Dopamine, Locus of Control, and the Exploration-Exploitation Tradeoff

Andrew S Kayser*[1,2,3], Jennifer M Mitchell[1,2], Dawn Weinstein[1,2] and Michael J Frank[4]

[1]Department of Neurology, University of California, San Francisco, CA, USA; [2]Ernest Gallo Clinic & Research Center, Emeryville, CA, USA; [3]Department of Neurology, VA Northern California Health Care System, Martinez, CA, USA; [4]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, USA

Whether to continue to exploit a source of reward, or to search for a new one of potentially greater value, is a fundamental and underconstrained decision. Recent computational studies of this exploration-exploitation tradeoff have found that variability in exploration across individuals is influenced by a functional polymorphism (Val158Met) in the catechol-O-methyltransferase (COMT) gene, whose protein product degrades synaptically released dopamine. However, these and other genotype–phenotype associations have rarely been causally tested. To directly test this association and to evaluate additional behavioral characteristics, including perceived locus of control (LOC), here we used the COMT inhibitor tolcapone in a randomized, double-blind, counterbalanced, within-subject study of 66 subjects genotyped for the Val158Met allele to assess the hypothesis that reducing COMT enzymatic activity interacts with genotype to increase uncertainty-driven exploration. In keeping with our initial hypothesis, tolcapone led to an increase in exploratory, but not exploitative, behavior in Met/Met rather than Val/Val subjects. Independent of genotype, those subjects with a more external LOC also showed increases in uncertainty-driven exploration on tolcapone relative to placebo. However, we did not replicate our previous finding that Met/Met subjects show greater exploration at baseline. Together these findings support a model in which exploration is hypothesized to have a dopaminergic basis. Moreover, in keeping with findings in other behavioral and cognitive domains, the response to an increase in presumptively frontal dopamine is dependent upon baseline dopamine tone.
Neuropsychopharmacology (2015) **40,** 454–462; doi:10.1038/npp.2014.193; published online 20 August 2014

INTRODUCTION

When decisions are uncertain, individuals can vary significantly in their strategies. In particular, the decision to continue to exploit a known source of reward, or to explore the environment for a potentially greater one, depends upon a balance of factors whose weighting is subjective. For example, when confronted with the choice to return to a restaurant where the food is reliably good, or to search for a new one where the food might be even better, some individuals may elect to return, whereas others may elect to explore. This underconstrained 'exploration-exploitation tradeoff' may in turn be modified by people's beliefs in their own ability to effectively find a new restaurant: those with an internal locus of control (LOC)—ie, the belief that they have control over events in their lives ('I'm good at finding the best chefs in town')—may make different choices than those with an external LOC—ie, the belief

that events are outside their control ('Finding a good place is a matter of luck').

Other biological factors may also be at play. Previously, we demonstrated that the amount of exploration that subjects undertake varies with a functional polymorphism (Val158Met) in the catechol-O-methyltransferase (*COMT*) gene, which degrades synaptically released dopamine. In particular, college undergraduates with two copies of the less active *COMT* enzyme (Met/Met) demonstrated greater exploration than those with either the Met/Val or Val/Val polymorphisms (Frank *et al*, 2009). Moreover, greater exploratory behavior has been correlated with correspondingly greater activity in the rostral prefrontal cortex (PFC), particularly when choice outcomes are uncertain and hence there is information to be gained (Badre *et al*, 2012; Cavanagh *et al*, 2012). Together these studies argue that exploratory behaviors are likely to have a neurochemical (dopaminergic) basis, and suggest that increasing dopamine tone in PFC may increase exploratory behavior. However, genetic studies are correlational by nature, and converging evidence from pharmacological manipulations can substantially enhance the confidence in the relevance of the underlying neuromodulators. Although such converging evidence has been established for the role of striatal dopamine measures in learning and exploitation (reviewed in Frank and Fossella, 2011), it is still lacking for the role of

*Correspondence: Dr AS Kayser, Department of Neurology, U.C. San Francisco, Ernest Gallo Clinic & Research Center, 5858 Horton Street, Suite 200, Emeryville, CA 94608, USA, Tel: +510 985 3100, Fax: +510 985 3101, E-mail: akayser@gallo.ucsf.edu

prefrontal dopamine in exploration. Such data could conceivably have clinical implications: pharmacological manipulations that augment exploration could potentially address motivational deficits such as anhedonia in schizophrenia, for example, which has been linked to reduced uncertainty-driven exploration (Strauss *et al*, 2011).

To causally test this association, one approach would be to augment prefrontal dopamine levels by taking advantage of the neuroanatomical distribution of *COMT*. In most brain regions, dopamine is inactivated by reuptake into nerve terminals through the dopamine transporter. However, because the density of dopamine transporter is low in PFC, *COMT* is thought to significantly influence cortical dopamine tone (Chen *et al*, 2004). In fact, mice with inactivated *COMT* genes show increased cortical but not subcortical (including striatal) dopamine levels (Gogos *et al*, 1998). Consequently, although increases in striatal dopamine may favor exploitation of a resource (or more specifically, shift the objective of exploitation to maximizing gains rather than minimizing losses (Collins and Frank, 2014)), selectively increasing dopamine tone in frontal cortex via *COMT* inhibition may enhance uncertainty-driven exploratory behaviors. We hypothesized that a single 200 mg dose of the brain-penetrant *COMT* inhibitor tolcapone, by preferentially increasing dopamine tone in frontal cortex, would increase exploration in a genotype-specific manner, in contrast with influences of striatal dopamine manipulations on reward learning and exploitation in the same task (Moustafa *et al*, 2008; Frank *et al*, 2009).

Of course, this increase in exploration may be contingent upon factors other than *COMT* genotype. For example, it has previously been hypothesized that LOC, which might impact decisions between exploration and exploitation, depends upon dopamine tone (De Brabander and Declerck, 2004; Declerck *et al*, 2006). As recently reviewed (De Brabander and Declerck, 2004; Declerck *et al*, 2006), an internal LOC appears to correlate positively with greater incentive motivation and enhanced executive functions such as working memory, both of which are thought to be influenced by dopamine tone. Furthermore, the decision to explore requires cognitive control to override a prepotent tendency to choose based on external reinforcement statistics alone. Those subjects with low PFC dopamine and an external LOC may therefore be less likely to forego available rewards by exploring. Thus, we also investigated the secondary hypothesis that LOC should modulate exploratory behaviors—specifically, that subjects with a more external LOC would be more likely to respond to *COMT* inhibition with greater exploration.

METHODS AND MATERIALS

Study Population

We enrolled 70 healthy subjects who were without a history of neurological and psychiatric illnesses. After three subjects were excluded who failed to follow task instructions or whose data were incomplete, 67 subjects who completed the task remained. Following exclusion of one outlier subject ($p < 0.05$ *via* Rosner's extreme studentized deviate test), we included 66 subjects in all analyses for which genotype information was not needed. After efforts to reevaluate

10 subjects whose genetic data were initially unavailable, ultimately only two participants did not have *COMT* genotyping because of technical failures, leaving 64 subjects for genotype-specific analyses. All subjects gave written informed consent in accordance with the Committee for the Protection of Human Subjects at the University of California, San Francisco and University of California, Berkeley and were paid for their participation. Subjects first underwent a history and physical exam, as well as blood testing for liver function, to ensure that there were no medical contraindications to tolcapone use or MRI scanning. They then completed a number of screening questionnaires including the Alcohol Use Disorders Identification Test (AUDIT), the State-Trait Anxiety Index (STAI) (Spielberger, 1983), Rotter's LOC scale (Rotter, 1966), and the Barratt Impulsiveness Scale (BIS) (Patton *et al*, 1995), as per our previous work (Mitchell *et al*, 2005). Of note, in this version of the LOC scale, higher scores corresponded to a more internal LOC. For a separate study, subjects also completed a delay discounting task, not further discussed here; data for a subset of these subjects and information on the discounting task have been published previously (Kayser *et al*, 2012).

Experimental Paradigm

Subjects received either placebo or a single 200 mg dose of tolcapone on their first visit and the alternative treatment on their second visit in randomized, double-blind, counterbalanced fashion. As in our previously published studies (Frank *et al*, 2009; Badre *et al*, 2012), on each trial participants observed a clock arm that completed a revolution over 5 s. They stopped the clock with a key press during that time in an attempt to win points. Rewards were delivered with a probability and magnitude that varied as a function of response time (RT); together, these factors defined the reward space for each condition. Importantly, although subjects knew that the underlying reward statistics for each condition would remain consistent across 50 trials, subjects were not cued to the nature of the reward space beforehand, requiring them to learn how reward probability and magnitude varied with duration from trial onset. Over the course of 50 trials, subjects explored each of four conditions, named in accordance with the change in expected value (probability * magnitude) with increasing RT: Increasing Expected Value (IEV), Decreasing Expected Value (DEV), Constant Expected Value (CEV), and CEV-Reversed (CEV-R). Figure 1 demonstrates reward frequency, reward magnitude, and expected value for each of the four conditions. CEV and CEV-R are distinguished by contrasting reward frequency and reward magnitude curves whose product gives rise to overlapping expected value curves.

A computational model was fit to each subject's behavioral data to estimate the magnitude of the exploration parameter ε. (Please note that the ε term used here to capture exploration is distinct from the ε found in ε-greedy reinforcement learning algorithms (Kalidindi and Bowman, 2007), in that the exploration considered here is strategically directed rather than just noisy—see Daw *et al*, 2006; Frank *et al*, 2009). Briefly, this model captures the degree to which uncertainty in the reward space drives individual subjects to explore (Figure 2). At the beginning of learning, subjects have little knowledge as to whether faster or slower
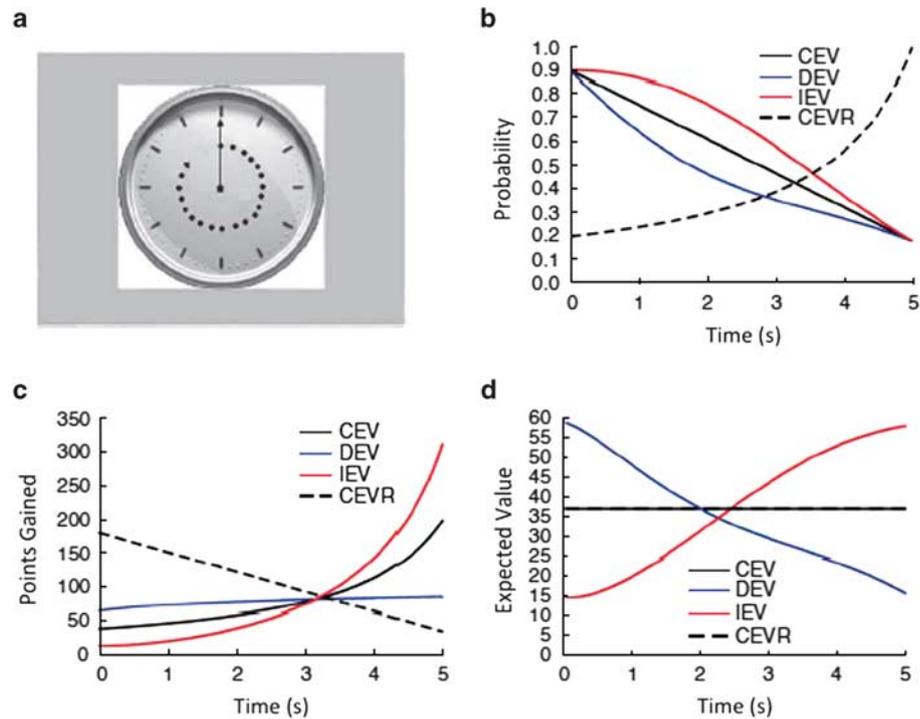
Figure I (a) Subjects viewed a clock face while deciding when to stop a hand that made one clockwise rotation over 5 s. (b) The probability of reward for each of the four conditions across the 5-s trial time. (c) The magnitude of reward for each of the four conditions across the 5-s trial time. (d) The expected value of the reward (probability × magnitude) for each of the four conditions. Note that CEV and CEV-R both have a constant expected value, but differ in the underlying probability and magnitude of reward with respect to time.
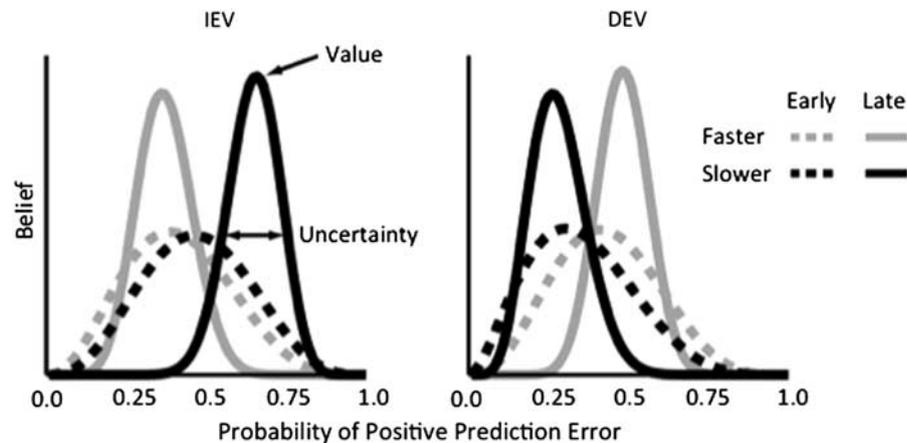
Figure 2 Intuition for the explore parameter. The strength of a subject's belief (*y* axis) about the probability of a better than expected outcome (*x* axis) depends upon both the stage of learning and the task condition (shown for IEV (left) and DEV (right)). In early stages (dashed lines) of the IEV condition (left panel), the subject has not yet learned whether a faster (in gray) or slower (in black) response is more likely to be rewarded, a state reflected in both a weaker belief strength (value on the *y* axis) and greater uncertainty (a broader belief distribution). Later in the task (solid lines), belief strength increases and the uncertainty about reward likelihood decreases—ie the subject learns that slower responses are more likely to yield reward (ie, a positive prediction error). The explore parameter indexes the degree to which subjects use the relative uncertainty between the faster and slower distributions to explore the reward space (ie, to reduce the variance in the more uncertain distribution). Similar qualitative changes in the probability of reward, with a different sign relative to fast and slow distributions, can be seen for the DEV condition (right panel).

responses will produce greater expected reward. In the absence of significant experience, the belief distributions about outcomes statistics are uncertain, reflecting the large variance in the set of possible values that can be expected for each action. After subjects experience more trials, the feedback about rewards allows them to reduce the uncertainty related to their prior beliefs, manifest as a reduction in the variance of the relevant belief distribution.

The explore parameter ε indexes the degree to which subjects are likely to guide exploration toward the more uncertain of these two (fast/slow) distributions to increase their understanding of the reward space. In other words, subjects with larger explore parameters will tend to use their lack of knowledge about a part of the reward space to explore it further, thereby reducing the associated uncertainty. In equation form,

$$\text{Exploration}(t) = \varepsilon * [\sigma_{\text{slow}}(t) - \sigma_{\text{fast}}(t)]$$

where the magnitude of Exploration in milliseconds at time t is equal to the difference between the standard deviations σ of the fast and slow distributions at that time weighted by the scale factor ε, derived from each subject's data. ε therefore represents, in units of milliseconds per unit standard deviation of the belief distributions, how strongly each subject weights the differences between the standard deviations to drive the contribution of exploratory behavior to reaction time. The relative magnitude of epsilon is meaningful because it is monotonic, but its absolute value is only relevant in the context of the other parameters translating from other aspects of the data (eg, go and no-go learning; see Supplementary Methods) to response time. This component of the model predicts that subjects will increase RTs if the outcome statistics are more uncertain for slow than fast responses, and vice versa—consistent with other conceptualizations of exploration (eg, (Daw et al, 2006)), though explaining action at the level of RT rather than categorical choice.

In contrast, the primary exploitation term in the model is defined as follows:

$$\text{Exploitation}(t) = \rho * [\mu_{\text{slow}}(t) - \mu_{\text{fast}}(t)]$$

where the magnitude of Exploitation at time t is equal to the difference between the mean rewards μ of the fast and slow distributions at that time, weighted by a scale factor ρ fit to each subject's data. This component of the model predicts that subjects will increase RTs if the average reward is greater for slow than fast responses, and vice versa. The exploitation factors therefore complement factors associated with exploration, in that reward depends on knowledge of the mean but exploration depends on the variance (see Supplementary Methods for further details). Finally, because the explore parameter reflects adjustment of RT (ie, changes to faster or slower responses) because of uncertainty, it is less sensitive to overall speed of responding.

Data Analysis

Three computational models of the exploration-exploitation tradeoff were evaluated. In the core model (Frank et al, 2009; Badre et al, 2012; Cavanagh et al, 2012), response times were modeled as a function of multiple factors, including uncertainty as well as sensitivity to gains and to losses (see Supplementary Methods). Additional models were tested that (i) explicitly modeled trial-to-trial changes in RT ('RT difference') rather than trying to capture raw RT on each trial (Frank et al, 2009; Badre et al, 2012), and (ii) allowed the explore parameter to attain negative values ('negative-permitting'). The former variant permitted us to directly investigate whether the large, trial-wise shifts in RT observed in this task are captured by uncertainty-based

exploration as subjects switch from one distribution (fast/slow) to the other (slow/fast), as opposed to the core model that attempts to model RT governed by other factors (including exploitation) and can operate across a longer time scale. The negative-permitting variant allowed us to explicitly determine whether accounting for perseverative choice and uncertainty aversion better explained the data, as negative explore parameters indicate a greater tendency to choose the more certain option. Modeling and statistical analyses were conducted using Matlab (The Mathworks, Natick, MA) and SPSS version 21.0 (IBM, Armonk, NY).

The primary analysis was performed using a two-way, repeated-measures ANOVA with factors of COMT genotype (three levels: Met/Met, Met/Val, and Val/Val) and the repeated pharmacological measure (two levels: tolcapone vs placebo). Subjects were included as a random effect. *Post hoc* tests consisted primarily of T tests. For the analyses of LOC effects, a median split was applied because we did not have strong *a priori* reasons to believe either that a predefined cutoff for the value of LOC would distinguish explorers from non-explorers, or that LOC values would necessarily scale linearly with exploration.

Genetic Analysis

Standard methods for DNA extraction and analysis were conducted on blood samples obtained from all subjects who gave informed consent ($N = 67$). As noted previously, technical failures prevented genetic data for two of these subjects from being obtained. Genotyping for the COMT polymorphism Val158Met was performed by the UCSF genetics core using a PCR-based TaqMan prep (Applied Biosystems, Foster City, CA). Sixteen subjects were homozygous for the Met allele, 27 subjects were Met158Val heterozygotes, and 22 subjects were homozygous for the Val allele (see also Table 1 and Supplementary Methods).

RESULTS

Sixty-seven subjects completed at least one drug session for the study and were eligible for inclusion (Table 1). Of these subjects, 64 could ultimately be included in genotype-specific analyses (see Materials and Methods). Before examining exploration explicitly, we first ensured that subjects learned the task well, as evidenced by the fact that their performance could be readily distinguished via their mean RTs over the second half of each of the different task conditions (Figure 3a). In a three-way ANOVA including *COMT* genotype and repeated measures of task condition and drug, with subjects as a random effect, there was a strongly significant effect of task condition on RT ($F(2.6, 165.1) = 7.8$, $p \ll 1.0 \times 10^{-3}$, Greenhouse-Geiser corrected). This effect of task condition was manifest as a strong linear increase in mean reaction time ($F(1, 63) = 16.5$, $p \ll 1.0 \times 10^{-3}$) from the DEV (RT = 1803 ± 109 ms (SEM)) through the CEV (RT = 2059 ± 131 ms), CEV-R (2291 ± 103 ms), and IEV (IEV: 2330 ± 110 ms) conditions. As expected, there was no effect of tolcapone on mean RT ($F(1, 63) = 2.47$, $p = 0.12$ (n.s.)), and there were no significant higher-order interactions between genotype, task condition, and drug. The effects of the different conditions on RT were well fit by the

computational model (Figure 3b). These results confirm our previous experience that subjects rapidly learned the task structure (Moustafa et al, 2008), and that tolcapone does not exert its effects by simply speeding motor responses (Kayser et al, 2012), or—unlike striatal DA manipulations in this task—by affecting the balance of learning to speed up/slow down as a function of positive and negative outcomes (Moustafa et al, 2008; Frank et al, 2009).

We next evaluated the effects of tolcapone specifically on the magnitude of the explore parameter, which previous

Table 1 Subject Demographics.

|	Mean/count	Range/percentage
Age	30.0 ± 8.4	19–50
Gender		
Male	34	50.8%
Female	33	49.2%
Ethnicity		
Caucasian	42	62.7%
Hispanic	9	13.4%
African American	8	11.9%
Asian	5	7.5%
Mixed	3	4.5%
COMT Genotype		
Met 158 Met	16	23.9%
Met 158 Val	27	40.3%
Val 158 Val	22	32.8%
Unknown	2	3.0%
AUDIT	9.3 ± 4.9	2–24
BIS	59.6 ± 9.3	35–82
LOC	13.0 ± 3.6	3–20
STAI—Trait	34.6 ± 8.1	20–57

Abbreviations: AUDIT, alcohol use disorders identification test; BIS, Barratt impulsiveness scale; COMT, catechol-O-methyltransferase; LOC, Rotter's locus of control; STAI-Trait, State-trait anxiety scale, trait subscale.
Table data include one outlier subject who was excluded from subsequent analyses (see Materials and Methods section).

work has demonstrated is sensitive to *COMT* genotype (Frank et al, 2009). A two-way repeated-measures ANOVA with factors of genotype and drug identified a significant drug x genotype interaction ($F_{(2,61)} = 3.68$, $p = 0.03$) and a trend main effect of drug ($F_{(1,61)} = 3.15$, $p = 0.08$). No main effect of genotype on exploration was seen ($F_{(2,61)} = 0.13$, $p = 0.88$ (n.s.)). *Post hoc* analysis indicated that the explore parameter for the Met/Met subjects was significantly greater on tolcapone than placebo ($T_{(15)} = 3.0$, $p = 0.009$; Figure 4a), and that Met/Met subjects showed a differential increase in exploration on tolcapone at trend-level significance relative to Met/Val subjects ($T_{(41)} = 1.98$, $p = 0.054$), and significantly so relative to Val/Val subjects ($T_{(35)} = 2.5$, $p = 0.016$). Finally, when the tolcapone condition was excluded from the ANOVA analyses, a trend-level effect of genotype on exploration in the placebo condition ($F_{(2,61)} = 3.1$, $p = 0.082$) favored greater exploration for Val/Val rather than Met/Met subjects (Figure 4a).

To evaluate the robustness of these genotype-related findings, we also assessed the effect of tolcapone for different instantiations of the computational model. Quantitatively similar results were found with the alternative model variants. The RT difference model confirmed a significant interaction between drug and genotype ($F_{(2,61)} = 5.0$, $p = 0.009$). As for the primary model, Met/Met subjects demonstrated both a significant increase in exploration on tolcapone *vs* placebo ($T_{(14)} = 3.4$, $p = 0.004$), and a significant differential increase relative to Val/Val subjects ($T_{(35)} = 3.16$, $p = 0.003$). In contrast, the negative-permitting model showed a non-significant interaction between drug and genotype ($F_{(2,61)} = 1.38$, $p = 0.26$ (n.s.)), arguing that perseverative responding did not explain these data. Neither the increase in exploration on tolcapone *vs* placebo for the Met/Met subjects ($T_{(15)} = 1.3$, $p = 0.21$ (n.s.)), nor the differential increase in exploration for the Met/Met subjects *vs* the Val/Val subjects ($T_{(35)} = 1.3$, $p = 0.19$ (n.s.)), reached significance, although the changes remained consistent with the other two models in direction; see Materials and Methods.

In contrast, evaluating these same comparisons in the primary model for a parameter sensitive to exploitation (ρ; see Materials and Methods) identified no significant results (ANOVA drug x genotype interactions and main effects of drug: p values > 0.24 (n.s.)). Consistent with the difference in tolcapone's effects on exploration and exploitation, following conversion of the tolcapone-placebo difference
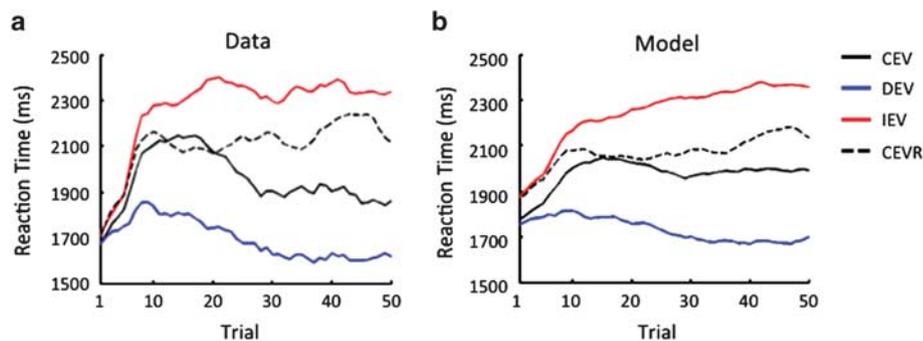
Figure 3 (a) Shown are the average reaction times across all 64 genotyped subjects for each of the four conditions, smoothed by a 10-trial weighted average. As evident in the plots, subject behavior strongly differentiated the four conditions ($F_{(2.6,165.1)} = 7.8$, $p \ll 1.0 \times 10^{-3}$, Greenhouse-Geiser corrected). (b) Model predictions for the average reaction time across trials for all 64 genotyped subjects.
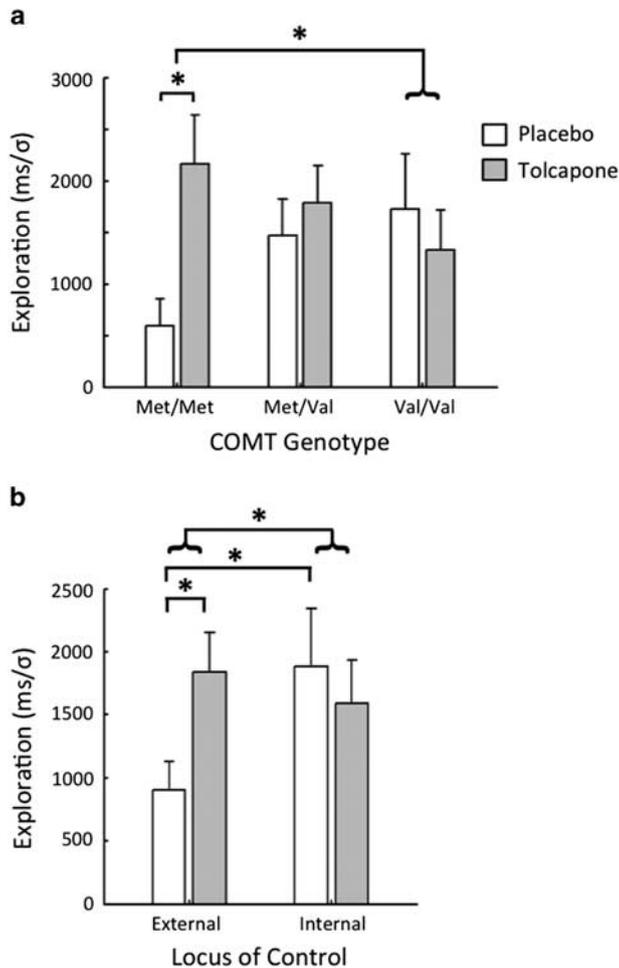
Figure 4 (a) Changes in the exploration parameter, in units of milliseconds per unit standard deviation of the belief distributions, on tolcapone vs placebo, segregated by genotype (mean ± SEM; $N_{Met/Met} = 16$, $N_{Met/Val} = 27$, $N_{Val/Val} = 21$). There is a significant drug × genotype interaction ($F_{(2,61)} = 3.68$, $p = 0.03$) that is driven by a significant differential increase in exploration for the Met/Met subjects relative to the Val/Val subjects. (b) When a median split is performed on locus of control (LOC) scores, the value of the exploration parameter is significantly larger in those with an internal, as opposed to external, LOC (mean ± SEM; $N_{External} = 37$, $N_{Internal} = 29$). Additionally, in subjects with an external LOC, exploration increased significantly on tolcapone; and this increase was significantly greater than that in subjects with an internal LOC. (*$p < 0.05$).

to Z scores for both the exploration and exploitation parameters, a two-way ANOVA with factors of parameter and genotype identified a significant parameter–genotype interaction ($F_{(2,122)} = 4.8$, $p = 0.01$) that was driven by the significant genotype–drug interaction for the explore parameter.

Because factors other than *COMT* genotype are likely to reflect underlying dopamine tone, we also assessed whether a behavioral factor, LOC, might predict response to tolcapone, as LOC has been hypothesized to indirectly reflect dopamine tone within frontostriatal circuitry (De Brabander and Declerck, 2004; Declerck *et al*, 2006). We performed a median split on our subjects, including those two participants whose genotype was unknown, and compared the change in exploration on tolcapone vs placebo in subjects with a more external as opposed to internal LOC

(Figure 4b). Consistent with previous theories, subjects with a more external LOC showed significantly lower exploration relative to subjects with a more internal LOC in the placebo condition ($T(64) = -2.14$, $p = 0.0185$ (one-tailed)). (Of note, these *p* values are not corrected for multiple comparisons, though they remain at trend significance or greater when Bonferroni-corrected for concomitant testing of alcohol use with the AUDIT, impulsivity with the BIS, and anxiety with the STAI (see Supplementary Methods)). Additionally, subjects with a more external LOC showed a significant increase in exploration on tolcapone ($T(36) = 2.43$, $p = 0.010$ (one-tailed)), and this increase was significantly greater than in subjects with a more internal LOC ($T(64) = 2.23$, $p = 0.0145$ (one-tailed)). (The different model variants gave qualitatively similar results although only the RT difference model reached significance for any comparisons: external *vs* internal LOC, placebo condition ($T(64) = -1.33$, $p = 0.095$ (one-tailed, *n.s.*)); external LOC, tolcapone *vs* placebo ($T(37) = 1.30$, $p = 0.10$ (one-tailed, *n.s.*)); and change on tolcapone *vs* placebo, external *vs* internal LOC ($T(64) = 2.00$, $p = 0.025$ (one-tailed)). For the negative-permitting model, all comparisons were non-significant (all *p*'s > 0.06, one-tailed), again suggesting that perseverative responding could not explain these data.). Importantly, no difference in LOC was seen between the three COMT genotypes ($F_{(2,62)} = 0.88$, $p = 0.42$ (*n.s.*)), arguing that genotype did not drive the LOC effects. Finally, these effects were also specific to exploration: LOC did not influence the exploitation parameter ρ (all *p*'s > 0.4 (*n.s.*)). In particular, differences in ρ were not seen for the response to tolcapone *vs* placebo between subjects with more external and internal LOC values ($T(64) = -0.04$, $p = 0.97$ (*n.s.*)).

DISCUSSION

In this study of the exploration-exploitation tradeoff, the randomized, double-blind, counterbalanced, within-subject administration of the *COMT* inhibitor tolcapone significantly increased exploration, particularly in Met/Met subjects relative to placebo, and the drug effect was significantly greater than in Val/Val subjects. Likewise, scores on a behavioral measure, Rotter's LOC scale, correlated with the magnitude of exploration effects: relative to subjects with a more internal LOC, subjects with a more external LOC showed both reduced exploration in the placebo condition, and increased exploration on tolcapone *vs* placebo. Although these effects will certainly benefit from replication in independent data sets, the influence of tolcapone in these subjects supports both the role of presumptively frontal dopamine in these behaviors and the structure of the computational models used to study them. We address each of these issues in turn.

The conceptual basis for exploratory behaviors lies in the evaluation of relative uncertainty. When the spectrum of outcomes associated with each of two competing actions is unknown, the brain must weigh the advantages of exploiting the action associated with a more certain outcome against exploring an action whose payoff is more unspecified. Intrinsically, the decision to explore represents a belief that neglecting more certain rewards now will produce new knowledge that may lead to greater gains later. Thus, there

is a potential distinction between the representation of knowledge about relative uncertainty, and the decision to act on it. Subjects' motivations to act on uncertainty may vary for many reasons: because the salience of the immediate rewards is great, because the prospect of more distant rewards is discounted, or because they feel that their agency—their ability to capitalize upon those rewards—is reduced.

In our data, we were able to at least partially address these possibilities. With respect to reward salience, our group as a whole was sensitive to reward probability. As shown in Figure 3, there was a shorter mean RT for the CEV relative to the CEV-R condition, which is distinguished from CEV by an increasing reward probability (and decreasing reward magnitude). By slowing in the CEV-R condition, subjects may be demonstrating a form of risk aversion (ie, they prefer frequent small rewards to infrequent large rewards), a result that we previously demonstrated is linked to striatal genes (Frank *et al*, 2009). However, this group sensitivity to reward probability would not explain the individual data; individual differences in mean RT between the CEV-R and CEV conditions did not correlate with exploration in placebo, tolcapone, and tolcapone-placebo conditions (all p's > 0.2 (*n.s.*)), nor were exploitation-related factors associated with exploratory behaviors (or the lack thereof). Alternatively, if subjects tend to discount the possibility of later rewards within a set of 50 trials, it may also be that they discount on longer time scales, as captured by traditional delay discounting tasks (Kayser *et al*, 2012). All of our subjects also completed a delay discounting task (see Methods; not further described in this paper), and there were no correlations between impulsive choice and exploration in placebo, tolcapone, and tolcapone-placebo comparisons (all r's < 0.07; all p's > 0.6 (*n.s.*)). It is possible that the different time scales and reward structures of these tasks render them incommensurate; regardless, we do not have evidence that discounting behavior is contributing here.

The third point is directly addressed by LOC. Subjects who have a greater sense of agency—who believe that their choices can determine future events—showed a greater tendency to explore in the placebo condition than subjects who had a more external LOC (Figure 4). In contrast, subjects with an external LOC showed a greater response to tolcapone. As an important negative control, these effects of tolcapone were not seen for a parameter that was more strongly related to exploitative behaviors. This finding directly demonstrates that differences in this behavioral trait may predict who is likely to act on differences in relative uncertainty.

Why should tolcapone have this effect? Underlying the motivation for studying tolcapone's influence on exploration is a theoretical framework in which increases in dopamine within PFC lead to improved cognitive control. This hypothesis has two components, one about the mediator (dopamine) and one about brain localization (PFC). Causal pharmacological interventions in working memory paradigms have shown that the effect of dopamine agonists depends on basal dopamine tone (Kimberg *et al*, 1997; Cools and D'Esposito, 2011): dopamine agonists improve working memory performance in those with low spans but worsen performance in those with high spans,

consistent with an inverted U-shaped effect of dopamine (Kimberg *et al*, 1997; Cools and D'Esposito, 2011; Vytlacil *et al*, 2014). Our findings here are in keeping with these ideas, albeit in a different cognitive domain. Specifically, those subjects with a more external LOC showed reduced exploration at baseline, but they differentially improved when tolcapone augmented dopamine tone. In contrast, subjects with a more internal LOC showed no improvement on drug. Previous authors have argued that dopamine tone may influence self-regulatory functions such as LOC (De Brabander and Declerck, 2004; Declerck *et al*, 2006), and the current findings suggest that an internal LOC, indicative of greater cognitive control, may be associated with greater dopamine tone in relevant frontostriatal circuits.

More unexpected are the current findings with respect to *COMT* alleles. In particular, our placebo data fail to replicate our previous non-pharmacological work in a different subject population, in which Val/Val individuals demonstrated less exploration than Met/Met subjects (Frank *et al*, 2009). Because subjects with the Val/Val genotype have a more active *COMT* enzyme, synaptically-released dopamine is presumably degraded more quickly, and therefore dopamine tone would be expected to be lower in Val/Val subjects than in subjects with the Met/Met or Met/Val genotypes. Thus, subjects with the Val/Val genotype might have been expected to show less exploratory behavior in the placebo condition. In contrast, we saw no significant differences between *COMT* Val158Met genotypes on placebo, and, if anything, there instead existed a trend toward reduced exploration in the Met/Met individuals.

One important potential explanation relates to the age of the study population here and in our previous work (Frank *et al*, 2009). Our first study was performed in college undergraduates (19 ± 1.7 years), whereas the current study participants encompassed a much broader age range (29.5 ± 8.2 years). Given the association of exploratory behaviors with rostrolateral PFC (Badre *et al*, 2012; Daw *et al*, 2006) and the ongoing maturation of rlPFC through late adolescence (Dumontheil *et al*, 2008), it is possible that the effects of COMT genotype on exploratory behaviors vary between late adolescence and adulthood. Thus, dopaminergic modulation of activity in rlPFC may not effectively influence other brain systems until later in adulthood. This idea has a precedent in other paradigms: in a delay discounting task, Smith and Boettiger (Smith and Boettiger, 2012) found that Val/Val subjects in late adolescence (ages 18–21 years in their study) were less impulsive than Val/Val adults and Met/Met adolescents, whereas older subjects showed the opposite pattern—ie, Met/Met adults were less impulsive than Met/Met adolescents and Val/Val adults. These data amplify the possibility that the impact of the drug is moderated by developmental changes in the dopamine pathway, frontostriatal circuitry, or both.

As discussed previously, a second possibility is that the Met/Met genotype endows subjects with a greater tendency to compute uncertainty, but the tendency to use it to guide behavior varies with other factors such as motivational state. More specifically, even if identifying *COMT* Val158-Met genotype reliably distinguishes the fidelity with which different subjects encode relative uncertainty, independent differences in factors such as LOC between this study and the previous study might moderate how likely subjects are

to use this information to change behavior. In support of this idea, subjects with a more internal LOC were more likely to explore in the placebo condition than were subjects with a more external LOC. Moreover, LOC did not differ between COMT genotypes in these subjects, consistent with the idea that it indexes an independent motivational component of task behavior.

Finally, one cannot exclude the possibility that epistatic dopaminergic compensations in other components of the dopamine pathway during development in Met/Met subjects (Stelzel et al, 2009) render this genotype more sensitive to even small changes in frontal dopamine. However, given the reduced activity of the enzyme associated with the Met/Met polymorphism, such homeostatic changes over development might be expected to result in an overall reduction, not increase, in sensitivity to tolcapone's effects. Similarly, such changes are also less likely to result from unanticipated activity of the *COMT* enzyme itself. Apud and colleagues (Apud *et al*, 2007), among others, have previously confirmed that COMT genotype correlates with COMT activity. Specifically, they demonstrated that COMT activity in peripheral blood samples is greater in Val/Val subjects and shows a greater decline when tolcapone is administered, with all three genotypes (Val/Val, Met/Val, and Met/Met) reaching a similar level of enzyme activity post drug. Their findings are also consistent with a PET study demonstrating that, in Parkinsonian patients administered ^{18}F-dopa, tolcapone relative to placebo increased the ^{18}F-dopa signal between 180 and 240 min after injection, suggesting that tolcapone was inhibiting ^{18}F-dopa breakdown by COMT (Ceravolo et al, 2002).

Despite this failure to replicate our previous work, the genotype-related findings for tolcapone nonetheless conform with the broader hypothesis that augmenting dopamine tone in rlPFC should enhance exploration for that genotype with lower baseline exploration. The suspected role of the rlPFC here (Badre *et al*, 2012) would be strongly consistent with its hypothesized function in other tasks that include exploratory components (Daw *et al*, 2006), or that assess behavioral context more generally (Ramnani and Owen, 2004); and it could provide a common neural basis for both exploratory behaviors and the hypothesized site of action of tolcapone.

In conclusion, our data demonstrate that subjects with the Met/Met genotype at the Val159Met allele, as well as subjects with a more external LOC measure, showed significant increases in exploration on tolcapone *vs* placebo. One speculation is that genotype and LOC may, respectively, influence how well subjects track relative uncertainty, and how likely they are to then use it to guide decisions. Regardless, the more general idea that exploratory behaviors can be captured in a complementary manner by both behavioral and genetic measures (as well as by intermediate phenotypes determined by neuroimaging, for example) would be important to evaluate in future studies that include other behaviors. This study also demonstrates the more general importance of causally assessing genotype–phenotype association studies. Even when previous data do not replicate, larger principles about (in this case, drug) mechanism can be assessed. Ultimately, such causal tests may also have relevance to the understanding and treatment of patient groups. In this case, our findings suggest that tolcapone may modulate cognitive processes associated with the exploration-exploitation tradeoff in patients with suspected dopamine-related dysregulation (Moustafa *et al*, 2008; Maia and Frank, 2011), including patients with schizophrenia (Strauss *et al*, 2011) and substance-use disorders; and future work might therefore investigate whether this modulation has beneficial effects in selected patient groups.

REFERENCES

Apud JA, Mattay V, Chen J, Kolachana BS, Callicott JH, Rasetti R et al (2007). Tolcapone improves cognition and cortical information processing in normal human subjects. *Neuropsychopharmacology* 32: 1011–1020.

Badre D, Doll BB, Long NM, Frank MJ (2012). Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. *Neuron* 73: 595–607.

Cavanagh JF, Figueroa CM, Cohen MX, Frank MJ (2012). Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation. *Cereb Cortex* 22: 2575–2586.

Ceravolo R, Piccini P, Bailey DL, Jorga KM, Bryson H, Brooks DJ (2002). 18F-dopa PET evidence that tolcapone acts as a central COMT inhibitor in Parkinson's disease. *Synapse* 43: 201–207.

Chen J, Lipska BK, Halim N, Ma QD, Matsumoto M, Melhem S et al (2004). Functional analysis of genetic variation in catechol-O-methyltransferase (COMT): effects on mRNA, protein, and enzyme activity in postmortem human brain. *Am J Hum Genet* 75: 807–821.

Collins AG, Frank MJ (2014). Opponent Actor Learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychol Rev* 121: 337–366.

Cools R, D'Esposito M (2011). Inverted-U-shaped dopamine actions on human working memory and cognitive control. *Biol Psychiatry* 69: e113–e125.

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006). Cortical substrates for exploratory decisions in humans. *Nature* 441: 876–879.

De Brabander B, Declerck CH (2004). A possible role of central dopamine metabolism associated with individual differences in locus of control. *Personality and Individual Differences* 37: 735–750.

Declerck CH, Boone C, De Brabander B (2006). On feeling in control: a biological theory for individual differences in control perception. *Brain Cogn* 62: 143–176.

Dumontheil I, Burgess PW, Blakemore SJ (2008). Development of rostral prefrontal cortex and cognitive and behavioural disorders. *Dev Med Child Neurol* 50: 168–181.

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.

Frank MJ, Fossella JA (2011). Neurogenetics and pharmacology of learning, motivation, and cognition. *Neuropsychopharmacology* **36**: 133–152.

Gogos JA, Morgan M, Luine V, Santha M, Ogawa S, Pfaff D 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.

Kalidindi K, Bowman H (2007). Using epsilon-greedy reinforcement learning methods to further understand ventromedial prefrontal patients' deficits on the Iowa Gambling Task. *Neural Netw* **20**: 676–689.

Kayser AS, Allen DC, Navarro-Cebrian A, Mitchell JM, Fields HL (2012). Dopamine, corticostriatal connectivity, and intertemporal choice. *J Neurosci* **32**: 9402–9409.

Kimberg DY, D'Esposito M, Farah MJ (1997). Effects of bromocriptine on human subjects depend on working memory capacity. *Neuroreport* **8**: 3581–3585.

Maia TV, Frank MJ (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nat Neurosci* **14**: 154–162.

Mitchell JM, Fields HL, D'Esposito M, Boettiger CA (2005). Impulsive responding in alcoholics. *Alcohol Clin Exp Res* **29**: 2158–2169.

Moustafa AA, Cohen MX, Sherman SJ, Frank MJ (2008). A role for dopamine in temporal decision making and reward maximization in parkinsonism. *J Neurosci* **28**: 12294–12304.

Patton JH, Stanford MS, Barratt ES (1995). Factor structure of the Barratt impulsiveness scale. *J Clin Psychol* **51**: 768–774.

Ramnani N, Owen AM (2004). Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nat Rev Neurosci* **5**: 184–194.

Rotter JB (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychol Monogr* **80**: 1–28.

Smith CT, Boettiger CA (2012). Age modulates the effect of COMT genotype on delay discounting behavior. *Psychopharmacology (Berl)* **222**: 609–617.

Spielberger CD (1983). *Manual for the state-trait anxiety inventory*. Consulting Psychologists Press.

Stelzel C, Basten U, Montag C, Reuter M, Fiebach CJ (2009). Effects of dopamine-related gene-gene interactions on working memory component processes. *Eur J Neurosci* **29**: 1056–1063.

Strauss GP, Frank MJ, Waltz JA, Kasanova Z, Herbener ES, Gold JM (2011). Deficits in positive reinforcement learning and uncertainty-driven exploration are associated with distinct aspects of negative symptoms in schizophrenia. *Biol Psychiatry* **69**: 424–431.

Vytlacil J, Kayser A, Miyakawa A, D'Esposito M (2014). An approach for identifying brainstem dopaminergic pathways using resting state functional MRI. *PLoS One* **9**: e87109.

Supplementary Information accompanies the paper on the Neuropsychopharmacology website (http://www.nature.com/npp)