Wunderlich K, Smittenaar P, Dolan RJ (2012) Dopamine enhances model-based over model-free choice behavior. *Neuron* 75(3):418–424.
5. Gläscher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66(4):585–595.
6. Lee SW, Shimojo S, O'Doherty JP (2014) Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 81(3):687–699.
7. Huys QJ, et al. (2011) Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS Comput Biol* 7(4):e1002028.
8. Huys QJ, et al. (2012) Bonsai trees in your head: How the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput Biol* 8(3):e1002410.
9. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. *Neuroimage* 46(4):1004–1017.

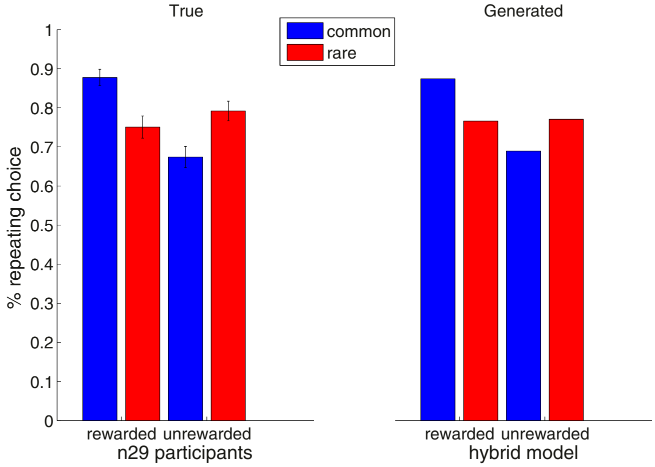

Fig. S1. (*Left*) Observed choice behavior showing a significant main effect of reward ($F = 11.92$; $P = 0.01$) and a significant reward by state interaction ($F = 21.97$; $P < 0.001$) but no main effect of state ($F = 0.0183$; $P = 0.8933$). (*Right*) Choices generated by the hybrid model with the best-fitting parameters.

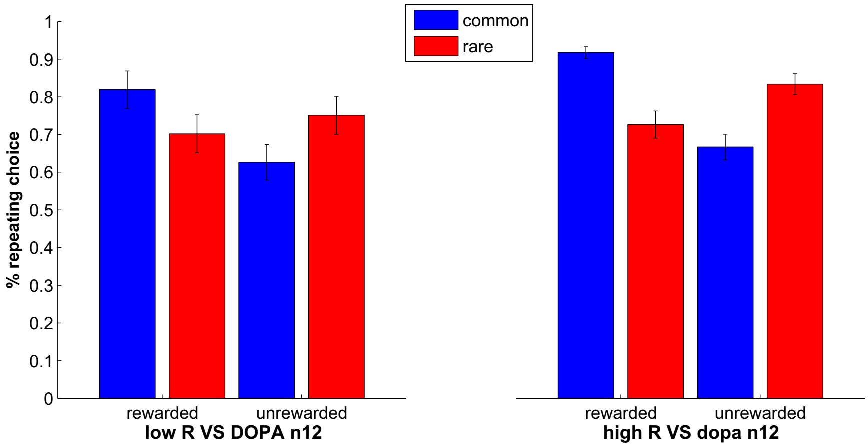

Fig. S2. Stay–switch behavior at the first stage as a function of reward, state, and high or low dopamine groups (R, right; VS, ventral striatum). The reward by state by dopamine group interaction approached significance ($F = 1.82$; $P = 0.09$).

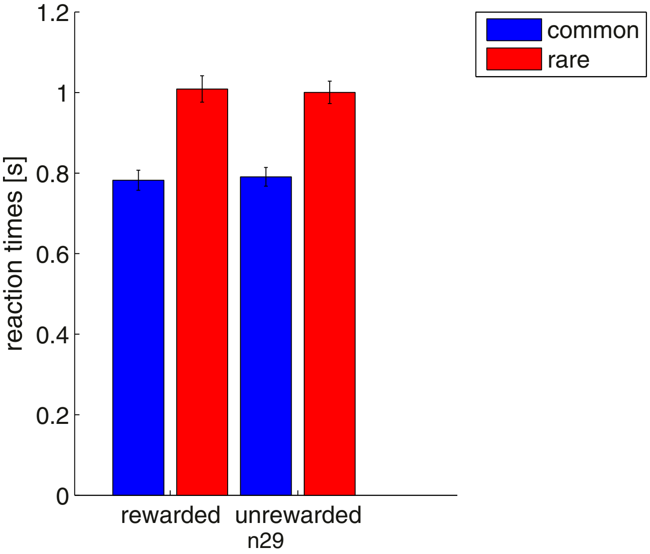

Fig. S3. Second-stage reaction times showing a significant main effect of state ($F = 50.45$; $P < 0.001$) but no main effect of reward ($F = 0.00$; $P = 0.99$) or reward by state interaction ($F = 7.83$; $P = 0.37$). Mean overall reaction times at the second stage were 859 ± 117 ms SD.

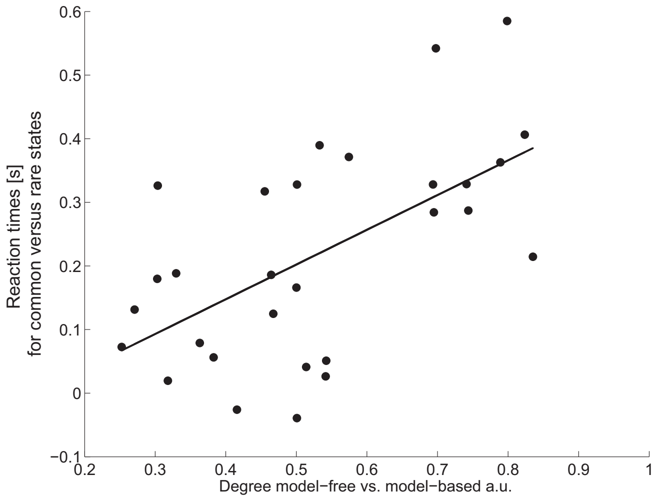

Fig. S4. Correlation between reaction time difference for rare versus common states and the parameter ω ($r = 0.59$; $P = 0.001$). Mean overall reaction times at the second stage were not correlated with the parameter ω or K_i from the ventral striatum ($r < 0.15$; $P > 0.45$).

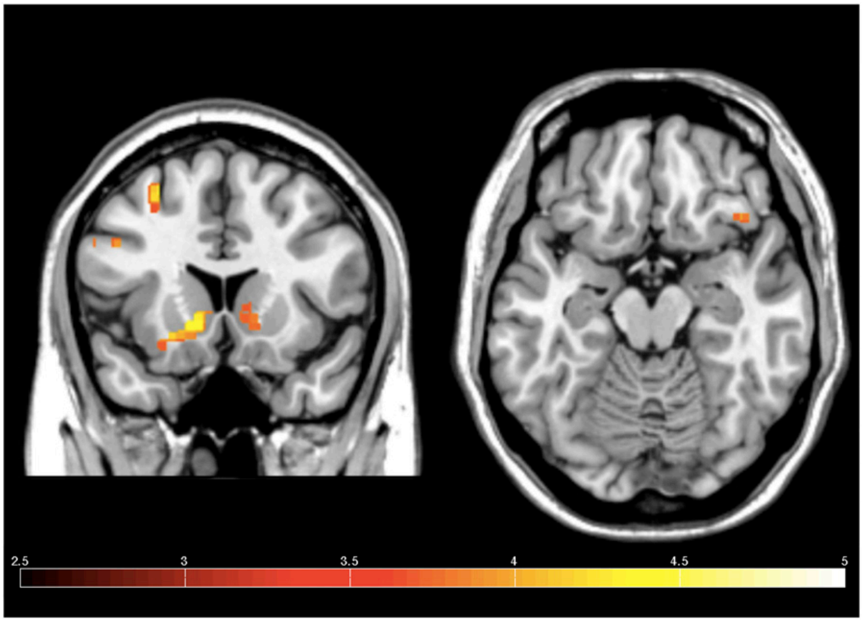

Fig. S5. Statistical parametric maps of fMRI-PET correlations. Ventral striatal model-free prediction errors and right ventral striatal presynaptic dopamine (*Left*); right lateral prefrontal model-based signals and right ventral striatal presynaptic dopamine (*Right*). For display purposes, all statistical maps are thresholded at a minimum T value of ≥2.5 and were implicitly masked for voxels showing an effect of model-free prediction errors or additional model-based signals, respectively.

Table S1. Model comparison

Model	−LL	BIC	BIC$_{\text{int}}$	XP
Full hybrid model	5,199.82	10,465.05	10,946.94	0.9755
	Δ −LL hybrid	Δ BIC hybrid	Δ BIC$_{\text{int}}$ hybrid	
$\lambda = 0$	−121.85	234.35	148.57	0.0022
$\omega = 1$	−179.83	331.63	185.29	0.0009
$\omega = 0$	−156.63	303.91	221.16	0.0215
$\omega = 0$, $\lambda = 0$	−321.46	624.23	458.21	1e-05

BIC, Bayesian information criterion (the subscript "int" refers to integrating out the free parameters); −LL, negative log-likelihood; XP, exceedance probability.

Table S2. Distribution of best-fitting parameters (hybrid model) and the negative log-likelihood

Percentile	β_1	β_2	α_1	α_2	λ	ω	ρ	−LL
25th	4.50	2.65	0.10	0.52	0.67	0.37	0.09	200.10
50th	6.39	3.38	0.35	0.58	0.69	0.50	0.16	179.08
75th	7.73	4.54	0.55	0.70	0.77	0.70	0.21	156.07

α_1, α_2, learning rates at the first and second stage; β_1, β_2, stochasticity of first- and second-stage choices; λ, stage-skipping update; ρ, repetition parameter; ω, weighting of model-free and model-based values; −LL, negative log-likelihood.

Table S3. fMRI results for model-free and mode-based effects and their conjunction

Region	Coordinates	*t* value	p-FWE in ROI	*k*
Model-free				
Lateral PFC				
Inferior frontal gyrus	42, 24, −14	6.13	<0.001	1,027
Inferior frontal gyrus	−24, 24, −18	7.46	<0.001	1,720
Superior frontal gyrus	−42, 14, 52	5.26	<0.001	738
Medial PFC				
Superior frontal gyrus	0, 62, 32	5.80	<0.001	3,487
Anterior cingulate cortex	−4, 8, 28	4.95	<0.001	72
Ventral striatum				
	10, 10, −10	10.87	<0.001	77
	−12, 6, −12	9.16	<0.001	55
Model-based				
Lateral PFC				
Inferior frontal gyrus	42, 24, −14	4.73	0.03	73
Medial PFC				
Medial frontal gyrus	12, 54, 20	4.26	0.053	25
Ventral striatum				
	14, 8, 8	3.69	0.007	29
	−14, 8, −12	3.66	0.008	21
Conjunction				
Lateral PFC				
Inferior frontal gyrus	42, 24, −14	4.73	0.031	73
Medial PFC				
Anterior cingulate cortex	2, 38, 16	3.22	0.525	
Ventral striatum				
	14, 8, 8	3.69	0.007	29
	−14, 8, −12	3.66	0.008	21

Shown are significant effects in the bilateral region of interests: lateral and medial prefrontal cortex and ventral striatum; *k* refers to the cluster size at $P < 0.001$, uncorrected. Peak coordinates of model-free learning signals were also significant at FWE < 0.05 corrected for the whole brain.