Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions

Lea K. Krugel[a,b,c,1], Guido Biele[a,b,c], Peter N. C. Mohr[a,b,c], Shu-Chen Li[a,c], and Hauke R. Heekeren[a,b,c,1]

[a]Neurocognition of Decision Making, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany; [b]Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany; and [c]Neuroscience Research Center and Berlin NeuroImaging Center, Charité University Medicine Berlin, Schumannstrasse 20/21, 10117 Berlin, Germany

The ability to rapidly and flexibly adapt decisions to available rewards is crucial for survival in dynamic environments. Reward-based decisions are guided by reward expectations that are updated based on prediction errors, and processing of these errors involves dopaminergic neuromodulation in the striatum. To test the hypothesis that the *COMT* gene Val[158]Met polymorphism leads to interindividual differences in reward-based learning, we used the neuromodulatory role of dopamine in signaling prediction errors. We show a behavioral advantage for the phylogenetically ancestral Val/Val genotype in an instrumental reversal learning task that requires rapid and flexible adaptation of decisions to changing reward contingencies in a dynamic environment. Implementing a reinforcement learning model with a dynamic learning rate to estimate prediction error and learning rate for each trial, we discovered that a higher and more flexible learning rate underlies the advantage of the Val/Val genotype. Model-based fMRI analysis revealed that greater and more differentiated striatal fMRI responses to prediction errors reflect this advantage on the neurobiological level. Learning rate-dependent changes in effective connectivity between the striatum and prefrontal cortex were greater in the Val/Val than Met/Met genotype, suggesting that the advantage results from a downstream effect of the prefrontal cortex that is presumably mediated by differences in dopamine metabolism. These results show a critical role of dopamine in processing the weight a particular prediction error has on the expectation updating for the next decision, thereby providing important insights into neurobiological mechanisms underlying the ability to rapidly and flexibly adapt decisions to changing reward contingencies.

COMT | dopamine | functional MRI | learning rate | reinforcement learning

Human learning spans a wide range of functions from ancient evolutionary accomplishments [e.g., fast and intuitive learning from rewards (1, 2)], to complex executive processes that require high capacities of attention and working memory (3, 4). These functions are supported by interacting brain regions that comprise phylogenetically ancient structures (e.g., the striatum) as well as more recently evolved neocortical structures [e.g., the prefrontal cortex (PFC)]. Still, different learning processes rely on the same neuromodulators. Dopamine (DA) modulates synaptic efficacy in both the striatum and PFC (5) and is thus involved in different kinds of learning. Accordingly, interindividual differences in dopaminergic neuromodulatory mechanisms contribute to performance differences in both simple instrumental learning tasks and complex cognitive tasks (6–10).

One interindividual difference in dopaminergic neuromodulation that has been linked to PFC function is the uniquely human Val[158]Met polymorphism in the enzyme catechol-*O*-methyltransferase (COMT), which regulates extrasynaptical DA degradation (8–10). COMT activity is lower in individuals homozygous for the mutated allele (Met homozygotes) than in homozygotes of the phylogenetically ancestral allele (Val homozygotes). Therefore, Met homozygotes have higher levels of DA in the PFC, where the influence of COMT on DA degradation is largest (11). Whereas most prior studies focused on the direct effects of the Val[158]Met polymorphism in the PFC and emphasized a behavioral advantage for the phylogenetically emerging Met genotype in executive cognitive tasks (8–10) (but see ref. 12), two recent studies provided evidence for an influence of the *COMT* genotype on subcortical regions in gambling paradigms (13, 14).

Animal studies have shown that DA flux in the PFC indirectly affects downstream dopaminergic targets, particularly the striatum (15, 16) where DA activity is—unlike in the PFC—predominantly regulated by DA reuptake through presynaptic DA transporters (17, 18). Specifically, these studies indicate an inverse relationship between extrasynaptic DA levels in the PFC and synaptic DA activity in the striatum by means of an indirect subcortical effect. Whereas projections from the PFC to the midbrain directly contact dopaminergic cell groups that project back to the PFC to generate a positive feedback loop, they indirectly (via inhibiting intermediates) regulate dopaminergic cell groups that project to the striatum (15, 16). These previous findings raise the possibility that the *COMT* Val[158]Met polymorphism also has an indirect opposite effect on striatal DA activity. Several animal studies found evidence for a reciprocal relationship between cortical and subcortical DA systems (19–21), whereas a study on *COMT* knockout mice did not find a striatal effect on basal concentrations of DA (22). Importantly, a postmortem study in humans revealed higher DA synthesis rates in projections from the midbrain to the striatum of Val homozygotes (23). Furthermore, multimodal neuroimaging in humans showed greater DA synthesis in the dopaminergic midbrain in Val carriers than Met homozygotes (24). Although this was found for dopaminergic midbrain neurons in general (i.e., not specifically for those midbrain neurons that project to the striatum), an overall higher DA synthesis in Val carriers is likely to translate into a greater DA supply in the striatum [as opposed to the PFC (11)], where COMT expression is sparse. The greater supply of striatal DA in Val homozygotes has been hypothesized to result in greater striatal DA burst firing (phasic activity) and a subsequent advantage for Val homozygotes in tasks that demand flexibility (25); this contrasts with the better stabilization effects through higher levels of prefrontal DA (26) in Met homozygotes (8–10).

NEUROSCIENCE

A Experimental design

B Collected points
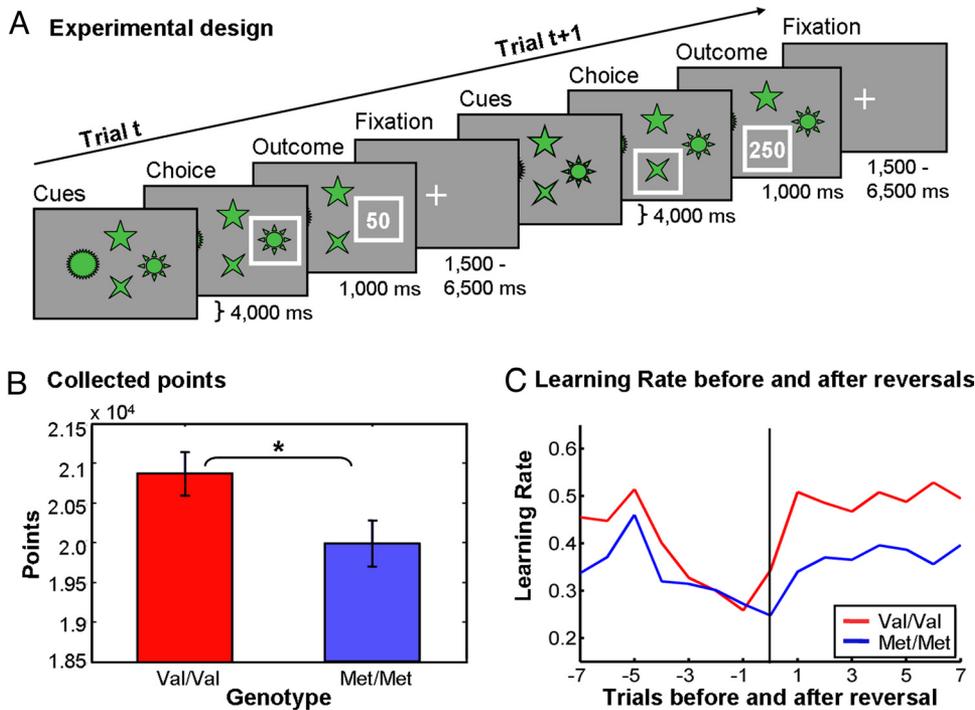

C Learning Rate before and after reversals

Fig. 1. Experimental design and behavioral results. (*A*) Participants repeatedly chose one of four cues associated with different positive amounts of points. The probability of yielding the highest possible outcome was 80% for the best option and 20% for the other options. After participants made a choice for one cue, the chosen option was highlighted and followed by an outcome display. Failure to make a choice led to 0 points. Points were converted into money after the experiment. Choice options were represented by four easily memorable symbols that stayed at the same screen location throughout the whole experiment. (*B*) Val homozygotes collected more points than Met homozygotes ($P = 0.038$). (*C*) Dynamic learning rate of Val and Met homozygotes before and after reversal (indicated by vertical line).

We hypothesized that the phylogenetically ancestral Val genotype, which constitutes the most frequent *COMT* genotype worldwide (27), is associated with an advantage in rapid and flexible learning from rewards through greater striatal burst firing of DA neurons. Phasic DA activity plays a prominent role in the reward-based learning that underlies the ability to adapt decisions to available outcomes (1, 2). Specifically, neurophysiological studies revealed that phasic DA activity signals the discrepancy between reward prediction and reward occurrence [prediction error (PE)], which is used as a teaching signal in the learning process to update the expectation about the next outcome (1, 2). A number of functional magnetic resonance imaging (fMRI) studies showed that striatal fMRI signals during reward-based learning correlate with PEs estimated by reinforcement learning models (28–30). A recent study demonstrated that pharmacological modulation of striatal DA availability affects performance and PE-related striatal fMRI responses during an instrumental learning task (30).

Thus, we predicted that Val homozygotes demonstrate a stronger and more differentiated influence of PEs than Met homozygotes, manifest as greater PE-related fMRI signals at the imaging level, a higher and more flexible learning rate at the computational level, and better performance at the behavioral level. Consistent with these predictions, we report here a behavioral advantage for the phylogenetically ancestral Val/Val genotype in instrumental reversal learning. Using a reinforcement learning model with a dynamic learning rate and fMRI, we discovered that a higher and more flexible learning rate, reflected by greater and more differentiated striatal fMRI responses to PEs, underlies the advantage. Independent of genotype, we found learning rate-dependent changes in the effective connectivity between these PE-representing striatal regions and the PFC. Moreover, learning rate-dependent changes in the effective connectivity were greater in Val than Met homozygotes, suggesting that the advantage results from a downstream effect of the PFC that is presumably mediated by differences in DA metabolism.

Results

Behavior. Participants carried out a probabilistic reversal learning task, in which they were monetarily encouraged to collect as many

points as possible by adapting their choices to ongoing changes in reward contingencies (Fig. 1*A*). In line with our prediction, Val homozygotes won significantly more points than Met homozygotes ($P = 0.038$, see Fig. 1*B*). Val homozygotes also reached the hidden criterion for a reversal more often. Neither Val homozygotes nor Met homozygotes became aware of the underlying pattern in the reward schedule described in *Methods*, indicating that the task relies on implicit learning. The behavioral advantage of Val homozygotes is unlikely to be caused by other single-nucleotide polymorphisms or differences in cognitive function, as we controlled for genetic differences in a wide range of other genetic polymorphisms as well as for differences in a large battery of psychometric tests (see *SI Text*, Figs. S1–S3, Tables S1 and S2 for further details).

Computational Modeling. We implemented a reinforcement learning model with a dynamic learning rate that accounts for the requirements of reversal learning in a dynamic environment. According to this model, increasing magnitudes of unsigned PEs signal that learning needs to be faster; more rapid learning is then accomplished through an increase of the learning rate (see *SI Text* for further details).

Averaged trial-wise learning rates were higher in Val homozygotes than Met homozygotes ($P < 0.001$). We used a repeated-measures ANOVA to investigate the influence of genotype on the adaptation of the learning rate before and after reversals. The significant main effect of time relative to reversal [$F(24) = 5.729$, $P = 0.025$] was qualified by a time × genotype interaction [$F(24) = 4.367$, $P = 0.047$], showing that the learning rate increased more strongly in Val homozygotes (0.37 to 0.47) than Met homozygotes (0.37 to 0.38) after a reversal (Fig. 1*C*).

fMRI Data. *Main effects of positive and negative PEs.* Model-based fMRI analysis in the striatum ($P < 0.01$, corrected) revealed greater fMRI signals in response to positive PEs in the left ventral striatum in Val homozygotes compared with Met homozygotes (Fig. 2*A*; maximum Z-score, 3.26; $x = -18$, $y = 6$, $z = -10$). The same region also showed greater fMRI responses to negative PEs in Val than Met homozygotes (Fig. 2*C*; maximum Z-score, 3.18; $x = -16$, $y = 4$, $z = -12$). Accordingly, a finite impulse response functions analysis
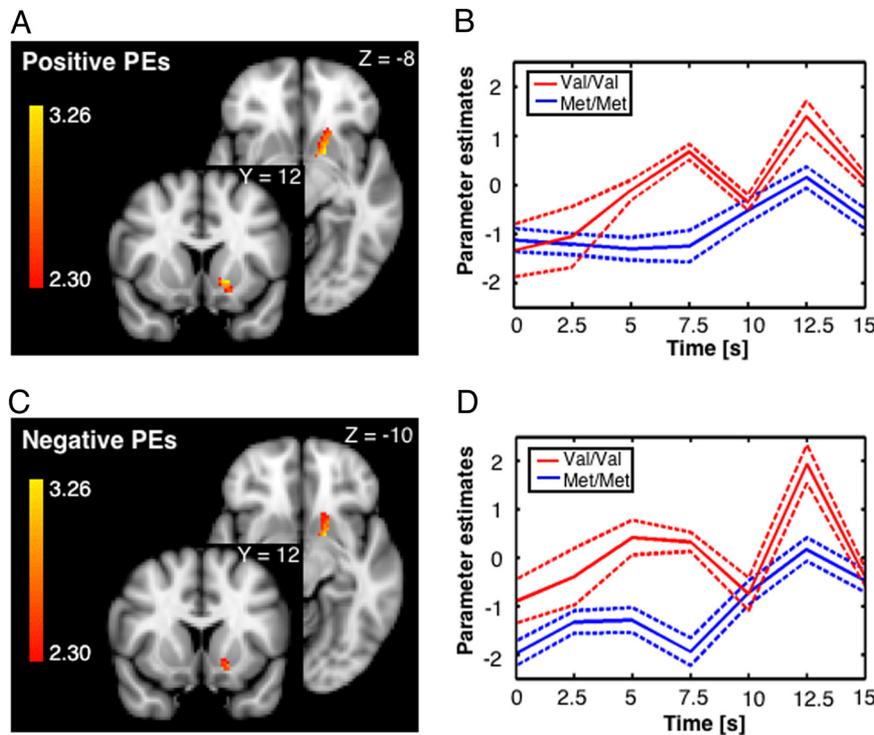
Fig. 2. Genotype-related differences in fMRI responses to positive and negative PEs. (*A*) Greater ventral striatal fMRI responses to positive PEs in Val than in Met homozygotes in the ventral striatum (coordinates of peak voxel: $x = -18$, $y = 6$, $z = -10$). (*B*) Averaged time courses ($\pm$SE) of parameter estimates for voxels within the cluster shown in *A* for Val and Met homozygotes. (*C*) Greater fMRI responses to negative PEs in Val than in Met homozygotes in the ventral striatum ($x = -16$, $y = 4$, $z = -12$). (*D*) Averaged time courses ($\pm$SE) of parameter estimates for voxels within the cluster shown in *C* for Val and Met homozygotes.

revealed that Val homozygotes had higher parameter estimates in response to both positive and negative PEs (Fig. 2 *B* and *D*, respectively).

Covariation with PE magnitude. To further investigate genotype-related differences predicted by our hypothesis, we compared striatal fMRI signal covariations with positive and negative PE magnitudes between the two genotypes. As compared with Met homozygotes, Val homozygotes showed a greater difference in the fMRI signal covariation between positive versus negative PE magnitudes in the right putamen (Fig. 3*A*; maximum *Z*-score, 2.68; $x = 30$, $y = 2$, $z = -8$). The same region showed a significant effect of the covariation difference when considering Val homozygotes alone (maximum *Z*-score, 2.60; $x = 30$, $y = 2$, $z = -8$). The underlying parameter estimates of this interaction effect (Fig. 3*B*) revealed that the fMRI signal covaried positively with PE magnitude in both Val and Met homozygotes for positive PEs, but this effect was stronger in Val homozygotes. For negative PEs, however, the fMRI signal covaried negatively with PE magnitude in Val homozygotes; in contrast, Met homozygotes showed again a positive effect. Importantly, Met homozygotes thus lacked a significant differentiation between positive and negative PEs.

Taken together, Val homozygotes showed greater fMRI responses to PEs in the ventral striatum as well as a more differentiated representation of PE valence in the putamen.

Covariation with learning rate. Based on previous results on learning from errors (7, 31), we expected a representation of the dynamic learning rate estimated by our reinforcement learning model in the posterior medial prefrontal cortex (pmPFC). In line with this prediction, the dynamic learning rate and fMRI response in the paracingulate cortex (which is part of the pmPFC) were significantly correlated (Fig. 4; maximum *Z*-score, 2.99; $x = -4$, $y = 30$, $z = 42$).

Connectivity analysis. To test whether the genotype effects on striatal representations of PEs are driven by the PFC, we performed effective connectivity analyses [implemented as psychophysiological interaction analyses (32)]. Specifically, we assessed whether coupling between the PFC and seed regions in the left ventral striatum [determined by the main effects of positive and negative PEs in Val homozygotes compared with Met homozygotes (Fig. 2 *A* and *C*)] changed depending on the dynamic learning rate. These analyses revealed regions of interaction across all participants in the PFC (Fig. S3; *A*: maximum *Z*-score, 2.86; $x = 24$, $y = 60$, $z = 10$; *B*: maximum *Z*-score, 2.77; $x = -30$, $y = 60$, $z = 10$). Furthermore, we found a greater change in the effective connectivity between the PFC and striatum depending on the learning rate for Val homozygotes than Met homozygotes (Fig. S3; *C*: maximum *Z*-score, 2.84; $x = 22$, $y = 58$, $z = 10$; *D*: maximum *Z*-score, 2.67; $x = 24$, $y = 54$, $z = 10$).
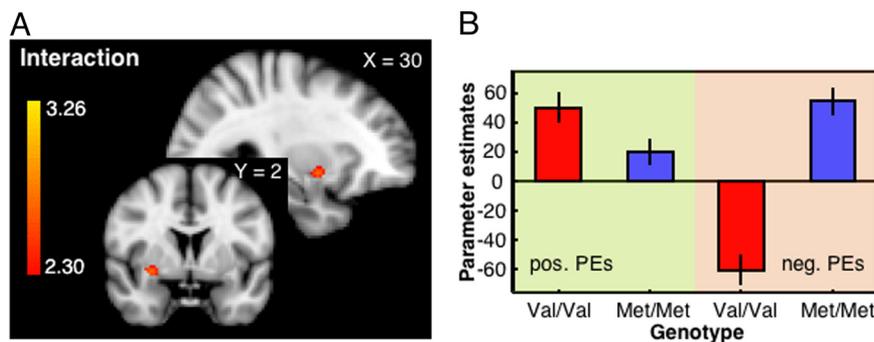

Fig. 3. Genotype-related differences in the covariation of fMRI responses with parametric positive and parametric negative PEs. (*A*) Genotype $\times$ PE interaction with only Val homozygotes showing positive fMRI signal covariation with positive PEs and negative covariation with negative PEs ($x = 30$, $y = 2$, $z = -8$). (*B*) Averaged changes in BOLD signal ($\pm$SE) in voxels within the cluster shown in *A* for covariation with positive and negative PEs and for Val and Met homozygotes, separately.
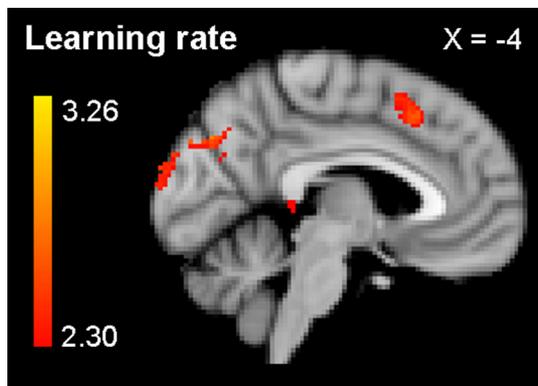
Krugel et al.

Fig. 4. Covariation of BOLD signal with magnitude of dynamic learning rate. BOLD response and learning rate were significantly correlated in the paracingulate cortex as part of the pmPFC ($x = -4$, $y = 30$, $z = 42$).

Discussion

As predicted, Val homozygotes performed better in a reward-based learning task, in which explicit declarative components were minimized and participants were challenged to adapt to ongoing changes in reward contingencies. Using a multilevel approach consisting of a behavioral comparison in combination with model-based fMRI analysis, we were able to relate the behavioral advantage of Val homozygotes to higher and more adaptive weights of PEs, as reflected by the greater and more differentiated striatal fMRI responses to PEs. These findings suggest that higher and more adaptive striatal DA activity underlies the behavioral advantage of Val homozygotes. For a given PE magnitude, Val homozygotes weighted the PE more aptly in the generation of the next reward expectation such that their decisions were better adapted to the current reward schedule. An analysis of effective connectivity provided evidence that a PFC-driven mechanism underlies the Val/Val advantage. First, regions in the PFC showed greater coupling with striatal seed regions for high relative to low values of the dynamic learning rate across all participants (i.e., independent of the genotype). This suggests that PE processing in our implicit reward-based learning task involves updating processes that are supported by a corticostriatal interaction. Additionally, learning rate-dependent changes in the connectivity between the PFC and striatum were greater in Val homozygotes than Met homozygotes. We consider this further evidence that the Val/Val advantage, which relies on a stronger and more differentiated weight of PEs, results from a downstream effect of prefrontal DA levels.

Similar to the effect of genetic variation on dopaminergic neuromodulation observed in the current study, pharmacological DA manipulation modulates performance and striatal fMRI responses in a reward-based learning task, in which enhanced DA availability quantitatively corresponds to increased effective reward values for the same objective outcomes (30). Here, we focused on the expectation updating itself and determined how striatal DA availability affected the weight of PEs on the updating process that preceded the next choice. We therefore implemented a reinforcement learning model with a dynamic learning rate to estimate the trial-wise influence of PEs on the updating process. In a recent study, participants' choices in a two-option probabilistic reversal task were predicted using a dynamic learning rate that is modulated by the volatility of the environment as estimated by a Bayesian learner (31). Similar in spirit but computationally different, we implemented a dynamic learning rate that was plainly modulated by the slope of unsigned PEs.

Crucially, our dynamic learning rate incorporates contextual information without demanding explicit strategic learning mechanisms. In that, we complement a different approach to overcome the challenge of simple reinforcement learning models to describe

human behavior in reversal learning (33). That study successfully deployed abstract models incorporating higher-order task structure to describe choice behavior in another two-option probabilistic reversal task. Focusing on abstract strategic parts of decision making, it is important to note that choice behavior in that study was characterized after participants were extensively and systematically trained with the payoff structure of the task to ensure that they had established an internal model of payoff contingencies before entering the scanner. In contrast, participants in our task implicitly learned to make rewarding choices, as reflected in the fact that none of the participants could describe the payoff structure or reversal pattern upon debriefing. We investigated choices and neural activity in the striatum during an early learning phase, in which participants were not familiar with the higher-order structure of the reversal paradigm. Accordingly, the advantage of Val homozygotes in reward-based decision making is of greatest behavioral relevance in learning environments, in which performance maximally depends on the ability to rapidly and flexibly adapt choices to changing reward contingencies (as it is the case for dynamic environments). Interindividual differences in expectation updating might be less pronounced in stable environments, in which the consistency between expected and received outcomes is high and PEs are therefore low after an initial learning phase. With regard to the approach of Hampton et al. (33) to investigate participants after they establish an internal model of payoff contingencies, we hold that using a dynamic learning rate and incorporating abstract models of task structure are two complementary approaches to enhance standard reinforcement learning to cope with dynamic and complex environments.

Differences in performance and learning rate were paralleled by greater striatal fMRI responses to positive and negative PEs and greater differentiation between the fMRI signal covariation with positive and negative PEs, respectively, in Val homozygotes. Neurophysiological data from single-unit recordings in monkeys show phasic increases in dopaminergic firing to unexpected rewards and phasic dips in the absence of expected reward (1, 2). These are mediated in the striatum by the transient activation of postsynaptic D1 receptors for phasic increases and transient deactivation of postsynaptic D2 receptors for phasic dips. Whereas activation of postsynaptic D1 receptors has been repeatedly correlated with enhanced striatal fMRI signals (34), the relationship between negative outcomes (or negative PEs) and the fMRI signal is less clear (35): Some studies report deactivations in response to negative outcomes (29, 30), whereas others find enhanced fMRI responses (36, 37). Individuals with lower D2 receptor density show lower fMRI responses to negative PEs and impaired learning from errors (7). Based on the neurobiological mechanisms involved in signaling negative outcomes, it has been argued that D2 receptor deactivation transiently activates a population of striatopallidal cells corresponding to a NoGo effect; this then causes an increase in the fMRI signal in response to negative outcomes (38). Our results thus indicate that both phasic increases to positive PEs and phasic dips to negative PEs were greater in Val homozygotes than Met homozygotes. This broader range of phasic activity was complemented by a higher sensitivity to positive and negative PE magnitudes as suggested by our fMRI result in the putamen (Fig. 3), where we observed a differentiation between positive and negative PE magnitudes in Val but not Met homozygotes.

Which neurobiological mechanisms lead to the behavioral advantage of Val homozygotes? Taken together, our behavioral, computational, and imaging results suggest that the advantage of Val homozygotes in the reversal task relies on higher and more flexible striatal DA activity resulting from an indirect downstream effect of the PFC. As described in the Introduction, a number of animal studies suggest that one possible neurobiological mechanism relates to a reciprocal relationship between DA levels in the PFC and dopaminergic burst firing in the striatum (15, 16, 19–21); note, however, that no such effect was found in *COMT* knockout mice

(22). With regard to the *COMT* Val[158]Met polymorphism in humans, results from a postmortem examination of striatal DA projections (23) and from an in vivo positron emission tomography study on the dopaminergic midbrain (24) suggest greater DA availability in the striatum of Val homozygotes. Higher striatal DA availability in Val homozygotes is thought to specifically enhance dopaminergic burst firing as opposed to tonic baseline firing (25). Dopaminergic burst firing results in transiently increased synaptic DA concentrations needed to encode reward PEs through their effect on postsynaptic sites (1, 2). This effect on activity patterns is thought to result from the translation of a lower PFC DA tone into a lower striatal DA tone in Val homozygotes. Inhibitory presynaptic D2 autoreceptors depending on this tone (39, 40) may thereby exert less inhibitory regulation on dopaminergic burst firing in Val homozygotes (25).

In addition to the effects described here, we emphasize that other neurobiological mechanisms for a downstream influence of COMT or for minor local effects of COMT on striatal DA [although predominantly located in the PFC (11)] are conceivable. Furthermore, when investigating interindividual differences in dopaminergic neuromodulation, it is important to consider the role of other mechanisms and interindividual differences [e.g., DA transporter reuptake (13, 14, 17, 18) or DA activity regulation at the postsynaptic site (6, 7, 41)] that determine dopaminergic activity. Recent interaction studies on interindividual differences in COMT and DA transporters very elegantly show that one particular metabolic pathway within the complement determinants of dopaminergic activity in the brain may not only directly affect DA function at its acting site but also indirectly affect other dopaminergic pathways via interaction effects (13, 14). These studies extended the influence of COMT to subcortical function, however, they used gambling paradigms that did not involve learning.

Studies on the differences between Val and Met homozygotes thus far highlighted the advantages of Met homozygotes in tasks that rely on the PFC as direct local effects of the *COMT* Val[158]Met polymorphism. We found an advantage for the ancestral Val genotype in rapid and implicit learning from rewards. The advantage was paralleled by stronger and more differentiated PE-related fMRI signals in the striatum, which suggests that higher and more flexible striatal DA activity underlies this effect. Finally, increased learning rate-dependent changes in effective connectivity between these PE-representing striatal regions and the PFC across both genotypes suggest that the PE updating process is supported by an interaction of the PFC and striatum. The greater changes in effective connectivity in Val than Met homozygotes indicate that the Val/Val advantage is likely to result from an indirect downstream effect of the PFC that is presumably mediated by differences in DA metabolism. Our study thus provides important insights into the mechanisms underlying the ability to rapidly and flexibly adapt decisions to available rewards.

Methods

Participants. Twenty-six healthy, young subjects (12 Val and 14 Met homozygotes) participated in our fMRI experiment after having given informed written consent according to the Declaration of Helsinki. The study was approved by the Ethics Committee of the Charité University Medicine, Berlin, Germany, and by the Max Planck Institute for Human Development. Participants were invited based on their *COMT* Val[158]Met polymorphism configuration from a larger sample ($n = 164$), which did not significantly differ from Hardy–Weinberg equilibrium [$(\chi^2(1) = 0.68, P > 0.10)$] (42). The 12 Val homozygotes (4 females, mean age, 25.7 ± 2.9 years) and 14 Met homozygotes (5 females, mean age, 25.2 ± 3.3 years) did not differ in their composition with regard to any of 52 other considered single-nucleotide polymorphisms (among them *BDNF, DRD1, DRD2, MAO A, MAO B*), as revealed by the exact Fisher–Freeman–Halton test ($P = 0.2$ to confirm the null hypothesis, corrected for multiple comparisons). In addition to the task, participants completed a battery of psychometric tests (*SI Text*). No significant differences in any of these tests were found between the two homozygous *COMT* genotypes.

Task. We used a probabilistic object-reversal task (43, 44), which investigates flexible reward-based learning while minimizing the influence of working memory. The task was to win as many points as possible by learning to choose the most profitable among four choice options (Fig. 1 A). Participants knew in advance that with each single choice, they would win 50, 150, or 250 points and that omitted choices would lead to 0 points. Additionally, participants were informed that the four choice options were not equally good and that the best stimulus might not remain the best choice over the whole experiment. Unknown to participants, the best option led to the highest outcome in 80% of the trials, whereas in 20% of the trials, its choice resulted in one of the lower feedbacks (50 or 150 points). The three bad options had opposite outcome probabilities (low feedback in 80%, high feedback in 20% of the trials). To simulate the situation of a dynamic environment, in which the individual is required to adapt to ongoing changes in reward contingencies, we introduced payoff reversals dependent on subjects' behavior: After the option currently associated with the highest average payoff had been chosen in six trials of seven, the payoff scheme was covertly shifted such that another option was assigned the highest average payoffs. This learning criterion for reversals was set based on results of a behavioral pilot study. We included a written debriefing procedure after the scanning session, in which we asked participants to rate the perceived difficulty in finding the best option and to indicate the switching pattern if they noticed any. See *SI Text* for further details.

Computational Modeling. To better understand the differences in learning between Val and Met homozygotes, we modeled participants' choices by using a reinforcement learning model (45, 46), optimized for reversal learning. Reinforcement learning models assign each option i an expected reward $q_i(t)$, which is used to derive choice probabilities $p_i(t)$ of choosing option i in trial t, here according to a softmax choice rule:

$$p_i(t) = \frac{\exp[\gamma \cdot q_i(t)]}{\sum_{j=1}^{n} \exp[\gamma \cdot q_j(t)]},$$

where n is the number of options, and γ is the sensitivity parameter that determines the influence of reward expectations on choice probabilities (47). After the decision for one option, the received reward $r_i(t)$ is compared with the expected reward $q_i(t)$, with the deviation formalized as prediction error (PE): $\delta_i(t) = r_i(t) - q_i(t)$. Reinforcement learning models assume that learning is driven by these deviations, such that PEs are used to update the expected rewards, thereby allowing for optimizing reward predictions about the different choice options at hand: $q_i(t) = q_i(t - 1) + \alpha \cdot \delta_i(t - 1)$. Here, α is the learning rate that determines the influence of the PE on the expectation updating process.

Traditionally, α is assumed to be a constant parameter. Reversal learning, however, is a challenge for conventional learning models as constant learning rates either do not allow for fast adaptation after the occurrence of a reversal, or do not allow for stabilization of behavior once the best option is found. We therefore implemented a dynamic learning rate $\alpha(t)$, replacing the constant learning rate α in the update equation, that was modulated by the slope m of the smoothed unsigned PEs:

$$\alpha(t) = \alpha(t - 1) + f(m(t)) \cdot (1 - \alpha(t - 1)), \quad \text{if } m > 0$$

$$\alpha(t) = \alpha(t - 1) + f(m(t)) \cdot \alpha(t - 1), \qquad \text{if } m < 0$$

Positive slopes in PE magnitude thus raised the learning rate and increased PEs' influence on the expectation updating, allowing for quick adaptation of behavior after reversals when uncertainty in stimulus–outcome associations increased. By contrast, negative slopes in PE magnitudes lowered the influence of PEs on expectation updating and thus facilitated stabilization of behavior once the best option had been found. Note that the slope of PEs, which modulated the learning rate, was calculated based on smoothed unsigned PEs and thus did not depend only on the last two PEs. As for successful learning in a real world dynamic environment, in which single outliers should not easily be behaviorally relevant, this prevented the model from being oversensitive to single outliers. See *SI Text* for further details.

fMRI Data and Analyses. Acquisition of imaging data and whole-brain analyses were carried out by using standard procedures described in *SI Text*. PEs were modeled by using four regressors derived from crossing positive versus negative and binary versus parametric PEs, where binary regressors corresponded to the intercept in the linear model, and parametric regressors corresponded to the

Krugel et al.

slope, that is, the covariation of PE magnitudes with the fMRI signal, and were orthogonalized with respect to binary PEs.

Because PEs have been repeatedly demonstrated to be represented in the striatum (28–30) and led by our a priori hypothesis about genotype-related differences in striatal fMRI signals, we used an anatomically determined region of interest that was based on the 25% probability for a voxel to belong to the striatum (see *SI Text*). We report clusters of maximally activated voxels that (*i*) survived statistical thresholding at $Z = 2.3$ and (*ii*) had a cluster size of at least 128 mm^3, resulting in a corrected mapwise $P < 0.01$ determined by using Monte Carlo simulations (see *SI Text*).

For the effective connectivity analysis implemented as psychophysiological interaction analysis (32), we assessed changes in effective connectivity between seed regions in the left ventral striatum determined by the contrasts ''Positive PEs, Val/Val > Met/Met'' and ''Negative PEs, Val/Val > Met/Met'' of our original

analyses, and other brain regions dependent on the dynamic learning rate, $\alpha(t)$. We extracted individual blood oxygenation level-dependent (BOLD) time series of voxels in the two striatal seed regions and conducted four separate analyses: two across all participants to reveal genotype-independent effects and two for the contrast ''Val/Val > Met/Met,'' for both striatal seed regions. The psychophysiological interaction was defined as the element-by-element product (interaction term) of the respective striatal time series and a vector coding for trial-to-trial changes in $\alpha(t)$. Group effects were computed by using the transformed contrast images in a mixed-effects model treating subjects as random.

1. Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275:1593–1599.
2. Hollerman JR, Schultz W (1998) Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat Neurosci* 1:304–309.
3. Seamans JK, Yang CR (2004) The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Prog Neurobiol* 74:1–57.
4. Williams GV, Goldman-Rakic PS (1995) Modulation of memory fields by dopamine D1 receptors in prefrontal cortex. *Nature* 376:572–575.
5. Wise RA (2004) Dopamine, learning and motivation. *Nat Rev Neurosci* 5:483–494.
6. Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE (2007) Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc Natl Acad Sci USA* 104:16311–16316.
7. Klein TA, et al. (2007) Genetically determined differences in learning from errors. *Science* 318:1642–1645.
8. Malhotra AK, et al. (2002) A functional polymorphism in the COMT gene and performance on a test of prefrontal cognition. *Am J Psychiatry* 159:652–654.
9. Egan MF, et al. (2001) Effect of COMT Val(108/158) Met genotype on frontal lobe function and risk for schizophrenia. *Proc Natl Acad Sci USA* 98:6917–6922.
10. Barnett JH, Jones PB, Robbins TW, Muller U (2007) Effects of the catechol-*O*-methyltransferase Val(158)Met polymorphism on executive function: A meta-analysis of the Wisconsin Card Sorting Test in schizophrenia and healthy controls. *Mol Psychiatry* 12:502–509.
11. Matsumoto M, et al. (2003) Catechol *O*-methyltransferase mRNA expression in human and rat brain: Evidence for a role in cortical neuronal function. *Neuroscience* 116:127–137.
12. Barnett JH, Scoriels L, Munafo MR (2008) Meta-analysis of the cognitive effects of the catechol-*O*-methyltransferase gene val158/108Met polymorphism. *Biol Psychiatry* 64:137–144.
13. Yacubian J, et al. (2007) Gene-gene interaction associated with neural reward sensitivity. *Proc Natl Acad Sci USA* 104:8125–8130.
14. Dreher JC, Kohn P, Kolachana B, Weinberger DR, Berman KF (2009) Variation in dopamine genes influences responsivity of the human reward system. *Proc Natl Acad Sci USA* 106:617–622.
15. Carr DB, Sesack SR (2000) Projections from the rat prefrontal cortex to the ventral tegmental area: Target specificity in the synaptic associations with mesoaccumbens and mesocortical neurons. *J Neurosci* 20:3864–3873.
16. Takahata R, Moghaddam B (2000) Target-specific glutamatergic regulation of dopamine neurons in the ventral tegmental area. *J Neurochem* 75:1775–1778.
17. Jaber M, Jones S, Giros B, Caron MG (1997) The dopamine transporter: A crucial component regulating dopamine transmission. *Mov Disord* 12:629–633.
18. Floresco SB, West AR, Ash B, Moore H, Grace AA (2003) Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission. *Nat Neurosci* 6:968–973.
19. Roberts AC, et al. (1994) 6-Hydroxydopamine lesions of the prefrontal cortex in monkeys enhance performance on an analog of the Wisconsin Card Sort Test—Possible interactions with subcortical dopamine. *J Neurosci* 14:2531–2544.
20. Kolachana BS, Saunders RC, Weinberger DR (1995) Augmentation of prefrontal cortical monoaminergic activity inhibits dopamine release in the caudate-nucleus—An in-vivo neurochemical assessment in the rhesus monkey. *Neuroscience* 69:859–868.
21. King D, Zigmond MJ, Finlay JM (1997) Effects of dopamine depletion in the medial prefrontal cortex on the stress-induced increase in extracellular dopamine in the nucleus accumbens core and shell. *Neuroscience* 77:141–153.
22. Gogos JA, et al. (1998) Catechol-*O*-methyltransferase-deficient mice exhibit sexually dimorphic changes in catecholamine levels and behavior. *Proc Natl Acad Sci USA* 95:9991–9996.
23. Akil M, et al. (2003) Catechol *O*-methyltransferase genotype and dopamine regulation in the human brain. *J Neurosci* 23:2008–2013.

24. Meyer-Lindenberg A, et al. (2005) Midbrain dopamine and prefrontal function in humans: Interaction and modulation by COMT genotype. *Nat Neurosci* 8:594–596.
25. Bilder RM, Volavka J, Lachman HM, Grace AA (2004) The catechol-*O*-methyltransferase polymorphism: Relations to the tonic-phasic dopamine hypothesis and neuropsychiatric phenotypes. *Neuropsychopharmacology* 29:1943–1961.
26. Durstewitz D, Seamans JK, Sejnowski TJ (2000) Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. *J Neurophysiol* 83:1733–1750.
27. Palmatier MA, Kang AM, Kidd KK (1999) Global variation in the frequencies of functionally different catechol-*O*-methyltransferase alleles. *Biol Psychiatry* 46:557–567.
28. O'Doherty J, et al. (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304:452–454.
29. Schonberg T, Daw ND, Joel D, O'Doherty JP (2007) Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *J Neurosci* 27:12860–12867.
30. Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD (2006) Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* 442:1042–1045.
31. Behrens TE, Woolrich MW, Walton ME, Rushworth MF (2007) Learning the value of information in an uncertain world. *Nat Neurosci* 10:1214–1221.
32. Friston KJ, et al. (1997) Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* 6:218–229.
33. Hampton AN, Bossaerts P, O'Doherty JP (2006) The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J Neurosci* 26:8360–8367.
34. Knutson B, Gibbs SEB (2007) Linking nucleus accumbens dopamine and blood oxygenation. *Psychopharmacology* 191:813–822.
35. Dayana P, Niv Y (2008) Reinforcement learning: The Good, The Bad and The Ugly. *Curr Opin Neurobiol* 18:185–196.
36. Cools R, Clark L, Owen AM, Robbins TW (2002) Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *J Neurosci* 22:4563–4567.
37. Rodriguez PF, Aron AR, Poldrack RA (2006) Ventral-striatal/nucleus-accumbens sensitivity to prediction errors during classification learning. *Hum Brain Mapp* 27:306–313.
38. Frank MJ (2005) Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated parkinsonism. *J Cogn Neurosci* 17:51–72.
39. Grace AA (1991) Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: A hypothesis for the etiology of schizophrenia. *Neuroscience* 41:1–24.
40. Moore H, West AR, Grace AA (1999) The regulation of forebrain dopamine transmission: Relevance to the pathophysiology and psychopathology of schizophrenia. *Biol Psychiatry* 46:40–55.
41. Meyer-Lindenberg A, et al. (2007) Genetic evidence implicating DARPP-32 in human frontostriatal structure, function, and cognition. *J Clin Invest* 117:672–682.
42. Nagel IE, et al. (2008) Human aging magnifies genetic effects on executive functioning and working memory. *Front Hum Neurosci* 2:1.
43. Mell T, et al. (2005) Effect of aging on stimulus-reward association learning. *Neuropsychologia* 43:554–563.
44. Heekeren HR, et al. (2007) Role of ventral striatum in reward-based decision-making. *NeuroReport* 18:951–955.
45. Rescorla RA, Wagner AR (1972) in *Classical Conditioning II: Current Research and Theory*, eds Black AH, Prokasy WF (Appleton–Century Crofts, New York), pp 64–99.
46. Sutton RS, Barto AG (1998) *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, MA).
47. Luce RD (1959) *Individual Choice Behavior: A Theoretical Analysis* (Wiley, New York).