Dopamine modulates reward-related vigor

Beierholm, U.; Guitart-Masip, M.; Economides, M.; Chowdhury, R.; Dolan, R.; Düzel, E.; Dayan, P.

Document Version
Early version, also known as pre-print

[Link to publication on Research at Birmingham portal](#)

Dopamine modulates reward related vigor

Authors:
Ulrik Beierholm* (1), PhD
Marc Guitart-Masip* (2), PhD
Marco Economides (2), MsC
Rumana Chowdhury (2,3), MD
Emrah Duzel (3), PhD
Ray Dolan (2), PhD
Peter Dayan (4), PhD

*Equal contribution

1 Centre for Computational Neuroscience & Cognitive Robotics, University of Birmingham, UK.
2 Wellcome Trust Centre for Neuroimaging, UCL, UK.
3 Institute of Cognitive Neuroscience, UCL, UK.
4 Gatsby Computational Neuroscience Unit, UCL, UK.

Running Title: Dopamine modulates reward related vigor

Corresponding author:

Ulrik Beierholm, PhD
Lecturer in Computational Neuroscience
University of Birmingham
School of Psychology
Edgbaston
Birmingham, B15 2TT
UK
Ph: (+44) 121 414 5534
Fax: (+44) 121 414 4897
u.beierholm@bham.ac.uk

Dopamine modulates reward related vigor

Abstract

Subjects routinely control the vigor with which they emit motoric responses. However the bulk of formal treatments of decision-making ignores this dimension of choice. A recent theoretical study suggested that action vigor should be influenced by experienced average reward rate, and that this rate is encoded by tonic dopamine in the brain. We previously examined how average reward rate modulates vigor as exemplified by response times, and found a measure of agreement with the first suggestion. In the current study we examined the second suggestion, namely the potential influence of dopamine signaling on vigor. Ninety healthy subjects participated in a double-blind experiment in which they received one of the following: placebo, L-DOPA (which increases dopamine levels in the brain) or citalopram (which affects brain serotonin levels). Subjects performed multiple trials of a rewarded odd-ball discrimination task in which we varied the potential reward over time in order to exercise the putative link between vigor and average reward rate. As before, we found a significant fraction of the variance in subjects' responses could be explained by our experimentally manipulated changes in average reward rate. Crucially, this relationship was significantly stronger under L-Dopa than under Placebo, suggesting the impact of average reward levels on action vigor is indeed subject to a dopaminergic influence.

INTRODUCTION

Whenever subjects perform actions, they face two fundamental classes of choice. One concerns which of a number of available actions to perform. The empirical literature in psychology and neuroscience on this topic has been the subject of powerful and illuminating theoretical treatments, based on normative decision-theoretic principles. The other class of choice concerns when, or how vigorously, one should perform an action. This is actually of broader significance because in many paradigms in animal learning, such as free operant tasks, vigor is the only dependent variable. Much is known about how subjects behave in such tasks but there has been little theoretical work examining and explaining these data.

One line of theoretical investigation has considered instrumental aspects of vigor in free-operant tasks (Niv *et al*, 2007). This account starts with two key assumptions: the first is that subjects seek to maximize the average rate of net utility per unit time; the second is that utility is decreased according to the cost of performing an action, and that performing an action more quickly (i.e., more vigorously) is more costly. Given a hyperbolic functional form for this increasing cost, vigor turns out to be determined by the opportunity cost of being slothful, where this opportunity cost is just the average rate of net utility.

In a recent study (Guitart-Masip *et al*, 2011) we tested this prediction in a qualitative manner by modulating the monetary rewards subjects could receive for making appropriate responses. We showed that subjects indeed modulated their response times based on the local average reward rate, i.e. the average amount of reward they had received over the last few minutes in the task. On the other hand, we found that the size of the instantaneously available reward for the immediate choice had less of an effect, with larger offers being, if anything, antagonistic to fast performance. Although this finding is perhaps surprising given the, coarser scale, Pavlovian effects on vigor examined in paradigms such as the monetary incentive delay task (Knutson *et al*, 2000), it is in keeping with the actual model that inspired the experiment (Niv *et al*, 2007).

Based on various sources of evidence (Evenden and Robbins, 1983; Floresco *et al*, 2003; Grace, 1991; Salamone and Correa, 2002), it was also predicted that the opportunity cost of time, i.e., the average rate of net appetitive utility, would be conveyed by tonic levels of the neuromodulator dopamine (Niv *et al*, 2007). This idea is supported by a large literature showing that dopamine manipulations have specific effects on the vigor of motivated behavior (Aberman and Salamone, 1999; Correa *et al*, 2002; Mingote *et al*, 2005; Salamone *et al*, 2001; Sokolowski *et al*, 1998). For example, dopamine depletion in rat nucleus accumbens leads to less responding in a reward schedule requiring a large number of lever presses, but not for a small number of lever presses (Aberman and Salamone, 1999). Indeed, the latter experimental data was a key influence on the computational model of (Niv *et al*, 2007) Other experiments have shown that dopamine depletion reverses animals' preference from a high-cost/high-reward to a low-cost/low-reward option in multiple experimental settings (e.g. (Salamone *et al*, 1991; Sokolowski and Salamone, 1994); for a recent review on these issues see (Salamone and Correa, 2012)).

A more speculative possibility is that tonic levels of another neuromodulator, namely serotonin, could be involved in reporting the average rate of net aversive utility, and, by symmetry, be involved in behavioral sloth. The reason for this is the (still somewhat contentious) notions of opponency between dopamine and serotonin (Boureau and Dayan, 2011; Cools *et al*, 2011; Daw *et al*, 2002; Deakin and Graeff, 1991), and indeed serotonin's known involvement in behavioral inhibition (Boureau and Dayan, 2011; Cools *et al*, 2011;

Crockett *et al*, 2009; Huys and Dayan, 2009; Soubrié, 1986). This leads to a possibility that boosting serotonin levels might have opposite effects on vigor to that seen when boosting dopamine levels. Here, we set out to test the effect on vigor of manipulating dopamine and serotonin in healthy human subjects.

Participants were assigned to receive placebo, levodopa (150mg), or citalopram (24mg in oral drops, equivalent to 30mg in tablets) and performed the exact task described in our previous paper (Guitart-Masip *et al*, 2011). The pharmacological agents are assumed to increase postsynaptic levels of dopamine and serotonin, respectively. We predicted that an increase in dopamine after levodopa administration would lead to a stronger modulation of the response times due to the influence of average reward rate. An additional possibility would be that an increase in serotonin after citalopram would weaken this modulation.

METHODS

Subjects

90 healthy volunteers were recruited for our pharmacological experiment (*pharmacological subjects*) using the subject pool associated with University College London's Psychology Department. A further 25 healthy volunteers were recruited for a control experiment (*tired subjects*). They received full written instructions, and provided written consent in accordance with the provisions of the University College London Research Ethics Committee.

"Pharmacological" subjects

Participants were randomly assigned to one of three treatment groups: 30 participants received levodopa (13 female; age range 17 years; mean 24.07, SD = 4.08 years), 30 participants received citalopram (17 female; age range 15 years; mean 23.6, SD = 4.2 years), and 30 participants received placebo (13 female; age range 11 years; mean 24.23, SD = 3.18 years). The study was double blind. All participants were right-handed and had normal or corrected-to-normal visual acuity. None of the participants reported a history of neurological, psychiatric or any other current medical problems.

"Tired" subjects

The pharmacodynamics of levodopa and citalopram imply that a waiting period is necessary before they exert potent effects on dopamine and serotonin. Further, before performing the vigor task, subjects participated in an unrelated task that also yielded monetary reward. These factors may potentially cause fatigue and a lower interest in the task for the subjects participating in the current experiment, and so temper the linkage to results of our previous study (Guitart-Masip et al., 2011). Consequently, to address this we examined the potential role of exhaustion and lowered interest by recruiting an additional twenty-five subjects (14 female; age range 15 years; mean 24.2, SD = 4.3 years) as above, who performed an unrelated task (reward based decision-making) for an average of 150 minutes before performing the vigor task. We intended their performance to mimic the behavior of the subjects in our pharmacological sample, but in the absence of any pharmacological manipulation.

Experimental procedure for the drug study

Participants completed the vigor task (see below) 100 minutes after receiving levodopa (150mg + 37.5mg benserazide) or 220 minutes after receiving citalopram (24mg in drops

which is equivalent to 30 mg in tablet). To ensure participants and researchers were blind to condition, on the day of the experiment each participant was allocated to one treatment group and received one glass containing either citalopram or placebo. Two hours later they received another glass containing either placebo or levodopa, and waited for another hour before engaging for 40 minutes in a go/no-go learning task (Guitart-Masip *et al*, 2012b) reported elsewhere. On the go/no-go task participants earned a minimum of £10 and a maximum of £35, based on their performance. Therefore, for all treatments groups, participants participated in the vigor task 220 minutes after arriving at the laboratory and after receiving a substantial monetary incentive. Participants completed a subjective state analogue-scales questionnaire on three occasions. We did not detect any difference in subjective ratings between treatment groups (see Supplementary Online Material for details).

Behavioral Paradigm (vigor task)

The behavioral paradigm was presented using a regular PC monitor and keyboard, exactly as described in Guitart-Masip et al. (2011). The layout of a trial is depicted in Figure 1A. In each trial subject could receive a potential payout in the range 1-100 pence, as presented visually on the screen at the beginning of the trial. The potential payouts, R_t were varied across trials according to a pre-specified function of trial number. This function was fixed across subjects and designed to vary over time in a way to minimize the correlation between the available reward, averaged reward rate and the linear component (see below). The potential payout function used is shown in Figure 1B. After a variable period (750–1250 msec, later referred to as the *Intertrial interval*), subjects were shown three visual figures and had to indicate the "odd one out" by pressing a button. If subject responded within 500 msec by pressing the button corresponding to the deviant stimulus the trial was considered successful. To keep participants engaged throughout the task we induced unexpected misses by lowering the time constraint to 400 msec in 20% of the trials. Subjects were informed as to their success on the trial after being shown a blank screen for 500 msec. Feedback was followed by another blank screen and the beginning of the next trial.

Subjects performed 458 to 472 trials within the 27 minutes time limit allocated. For subject payments 10% of the trials were chosen randomly, and subjects were paid the sum of the value of the successful subset of those trials, plus a fixed fee of £5 that was added to the amount of money that they had obtained on the previous, unrelated, task (see experimental procedure). Critically, this incentive structure implies that the faster they make correct responses, the more money they could potentially make.

Table 1 shows the average (and standard deviation) money made by each group on the vigor task. Four subjects amongst the 'pharmacology' participants, and one from the 'tired' participants, managed fewer than 200 correct trials within the available experimental time limits and were thus discarded from further analysis.

Data Analysis

We fitted a log normal distribution to each individual's reaction times (RTs) and removed any data points that were more than three standard deviations from the individual mean. We then recalculated the mean and standard deviations, and repeated this procedure. Missed trials (trials without any behavioral response) were not included in the analysis. To allow subjects to get used to the task we only analyzed trials 21 to 460. Participants who

managed fewer than 200 complete trials (correct button press within the time limit) were subsequently omitted from further analysis.

<Insert Figure 1 around here>

Given the log-normalized data, we assumed a linear model for the contribution of different factors on the z-scored response times for each subject *i*.

$$\log\overrightarrow{RT_i} = \bar{\bar{X}}_i * \vec{\beta}_i + \vec{\varepsilon}_i$$

where the columns in the $\bar{\bar{X}}$ matrix were given by the following variables which were chosen in the light of the results of our previous study (Guitart-Masip *et al*, 2011):

R_t: available reward for the subjects to win in a given trial.

$\bar{r}_t$: average reward signal, as given by $\bar{r}_t = \bar{r}_{t-1} + a(r_{t-1} - \bar{r}_{t-1})$, where r_{t-1} is the reward achieved in the previous trial. This update rule is equivalent to the Rescorla-Wagner rule, which is used routinely in learning approaches to average reward reinforcement learning. The update or learning rate for the average reward α was a free parameter of a random effects model fitted to each subject's responses according to the algorithm described below. The update rate could range between 0 (equivalent to no learning) and 1 (equivalent to only using the reward in the previous trial).

Repetition of stimulus (RS)*:* binary vector indicating whether the stimulus in the last trial was the same as in this trial.

Linear: linear function.

Too late: binary return indicating whether the response was too late in the previous trial.

Intertrial interval (ITI): pretrial interval while waiting for the stimulus to be presented.

Our key variables of interest were the available reward, R_t and the average reward signal $\bar{r}_t$.

The model is similar to a linear model for linear regression, apart from the effect of the individualized learning rates α. We treated it as a random effects model with a top level, Gaussian, prior $N(\mu_{prior}, \Sigma_{prior})$ for the β parameters and the learning rates α (with the latter being transformed through a sigmoid to restrict their range to the interval [0,1] so they can be treated uniformly). We fit the values of μ_{prior} and Σ_{prior} using a Bayesian Expectation-Maximization method, using regular linear regression as the inner loop for maximizing the likelihood with respect to β. We made a Laplace approximation about this maximum to realize an approximately normally distributed (but unnormalized) likelihood proportional to $N(m^i_{like}, S^i_{like})$, which was multiplied by $N(m_{prior}, S_{prior})$ and normalized to create the posterior estimate of each β value $N(m^i_{post}, S^i_{post})$. This can be easily done analytically as $m^i_{post} = S^i_{post}((S^i_{like})^{-1} m^i_{like} + (S_{prior})^{-1} m_{prior})$ and $S^i_{post} = ((S^i_{like})^{-1} + (S_{prior})^{-1})^{-1}$.

In the M-step the parameters for the prior were optimized as

$$[\mu_{priorNew}, \Sigma_{priorNew}] = \underset{\mu_0, \Sigma_0}{\text{argmax}} \sum_{i=1}^{m} \int N(\mu^i_{post}, \Sigma^i_{post}) \log N(\mu_0, \Sigma_0) dx$$

Dopamine modulates reward related vigor

$$\approx \underset{\mu_0, \Sigma_0}{\operatorname{argmax}} \sum_{i=1}^{m} \int (2\pi)^{-k/2} \left|\Sigma_{post}^i\right|^{-1/2} e^{\left(-\frac{1}{2}(x-\mu_{post}^i)\,\Sigma_{post}^{i\,-1}(x-\mu_{post}^i)\right)} \left(-\frac{1}{2}(x-\mu_0)\Sigma_0^{-1}(x-\mu_0) - \frac{1}{2}\log\left|\Sigma_0\right|\right) dx$$

where the dimensionality *k*=7 and *m* is the number of subjects.

There is an analytical solution to this maximization with $m_{priorNew} = \frac{1}{m}\sum_{i} m_{post}^i$ and

$$S_{priorNew} = \frac{1}{m}\sum_{i}\left((m_{post}^i)^2 - (m_{priorNew})^2\right) + S_{post}^i .$$

The E and M steps were repeated until the change in estimated variables between two E-steps were less than 0.001, signifying convergence.

Notice that our approach was not fully Bayesian in that we did not assume a 'hyper-prior' over the parameters for the prior. Having estimated the β-values, we then asked whether any of them explained a significant amount of variability of the data.

Re-analysis of original data

In our previous study (Guitart-Masip *et al*, 2011), we used the same experimental method, but no pharmacological treatment and also no waiting or intermediate task (the go/no-go task mentioned above). We had also employed a simpler analysis method. In order to compare our current results with the earlier ones, we reanalyzed the results of that prior experiment using the statistical methods described above. Out of the thirty-nine subjects we excluded one subject due to failure of the recording software.

RESULTS

We tested the effects of manipulating dopamine (using levodopa) on the vigor displayed by human subjects using a task developed to test how reward modulates reaction times. This task was the same as in our previous published study (Guitart-Masip *et al*, 2011), requiring speeded responses in an 'odd-one-out' task, where subjects were rewarded for being both accurate and suitably fast (see Methods).

For completeness, we report results and comparisons among three different sets of data. Set D_{orig} comes from reanalyzing data from (Guitart-Masip *et al*, 2011) using a more sophisticated statistical method (a random effects model) that we adopted for our new data, in view of an anticipated need to make comparisons between different groups of subjects.

Dataset D_{pharm} was the main experimental focus in the present paper. Subjects were administered placebo, levodopa or citalopram and after a fixed time period (which included participation in an unrelated reward task) performed the vigor task. D_{pharm} comprises data from the three groups D_{plac}, D_{ldopa}, D_{cit}.

Finally, we collected an additional dataset D_{tired} to assess the effect of the key experimental difference between the paradigm underlying D_{orig} and D_{pharm}, namely the requirement for

participants to remain on our premises for 220 minutes and engage in a learning task before performing the vigor task. This waiting period may have caused fatigue and decreased motivation in participants. To assess the effect of the waiting period, if any, we tested twenty-five subjects using a similar design: in effect these subjects participated in the vigor task 150 minutes after arrival for the study and after engaging in an unrelated reward based decision-making task with a pay-off ranging between £5 and £20.

Table 1 shows for each data set the means (and standard deviations) of the number of performed trials, the number of correct responses within the trial time limit, the number of trials with too late responses, and the number of wrong button presses. There was no significant difference across groups on any of these measures (Wilcoxon signed-rank test, p>0.05). Table 1 also includes the amount of money won and the average response times for each experiment, which also did not differ significantly across groups (1 way ANOVA, F(147)=0.62, p>0.05 and F(147)=1.08, p>0.05 respectively).

To examine the factors influencing the subjects' reaction times, we used an Expectation Maximization (EM) variant of linear regression (see Methods). The regressors of interest were the available reward and the average reward history (vigor signal), and several nuisance variables (see Methods).

The most significant formal difference to our previous analysis was that here, we fit the learning rate α used to calculate the average reward on a subject-by-subject basis as part of the random effects model, rather than using a single value for α across subjects estimated based on a maximum likelihood fit to pilot data (Guitart-Masip *et al*, 2011). Learning rates were individually fit in order to rule out the possibility that any effects of the pharmacological treatments on vigor were caused by undetected differences in learning rates. Our results, however, showed no difference in learning rate across data sets (1-way ANOVA F(147)=1.38, p>0.05) nor across the three pharmacological groups (F(85)=2.09, p>0.05)) .

<Insert Figure 2 around here>

	Original	Tired	Placebo	L-Dopa	Citalopram
# performed trials	459.9(3.0)	465.9 (3.1)	466.3 (2.4)	465.6 (2.2)	466.2 (2.4)
# correct trials	339.5 (49.2)	351.0 (58.6)	355.9 (46.0)	345.1(47.1)	347.6 (44.8)
# too late trials	86.8 (40.2)	86.8(45.1)	78.8(37.0)	86.7(40.2)	82.3 (39.9)
# wrong trials	32.9 (25.0)	28.1 (20.4)	31.6(19.7)	33.8(20.4)	36.4(18.5)
Money made	£7.17(1.39)	£7.21 (1.64)	£7.44 (1.29)	£6.82 (1.63)	£7.09 (1.78)
Mean Indiv. RT, all trials	415.0 (24.5) ms	408.7 (29.7) ms	403.9(22.9) ms	410.0(23.5) ms	404.9(23.4) ms
Mean Indiv. RT, correct trials	402.1 (18.8) ms	395.3(21.5) ms	393.1(17.2) ms	398.2(16.3) ms	392.3(15.9) ms
Mean Indiv. RT, too late trials	468.2 (25.4) ms	471.6(29.8) ms	466.9 (20.8) ms	468.0 (24.5) ms	473.0(22.1) ms
Mean Indiv. RT, wrong trials	365.6 (47.5) ms	360.1(43.6) ms	349.9 (30.2) ms	368.8 (24.7) ms	351.7 (29.8) ms
Alpha, learning rate per trial	0.1133 (0.1176)	0.1190 (0.0810)	0.1458 (0.1505)	0.1543 (0.1453)	0.0870 (0.0942)

Table 1: Comparing behavioral responses across all 5 data sets, mean and standard deviations in parenthesis.

'Original' experiment
The blue values in Figure 2 show the mean β weights (with associated standard errors) for the six regressors for dataset D_{orig}. As in the original study, we found that the average reward rate (vigor signal) had a significantly negative influence on reaction times (t=-6.91, p<0.001); the higher the average reward the quicker the responses. Note that the effect of average reward rate is similarly strong in the current analysis despite the fact that the mean learning rate across subjects is an order of magnitude higher, and the resulting regressor implies the integration of rewards over a much shorter time window (see figure 1b). This effect of average reward is very different from the effect of the immediate reward that would be available given correct performance on the current trial. Similar to our previous analysis we found that immediate reward had no significant impact on the reaction time; the β parameter for the available reward regressor was not significantly positive (one sample t-test, t=1.82, p>0.05). For the nuisance regressors we found a significant negative effect of the *Repetition of Stimulus*, *Linear*, and *Intertrial Interval* regressors (t=[-8.72,-9.23,-2.98,-7.58], p<0.01) in keeping with our previous report.

<Insert Figure 3 around here>

'Tired' experiment
We expected that after waiting for 150 minutes and participating in an unrelated reward-based decision-making task, the impact of average reward on vigor would be reduced, due to either fatigue or devaluation of reward. Indeed, although we found that the value of β associated with the average reward signal in D_{tired} was significantly negative (1 sample t-test, p<0.05, t=-2.21; red points in Fig 2), the value was significantly less than in D_{orig} (2-sample t-test, p<0.05, t=2.58). Effects of other regressors were similar in the two data sets, with only *Repetition of Stimulus* having a significantly stronger effect (2-sample t-test, t=-2.50, p<0.05). These results led us to expect a weaker vigor signal across subject groups in D_{pharm}.

Effects of pharmacological manipulations

Having examined the effect of fatigue we turn to D_{pharm}, with the results for placebo (D_{plac}, blue), L-Dopa (D_{ldopa}, red) and citalopram (D_{cit}, green) shown in Fig 3. We first checked the effects on the D_{plac}. For the average reward signal, $\overline{r}_t$, we again found a negative weight implying that the average reward signal causes subjects to speed up, although this was now only borderline significant (t=-1.96, p~=0.06) for the control group. This is in contrast to D_{orig}, but consistent with D_{tired}. For the available reward, R_t, we found a small positive effect (not significant using a t-test, p>0.05, t=0.92), similar to our previous study. As in our previous data sets, this implies that there was a weak tendency for subjects to slow down as the instantaneously available reward increased. Regarding the nuisance parameters, the β coefficients of the D_{plac} were nearly identical to D_{orig}, exhibiting no significant difference (2-sample t-test, p>0.05 for all regressors, t=[0.69,-1.02, 1.38,-0.26]).

<Insert Figure 4 around here>

Comparing 'Placebo' with 'L-Dopa' subjects
Our main interest in this study was to compare D_{ldopa} with D_{plac}. As predicted by our hypothesis, and the original model (Niv *et al*, 2007), we found the L-Dopa group had a

significantly stronger (more negative) effect of average reward rate compared to the control group (t=-2.28, p<0.05). That is, subjects receiving L-Dopa modulated their response times more strongly based on the recent reward history than control subjects (see Figure 3).

There was also a significant difference between the two groups for the *Intertrial Interval* regressor, for which L-Dopa subjects showed less effect of a long preparation time (2 sample t-test, p<0.05, t=-2.17). We did not find any other significant difference for the available reward regressor, R_t (2-sample t-test, t=-1.86, p>0.05), the linear regressor (t=-1.10,p>0.05), or the binary regressors indicating the repetition of stimulus (t=0.21,p>0.05), or a too late response in the previous trial (t=-0.73,p>0.05).

Comparing 'Placebo' with 'Citalopram' subjects
There was no significant difference between D_{cit} and D_{plac} for any regression coefficient (2 sample t-test, p>0.05), and, indeed, all β values were very similar (see Fig. 3). Comparing the coefficients for the average reward rate (vigor signal) across all groups, we found that the values for D_{plac} and D_{cit} and D_{tired} were all significantly different from that for D_{orig} (2-sample t-test, t=[3.25, 2.31, 2.58], p<0.05), while that for D_{ldopa} was not (D_{orig} versus D_{ldopa}, t=0.94, p>0.05), see Figure 4.

Comparing 'L-Dopa' with 'Citalopram' subjects
There was no significant difference between D_{ldopa} and D_{cit} for the effect of the available reward or the average reward rate (2 sample t-test, p>0.05, t=[0.65, -1.37]), nor for the *Repetition of Stimulus* or *Linear* regressors (2 sample t-test, p>0.05, t=[0.39, -0.79]). The only regressors to show a significant effect were the *Too Late* and *Inter-trial Interval* (2-sample t-test, t=[-2.49, 2.08], p<0.05, see Fig. 3). This implies that the Citalopram group was less spurred on by being too late than the L-Dopa group, while the L-Dopa group was less able to use preparation time to speed up their responses.

DISCUSSION

The relationship between incentive motivation and response vigor has been examined in a number of experimental studies (Cools *et al*, 2005; Satoh *et al*, 2003; Wittmann *et al*, 2005) showing that for humans and animals alike instrumental actions are influenced by the subjective value of rewards in the environment. From a formal standpoint, at least three factors should determine the vigor or latency of a response. First, the task itself could demand an appropriately quick response, as indeed was the case in the present task, and as studied more systematically in active avoidance paradigms (McCleary, 1961). Second, subjects may exhibit a trade-off between speed and accuracy, slowing down in order to perform more competently. Third, subjects may be able to minimize the opportunity cost associated with rewards that are postponed if actions are slothful.

In the context of our task, the first of these effects, and indeed any Pavlovian counterpart, such as preparatory or consummatory appetitive impulsivity associated with the prediction or presence of a potential reward (Evenden and Robbins, 1983), or appetitive Pavlovian to Instrumental Transfer (Talmi *et al*, 2008), should depend on the immediately offered reward R_t. The same dependence would be expected for a speed-accuracy trade-off. We discuss the first and second factors later as the theory we set out to test (Niv *et al*, 2007) considers the influence of the third factor, with the opportunity cost being the average rate of reward.

Dopamine modulates reward related vigor

We replicated our previous finding (Guitart-Masip *et al*, 2011) showing that the local average rate of reward influences the vigor of the instrumental responses of healthy human volunteers. Here, because of the exigencies of the pharmacokinetics of the drugs used, and an intermediate task, our subjects were prone to tiredness. Thus, as any finding that reward related vigor was reduced could be subject to alternative interpretations we collected additional control data to assess the effects of tiredness itself.

Our main finding was that the effect of this fatigue was reversed by the administration of levodopa – that is, boosting dopamine boosted reward-related vigor. This effect was specific to reward related vigor and not a general effect on arousal as shown by the lack of effect on the subjective analogue scale ratings. Moreover, among all the nuisance variables included in our regression analysis the only significant difference between placebo and L-Dopa was observed on the inter-trial interval regressor, whereby L-Dopa decreased the effect of longer preparation times between the display of the available reward and the appearance of the target stimulus. If anything this suggests that participants were less aroused after receiving the drug. These results confirm a theoretical suggestion that dopamine plays a critical role in response vigor (Guitart-Masip *et al*, 2011), an hypothesis based on a wealth of experimental data showing enhanced dopamine levels increase movement vigor in both animals (Lex and Hauber, 2008; Salamone and Correa, 2002; Taylor and Robbins, 1986; Ungerstedt, 1971) and humans (Guitart-Masip *et al*, 2012a), and had previously been tested less directly (Mazzoni *et al*, 2007; Salamone and Correa, 2002). These studies implicate the nucleus accumbens as at least one relevant site for the invigorating action of dopamine (Balleine, 2005; Lex and Hauber, 2008; Parkinson et al, 2002).

According to the original theory, the coupling of dopamine to vigor is instrumental in nature (maximizing the average rate of reward) and depends on tonic levels of this neuromodulator. However, Pavlovian effects may also play a key part, e.g., a direct coupling between reward rate and vigor even in tasks in which this does not actually increase the average reward rate. Such a direct coupling can in extreme cases lead to detrimental behavior as described in animals (Breland and Breland, 1961) and humans (Guitart-Masip *et al*, 2012b), where vigorous behavior in the light of immediately available reward in fact can lead to less overall reward. In our task an effective decoupling of the Pavlovian and instrumental systems could be achieved if the inter-trial interval had been increased to compensate for trials in which the subjects responded quickly. This manipulation would place Pavlovian and instrumental behavioral tendencies in opposition. It would be straightforward to test for this, for instance using a reward scheme based on differential reinforcement of low rates of responding (for an example of the use of this technique, see (Sokolowski and Salamone, 1994)). Our experiment tested a mild form of this in that it violated the conditions of the original theory by making the attainment of reward contingent upon a response being executed within a fixed time (either 400 or 500 ms). The fact that it is the rate of reward based on historical trials rather than the actual offer on the current trial that increases vigor can be seen in such Pavlovian terms.

The role of the instantaneously available reward, R_t, was puzzling. Using our previous analysis, as well as the more sensitive analysis in this study, we found that it had no significant effect on the reaction time (although, as before, the coefficient was numerically positive, i.e., associated if anything with slowing down). This is in stark contrast with the clear and repeatable effect of the average reward rate $\bar{r}_t$. One issue for the lack of effect of the instantaneous reward is the possibility that it arises through interaction with a speed-accuracy tradeoff (for instance, with a tendency for subjects to speed up because of the larger reward being tempered by the fact that this might make them less accurate).

However, as in our previous experiment, we found no between-subjects correlation between the available reward and the proportion of correct responses (calculated across subjects, r=0.047, p>0.05). This suggests that our main pharmacological effect is independent of such tradeoffs, in agreement with a recent study (Winkel *et al*, 2012). It is notable that Parkinson's patients, with impairments to dopamine signaling, show normal speed-accuracy tradeoffs, and merely a propensity to slower actions (Mazzoni *et al*, 2007). Together with our finding of dopaminergic modulation of reward related vigor, it is tempting to suggest that dopamine is *selectively* involved in a coupling between average reward rates and vigor.

Our results are apparently not in agreement with studies using the Monetary Incentive Delay task or related tasks (e.g. (Cools *et al*, 2005; Knutson *et al*, 2001; Wittmann *et al*, 2005)). In these experiments participants must perform fast button presses upon receiving a go- signal so as to obtain monetary rewards of different magnitudes. Responses are typically faster for trials on which larger rewards are available. However, these tasks involve a categorical comparison between different levels of reward magnitude, usually involving differences of one order of magnitude and where local fluctuations in average reward are likely to be small. Future research with a variant of the MID task involving both categorical levels of reward magnitude with systematic manipulations of average reward rate are needed to understand the exact relationship of available reward and average reward on response vigor.

The computational model by Niv and colleagues (2007) suggested that average reward was coded by tonic dopamine signals. However, in this theory, the underlying decision problem is stationary, which is not true in our experimental test, and so the exact timescale at which the average reward is realized in dopaminergic signaling is not completely clear. A variety of ways for measuring dopamine concentrations across various timescales are available, including cyclic voltammetry (Gan *et al*, 2010) or microdialysis (Westerink *et al*, 1996). Indeed, experiments using microdialysis have shown that in the ventral striatum dopamine increases in a time scale of minutes when animals perform instrumental responses in free operant tasks (Ostlund *et al*, 2011; Segovia *et al*, 2011). Interestingly, satiation induced a decrease in response vigor which was correlated with the decrease in dopamine efflux in the shell of the nucleus accumbens (Ostlund et al 2011).

One important complexity concerns phasic dopamine signals, as tonic and phasic aspects of dopamine are directly linked (Grace, 1991) and both are affected by levodopa (Cools, 2006). Reward prediction errors induced by cues, potentially including the indication of the immediately available reward, are tightly coupled to the phasic activity of dopamine neurons (Schultz *et al*, 1997) as well as with the extracellular concentration of dopamine (Gan *et al*, 2010). Further, reaction times are negatively correlated with phasic activity of dopamine neurons (Satoh *et al*, 2003). Therefore, one would expect a negative effect of the available reward on the reaction times. However, opposite to this we found a neutral or positive correlation, making it unlikely that these short-term dopaminergic signals are responsible for the observed dopaminergic modulation of average reward.

Some evidence about the appropriate timescale comes from the learning rates associated with average reward that we found through fitting the model. It is important to note that there was no significant difference in learning rates between groups, thus this does not appear to be a route by which levodopa could affect vigor. However, whereas in the current experiment we found that the learning rate ranged between 0.113 and 0.154 per trial, implying a time window for averaging of about 30s, in the previous study we used a single

learning rate for all subjects fit on pilot data that was 0.012 per trial, implying averaging over 5 minutes. Due to limitations in sensitivity of the analysis used in the original study we generated a single regressor for the averaged reward, with the learning rate fitted to the averaged response times across all subjects. In the current, more sensitive, analysis we fit the learning rate individually to each subject. As shown in table 1, calculating the learning rate in this way for subjects in the original study leads to a mean value consistent with all the other learning rates that we found in the current data set. We therefore suspect our previous procedure underestimated the learning rates. Nevertheless, we show here that the critical conclusions about the influence of average reward and immediately offered reward remain true.

We did not find any effects of citalopram on the impact of average reward on response vigor, nor did we find a significant difference from the levodopa group. One major pillar of the current version of the computational proposal that serotonin acts as an opponent to dopamine (Boureau and Dayan, 2011; Deakin and Graeff, 1991), is that serotonin is directly implicated in behavioral inhibition, behavioral quiescence and waiting (Huys and Dayan, 2009; Miyazaki *et al*, 2012) as a contrast to dopamine's involvement in behavioral activation (Cools *et al*, 2011; Guitart-Masip *et al*, 2012a). However, it is a possibility that the effects of serotonin on response inhibition are only observed when actions are taken in a context that includes punishments (Crockett *et al*, 2009, 2012). Furthermore, the involvement of serotonin in inhibition is typically complicated (Cools *et al*, 2011), and even the regional effects on serotonin concentration of single doses of citalopram are controversial (Bari *et al*, 2010). As a selective serotonin reuptake inhibitor citalopram's direct effect is via locally increased serotonin availability. However, acute citalopram administration can result in decreased total postsynaptic serotonin availability, at least at the cortical level (Selvaraj *et al*, 2012), possibly through a presynaptic inhibitory mechanism (Artigas *et al*, 1996; Hajós *et al*, 1995). Despite these uncertainties on the effects of citalopram on serotonin levels, the inclusion of this drug may highlight selective involvement of the serotonergic system in specific cognitive functions. One possibility that can be tested in future experiments is whether serotonin is involved in coupling the average rate of punishment into sloth (Dayan, 2012). More complex predictions have been made about the effect of dopamine and serotonin in active avoidance paradigms where appropriately early responses are necessary to obviate punishments (Dayan, 2012).

One possible limitation of the current experiment relates to the fact that dopamine fluctuates with the menstrual cycle (Czoty *et al*, 2009; Jacobs and D'Esposito, 2011; Ossewaarde *et al*, 2011). This may result in increased variability of the effects of L-Dopa, and may have had a deleterious effect when trying to assess cognitive effects of a pharmacological manipulation. Importantly, although increased noise could certainly have been problematic in the light of a negative result, the significant effect of L-Dopa that we found across what was a mixed sample can be seen as more strongly suggestive of the involvement of the dopaminergic system in the regulation of response vigour.

In sum, we show that not only is the vigor of human movement modulated by average reward rate but that this signal is likely to be encoded by the dopaminergic system in the central nervous system. This adds to our understanding of the motivational aspects of dopamine, to complement the vastly more extensive investigations of its role in learning about rewards.

CONFLICTS OF INTERST

ACKNOWLEDGMENTS

REFERENCES

Aberman JE, Salamone JD (1999). Nucleus accumbens dopamine depletions make rats more sensitive to high ratio requirements but do not impair primary food reinforcement. *Neuroscience* **92**: 545–52.

Artigas F, Romero L, Montigny C de, Blier P (1996). Acceleration of the effect of selected antidepressant drugs in major depression by 5-HT1A antagonists. *Trends in neurosciences* **19**: 378–83.

Balleine B (2005). Neural bases of food-seeking: Affect arousal and reward in corticostriatolimbic circuits. *Physiology & Behaviour* **86**: 717–730.

Bari A, Theobald DE, Caprioli D, Mar AC, Aidoo-Micah A, Dalley JW, *et al* (2010). Serotonin Modulates Sensitivity to Reward and Negative Feedback in a Probabilistic Reversal Learning Task in Rats. *Neuropsychopharmacology* **35**: 1290–1301.

Boureau Y-L, Dayan P (2011). Opponency revisited: competition and cooperation between dopamine and serotonin. *Neuropsychopharmacology* **36**: 74–97.

Breland K, Breland M (1961). The misbehavior of organisms. *American Psychologist* **16**: 681–84.

Cools R (2006). Dopaminergic modulation of cognitive function-implications for L-DOPA treatment in Parkinson's disease. *Neuroscience and biobehavioral reviews* **30**: 1–23.

Cools R, Blackwell A, Clark L, Menzies L, Cox S, Robbins TW (2005). Tryptophan depletion disrupts the motivational guidance of goal-directed behavior as a function of trait impulsivity. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* **30**: 1362–73.

Cools R, Nakamura K, Daw ND (2011). Serotonin and dopamine: unifying affective, activational, and decision functions. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* **36**: 98–113.

Correa M, Carlson BB, Wisniecki A, Salamone JD (2002). Nucleus accumbens dopamine and work requirements on interval schedules. *Behavioural brain research* **137**: 179–87.

Crockett MJ, Clark L, Apergis-Schoute AM, Morein-Zamir S, Robbins TW (2012). Serotonin Modulates the Effects of Pavlovian Aversive Predictions on Response Vigor. *Neuropsychopharmacology* **37**: 2244–2252.

Crockett MJ, Clark L, Robbins TW (2009). Reconciling the Role of Serotonin in Behavioral Inhibition and Aversion: Acute Tryptophan Depletion Abolishes Punishment-Induced Inhibition in Humans . *The Journal of Neuroscience* **29** : 11993–11999.

Czoty PW, Riddick N V, Gage HD, Sandridge M, Nader SH, Garg S, *et al* (2009). Effect of menstrual cycle phase on dopamine D2 receptor availability in female cynomolgus monkeys. *Neuropsychopharmacology* **34**: 548–54.

Daw ND, Kakade S, Dayan P (2002). Opponent interactions between serotonin and dopamine. *Neural networks : the official journal of the International Neural Network Society* **15**: 603–16.

Dayan P (2012). Instrumental vigour in punishment and reward. *European Journal of Neuroscience* **35**: 1152–1168.

Deakin JF, Graeff FG (1991). 5-HT and mechanisms of defence. *Journal of psychopharmacology (Oxford, England)* **5**: 305–15.

Evenden JL, Robbins TW (1983). Dissociable effects of d-amphetamine, chlordiazepoxide and alpha-flupenthixol on choice and rate measures of reinforcement in the rat. *Psychopharmacology* **79**: 180–6.

Floresco SB, West AR, Ash B, Moore H, Grace AA (2003). Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission. *Nature neuroscience* **6**: 968–73.

Gan JO, Walton ME, Phillips PEM (2010). Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. *Nature neuroscience* **13**: 25–7.

Grace AA (1991). Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: a hypothesis for the etiology of schizophrenia. *Neuroscience* **41**: 1–24.

Guitart-Masip M, Beierholm U, Dolan R, Duzel E, Dayan P (2011). Vigor in the Face of Fluctuating Rates of Reward: An Experimental Examination. *Journal of Cognitive Neuroscience* **23**: 1–6.

Guitart-Masip M, Chowdhury R, Sharot T, Dayan P, Duzel E, Dolan RJ (2012a). Action controls dopaminergic enhancement of reward representations. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 7511–6.

Guitart-Masip M, Huys QJM, Fuentemilla L, Dayan P, Duzel E, Dolan RJ (2012b). Go and no-go learning in reward and punishment: Interactions between affect and effect. *NeuroImage* **62**: 154–66.

Dopamine modulates reward related vigor

Hajós M, Gartside SE, Sharp T (1995). Inhibition of median and dorsal raphe neurones following administration of the selective serotonin reuptake inhibitor paroxetine. *Naunyn-Schmiedeberg's archives of pharmacology* **351**: 624–9.

Huys QJM, Dayan P (2009). A Bayesian formulation of behavioral control. *Cognition* 1–15doi:10.1016/j.cognition.2009.01.008.

Jacobs E, D'Esposito M (2011). Estrogen shapes dopamine-dependent cognitive processes: implications for women's health. *The Journal of Neuroscience* **31**: 5286–93.

Knutson B, Adams CM, Fong GW, Hommer D (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *Journal of Neuroscience* **21**: RC159.

Knutson B, Westdorp A, Kaiser E, Hommer D (2000). FMRI visualization of brain activity during a monetary incentive delay task. *Neuroimage* **12**: 20–27.

Lex A, Hauber W (2008). Dopamine D1 and D2 receptors in the nucleus accumbens core and shell mediate Pavlovian-instrumental transfer. *Learning & memory (Cold Spring Harbor, NY)* **15**: 483–91.

Mazzoni P, Hristova A, Krakauer JW (2007). Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **27**: 7105–16.

McCleary RA (1961). Response specificity in the behavioral effects of limbic system lesions in the cat. *Journal of Comparative and Physiological Psychology* **54**: 605–613.

Mingote S, Weber SM, Ishiwari K, Correa M, Salamone JD (2005). Ratio and time requirements on operant schedules: effort-related effects of nucleus accumbens dopamine depletions. *The European journal of neuroscience* **21**: 1749–57.

Miyazaki K, Miyazaki KW, Doya K (2012). The role of serotonin in the regulation of patience and impulsivity. *Molecular neurobiology* **45**: 213–24.

Niv Y, Daw ND, Joel D, Dayan P (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology* **191**: 507–20.

Ossewaarde L, Wingen GA van, Kooijman SC, Bäckström T, Fernández G, Hermans EJ (2011). Changes in functioning of mesolimbic incentive processing circuits during the premenstrual phase. *Social cognitive and affective neuroscience* **6**: 612–20.

Ostlund SB, Wassum KM, Murphy NP, Balleine BW, Maidment NT (2011). Extracellular dopamine levels in striatal subregions track shifts in motivation and response cost during instrumental conditioning. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **31**: 200–7.

Parkinson JA, Dalley JW, Cardinal RN, Bamford A, Fehnert B, Lachenal G, *et al* (2002). Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive Pavlovian approach behaviour: implications for mesoaccumbens dopamine function. *Behavioural brain research* **137**: 149–63.

Salamone JD, Correa M (2002). Motivational views of reinforcement: implications for understanding the behavioral functions of nucleus accumbens dopamine. *Behavioural brain research* **137**: 3–25.

Salamone JD, Correa M (2012). The Mysterious Motivational Functions of Mesolimbic Dopamine. *Neuron* **76**: 470–485.

Salamone JD, Steinpreis RE, McCullough LD, Smith P, Grebel D, Mahan K (1991). Haloperidol and nucleus accumbens dopamine depletion suppress lever pressing for food but increase free food consumption in a novel food choice procedure. *Psychopharmacology* **104**: 515–21.

Salamone JD, Wisniecki A, Carlson BB, Correa M (2001). Nucleus accumbens dopamine depletions make animals highly sensitive to high fixed ratio requirements but do not impair primary food reinforcement. *Neuroscience* **105**: 863–70.

Satoh T, Nakai S, Sato T (2003). Correlated coding of motivation and outcome of decision by dopamine neurons. *Journal of Neuroscience* **23**: 9913–9923.

Schultz W, Dayan P, Montague PR (1997). A neural substrate of prediction and reward. *Science* **275**: 1593.

Segovia KN, Correa M, Salamone JD (2011). Slow phasic changes in nucleus accumbens dopamine release during fixed ratio acquisition: a microdialysis study. *Neuroscience* **196**: 178–88.

Selvaraj S, Turkheimer F, Rosso L, Faulkner P, Mouchlianitis E, Roiser JP, *et al* (2012). Measuring endogenous changes in serotonergic neurotransmission in humans: a [(11)C]CUMI-101 PET challenge study. *Molecular psychiatry* **17**: 1254–60.

Sokolowski JD, Conlan AN, Salamone JD (1998). A microdialysis study of nucleus accumbens core and shell dopamine during operant responding in the rat. *Neuroscience* **86**: 1001–9.

Sokolowski JD, Salamone JD (1994). Effects of dopamine depletions in the medial prefrontal cortex on DRL performance and motor activity in the rat. *Brain Research* **642**: 20–28.

Soubrié P (1986). Reconciling the role of central serotonin neurons in human and animal behavior. *Behavioral and Brain Sciences* **9**: 319–364.

Talmi D, Seymour B, Dayan P, Dolan RJ (2008). Human pavlovian-instrumental transfer. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **28**: 360–8.

Taylor JR, Robbins TW (1986). 6-Hydroxydopamine lesions of the nucleus accumbens, but not of the caudate nucleus, attenuate enhanced responding with reward-related stimuli produced by intra-accumbens d-amphetamine. *Psychopharmacology* **90**: 390–7.

Ungerstedt U (1971). Adipsia and aphagia after 6-hydroxydopamine induced degeneration of the nigro-striatal dopamine system. *Acta physiologica Scandinavica Supplementum* **367**: 95–122.

Westerink B, Kwint H, deVries J (1996). The pharmacology of mesolimbic dopamine neurons: a dual-probe microdialysis study in the ventral tegmental area and nucleus accumbens of the rat brain. *J Neurosci* **16**: 2605–2611.

Winkel J, Maanen L van, Ratcliff R, Schaaf ME van der, Schouwenburg MR van, Cools R, *et al* (2012). Bromocriptine does not alter speed-accuracy tradeoff. *Frontiers in neuroscience* **6**: 126.

Wittmann BC, Schott BH, Guderian S, Frey JU, Heinze H-J, Düzel E (2005). Reward-related FMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron* **45**: 459–67.

FIGURE LEGENDS

Figure 1
Structure of one trial of the behavioral task. Subjects are shown their potential reward, followed by an odd-one-out task to be completed within 500 msec (400 msec for 20% of trials). After a further 500 msec, they were shown their received reward. (B) The induced fluctuation in available reward (blue) and averaged reward (for learning rate $\alpha = 0.012$ in red, the fixed value used in Guitart-Masip et al 2011, and $\alpha = 0.113$ in green, the mean value across subjects found in the current analysis) varying over time.

Figure 2
β values for data sets D_{orig} (blue; Guitart-Masip *et al.* 11), and D_{tired} (red), estimated through the expectation maximization algorithm. Error bars are standard errors and '*' indicates significant difference in means at p<0.05 based on a 2-sample t-test.

Figure 3
β values for subjects given placebo (blue), L-Dopa (red) and citalopram (green) estimated through expectation maximization algorithm. Error bars are standard errors and '*' indicates significant difference in means at p<0.05 based on a 2-sample t-test.

Figure 4
Beta values for average reward rate regressor across conditions and experiments. Error bars are standard errors and '*' indicates significant difference in means at p<0.05 based on a 2-sample t-test.