Behavioral/Systems/Cognitive

Dopaminergic Genes Predict Individual Differences in Susceptibility to Confirmation Bias

Bradley B. Doll,[1] Kent E. Hutchison,[4] and Michael J. Frank[1,2,3]

Departments of [1]Cognitive, Linguistic, and Psychological Sciences and [2]Psychiatry and Human Behavior and [3]Brown Institute for Brain Science, Brown University, Providence, Rhode Island 02912, and [4]Department of Psychology, University of New Mexico, Albuquerque, New Mexico 87131

The striatum is critical for the incremental learning of values associated with behavioral actions. The prefrontal cortex (PFC) represents abstract rules and explicit contingencies to support rapid behavioral adaptation in the absence of cumulative experience. Here we test two alternative models of the interaction between these systems, and individual differences thereof, when human subjects are instructed with prior information about reward contingencies that may or may not be accurate. Behaviorally, subjects are overly influenced by prior instructions, at the expense of learning true reinforcement statistics. Computational analysis found that this pattern of data is best accounted for by a confirmation bias mechanism in which prior beliefs—putatively represented in PFC—influence the learning that occurs in the striatum such that reinforcement statistics are distorted. We assessed genetic variants affecting prefrontal and striatal dopaminergic neurotransmission. A polymorphism in the COMT gene (rs4680), associated with prefrontal dopaminergic function, was predictive of the degree to which participants persisted in responding in accordance with prior instructions even as evidence against their veracity accumulated. Polymorphisms in genes associated with striatal dopamine function (DARPP-32, rs907094, and DRD2, rs6277) were predictive of learning from positive and negative outcomes. Notably, these same variants were predictive of the degree to which such learning was overly inflated or neglected when outcomes are consistent or inconsistent with prior instructions. These findings indicate dissociable neurocomputational and genetic mechanisms by which initial biases are strengthened by experience.

Introduction

Overwhelming evidence across species indicates that dopaminergic neurons signal reward prediction errors (Schultz, 1998; Roesch et al., 2007; Zaghloul et al., 2009). These phasic dopamine (DA) responses facilitate corticostriatal synaptic plasticity (Centonze et al., 2001; Reynolds et al., 2001; Shen et al., 2008) that is necessary and sufficient to induce behavioral reward learning from experience (Tsai et al., 2009; Zweifel et al., 2009). However, reward is better harvested from some environments through strategies that dismiss individual instances of feedback. For example, in the stock market, a standard reinforcement learning (RL) mechanism (Sutton and Barto, 1998) would employ a ruinous "buy high, sell low" scheme, whereas a cleverer investor would buy and hold over the ups and downs in the interest of long-term gains.

Behavioral and computational work suggests that verbal instruction (such as that from a financial planner) can powerfully control choice (Hayes, 1989; Waldmann and Hagmayer, 2001; Biele et al., 2009; Doll et al., 2009), often leading to a confirmation bias (Nickerson, 1998) whereby subjects behave in accordance with contingencies as they are described, rather than as they are actually experienced. Functional magnetic resonance imaging evidence suggests that verbal information might exert control by altering activation in brain areas implicated in reward learning (Plassmann et al., 2008; Engelmann et al., 2009; Li et al., 2011), though the mechanism for such modulation remains unclear. Our previously developed computational models (Doll et al., 2009) propose two biologically viable alternative explanations for this effect, entailing different versions of guided-activation theory whereby rule representations in prefrontal cortex (PFC) modulate downstream neural activation (Miller and Cohen, 2001).

In the first (override model), the striatum learns objective reinforcement probabilities as experienced, but is overridden by the PFC at the level of the decision output. In the second (bias model), PFC instruction representations bias striatal action selection and learning. These models (see Figs. 1*B*, 4*A*) describe competitive and cooperative relationships between striatum and PFC, and make opposite predictions regarding the striatal valuation of instructed stimuli.

Probabilistic reinforcement learning is modulated by striatal dopamine manipulation (Frank et al., 2004; Pessiglione et al., 2006; Bódi et al., 2009). Striatal activation and dopaminergic genes are predictive of individual differences in such learning (Cohen et al., 2007; Frank et al., 2007; Klein et al., 2007; Schönberg et al., 2007; Jocham et al., 2009). However, no studies have investigated neurochemical or genetic influences on learning when subjects are given explicit information about the learning environment. We reasoned that the directionality of any striatal

Received Dec. 13, 2010; revised Feb. 27, 2011; accepted March 4, 2011.

Author contributions: B.B.D. and M.J.F. designed research; B.B.D. performed research; K.E.H. contributed unpublished reagents/analytic tools; B.B.D. analyzed data; B.B.D., M.J.F., and K.E.H. wrote the paper.

This work was supported by National Institutes of Health Grant R01MH080066-01. We thank Christina Figueroa for assistance with data collection.

Correspondence should be addressed to Bradley B. Doll, Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI 02912. E-mail: Bradley_Doll@brown.edu.

DOI:10.1523/JNEUROSCI.6486-10.2011

 0270-6474/11/316188-11$15.00/0

DA-related genetic effects on instructed learning could help discriminate between our candidate models. According to the override model, enhanced striatal DA function should support better learning from environmental contingencies. In contrast, according to the bias model, striatal DA genetic markers would predict the degree to which learning is modulated by prior instructions.

We collected DNA from subjects performing an instructed probabilistic RL task. We find that genotypes associated with better learning from rewards [dopamine- and cAMP-regulated phosphoprotein of molecular weight 32 kDa (DARPP-32)] and punishments (DRD2) are predictive of distorted learning of reinforcement statistics in accordance with prior information. We further report that a catechol-*O*-methyl transferase (COMT) genotype typically associated with better prefrontal function impedes the discovery of true reinforcement contingencies. These data support the bias model, where PFC rule-like representations distort the effect of reinforcement in the striatum, confirming prior beliefs.

Materials and Methods

Theoretical framework

In this section, we describe the assumptions that have been simulated in our a priori neural network and algorithmic models of instructional control over learning, which allow us to derive differential predictions regarding the role of dopaminergic genes in capturing individual differences. The computational models have been described in detail previously (Doll et al., 2009); here we highlight the key features. These models build on previous basal ganglia network models (Frank, 2005, 2006), with the addition of PFC architecture to represent instructions and permit instruction following. In these models, the striatum learns the positive and negative values of stimulus–action combinations via dopaminergic modulation of synaptic plasticity. In particular, DA bursts modulate reward-driven activity and plasticity along the direct pathway via D_1 receptors, whereas DA dips modulate punishment-driven activity and plasticity along the indirect pathway via disinhibition of D_2 receptors (for review, see Doll and Frank, 2009). In both models, a population of PFC units represents the instructed rule in working memory. There are two routes with which PFC information can influence action selection, due to anatomical connections to multiple striatal regions and also directly to premotor/motor cortex (Haber, 2003; Wallis and Miller, 2003; Draganski et al., 2008). The roles of these different pathways in instructed choice are qualitatively different and are assessed in two separate models.

When a stimulus that had been associated with instructions is presented, it activates a corresponding population of PFC units that encode the rule. In the override model, this representation directly activates the instructed response in motor cortex. Here the striatum is free to represent the learned value of alternative responses, and to attempt to select a (possibly different) response via gating of motor cortex, but it must compete with the instructed response.

In the bias model, PFC influences action selection by sending glutamatergic projections to striatum, thereby biasing striatal representations to be more likely to gate actions consistent with the instructed rule. This modulation of striatal activity is in turn associated with modulation of activity-dependent plasticity, and hence learning. This mechanism amplifies the effects of instruction-consistent outcomes while diminishing the effects of inconsistent ones, such that the PFC "trains" the basal ganglia (BG) to learn in accordance with prior beliefs, while distorting the true observed statistics. The effects of DA bursts accompanying rewards for the instructed response are exaggerated due to top-down PFC influence, further promoting long-term potentiation (LTP). Similarly, the effects of DA dips accompanying instructed punishments are reduced, with top-down PFC input onto D_1 units biasing the activation landscape such that D_2 units do not learn as much as they would otherwise from DA dips. In this model, in contrast to the override model, PFC and striatum cooperate in the distortion of learning to confirm the instructions.

These two models make distinct predictions regarding the effects of individual genetic variation in striatal and PFC DA function on instructional control of reinforcement learning (see Fig. 4*A*). In particular, the bias model suggests that the same striatal genetic variations that predict better learning from positive and negative outcomes should be predictive of the extent to which such learning is distorted by instructional influence. Increases in striatal efficacy enhance the cooperative confirmation bias learning mechanism implemented by PFC and striatum. Thus, genotypes associated with better uninstructed learning should be associated with the opposite pattern (worse learning) when given an invalid instruction. In contrast, the override model suggests that these genes should predict the degree to which subjects learn the true programmed task contingencies regardless of instruction. Here, increases in striatal efficacy increase the ability of this region to compete with PFC at motor cortex for action selection. Our experiment sought to arbitrate between these alternatives with behavioral manipulations. Our computational models described below provide an estimate of the degree to which outcomes consistent and inconsistent with instructions are amplified or distorted.

Sample

Eighty subjects (51 females, mean age 22.5, SE 0.5) were recruited from Brown University and the Providence, Rhode Island, community and were paid $10 for completing the study. Subjects provided saliva samples before completing two iterations of a cognitive task. Of these, four computer crashes eliminated two subjects entirely from the dataset and two subjects from the second iteration dataset (due to the crashes occurring at the start and end of the experiment, respectively). Finally, we eliminated subjects unable to demonstrate task learning (details below). By our filtering scheme, six subjects were eliminated from the first and six from the second iteration of the task (with one subject performing below chance on both iterations). Results did not differ if these subjects were included in the analyses. We obtained genotype data on three polymorphisms associated with DA function: the val158met single-nucleotide polymorphism (SNP) of the COMT gene (rs4680), a SNP of the PPP1R1B (DARPP-32) gene (rs907094), and a SNP of the DRD2 gene (rs6277).

For four subjects, we were unable to obtain any genotype data from the samples. For another two subjects, DRD2 genotyping failed. Of successfully genotyped subjects, frequencies per allele were COMT—28:33:13 (Val/Val:Val/Met:Met/Met), DRD2—22:38:12 (C/C:C/T:T/T), and DARRP-32—10:37:27 (C/C:C/T:T/T). All three SNPs were in Hardy–Weinberg equilibrium (χ^2 values < 1, p values > 0.05). No correlations between categorical gene groups (see below, Data analysis) were significant (p values >0.15), though DRD2 T and COMT Met alleles correlated ($r_{(72)} = 0.23$, $p = 0.05$). The behavioral and computational effects reported for each SNP remain significant when controlling for this correlation.

The majority of the sample (58 subjects) classified themselves as Caucasian, 10 as Asian, 7 as African-American, and 4 as "other" (1 subject declined classification). Eight individuals classified themselves as Hispanic. Because population stratification represents a potential confound for the observed genetic effects, several additional measures were taken to verify that the effects reported herein were not due to admixture in the sample. Behavioral and computational results remained significant for the entire sample when minority subgroups were excluded. In addition, allele frequencies did not differ from Hardy–Weinberg equilibrium in any subgroup when analyzed independently ($\chi^2 < 2.9$, p values >0.05). In sum, there was little evidence to suggest that the genetic effects observed in the present study were due to population admixture in the sample.

Genotyping method

DNA was collected via 2 ml salivettes (DNA Genotek). Samples were genotyped using TaqMan primer and probe pairs; the probes are conjugated to two different dyes (one for each allelic variant). Taqman assays

are designed and selected using the SNPBrowser program (Applied Biosystems). The PCR mixture consists of 20 ng of genomic DNA, 1× Universal PCR Master Mix, a 900 nM concentration of each primer, and a 200 nM concentration of each probe in a 15 μl reaction volume. Amplification was performed using the TaqMan Universal Thermal Cycling Protocol, and fluorescence intensity was measured using the ABI Prism 7500 Real-Time PCR System. Genotypes were acquired using the 7500 system's allelic discrimination software (SDS version 1.2.3).

Cognitive task

To assess the effect of instructions on learning, we administered a modified version of the same probabilistic selection task previously shown to be sensitive to striatal DA function by numerous measures, including genetic variability (Frank et al., 2004, 2007; Frank and O'Reilly, 2006). In this task (Fig. 1*A*), subjects select repeatedly among three randomly presented stimulus pairs (AB, CD, and EF), and receive corrective feedback after each choice. Before completing the instructed probabilistic selection task, subjects read the task instructions that appeared on screen:

> Please read through these instructions carefully. It is important that you understand them before beginning the task. Two black symbols will appear simultaneously on the computer screen. One symbol will be "correct" and the other will be "incorrect", but at first you won't know which is which. Try to guess the correct figure as quickly and accurately as possible. There is no ABSOLUTE right answer, but some symbols will have a higher chance of being correct than others. Try to pick the symbol you find to have the highest chance of being correct. This symbol will have the highest probability of being correct: [symbol shown]. You'll have to figure out which of the other symbols you should select by trying them out. Press the "0" key to select the stimulus on the right, and the "1" key to select the symbol on the right. Now you will be tested on these instructions to make sure you have understood them fully.

After reading the instructions, subjects were quizzed to make sure they understood the instructions. They were (1) asked how many stimuli would appear on screen at a time, (2) asked how to select the left and right stimuli, and (3) shown all of the stimuli and asked to select the stimulus they were told would have the highest chance of being correct. Incorrect answers on any questions restarted the instructions and subsequent test. The training phase consisted of a minimum of two blocks of 60 trials (20 of each type: AB, CD, and EF, presented randomly). Subjects completed the training phase if, after two blocks, they were at or above accuracy criteria for each uninstructed stimulus pair (65% A in AB, 60% C in CD, and 50% E in EF); otherwise, training blocks recurred until criteria were met. The test phase was then completed after reading the following instructions:

> It's time to test what you've learned! During this set of trials you will NOT receive feedback ("correct" or "incorrect") to your responses. If you see new combinations of symbols in the test, please choose the symbol that "feels" more correct based on what you learned during the training sessions. If you're not sure which one to pick, just go with your gut instinct!

All possible stimulus combinations were presented, with each unique pair appearing six times. This test phase enables us to estimate the value assigned to each stimulus as compared to all other stimuli, using both traditional behavioral analysis (proficiency in selecting/avoiding stimuli that are statistically superior/inferior options) and computational model fits to these choices. Subjects completed two iterations of training and testing, and were instructed on one randomly selected stimulus before each training run (thereby ensuring that most participants would be instructed at least once with a suboptimal stimulus).

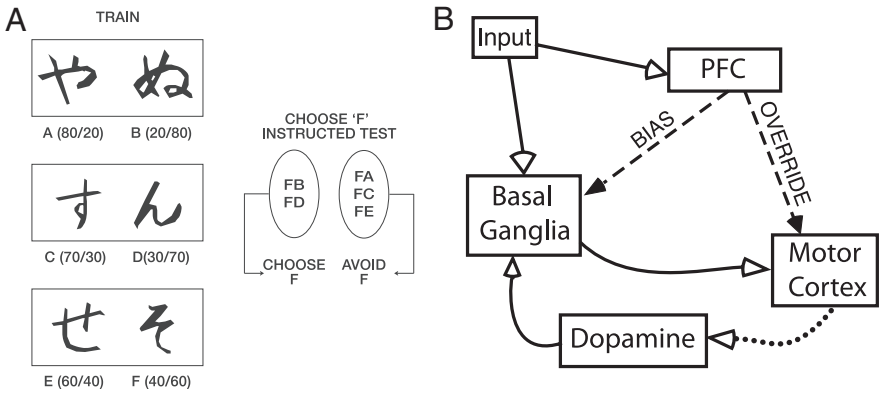

Figure 1. ***A***, Instructed probabilistic selection task. Example stimulus pairs, which minimize explicit verbal encoding. Correct choices are determined probabilistically, with percentage positive/negative feedback shown in parentheses for each stimulus. Before training, subjects were shown one randomly selected stimulus and told it "will have the highest probability of being correct." Instructions on stimuli in the left column are accurate in training pairs. Instructions for those on the right are inaccurate. After training, subjects completed a test phase in which all stimulus combinations were presented. Instructional control is assessed by choices of the instructed stimulus *i* on test choose-*i* trials when it is the statistically superior option and avoid-*i* trials when it is statistically inferior. The figure shows test trials for a subject instructed to choose "F." ***B***, Diagram depicting neural network accounts of instructional control over learning. Dashed lines indicate anatomical projections with differing computational roles. Instruction representations either directly bias the striatal valuation, selection, and learning (bias model), or simply override the otherwise accurate striatal learning of probabilities via competition at motor cortex (override model). The dotted line indicates the time course in the evaluative loop, not an anatomical projection.

In the second task iteration, the instructions were repeated and the task was completed with new stimuli. After completion of the second task iteration, subjects were queried on the instructed stimulus for both iterations. We removed subjects from the dataset who did not show evidence in the test phase of having learned the reinforcement statistics in training. Specifically, subjects at or below chance on the AB test pair were removed. For subjects instructed on A or B, above chance performance on the CD test pair was used.

Data analysis

Due to variability in allele frequencies, for all categorical genotype analyses we compared the largest homozygote group to the other two (i.e., the smallest homozygote group was combined with the heterozygotes). This produced the following groups: COMT—Val/Val, Val/Met+Met/Met; DRD2—C/C, C/T+T/T; DARRP-32—C/C+C/T, T/T.

Behavioral data were analyzed using mixed models with gene groups as categorical variables, and, in the gene-dose analyses, with allele numbers as continuous variables. Tests on parameter estimates from computational models were conducted with logistic regression.

Degrees of freedom listed in the statistical tests we present vary somewhat due to available trials for each analysis (e.g., subjects instructed on stimulus A or B were dropped from choose-A and avoid-B replication analysis) and available subjects meeting accuracy criteria in each iteration.

Models

Prior theoretical and empirical work identified several candidate models for instructional control over reinforcement learning in this task (Doll et al., 2009). We applied the same models to the larger sample of data here to determine which is the best behavioral fit, by maximizing the log-likelihood (Burnham and Anderson, 2002) of each individual subject's responses (Matlab optimization toolbox), using multiple random starting points for each model fit. We then analyzed genetic effects on parameters of interest in only the models providing the best behavioral fit (preventing multiple genetic comparisons that could be done across models).

Instructed learning (bias) model. We assume subjects learn the value of choosing each stimulus s in the training phase as a function of each outcome. As in prior algorithmic models of this task (Frank et al., 2007; Doll et al., 2009), we allow for asymmetric learning rates in accord with the posited dissociable neural pathways supporting learning from positive and negative reward prediction errors (Frank, 2005). Specifically, on each trial we compute the action value Q of choosing stimulus s as follows:

$$Q_s(t+1) = Q_s(t) + [\alpha_G \times \delta(t)]_+ + [\alpha_L \times \delta(t)]_-, \quad (1)$$

where $\delta(t) = r(t) - Q_s(t)$ is the reward prediction error on trial t computed as the difference between delivered reward r (1 for gains 0 for losses) and expected reward for the selected action. Free parameters α_G and α_L are the learning rates applied for gain (δ_+) and loss (δ_-) trials, respectively. These learning rates estimate the impact of each positive or negative outcome on the updating of action values for uninstructed trials.

The critical feature of the bias model is that when the instructed stimulus i is present and chosen, learning is distorted to confirm the instructions:

$$Q_i(t+1) = Q_i(t) + \alpha_{IA}[\alpha_G \times \delta(t)]_+ + \left[\frac{\alpha_L}{\alpha_{ID}} \times \delta(t)\right]_-, \quad (2)$$

where α_{IA} and $\alpha_{ID} \geq 1$ are free parameters that respectively amplify and diminish the impact of reinforcement that follows the selection of the instructed stimulus. These parameters capture the proposed top-down PFC–BG confirmation bias, increasing striatal evaluation of gains and reducing that of losses. Choice on each trial is modeled with the "softmax" choice rule commonly used to model striatal action selection: the probability of selecting stimulus s1 over s2 is computed as follows:

$$P_{s1}(t) = \frac{e^{\frac{Q_{s1}(t)}{\zeta}}}{e^{\frac{Q_{s1}(t)}{\zeta}} + e^{\frac{Q_{s2}(t)}{\zeta}}}, \quad (3)$$

where ζ is the temperature parameter controlling the gain with which differences in Q values produce more or less deterministic choice.

Standard Q-learning model. We compared the fit of the instructed learning model with several other models of subject behavior. First, we fit a basic uninstructed Q-learning model with no provisions for incorporating prior beliefs into outcome evaluations. This model computes Q values for each stimulus identically to those computed by the instructed learning model for uninstructed trials (Eq. 1) and selects between them according to the softmax choice rule (Eq. 3). Thus this model is equivalent to fixing the confirmation bias parameters α_{IA} and α_{ID} to 1 for all subjects, and serves as a control against which to compare models attempting to account for the effect of instructions.

Bayesian "strong prior" model. Next, we fit a Bayesian Q-learning model to subject choices, which computes a distribution of Q values for each stimulus (Dearden et al., 1998; Doll et al., 2009). Given the binomial outcomes in this task, Q values were represented with the conjugate prior of the binomial distribution, the beta distribution, which has the following density:

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du}. \quad (4)$$

On each outcome, the posterior distribution is computed by updating the hyperparameters α and β, which store a running total of gain and loss outcomes, respectively. These counts were allowed to decay back toward uniform with two free parameters γ_G and γ_L, which were applied to gain and loss trials, consistent with separable neural mechanisms for learning from rewards and punishments. Choices were modeled by submitting the mean of the Q distributions $[\mu(Q) = \alpha/(\alpha+\beta)]$ for each stimulus to the softmax choice rule (Eq. 3).

In this model, the prior probability distribution of the instructed stimulus on the first trial was allowed to vary (α was free, $\beta = 1$), whereas the priors for uninstructed stimuli were all uniform (flat). Allowing α to vary captures the initial biasing effect of the instructions on each subject, testing the possibility that subjects simply have a "strong prior" belief in the veracity of the instructions, but one that could be eventually overridden with sufficient experience. Greater priors (higher α) indicate that subjects require greater evidence of true contingencies before abandoning the instructions. In this case, instructions do not modulate the learning process.

Bayesian "hypothesis testing" model. In all of the models described above, we estimated the parameters of the learning models that best explain participants' choices during the final test phase, which is most diagnostic for assessing the learned values of each stimulus (see fitting details below) (Frank et al., 2007; Doll et al., 2009; Cavanagh et al., 2010b; Chase et al., 2010). However, we also sought to model the dynamics of the learning phase in which participants may at some point decide to abandon the instructions. Models of this learning phase are of particular interest in elucidating the behavioral effects of COMT on the persistence of instructional control.

We investigated the possibility that during training, participants undergo "hypothesis testing" with respect to the veracity of the instructions, and that they abandon the instructed rules when they are sufficiently confident that they are incorrect. To this end, we fit a Bayesian model similar to the "strong prior" model described above to the training phase data for inaccurately instructed subjects (Doll et al., 2009).

As in the model discussed above, we modeled the expected Q value of each stimulus as a beta distribution, with the added assumption that participants would choose the instructed stimulus and then abandon it only when they were confident that its mean Q value was less than that expected by chance. The probability of abandoning the instructed stimulus on trial t was computed as follows:

$$P_{abandon}(t) = \frac{e^{\frac{1}{2\zeta}}}{e^{\frac{1}{2\zeta}} + e^{\frac{\mu_{inst}(t) + \phi(\sigma_{inst}(t))}{\zeta}}}, \quad (5)$$

where μ_{inst} is the mean of the instructed Q distribution, and σ_{inst} is the standard deviation (uncertainty) of that distribution,

$$\sigma_{inst} = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}. \quad (6)$$

Thus this model predicts that subjects will only abandon the instructions if this mean is at least ϕ standard deviations below chance (0.5; or equivalently, if the mean is confidently below that for the alternative stimulus, since they know that one is correct and the other is not). The critical parameter ϕ determines how much confidence (certainty) is needed before abandoning instructions. The probability of selecting the instructed stimulus is simply $P_{inst}(t) = 1 - P_{abandon}$.

To capture the dynamics of initially following instructions before rejecting them, we restricted these model fits to those subjects who were inaccurately instructed in the first iteration of the task, due to the additional complication of trying to account for potential changes in criteria for abandoning instructions between task iterations.

Model fitting details

As in prior modeling efforts, parameters were optimized separately to fit choices in the training and testing phases of the task (Frank et al., 2007; Doll et al., 2009; Cavanagh et al., 2010b; Chase et al., 2010). In the training phase, parameter fits are optimized to account for choice dynamics as a function of trial-to-trial changes in reinforcement, working memory, hypothesis testing, etc. Such putatively PFC and hippocampal-dependent strategizing requires explicit recall of recent trial outcomes. When optimized to the test phase (in which there is no reinforcement feedback and therefore no working memory/hypothesis testing, and participants are asked to select among novel combinations of stimuli having a range of reinforcement probabilities), parameter fits provide an estimate of the learning process that yields final stabilized stimulus–action values that best explain participants' allocation of choices (Frank et al., 2007; Doll et al., 2009). Thus the test phase provides a relatively more pure measure of incrementally learned (putatively striatal) values than

the training phase. Supporting these distinctions, dopaminergic drugs, Parkinson's disease, and DARPP-32/DRD2 genotypes all affect test phase performance (Frank et al., 2004; Frank and O'Reilly, 2006), including parameter estimates (Frank et al., 2007), whereas trial-to-trial adaptations and associated parameter estimates during the training phase are predicted by frontal EEG and COMT genotype (Frank et al., 2007; Cavanagh et al., 2010a). Pharmacological challenge with the benzodiazepine midazolam, which reduces cerebral blood flow in the PFC and hippocampus but not the striatum (Bagary et al., 2000; Reinsel et al., 2000), further supports this view, impairing training but not test phase performance (Frank et al., 2006).

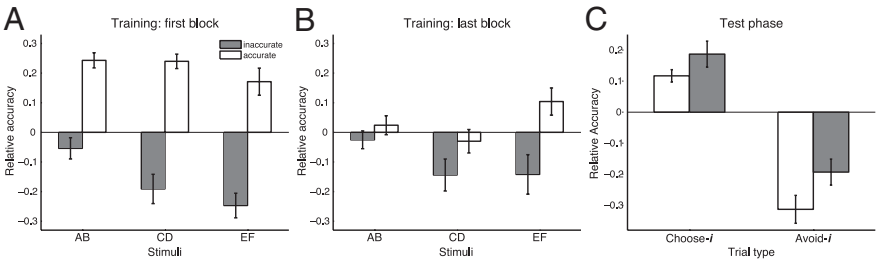

Figure 2. Effects of instruction on accuracy, plotted as relative accuracy: the difference in accuracy on instructed stimulus pairs relative to accuracy on uninstructed pairs of the same type (e.g., performance on instructed pair CD was compared with that of subjects in CD who had not been instructed on this pair). Error bars here and throughout reflect SE. ***A***, Accuracy relative to control is shifted in the direction of the instructions for all stimuli (p values <0.0001, uncorrected; all remain significant after correction) except "B," the worst stimulus statistically ($p = 0.16$). ***B***, By the last block of training, "D" and "F" instruction continue to affect choice (p values <0.03, uncorrected, only "D" survives correction). Effect of "E" is marginal ($p = 0.07$). ***C***, Instructional control at test. Choice of instructed stimuli increases accuracy on choose-*i* trials (where the instructed stimulus is statistically better than its paired alternative) and decreases accuracy on avoid-*i* trials (where it is statistically worse).

Model comparisons were made with Akaike information criterion (AIC) weights (Burnham and Anderson, 2002; Wagenmakers and Farrell, 2004), computed from Akaike information criterion values (Akaike, 1974), and with hierarchical Bayesian model comparison for group studies, which, given AIC values for model fits to each individual, provides an exceedance probability estimating the likelihood that each of the considered models provides the best fit (Stephan et al., 2009). Each of these measures is interpretable as the probability that a given model is the best among the candidates (see Tables 2, 3). We also report mean AIC values, where lower numbers indicate better fits.

Identifiability of confirmation bias parameters

We estimated α_{IA} and α_{ID} in separate models, which provided very stable parameter estimates across multiple runs. We note that α_{ID} and α_{IA} are separately estimable in principle, due to differences in the number of trials in which instruction-inconsistent losses could be discounted and instruction-consistent gains could be amplified. The relative proportion of these trials depends on the value of the instructed stimulus, which is randomized across subjects (and gene groups). Thus, this procedure allowed us to estimate the degree to which each behavioral genetic effect is best accounted for by modulation of α_{ID} versus α_{IA}. Note also that the estimation of each of these confirmation bias parameters is constrained by the best-fitting standard learning rates α_G and α_L for each subject, which were themselves largely determined by performance on (the majority of) uninstructed trials. Thus, confirmation bias parameters reflect the extent to which an individual's learning from gains or losses is altered due to instructions. We computed the log-likelihood of this five parameter model as the average of the best fits estimating the amplifying and diminishing parameters.

Results

Behavioral results

Training data showed that subjects tended to choose in accordance with the instructions (effect of instruction on choice, $F_{(1,76)} = 95.2$, $p < 0.0001$). This tendency waned over training as experience about the true contingencies accumulated (effect of training block, $F_{(1,76)} = 6.6$, $p = 0.01$). Notably, despite this learning effect across training, instructional control was strong in the test phase, with subjects choosing the instructed stimulus more often than in control conditions regardless of whether it was accurate to do so ($F_{(1,76)} = 92.4$, $p < 0.0001$) (Fig. 2). This result is consistent with our previous report (Doll et al., 2009). Task iteration did not significantly impact the training or test phase results. We next examine the genetic correlates of this effect.

Behavioral genetic effects

To test the predictions of the candidate models (Fig. 1*B*, see Fig. 4*A*) (Doll et al., 2009), we assessed SNPs of three genes primarily influencing prefrontal and striatal dopaminergic function. These SNPs were chosen both for data supporting their specific functional involvement in BG or PFC efficacy (see below), and for their prior associations with uninstructed RL measures (Frank et al., 2007, 2009).

First, we assessed the Val158Met polymorphism within the COMT gene, which codes for an enzyme that breaks down extracellular DA and affects individual differences in prefrontal DA levels (Gogos et al., 1998; Huotari et al., 2002; Matsumoto et al., 2003), predicting D_1 receptor availability in prefrontal cortex (Slifstein et al., 2008). In particular, carriers of the Met allele have reduced COMT efficacy, and therefore greater persistence of prefrontal DA, promoting sustained PFC cellular activation and working memory for abstract rules (Durstewitz and Seamans, 2008; Durstewitz et al., 2010). These molecular effects are supported by behavioral and neuroimaging observations that COMT influences prefrontal working memory, executive function, and higher-order cognitive faculties in RL environments, such as directed exploration and lose-switch strategizing (Egan et al., 2001; Tunbridge et al., 2004; Frank et al., 2007, 2009; de Frias et al., 2010). Thus, we reasoned that COMT Met alleles would index increasing stability of prefrontal working memory representations (Durstewitz and Seamans, 2008), and could therefore predict individual differences in decision making based on prior instructions. COMT effects on striatal DA are negligible, apparently due to the presence of the more efficient DA transporters in that region (Gogos et al., 1998; Sesack et al., 1998; Huotari et al., 2002; Matsumoto et al., 2003; Tunbridge et al., 2004). Nevertheless, COMT may modulate striatal activity indirectly, by influencing prefrontal neurons that project to striatum (Krugel et al., 2009).

Second, we assessed a polymorphism within the PPP1R1B gene coding for DARPP-32. DARPP-32 is an intracellular protein abundant in striatum that, when phosphorylated by D_1 receptor stimulation, inhibits protein phosphatase-1, thereby facilitating corticostriatal synaptic plasticity and behavioral reward learning (Ouimet et al., 1984; Calabresi et al., 2000; Svenningsson et al., 2004; Valjent et al., 2005; Stipanovich et al., 2008). Prior studies showed that a SNP within PPP1R1B influenced learning from positive relative to negative reward prediction errors in humans, with learning increasing as a function of T alleles (Frank et al., 2007, 2009). This result is consistent with the aforementioned theoretical models suggesting that phasic DA bursts promote positive prediction error learning via D_1-mediated modulation of synaptic potentiation in the striatonigral "Go" pathway, together

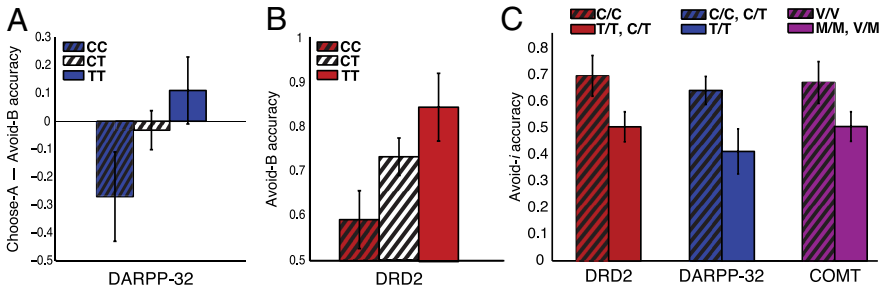

Figure 3. Gene effects on behavior. ***A***, Gene-dose effect of DARPP-32 T alleles on choose-A (the most positive stimulus) relative to avoid-B accuracy. ***B***, Gene-dose effect of DRD2 T alleles on avoiding the most negative stimulus (avoid-B). ***C***, Effect of inaccurate instructions on test choices is modulated by striatal genotype. When reinforcement statistics conflicted with prior instructions, efficacy of both striatal (DARPP-32 and DRD2) and prefrontal (COMT) genotypes modulated proportion of choices in accordance with prior instructions. Accuracy is plotted in terms of avoid-*i* when the instructed stimulus is suboptimal in the test phase.

with concurrent synaptic depression in the striatopallidal "NoGo" pathway (Reynolds and Wickens, 2002; Frank, 2005; Shen et al., 2008; Hikida et al., 2010). Note that D_1 and D_2 receptor stimulation leads to phosphorylation and dephosphorylation of DARPP-32, respectively (Svenningsson et al., 2004), and because the SNP we assess affects overall DARPP-32 expression (Meyer-Lindenberg et al., 2007), it should presumably affect both phosphorylation sites (Thr-34 and Thr-75), and thus modulate dopamine effects on synaptic plasticity in both pathways in opposite directions (Bateup et al., 2008). Together these effects should emphasize rewards relative to punishments, as observed here and in previous studies (Frank et al., 2007, 2009). We therefore reasoned that, if the override model is correct, enhanced DARPP-32 function should support learning the true reinforcement statistics via positive prediction errors. If the bias model is correct, individual differences in DARPP-32 function would modulate not only positive learning, but the extent to which such learning is amplified when outcomes are consistent with prior beliefs.

Finally, we examined the C957T polymorphism within the DRD2 gene, where T alleles are associated with decreased affinity of D_2 receptors in striatum (Hirvonen et al., 2005, 2009), where these receptors are primarily expressed (Camps et al., 1989). Models and experimental data suggest that D_2 receptors in the striatopallidal pathway are necessary for detecting when DA levels drop below baseline (as is the case during negative reward prediction errors). Thus, genetic modulation of D_2 receptor affinity with increasing T alleles should facilitate synaptic plasticity and therefore avoidance learning (Frank, 2005; Shen et al., 2008; Hikida et al., 2010). In humans, numerous studies show that low striatal DA levels, pharmacological manipulation of striatal D_2 receptors, and DRD2 genetic variation modulate learning from negative reward prediction errors (Frank and O'Reilly, 2006; Frank et al., 2007, 2009; Klein et al., 2007; Bódi et al., 2009; Cools et al., 2009; Frank and Hutchison, 2009; Jocham et al., 2009; Palminteri et al., 2009). In some contexts, DRD2 genotype also modulates functional connectivity between striatum and prefrontal cortex (Cohen et al., 2007; Jocham et al., 2009). We reasoned that if the override model is correct, D_2 receptor function would also predict learning from negative prediction errors when instructions were incorrect. Conversely and counterintuitively, the bias model suggests that striatal D_2 function should predict learning from negative prediction errors, but also the degree to which these negative outcomes are neglected when outcomes are inconsistent with prior beliefs—thereby leading to a confirmation bias.

We first analyzed the data for effects of genes on uninstructed learning. Consistent with prior reports, DARRP-32 and DRD2 T alleles (0, 1, or 2), indicative of enhanced striatal function, showed significant test phase associations with uninstructed approach and avoidance learning, respectively. DARRP-32 T alleles were associated with a relative performance advantage on choose-A compared to avoid-B trials ($F_{(1,45)} = 4.09$, $p = 0.05$), while DRD2 T alleles were associated with improved avoid-B accuracy ($F_{(1,54)} = 7.58$, $p = 0.008$) (Fig. 3*A*,*B*). In line with our previous findings, we found no evidence of a role for COMT in these putative striatal DA measures (p values >0.18). We did not find evidence for these effects to persist into the second task iteration (p values >0.11), possibly due to practice effects. (Experience in the first task iteration may have allowed subjects to learn the structure of the task, potentially changing the strategies they used and overwhelming individual differences in the basic reward-learning functions that are sensitive to the genes we investigate.)

Consistent with the above predictions, COMT genotype modulated the extent to which inaccurately instructed individuals persisted in selecting the instructed stimulus during the training phase of the first task iteration (effect of COMT group $F_{(1,36)} = 8.77$, $p = 0.005$ and marginal block by genotype interaction $F_{(2,36)} = 3.05$, $p = 0.07$) (see Fig. 5*A*). *Post hoc* tests showed that while both groups tended to abandon inaccurate instructions from first to last block (p values <0.05), Val/Val homozygotes (with putatively lower PFC DA) were more adept than Met carriers at doing so by the second block of training ($p = 0.006$ corrected for multiple comparisons). There was no difference between gene groups in instructional control in the first or last blocks (p values >0.15 corrected). These results indicate that, despite differences in the persistence of instruction following (see below for a model account of this effect based on Bayesian hypothesis testing), the COMT effect was not due to differences in initially recalling or trusting the instructions, and that the proportion of instructed choices did not differ significantly by the end of training. There were no DARPP-32 or DRD2 effects on these training measures (p values >0.14), nor of any SNP on performance in subjects given accurate instructions (p values >0.15).

The critical measures are from the test phase, which permits us to infer the accumulated learned value associated with each of the stimuli by having participants choose among all previously untested combinations of these stimuli (without feedback). As mentioned, our previous computational work (Doll et al., 2009) explored two plausible neural system mechanisms for instruction effects on test performance (Figs. 1*B*, 4*A*). The override model postulates that striatal and prefrontal systems compete for control of behavior, such that enhanced striatal function should lead to improved sensitivity to the true reinforcement probabilities. In contrast, the bias model posits that prefrontal instruction representations bias activity and learning in the striatum, increasing learning about outcomes consistent with instructed beliefs and discounting outcomes inconsistent with such beliefs.

In accordance with the bias model, we found that polymorphisms indicative of enhanced striatal DA function were associated with impaired accuracy on inaccurately instructed test trials (Fig. 3*C*). Strikingly, DRD2 T allele carriers, who demonstrated

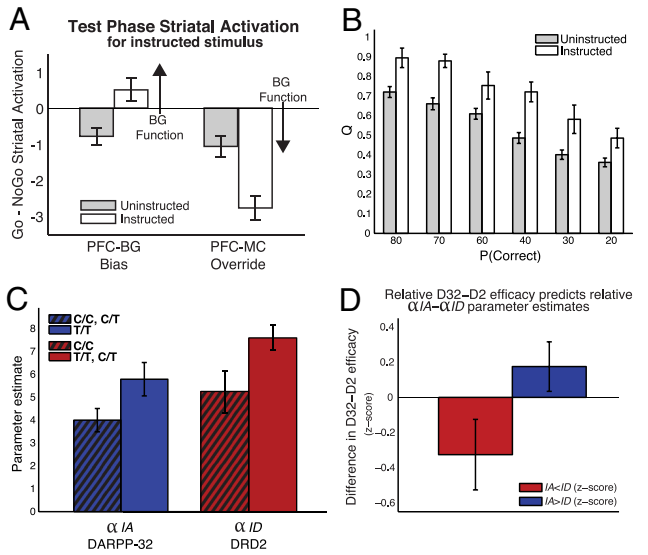

Figure 4. Model predictions and empirical results supporting bias model. ***A***, Neural network model predictions for BG efficacy on inaccurately instructed stimuli (Doll et al., 2009). "Go-NoGo" striatal activation indexes the extent to which the striatum assigns a relatively more positive than negative value to the action associated with selecting the instructed stimulus. "Go-NoGo" evaluations are initially neutral and are learned through training experience. The bias model predicts that increased BG function should produce greater distortion of striatal action values while the override model predicts greater striatal learning of the objective contingencies. ***B***, Algorithmic RL model showing final learned *Q* values after training that best explain test phase choices for instructed and uninstructed stimuli. Each unfilled bar reflects the learned *Q* value for the subset of participants who had been instructed on the given stimulus. Uninstructed (filled) bars reflect learned values in subjects not instructed on that stimulus. (Any individual subject contributes to one unfilled bar and five filled bars.) *Q* values are uniformly increased by instruction across stimulus probabilities. ***C***, Supporting the bias model, increased striatal efficacy was associated with increased confirmation bias parameter estimates leading to these skewed *Q* values. DARPP-32 genotype modulated the amplification of instruction-consistent positive outcomes (α_{IA}), whereas DRD2 genotype modulated the discounting of inconsistent negative outcomes (α_{ID}). ***D***, Striatal genotypes differentially predict α_{IA} and α_{ID}. Subjects fit by relatively greater than average differences in α_{IA} than α_{ID} (in *z*-scores), had relatively greater than average differences in DARPP-32 than DRD2 efficacy as indexed by *z*-score difference in number of T alleles (DARPP-32 − DRD2).

better uninstructed learning from negative outcomes, were actually less likely than C/C homozygotes to avoid inaccurately instructed (negative) stimuli ($F_{(1,70)} = 8.92$, $p = 0.0039$), suggesting that learning in these subjects was biased by top-down prior expectations. Similarly, DARPP-32 T/T homozygotes, who showed evidence of relatively better uninstructed reward learning than C carriers, also showed an impairment in avoiding inaccurately instructed stimuli ($F_{(1,72)} = 16.9$, $p = 0.0001$) (potentially due to amplification of positive learning and/or reduction of negative learning when selecting the instructed stimulus). Finally, COMT Met carriers were impaired relative to Val/Val homozygotes in these trials ($F_{(1,72)} = 7.31$, $p = 0.008$), consistent with the hypothesis that COMT modulates the initial maintenance and top-down modulation of striatal learning. These results and the computational results below hold over both task iterations and in the first iteration alone.

Neurocomputational results

To more directly assess the process-level mechanism suggested by the bias model to account for these effects, we fit subject data with an RL algorithm that captures some of the key computational attributes of our network model (Doll et al., 2009). This approach holds an advantage over the behavioral analysis above in that it takes into account the trial-to-trial experiences of each subject, and extracts parameters that are not directly measurable in the data but provide a plausible explanation of subjects' choices.

Replicating our prior work (Doll et al., 2009), model comparison revealed that subject test choices were best captured by allowing for the confirmation bias mechanism (instructed learning model: Eqs. 1, 2, and 3; see Table 2). We assessed the final *Q* values (i.e., at the end of training, given the reinforcement history), which best explain participants' test phase choices among all stimulus combinations (Frank et al., 2007; Doll et al., 2009). These *Q* values are estimates of the final learned values of each stimulus.

We first assessed whether the final *Q* values reflected the relative reinforcement probabilities of the stimuli, and whether those of the instructed stimulus were inflated. Indeed, there was a main effect of both stimulus-reinforcement probability and instruction on final *Q* values (p values <0.0001). Furthermore, there was no interaction ($p = 0.6$) (Fig. 4*B*), indicating that the slopes of the regression lines for instructed and uninstructed *Q* values across stimulus probabilities did not differ. Thus, although the learned *Q* value for the instructed stimulus was inflated relative to its objective reinforcement history, it was still graded (i.e., it was still sensitive to value). This result further supports the bias model, where instructions produce a distortion of learned values, but those values remain sensitive to relative differences in objective value. If the instructions were implemented as in the override model, one might expect that the probability of choosing the instructed stimulus would be insensitive to the value of the alternative stimulus with which it is paired, and thus instructed *Q* values would be equivalent for all reinforcement probabilities. Further analyses confirmed that decisions for instructed trials were best fit by assuming the same softmax choice function (which is sensitive to the value difference between stimuli) as for uninstructed trials, only with inflated values for the instructed stimuli. Worse fits were obtained when we modeled the instructed trials with a different choice function ("ϵ-greedy," in which choices are only sensitive to the sign of the difference in value between stimuli, and not to the relative differences in value; softmax: mean AIC = 176.08, AIC weight = 0.88, exceedance probability = 1; ϵ-greedy: mean AIC = 180.14, AIC weight = 0.12, exceedance probability = 0). Thus priors appear not to alter the choice process after learning, but rather to distort the initial learning of values.

Given these inflated values, we assessed the confirmation bias parameters α_{IA} and α_{ID} that were most likely to have generated them as a function of each participant's trial-by-trial sequence of choices and outcomes. We then used logistic regression to determine whether these parameters, which are distinguishable from one another (see Materials and Methods, Identifiability of confirmation bias parameters), were diagnostic of genotype. As predicted by the bias model, enhanced striatal DA efficacy was associated with greater influence of confirmation bias parameters on learning. Indeed, the DARPP-32 T/T genotype was associated with greater α_{IA} than C carriers (Wald $\chi^2 = 4.52$, $p = 0.03$), with no effect of α_{ID} ($p = 0.4$). Strikingly, the opposite pattern was observed for DRD2 T carriers, who had larger α_{ID} than C/C homozygotes ($\chi^2 = 5.04$, $p = 0.025$), with no α_{IA} effect ($p = 0.17$) (Fig. 4*C*, Table 1). Moreover, a follow-up test revealed that the relative difference between α_{IA} and α_{ID} is predicted by relative differences in the efficacy of DARPP-32 compared to DRD2 (difference in T alleles across SNPs) as indicated by an interaction between these measures ($F_{(1,70)} = 3.86$, $p = 0.05$) (Fig. 4*D*).

These results suggest that prior information supports confirmation bias learning by modulating the degree to which positive

Table 1. Parameter estimates

Gene	Alleles	α_{IA} (SE)	α_{ID} (SE)	α_{IA_train} (SE)	α_{ID_train} (SE)	ϕ (SE)
DARPP-32	T/T	5.79 (0.7)*	7.35 (0.7)	3.15 (0.5)	4.24 (0.8)	3.09 (1.3)
	C/C, C/T	3.99 (0.5)	6.52 (0.6)	3.15 (0.4)	4.13 (0.6)	1.30 (0.5)
DRD2	T/T, T/C	5.03 (0.5)	7.56 (0.5)*	3.32 (0.4)	4.86 (0.6)*	2.45 (0.8)
	C/C	3.67 (0.8)	5.25 (0.9)	2.96 (0.6)	2.84 (0.7)	1.00 (0.4)
COMT	Met/Met, Val/Met	4.43 (0.7)	7.02 (0.6)	3.73 (0.4)*	4.99 (0.6)*	2.74 (0.9)*
	Val/Val	4.78 (0.5)	6.50 (0.8)	2.19 (0.3)	2.83 (0.6)	0.66 (0.3)

*Difference in parameter estimate between allele groups ($p < 0.05$).

Table 2. Model comparisons

Model	Params	AIC_Tst	AIC_Tst_Wt	*P*(Exceed_Tst)	AIC_Trn	AIC_Trn_ Wt	*P*(Exceed_Trn)
Uninstructed	3	185.22	0.065	0.15	74.4	0.608	1
Instructed	5	177.59	0.915	0.83	77.71	0.136	0
Strong prior	4	182.88	0.02	0.02	76.13	0.256	0

Model comparisons of fits to all subject data. Params, Number of parameters. AIC_Tst, Akaike information criterion values for test phase fits (smaller values indicate better fit). AIC_Tst_Wt, AIC weights for test phase fits of candidate models. *P*(Exceed_Tst), Exceedance probability of test phase candidates. AIC_Trn, AIC_Trn_Wt, *P*(Exceed_Trn), Same measures for training phase fits.

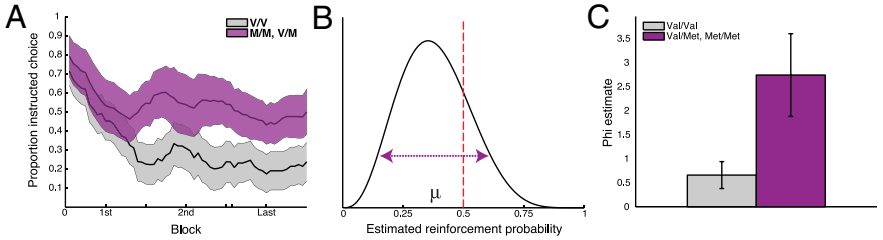

Figure 5. COMT hypothesis testing effects in training phase. ***A***, When instructions were inaccurate, COMT Met carriers showed greater persistence in instruction following than Val/Val homozygotes, who more rapidly abandoned inaccurate instructions. Some subjects completed additional training blocks to reach accuracy criteria before the test phase. Because the sample size decreases in these later blocks as subjects advance to the testing phase, only the first, second, and last blocks were assessed. Learning curve shading reflects SE. Data are smoothed over a 10-trial window. ***B***, Schematic of choice rule in Bayesian hypothesis testing model. For this example distribution, the mean μ (best guess) of the instructed *Q*-value is <0.5, but the belief distribution extends well above chance. Assuming a minimal temperature parameter (ζ) for illustrative purposes, subjects abandoning instructed stimulus selection at this point would be best described by a model with $\phi < 0.83$, given that the mean is below chance by 83% of a standard deviation. The red dashed vertical line indicates chance, and the purple dotted horizontal line indicates variability of distribution. ***C***, COMT Met carriers had greater ϕ parameter estimates than Val/Val homozygotes in the hypothesis testing model, indicating that they required more evidence on the inaccuracy of instructions before abandoning them.

and negative outcomes are consistent or inconsistent with this information, in the same neurogenetic pathways that modulate learning from positive and negative experiences. DARPP-32 genetic function supports relatively better learning from positive outcomes overall (here and in prior studies), and this same genetic factor appears to modulate the amplification of such learning when outcomes confirm prior biases. Conversely, and perhaps more counterintuitively, DRD2 predicts both enhanced learning from negative outcomes and the discounting of such outcomes when they betray prior beliefs.

COMT did not modulate the confirmation bias parameters that explain test phase performance (p values >0.5). We may, however, expect a gene–gene interaction, where DARPP-32 and DRD2 effects are dependent on COMT genotype, though our current sample lacks statistical power to assess this effect.

Our other studies suggest that COMT influences strategic exploration and acquisition during the early phases of learning (during which working memory and cognitive control demands are high) (Frank et al., 2007, 2009). Indeed, the behavioral effects of COMT in the training phase support this view. To further investigate this effect computationally, we fit models to account for the dynamics of the training phase, in which subjects initially chose according to instructions before eventually abandoning them. In contrast to the test phase fits, the training phase results were best explained by the basic uninstructed model [consistent with our prior report (Doll et al., 2009)] (Table 2). This result suggests that any advantage for models representing instructions only exists early in learning before subjects have "learned away" from the instructed response, and does not improve the fit beyond the penalization incurred for adding extra parameters. Nevertheless, prior work suggests that the hypothesis testing model described below provides a reasonable fit to a subset of participants who reliably chose the instructed stimulus and then abandoned it at some point (Doll et al., 2009). We tested for genetic effects across model parameters based on the a priori bias model prediction that PFC instruction representations modulate striatal reinforcement learning. Thus, the greater training phase adherence to instruction observed behaviorally in COMT Met carriers (Fig. 5*A*) should be detectable in more fine-grained, mechanistically precise computational model parameter estimates. (We only tested for these effects in the first iteration of training, due to the additional complications potentially introduced by learning of instruction quality over task repetitions.)

The best of the models incorporating instructions into the training phase was the Bayesian "hypothesis testing" model (see Materials and Methods and Fig. 5*B*). This model produced a better fit of inaccurately instructed subject data than the instructed learning model (Table 3). Notably, COMT Met allele carriers had larger estimated ϕ parameters than Val/Val homozygotes (Wald $\chi^2 = 3.73$, $p = 0.05$) (Fig. 5*C*). There were no effects of either striatal gene on ϕ (p values >0.2). This result provides support to the notion that during training, subjects follow instructions until sufficient evidence has accumulated that they are inaccurate, and that Met allele carriers required more evidence than Val/Val homozygotes before they were willing to abandon the instructions. This finding is consistent with the behavioral result that increased PFC efficacy increases the persistence in instruction following during learning (Fig. 5*A*), and contrasts with the no-

Table 3. Training phase model comparisons for iteration 1 inaccurately instructed subjects

Model	Params	AIC_Trn	AIC_Trn_Wt	*P*(Exceed_Trn)
Uninstructed	3	71.72	0.6008	1
Instructed	5	74.66	0.1386	0
Hypothesis testing	3	74.31	0.1648	0
Strong prior	4	75.4	0.0957	0

Abbreviations are as in Table 2.

tion that subjects with better working memory should be better able to falsify the hypothesis that the instructions are correct. Other studies have shown an advantage for Val alleles in the flexible gating of alternative hypotheses (Krugel et al., 2009; de Frias et al., 2010), consistent with predictions of computational models of COMT and PFC (Durstewitz and Seamans, 2008).

Discussion

Psychological experiments provide many examples of confirmation biases. Individuals with divergent prior beliefs are likely to interpret the same evidence as support for different conclusions (Lord et al., 1979; Nickerson, 1998). Evidence challenging preexisting views or beliefs is discounted, while evidence supporting these views is overemphasized. These effects apply to many domains of belief, including science, politics, and astrology, and may explain the persistent tendency to hold seemingly irrational beliefs in the face of contradictory evidence (Nickerson, 1998).

We found that when the evidence consists of reinforcement probabilities, individual differences in confirmation bias are predicted by the same dopaminergic genes involved in the learning process. Although there is increasing evidence that dopaminergic genes are predictive of individual differences in reinforcement learning (Cohen et al., 2007; Frank et al., 2007, 2009; Klein et al., 2007; Jocham et al., 2009; Krugel et al., 2009), the role such genes play in integrating explicit prior beliefs (here in the form of computerized instructions) with experience has not been investigated. We tested between two candidate neural models of instructional control based on prior behavioral and theoretical work (Biele et al., 2009; Doll et al., 2009). Results support the view that representations of prior information, maintained by PFC, exert their influence by modifying the striatal learning process in accord with a confirmation bias. Thus, individuals with enhanced statistical reinforcement integration processes are paradoxically hindered by their own strengths when reinforcing or punishing actions for which they have strong, but invalid, prior beliefs.

Parameter estimation revealed that the tendencies to overweight positive and to discount negative evidence are separately modulated by DARPP-32 and DRD2 genotypes, respectively. Our network model (Doll et al., 2009), building on previous models of uninstructed BG reinforcement learning (Frank, 2005), suggests that top-down biases operate to modify learning in the same neural pathways as uninstructed learning. The DARPP-32 findings we present fit similarly with the previously reported role for this SNP in reward learning (Frank et al., 2007, 2009), with T/T homozygotes having a performance advantage over C carriers in uninstructed tasks. We submit that this polymorphism should exert its effects via allele-dependent differences in the abundance of intracellular DARPP-32 mRNA (Meyer-Lindenberg et al., 2007). Available evidence suggests that DARPP-32 accumulates in the nucleus following reward, and facilitates reward learning via modulation of synaptic plasticity in the D_1-mediated striatonigral pathway (Calabresi et al., 2000; Svenningsson et al., 2004; Shen et al., 2008; Stipanovich et al., 2008), which is activated by phasic dopamine bursts that accompany positive prediction errors (Schultz, 1998). Our instructed neural models suggest that input from PFC to striatonigral/D_1 cells should bolster LTP along this pathway on trials with instruction-consistent feedback (Doll et al., 2009).

Similarly, the DRD2 SNP assessed here has been shown to affect learning from punishment in both forced choice and reaction-time tasks (Frank et al., 2007, 2009), and its effects are separable from those of other DRD2 SNPs (Frank and Hutchison, 2009). Our models posit that these effects are due to increased sensitivity of striatopallidal cells to DA pauses that accompany negative prediction errors (Schultz, 1998), disinhibiting D_2 cells and allowing for avoidance learning to occur (Frank, 2005). *In vitro* synaptic plasticity studies support this claim, demonstrating LTP facilitation in striatopallidal cells when D_2 receptors are not stimulated (Shen et al., 2008). Subjects carrying T alleles, exhibiting an advantage in avoidance learning from negative outcomes as in prior studies, actually showed an increased tendency to select inaccurately instructed stimuli. This finding is surprising given that inaccurate instruction produced the same negative outcomes DRD2 carriers are normally so adroit at integrating. Model parameter estimation suggests that these subjects were more likely to dismiss negative outcomes when they are inconsistent with prior beliefs. One possibility is that the top-down prefrontal input prevents the D_2 cells from being disinhibited by dopamine dips during negative prediction errors (Doll et al., 2009). Greater DRD2 function would then be associated with greater long-term depression, rather than potentiation, in the striatopallidal pathway (Shen et al., 2008). As a result, prior biases would promote unlearning along this pathway, allowing the positive evidence to dominate in other trials.

Given that the bias and override models both predict a role for PFC in maintaining instructions in working memory and facilitating instructional control, the COMT findings do not discriminate between the models as do those of striatal genotypes. Nevertheless, the findings demonstrate a rare advantage for the Val/Val genotype, which was associated with speeded ability to learn when the instructions were false during the learning phase. Met allele carriers showed greater susceptibility to general confirmation bias during this initial phase, requiring greater confidence that reinforcement probabilities of the instructed stimulus was below chance before abandoning it (Fig. 5*A*, *C*).

Despite our strong a priori hypotheses on the candidate neurocomputational substrates of instructional control, the correlational nature of these genetic data cannot rule out the possibility that the SNPs we assessed are unrelated to our interpretation of the behavioral effects. It is in principle possible that the SNPs discussed are associated with some unconsidered model of instructional control. For example, enhanced BG and PFC efficacy might support "obedience" to the experimenter (despite the fact that the instructions were given by the computer). We view such a hypothesis as unlikely for several reasons. First, previous work has shown effects of these genes on uninstructed learning tasks, which are unlikely to exert substantial demands of obedience. Second, final *Q* values show a relatively uniform distortion of value across stimulus probabilities (Fig. 4*B*). If subjects were simply obeying the instructions, best-fit *Q* values for instructed stimuli should be maximal, regardless of their probability of being correct. Third, the override model specifies a mechanism that might be termed "obedience," where behavior obeys the instructions as represented in PFC, despite evidence accumulated by the BG that the instructions are inaccurate. The pattern of genetic

results was the opposite of what would be expected if this model were correct.

Though our results suggest that better prefrontal and striatal function leads to a poorer assessment of "objective" reality, we submit that the ability to modulate evaluation of actions by explicit signals such as verbal instruction is typically adaptive, allowing the instruction follower to reap the benefits of others' experience. Further, the same frontal mechanisms for representing prior hypotheses may support hypothesis testing (Badre et al., 2010) and Bayesian statistical integration processes in which prior beliefs are fundamentally incorporated into the interpretation of evidence in the support of a hypothesis. These mechanisms might permit one to search among candidate causal models to determine that which best describes the environment (Waldmann and Martignon, 1998). However, it is more difficult to cast the striatal learning mechanisms explaining test performance as part of an optimal computation. First, Bayesian learning models do not fit the behavioral data as well as the reinforcement prediction error models used here [see Table 2 and Doll et al. (2009)]. Second, by the end of training, the majority of participants had effectively learned to stop choosing the instructed stimulus, implying that they abandoned the prior hypothesis. Nevertheless, these individuals responded in the test phase as if the value of the instructed stimulus had been inflated. Together with the genetic results, these findings suggest that the striatal learning process is modulated by prior expectations, and that the resulting associative weights cannot be easily "undone" after the prior is rejected. These findings may have implications for disorders of frontostriatal circuitry, which are often associated with irrational compulsions and maladaptive behaviors that are difficult to unlearn.

Many contingencies are difficult to learn by experience because they play out over extended temporal windows (e.g., the benefits of saving for retirement), are hard to observe (e.g., levels of harmful pesticides in food), or can provide rewards in the short term but punishments in the long term (e.g., the "contingency trap" laid by drugs of abuse). In such cases, altering the effects of feedback to confirm the validity of explicit signals is desirable. Future research into the neurocomputational mechanisms supporting such distortion of learning (and individual differences therein) should yield valuable insight into why explicit signals influence behavior in some cases, and do not in others.

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19:716–723.

Badre D, Kayser AS, D'Esposito M (2010) Frontal cortex and the discovery of abstract action rules. Neuron 66:315–326.

Bagary M, Fluck E, File SE, Joyce E, Lockwood G, Grasby P (2000) Is benzodiazepine-induced amnesia due to deactivation of the left prefrontal cortex? Psychopharmacology (Berl) 150:292–299.

Bateup HS, Svenningsson P, Kuroiwa M, Gong S, Nishi A, Heintz N, Greengard P (2008) Cell type-specific regulation of darpp-32 phosphorylation by psychostimulant and antipsychotic drugs. Nat Neurosci 11:932–939.

Bódi N, Kéri S, Nagy H, Moustafa A, Myers CE, Daw N, Dibó G, Takáts A, Bereczki D, Gluck MA (2009) Reward-learning and the novelty-seeking personality: a between- and within-subjects study of the effects of dopamine agonists on young Parkinson's patients. Brain 132:2385–2395.

Biele G, Rieskamp J, Gonzalez R (2009) Computational models for the combination of advice and individual learning. Cogn Sci 33:206–242.

Burnham KP, Anderson DR (2002) Model selection and multimodal inference. New York: Springer.

Calabresi P, Gubellini P, Centonze D, Picconi B, Bernardi G, Chergui K, Svenningsson P, Fienberg AA, Greengard P (2000) Dopamine and cAMP-regulated phosphoprotein 32 kDa controls both striatal long-term depression and long-term potentiation, opposing forms of synaptic plasticity. J Neurosci 20:8443–8451.

Camps M, Cortés R, Gueye B, Probst A, Palacios JM (1989) Dopamine receptors in the human brain: autoradiographic distribution of D_2 sites. Neuroscience 28:275–290.

Cavanagh JF, Frank MJ, Klein TJ, Allen JJB (2010a) Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. Neuroimage 49:3198–3209.

Cavanagh JF, Frank MJ, Allen JJB (2010b) Social stress reactivity alters reward and punishment learning. Soc Cogn Affect Neurosci. Advance online publication. doi:10.1093/scan/nsq041.

Centonze D, Picconi B, Gubellini P, Bernardi G, Calabresi P (2001) Dopaminergic control of synaptic plasticity in the dorsal striatum. Eur J Neurosci 13:1071–1077.

Chase HW, Frank MJ, Michael A, Bullmore ET, Sahakian BJ, Robbins TW (2010) Approach and avoidance learning in patients with major depression and healthy controls: relation to anhedonia. Psychol Med 40:433–440.

Cohen MX, Krohn-Grimberghe A, Elger CE, Weber B (2007) Dopamine gene predicts the brain's response to dopaminergic drug. Eur J Neurosci 26:3652–3660.

Cools R, Frank MJ, Gibbs SE, Miyakawa A, Jagust W, D'Esposito M (2009) Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. J Neurosci 29:1538–1543.

Dearden R, Friedman N, Russell S (1998) Bayesian q-learning. In: Proceedings of the 15th National Conference on Artificial Intelligence (AAAI) (Mostow J, Rich C, eds), pp 761–768. Menlo Park, CA: AAAI.

de Frias CM, Marklund P, Eriksson E, Larsson A, Oman L, Annerbrink K, Bäckman L, Nilsson LG, Nyberg L (2010) Influence of comt gene polymorphism on fmri-assessed sustained and transient activity during a working memory task. J Cogn Neurosci 22:1614–1622.

Doll BB, Frank MJ (2009) The basal ganglia in reward and decision making: computational models and empirical studies. In: Handbook of reward and decision making (Dreher J, Tremblay L, eds), pp 399–425. Oxford: Academic.

Doll BB, Jacobs WJ, Sanfey AG, Frank MJ (2009) Instructional control of reinforcement learning: a behavioral and neurocomputational investigation. Brain Res 1299:74–94.

Draganski B, Kherif F, Klöppel S, Cook PA, Alexander DC, Parker GJM, Deichmann R, Ashburner J, Frackowiak RSJ (2008) Evidence for segregated and integrative connectivity patterns in the human basal ganglia. J Neurosci 28:7143–7152.

Durstewitz D, Seamans JK (2008) The dual-state theory of prefrontal cortex dopamine function with relevance to catechol-o-methyltransferase genotypes and schizophrenia. Biol Psychiatry 64:739–749.

Durstewitz D, Vittoz NM, Floresco SB, Seamans JK (2010) Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. Neuron 66:438–448.

Egan MF, Goldberg TE, Kolachana BS, Callicott JH, Mazzanti CM, Straub RE, Goldman D, Weinberger DR (2001) Effect of COMT $Val^{108/158}$ Met genotype on frontal lobe function and risk for schizophrenia. Proc Natl Acad Sci U S A 98:6917–6922.

Engelmann JB, Capra CM, Noussair C, Berns GS (2009) Expert financial advice neurobiologically "offloads" financial decision-making under risk. PLoS One 4:e4957.

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.

Frank MJ (2006) Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. Neural Netw 19:1120–1136.

Frank MJ, Hutchison K (2009) Genetic contributions to avoidance-based decisions: striatal d2 receptor polymorphisms. Neuroscience 164:131–140.

Frank MJ, O'Reilly RC (2006) A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol. Behav Neurosci 120:497–517.

Frank MJ, Seeberger LC, O'Reilly RC (2004) By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306:1940–1943.

Frank MJ, O'Reilly RC, Curran T (2006) When memory fails, intuition reigns: midazolam enhances implicit inference in humans. Psychol Sci 17:700–707.

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 U S A 104:16311–16316.
Frank MJ, Doll BB, Oas-Terpstra J, Moreno F (2009) Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. Nat Neurosci 12:1062–1068.
Gogos JA, Morgan M, Luine V, Santha M, Ogawa S, Pfaff D, Karayiorgou M (1998) Catechol-O-methyltransferase-deficient mice exhibit sexually dimorphic changes in catecholamine levels and behavior. Proc Natl Acad Sci U S A 95:9991–9996.
Haber SN (2003) The primate basal ganglia: parallel and integrative networks. J Chem Neuroanat 26:317–330.
Hayes S, ed (1989) Rule-governed behavior: cognition, contingencies, and instructional control. New York: Plenum.
Hikida T, Kimura K, Wada N, Funabiki K, Nakanishi S (2010) Distinct roles of synaptic transmission in direct and indirect striatal pathways to reward and aversive behavior. Neuron 66:896–907.
Hirvonen M, Laakso A, Någren K, Rinne JO, Pohjalainen T, Hietala J (2005) C957T polymorphism of the dopamine D2 receptor (DRD2) gene affects striatal DRD2 availability *in vivo* (corrigendum). Mol Psychiatry 10:889.
Hirvonen MM, Laakso A, Någren K, Rinne JO, Pohjalainen T, Hietala J (2009) C957T polymorphism of dopamine D2 receptor gene affects striatal DRD2 in vivo availability by changing the receptor affinity. Synapse 63:907–912.
Huotari M, Gogos JA, Karayiorgou M, Koponen O, Forsberg M, Raasmaja A, Hyttinen J, Männistö PT (2002) Brain catecholamine metabolism in catechol-o-methyltransferase (comt)-deficient mice. Eur J Neurosci 15:246–256.
Jocham G, Klein TA, Neumann J, von Cramon DY, Reuter M, Ullsperger M (2009) Dopamine drd2 polymorphism alters reversal learning and associated neural activity. J Neurosci 29:3695–3704.
Klein TA, Neumann J, Reuter M, Hennig J, von Cramon DY, Ullsperger M (2007) Genetically determined differences in learning from errors. Science 318:1642–1645.
Krugel LK, Biele G, Mohr PNC, Li SC, Heekeren HR (2009) Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. Proc Natl Acad Sci U S A 106:17951–17956.
Li J, Delgado MR, Phelps EA (2011) How instructed knowledge modulates the neural systems of reward learning. Proc Natl Acad Sci U S A 108:55–60.
Lord C, Ross L, Lepper M (1979) Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence. J Pers Soc Psychol 37:2098–2109.
Matsumoto M, Weickert CS, Akil M, Lipska BK, Hyde TM, Herman MM, Kleinman JE, Weinberger DR (2003) Catechol O-methyltransferase mRNA expression in human and rat brain: evidence for a role in cortical neuronal function. Neuroscience 116:127–137.
Meyer-Lindenberg A, Straub RE, Lipska BK, Verchinski BA, Goldberg T, Callicott JH, Egan MF, Huffaker SS, Mattay VS, Kolachana B, Kleinman JE, Weinberger DR (2007) Genetic evidence implicating darpp-32 in human frontostriatal structure, function, and cognition. J Clin Invest 117:672–682.
Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24:167–202.
Nickerson RS (1998) Confirmation bias: a ubiquitous phenomenon in many guises. Rev Gen Psychol 2:175–220.
Ouimet CC, Miller PE, Hemmings HC Jr, Walaas SI, Greengard P (1984) DARPP-32, a dopamine- and adenosine 3′:5′-monophosphate-regulated phosphoprotein enriched in dopamine-innervated brain regions. III. Immunocytochemical localization. J Neurosci 4:111–124.
Palminteri S, Lebreton M, Worbe Y, Grabli D, Hartmann A, Pessiglione M (2009) Pharmacological modulation of subliminal learning in Parkinson's and Tourette's syndromes. Proc Natl Acad Sci U S A 106:19179–19184.
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.
Plassmann H, O'Doherty J, Shiv B, Rangel A (2008) Marketing actions can modulate neural representations of experienced pleasantness. Proc Natl Acad Sci U S A 105:1050–1054.
Reinsel RA, Veselis RA, Dnistrian AM, Feshchenko VA, Beattie BJ, Duff MR (2000) Midazolam decreases cerebral blood flow in the left prefrontal cortex in a dose-dependent fashion. Int J Neuropsychopharmacol 3:117–127.
Reynolds JN, Wickens JR (2002) Dopamine-dependent plasticity of corticostriatal synapses. Neural Netw 15:507–521.
Reynolds JNJ, Hyland BI, Wickens JR (2001) A cellular mechanism of reward-related learning. Nature 413:67–70.
Roesch MR, Calu DJ, Schoenbaum G (2007) Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nat Neurosci 10:1615–1624.
Schönberg 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.
Schultz W (1998) Predictive reward signal of dopamine neurons. J Neurophysiol 80:1–27.
Sesack SR, Hawrylak VA, Matus C, Guido MA, Levey AI (1998) Dopamine axon varicosities in the prelimbic division of the rat prefrontal cortex exhibit sparse immunoreactivity for the dopamine transporter. J Neurosci 18:2697–2708.
Shen W, Flajolet M, Greengard P, Surmeier DJ (2008) Dichotomous dopaminergic control of striatal synaptic plasticity. Science 321:848–851.
Slifstein M, Kolachana B, Simpson EH, Tabares P, Cheng B, Duvall M, Frankle WG, Weinberger DR, Laruelle M, Abi-Dargham A (2008) COMT genotype predicts cortical-limbic D1 receptor availability measured with [11C]NNC112 and PET. Mol Psychiatry 13:821–827.
Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. Neuroimage 46:1004–1017.
Stipanovich A, Valjent E, Matamales M, Nishi A, Ahn JH, Maroteaux M, Bertran-Gonzalez J, Brami-Cherrier K, Enslen H, Corbillé AG, Filhol O, Nairn AC, Greengard P, Hervé D, Girault JA (2008) A phosphatase cascade by which rewarding stimuli control nucleosomal response. Nature 453:879–884.
Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Cambridge, MA: MIT Press.
Svenningsson P, Nishi A, Fisone G, Girault JA, Nairn AC, Greengard P (2004) Darpp-32: an integrator of neurotransmission. Annu Rev Pharmacol Toxicol 44:269–296.
Tsai HC, Zhang F, Adamantidis A, Stuber GD, Bonci A, de Lecea L, Deisseroth K (2009) Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. Science 324:1080–1084.
Tunbridge EM, Bannerman DM, Sharp T, Harrison PJ (2004) Catechol-*O*-methyltransferase inhibition improves set-shifting performance and elevates stimulated dopamine release in the rat prefrontal cortex. J Neurosci 24:5331–5335.
Valjent E, Pascoli V, Svenningsson P, Paul S, Enslen H, Corvol JC, Stipanovich A, Caboche J, Lombroso PJ, Nairn AC, Greengard P, Hervé D, Girault JA (2005) Regulation of a protein phosphatase cascade allows convergent dopamine and glutamate signals to activate erk in the striatum. Proc Natl Acad Sci U S A 102:491–496.
Wagenmakers EJ, Farrell S (2004) Aic model selection using akaike weights. Psychon Bull Rev 11:192–196.
Waldmann MR, Hagmayer Y (2001) Estimating causal strength: the role of structural knowledge and processing effort. Cognition 82:27–58.
Waldmann MR, Martignon L (1998) A Bayesian network model of causal learning. In: Proceedings of the Twentieth Annual Conference of the Cognitive Science Society (Gernsbacher MA, Derry SJ, eds), pp 1102–1107. Mahwah, NJ: Erlbaum.
Wallis JD, Miller EK (2003) From rule to response: neuronal processes in the premotor and prefrontal cortex. J Neurophysiol 90:1790–1806.
Zaghloul KA, Blanco JA, Weidemann CT, McGill K, Jaggi JL, Baltuch GH, Kahana MJ (2009) Human substantia nigra neurons encode unexpected financial rewards. Science 323:1496–1499.
Zweifel LS, Parker JG, Lobb CJ, Rainwater A, Wall VZ, Fadok JP, Darvas M, Kim MJ, Mizumori SJY, Paladini CA, Phillips PEM, Palmiter RD (2009) Disruption of nmdar-dependent burst firing by dopamine neurons provides selective assessment of phasic dopamine-dependent behavior. Proc Natl Acad Sci U S A 106:7281–7288.