Dopamine restores reward prediction errors in old age

Rumana Chowdhury[1,2,10], Marc Guitart-Masip[2,3,10], Christian Lambert[2,4], Peter Dayan[5], Quentin Huys[2,5–7], Emrah Düzel[1,8,9] & Raymond J Dolan[2]

Senescence affects the ability to utilize information about the likelihood of rewards for optimal decision-making. Using functional magnetic resonance imaging in humans, we found that healthy older adults had an abnormal signature of expected value, resulting in an incomplete reward prediction error (RPE) signal in the nucleus accumbens, a brain region that receives rich input projections from substantia nigra/ventral tegmental area (SN/VTA) dopaminergic neurons. Structural connectivity between SN/VTA and striatum, measured by diffusion tensor imaging, was tightly coupled to inter-individual differences in the expression of this expected reward value signal. The dopamine precursor levodopa (L-DOPA) increased the task-based learning rate and task performance in some older adults to the level of young adults. This drug effect was linked to restoration of a canonical neural RPE. Our results identify a neurochemical signature underlying abnormal reward processing in older adults and indicate that this can be modulated by L-DOPA.

Aging in humans is associated with a range of changes in cognition. For example, older adults are particularly poor at making decisions when faced with probabilistic rewards, possibly because of impaired learning of stimulus-outcome contingencies[1,2]. Such findings raise two fundamental questions. Namely, what are the substrates for learning in these circumstances and what accounts for this aberrant decision-making?

One function that is critical for decision-making is learning to predict rewards. There is ample evidence from animal experiments that the neuromodulator dopamine encodes the difference between actual and expected rewards (so-called RPEs)[3,4]. In humans, there is compelling evidence that functional activation patterns in the nucleus accumbens, a major target region of dopamine neurons[5], report rewarding outcomes and associated prediction errors[6–9]. A more direct link to dopamine has been shown using pharmacological challenge with dopaminergic agents[10,11].

In terms of what might go wrong during aging, one important clue is the well-described age-related loss of dopamine neurons in the SN/VTA[12,13], evident both in histology and when using diffusion tensor imaging (DTI) as an indirect marker of structural degeneration[14,15]. However, the consequences for decision-making of this decline in dopamine are unclear, as there are functional interactions among the triplet of reward representations, representations of prediction errors associated with that reward and the learning of predictions that underpins the expression of these prediction errors. In older age, abnormal activity in the nucleus accumbens has been associated with suboptimal decision-making and reduced reward anticipation, but also with normal responses to rewarding outcomes[16–18]. This has led to the suggestion that, although older adults may maintain adequate representations of reward, they are unable to learn correctly from these representations.

We studied the effect of probabilistic rewarding outcomes on the separate reward and prediction components of a prediction error signal[19] in healthy older adults. We employed a simple probabilistic instrumental conditioning problem, the two-armed bandit choice task (**Fig. 1a**). Older adults underwent DTI and functional magnetic resonance imaging (fMRI) in combination with a pharmacological manipulation using the dopamine precursor L-DOPA in a within-subject, double-blind, placebo-controlled study. We collected behavioral data in a group of young adults to contextualize the effects of age on performance. We did not administer L-DOPA to these young adults, implying that the effects of L-DOPA could not be compared across age groups. By exploiting a reinforcement learning model, we were able to determine which component of the prediction error (the actual and/or expected reward representation) was impaired in older age. DTI enabled us to examine nigro-striatal structural connectivity strength, based on the hypothesis that individual differences in this structural measure would predict inter-individual differences in baseline functional RPE signaling. Crucially, L-DOPA administration has been associated with greater prediction errors in young adults[10] and higher learning rates in patients with Parkinson's disease[11]. Thus, we predicted that L-DOPA would increase the learning rate evident in behavior as well as boost the representation of an RPE in the nucleus accumbens of healthy older adults, specifically by increasing the component associated with the expected value.

RESULTS
Behavioral performance in young and older adults
We administered placebo and L-DOPA to 32 older adults (age = 70.00 ± 3.24 years, mean ± s.d.; **Supplementary Table 1**) and asked them,

---

[1]Institute of Cognitive Neuroscience, University College London, London, UK. [2]Wellcome Trust Centre for Neuroimaging, University College London, London, UK. [3]Aging Research Center, Karolinska Institute, Stockholm, Sweden. [4]Stroke and Dementia Research Centre, St. George's University of London, London, UK. [5]Gatsby Computational Neuroscience Unit, University College London, London, UK. [6]Translational Neuroimaging Unit, Department of Biological Engineering, ETH Zurich and University of Zurich, Zurich, Switzerland. [7]Department of Psychiatry, Psychotherapy and Psychosomatics, Zurich University Hospital of Psychiatry, Zurich, Switzerland. [8]Otto-von-Guericke-University Magdeburg, Institute of Cognitive Neurology and Dementia Research, Magdeburg, Germany. [9]German Center for Neurodegenerative Diseases (DZNE), Magdeburg, Germany. [10]These authors contributed equally to this work. Correspondence should be addressed to R.C. (rumana.neuro@gmail.com).

Figure 1 Two-armed bandit task design and performance in young and older adults. (**a**) On each trial, participants selected one of two fractal images, which were then highlighted in a red frame. This was followed by an outcome in which a green upward arrow indicated a win of £0.10 and a yellow horizontal bar indicated the absence of a win. If they did not choose a stimulus, the written message "you did not choose a picture" was displayed. The same pair of images was used throughout the task, although their position on the screen (left or right) varied. The task consisted of 220 trials separated into two sessions with a short break in between. Participants' earnings were displayed at the end of the task and given to them at the end of the test day. The probability of obtaining a reward associated with each image varied on a trial-by-trial basis according to a Gaussian random walk. Two different sets of probability distributions (set A and B) were used on the two testing days, counterbalanced across the order of L-DOPA or placebo administration. RT, reaction time. (**b**) Older adults ($n = 32$) in the placebo condition won less money than young adults ($n = 22$). When the same older adults ($n = 32$) received L-DOPA, performance was similar to young adults. *$P < 0.05$. Error bars indicate ±1 s.e.m.
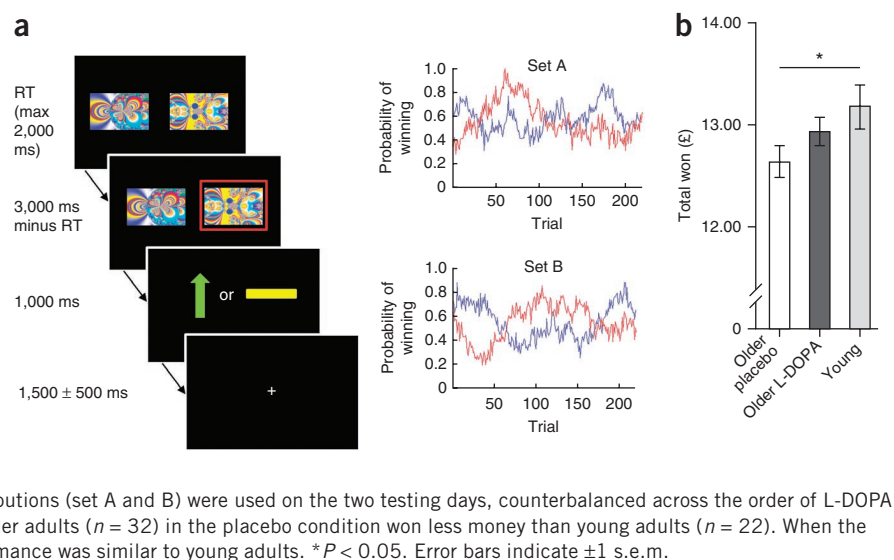
as well as 22 young adults (age = 25.18 ± 3.85 years), to perform a two-armed bandit choice task (**Fig. 1a**). Older adults completed a similar number of trials under both conditions (placebo: 218.16 ± 1.94; L-DOPA: 218.47 ± 1.74) as young adults (218.50 ± 2.44) (all $P > 0.4$). Older adults had similar choice reaction times after placebo (796.81 ± 152.89 ms) and L-DOPA (781.49 ± 140.17 ms) treatment (paired t test, $t_{31} = 1.01$, $P = 0.321$), but were slower overall under both conditions than young adults (629.69 ± 156.41 ms) (independent t tests, young versus old + placebo, $t_{52} = 3.91$; young versus old + L-DOPA, $t_{52} = 3.73$; both $P < 0.0005$).

Overall, the amount of money won by older adults performing the task did not differ following L-DOPA treatment (£12.94 ± 0.81) compared to placebo treatment (£12.64 ± 0.89) (paired t test, $t_{31} = 1.53$, $P = 0.137$). However, older adults on placebo won significantly less money than young adults (£13.17 ± 1.00; independent samples t test, $t_{52} = 2.05$, $P = 0.045$), whereas there was no difference in the amount won between older adults treated with L-DOPA and young adults ($t_{52} = 0.971$, $P = 0.336$) (**Fig. 1b**).

A more detailed examination of the behavioral data showed that only a proportion of older adults won more money on the task under L-DOPA compared to placebo. To examine this further, we performed a median split according to drug-induced changes in performance (Online Methods), creating a 'win less on L-DOPA' group (total won L-DOPA < placebo, $n = 17$) and a 'win more on L-DOPA' group (total won L-DOPA > placebo, $n = 15$). This analysis revealed that performance in older adults was consistent with an inverted U shape, whereby those with high baseline levels of performance on placebo performed less well on L-DOPA and, conversely, those with low baseline levels of performance improved following L-DOPA treatment (**Supplementary Fig. 1**). Performance in the win less on L-DOPA group on placebo and in the win more on L-DOPA group on L-DOPA was at a similar level to performance in young adults (young adults versus win less on L-DOPA group on placebo, $t_{37} = 0.19$, $P = 0.854$; young adults versus win more on L-DOPA group on L-DOPA, $t_{35} = -0.40$, $P = 0.690$), whereas performance in the win less on L-DOPA group on L-DOPA and in the win more on L-DOPA group on placebo was worse than performance in young adults (young adults versus win less on L-DOPA group on L-DOPA $t_{37} = 2.07$, $P = 0.045$; young adults versus win more on L-DOPA on placebo, $t_{35} = 3.53$, $P = 0.001$). This inverted U-shaped pattern of performance is consistent

with previous reports of the effects of dopamine on cognition[20] and suggests that variable performance across older adults is linked to individual differences in baseline dopamine status.

Reinforcement learning behavior

We analyzed trial-by-trial choice behavior using a standard reinforcement learning model with a fixed β parameter (**Fig. 2a**). Note that, by using this methodological approach, the learning rate reflects a summary measure of reinforcement learning strength (Online Methods). A model with a single fixed β = 1.27 across drug and placebo conditions, one single learning rate and one choice perseveration parameter provided the best model fit of older participants' choices among the range of models that we compared, indexed by the lowest Bayesian information criterion (BIC) values (**Supplementary Table 2**). When calculating the BIC, the log evidence was penalized using the number of data points associated with each parameter.

To further examine the effects of L-DOPA on older participants' behavior in the task, we used the Wilcoxon signed rank test to determine whether the learning rates (fitted using a single prior distribution including the drug and the placebo) differed between L-DOPA and placebo. We found that participants had a significantly higher learning rate under L-DOPA than placebo ($Z = -3.03$, $P = 0.002$; **Fig. 2b**). This effect was significant in the group of older adults who performed better under L-DOPA (win more on L-DOPA group, placebo versus L-DOPA: $Z = -2.90$, $P = 0.004$), but not among older adults who performed worse on L-DOPA (win less on L-DOPA group, placebo versus L-DOPA: $Z = -0.97$, $P = 0.332$), providing a direct link between the effects of L-DOPA and task performance (**Fig. 2c**). In contrast, choice perseveration was unaffected by L-DOPA ($Z = -0.58$, $P = 0.562$). In young adults, a model with a fixed β = 1.13 and single learning rate provided a better fit to participants' choices than when a choice perseveration parameter was added to the model (BIC = 4,348.15 and 4,361.01, respectively). The learning rate in young adults (median α = 0.62, range 0.01–0.94) was intermediate between, and not significantly different from, the learning rate of older adults with either placebo ($Z = -1.32$, $P = 0.187$) or L-DOPA ($Z = -1.25$, $P = 0.211$) (**Fig. 2b**).

L-DOPA and striatal prediction errors in older adults

We focused our imaging analysis on within-subject comparisons of reward predictions errors in the nucleus accumbens ($n = 32$ older adults). Using a functional region of interest (ROI) approach
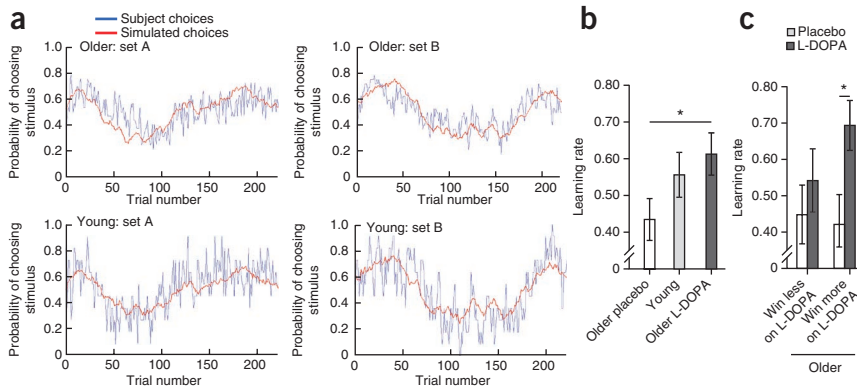
Figure 2 Reinforcement learning model and behavior. (**a**) For young and older adults, the predicted choices from the learning model (red) closely matched subjects' observed choices (blue). The red lines show the same time-varying probabilities, but evaluated on choices sampled from the model (Online Methods). Plots are shown for the two different sets of probability distributions used on the two test days. (**b**) Older adults ($n = 32$) had a higher learning rate under L-DOPA compared with placebo and did not differ from young adults ($n = 22$). *$P < 0.05$, two-tailed. Error bars represent ±1 s.e.m. (**c**) Older adults who won more on L-DOPA than placebo ($n = 15$) had a significantly higher learning rate under L-DOPA than placebo, whereas learning rates did not differ between placebo and L-DOPA for older adults who won less on L-DOPA than placebo ($n = 17$). *$P < 0.05$, two-tailed t test. Error bars represent ±1 s.e.m.

(**Supplementary Fig. 2**), we first defined voxels in the nucleus accumbens that signaled a 'putative' prediction error, namely voxels in which there was an enhanced response at the time of outcome to actual rewards that was greater than that to expected rewards ($R(t) > Q_{a(t)}(t)$; see Online Methods). Using this approach, we identified a cluster in the right nucleus accumbens (peak voxel MNI coordinates: x, y, $z = 15$, 11, -8; peak $Z = 4.45$, $P < 0.001$ uncorrected, 34 voxels; **Fig. 3a**). Note that this is a liberal definition of RPEs, as voxels showing a significant effect with this contrast may not satisfy all of the criteria to be considered for a canonical RPE, namely both a positive effect of reward and a negative effect of expected value[19,21]. We adopted this approach to test the hypothesis that canonical RPEs are not fully represented in old age and to test for the orthogonal effects of L-DOPA on the separate reward and expected value components of the prediction error signal.
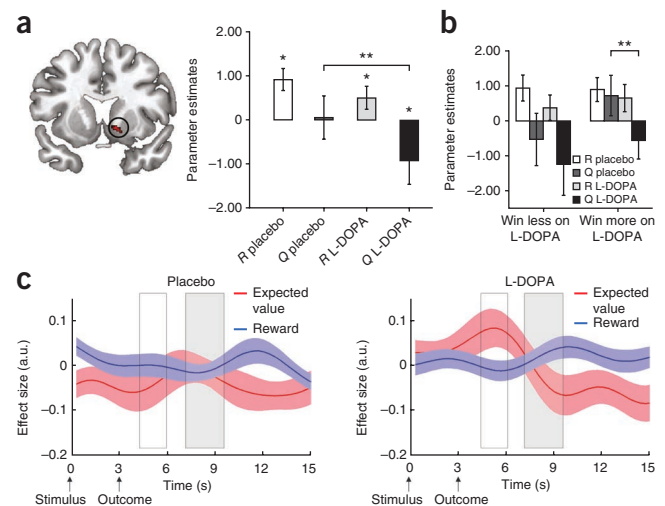
We used this anatomically constrained functional ROI to separately extract the parameter estimates for $R(t)$ and $Q_{a(t)}(t)$ in these activated

voxels. Our two (placebo or L-DOPA) by two ($R(t)$ or $Q_{a(t)}(t)$) repeated-measures ANOVA revealed a main effect of L-DOPA ($F_{1,31} = 5.712$, $P = 0.023$), suggesting administration of L-DOPA had an effect on the representations associated with the components of the RPE (**Fig. 3a**). Notably, blood oxygen level–dependent (BOLD) responses were only compatible with a canonical prediction error signal (positive correlation between BOLD and $R(t)$ along with a negative correlation between BOLD and $Q_{a(t)}(t)$) when participants were under L-DOPA (one-tailed one-sample t test: $R(t)$ L-DOPA, $t = 1.92$, $P = 0.033$; $Q_{a(t)}(t)$ L-DOPA, $t = -1.73$, $P = 0.047$; $R(t)$ placebo, $t = 3.72$, $P < 0.001$; $Q_{a(t)}(t)$ placebo, $t = -0.11$, $P = 0.455$). This was a result of a more negative representation of expected value $Q_{a(t)}(t)$ on L-DOPA compared with placebo (paired t test, $t_{31} = 2.37$, $P = 0.024$), whereas there was no difference in actual reward representation $R(t)$ between L-DOPA and placebo ($t_{31} = 1.38$, $P = 0.179$). These results indicate that canonical RPEs are not fully represented in older adults at baseline, whereby, under placebo, the nucleus accumbens responds to reward and not to expected value. Only after receipt of L-DOPA was a canonical RPE signal observed.

Under placebo, individual differences in the total amount won on the task correlated positively with the learning rate (Spearman's $\rho = 0.39$, $P = 0.027$) and task performance correlated negatively with the BOLD representation of expected value ($Q_{a(t)}(t)$, Pearson's $r = -0.42$, $P = 0.016$), although this was not the case with reward ($R(t)$, Pearson's $r = -0.07$, $P = 0.707$). Thus, better baseline performance was associated with a higher learning rate and more negative expected value representations in the nucleus accumbens. Across all 32 older participants, task performance on L-DOPA did not correlate with the learning rate or BOLD representations of reward or expected value (all $P > 0.15$; **Supplementary Table 3**).

However, subsequent analysis on the basis of a median split for the effects of drug on performance revealed that expected value ($Q_{a(t)}(t)$)

Figure 3 Reward prediction in the nucleus accumbens in 32 older adults. (**a**) A region in the right nucleus accumbens showed greater BOLD activity for reward (R) than for expected value (Q) at the time of outcome (putative RPE). However, the lack of a negative effect of expected value under placebo meant this prediction error signal was incomplete (*$P < 0.05$, one-sample t test, one-tailed). L-DOPA increased the negative effect of expected value (paired t test, **$P < 0.05$, two tailed), resulting in a canonical prediction error signal (both a positive effect of reward and negative effect of expected value). Error bars represent ±1 s.e.m. (**b**) Participants who won more on L-DOPA ($n = 15$) demonstrated a negative effect of expected value under L-DOPA and not under placebo. Reward and expected value parameter estimates did not differ between L-DOPA and placebo for participants who won less on L-DOPA ($n = 17$). **$P < 0.05$, paired t test. Error bars represent ±1 s.e.m. (**c**) Time course plots of the nucleus accumbens BOLD response to reward and expected value. White box corresponds with BOLD responses elicited at the time participants' made a choice; gray box corresponds with BOLD responses elicited when the outcomes were revealed. Under placebo, the only reliable signal observed was a reward response. Under L-DOPA, a canonical RPE was observed, involving a positive expectation of value at the time of the choice together with a positive reward response and a negative expectation of value at the time of the outcome. Reward anticipation (positive effect at the time of the choice) was only observed on L-DOPA. Solid lines are group means of the effect sizes and shaded areas represent ±1 s.e.m.
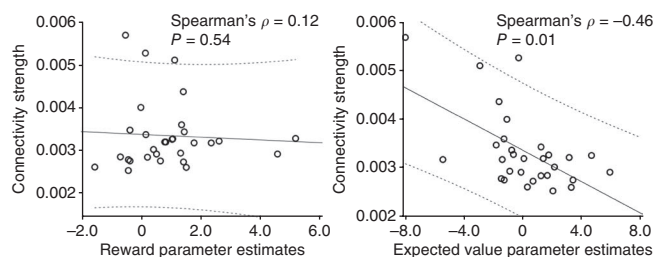
Figure 4 Nigro-striatal tract connectivity strength and functional prediction errors. Under placebo, individuals with higher white matter nigro-striatal tract connectivity strength (determined using DTI) had a more negative effect of expected value, whereas there was no correlation with functional parameters estimates of reward. Each dot on the plots represents one subject ($n = 30$, note two participants are overlapping on the plot on the left), the solid line is the regression slope, and the dashed lines represent 95% confidence intervals.

parameter estimates in older adults who performed better on L-DOPA were significantly more negative on L-DOPA than placebo (win more on L-DOPA group, $Q_{a(t)}(t)$, placebo versus L-DOPA, $t_{14} = 2.26$, $P = 0.040$; **Fig. 3b**). In contrast, L-DOPA did not affect expected value representation in the win less on L-DOPA group ($t_{16} = 1.18$, $P = 0.257$) or reward representation in either the win less on L-DOPA ($t_{16} = 1.56$, $P = 0.137$) or win more on L-DOPA groups ($t_{14} = 0.48$, $P = 0.637$). These results indicate that the restoration of a canonical prediction error signal, mediated by a more negative representation of expected value under L-DOPA, is associated with better task performance.

Although L-DOPA did not affect reward or expected value parameter estimates of those older participants in the win less on L-DOPA group, these participants continued to show a negative BOLD correlate of $Q_{a(t)}(t)$ under L-DOPA even though their performance was worse on L-DOPA (**Fig. 3b**). One possibility was a differential effect of L-DOPA on the noise in the representations of $R(t)$ and $Q_{a(t)}(t)$ for these participants. To address this, we measured individuals' standard error of the parameter estimates on L-DOPA and placebo. We found a significant negative correlation between the drug-induced change in total won on the task and the drug-induced change in the standard error of $R(t)$ (Spearman's $\rho = -0.62$, $P = 0.009$) and $Q_{(a(t)}(t)$ (Spearman's $\rho = -0.61$, $P = 0.009$) only in the win less on L-DOPA group (**Supplementary Fig. 3**). This suggests that, for participants with high baseline levels of performance, L-DOPA increased noise in their reward and expected value representations and this was associated with a worsening in performance. Notably, the increase in noise in the BOLD responses was not related to worse fits of the reinforcement learning models, as mean model likelihood did not differ between groups and did not correlate with standard error of $R(t)$ and $Q_{a(t)}(t)$ (**Supplementary Table 4**).

To visualize the effects of L-DOPA on reward prediction over the course of a trial, we extracted the BOLD time course from the nucleus accumbens functional ROI and performed a regression of this fMRI signal against $R(t)$ and $Q_{a(t)}(t)$. Typically, we would expect to see a pattern of a reward prediction (that is, anticipation) at the time of the choice indicated by a positive effect of $Q_{a(t)}(t)$ and an RPE at the time of the outcome, indicated by both a positive effect of $R(t)$ and negative effect of $Q_{a(t)}(t)$. Our time course analysis revealed exactly this expected pattern, but only in the L-DOPA condition (**Fig. 3c**). Thus, the abnormal response to the expected value observed among older adults on placebo (lack of reward anticipation at the time of the choice and absent negative expectation at the time of the outcome) was restored under L-DOPA. This analysis complements the aforementioned fMRI analysis, which showed that a canonical RPE was only present on L-DOPA, by revealing abnormal expected value representations throughout the course of a trial under placebo.

In addition, we performed further multiple regression analyses across all older adults to identify regions in the brain in which reward, expected value and putative RPEs correlated with task performance (total money won) separately for L-DOPA and placebo conditions. Of note, only a model examining negative correlations between expected value and performance identified regions that survived family-wise error whole-brain correction (**Supplementary Fig. 4**). We found a left superior parietal cluster in the placebo condition ($Z = 5.06$, peak voxel MNI coordinates: −26, −78, 50), and left inferior parietal ($Z = 5.34$, peak voxel MNI coordinates: −48, −49, 48) and right precuneus clusters ($Z = 5.02$, peak voxel MNI coordinates: 12, −72, 59) in the L-DOPA condition. This suggests that extra-striatal regions also influenced task performance, whereby individuals with a more negative representation of expected value in parietal regions won more money on the task.

Anatomical connectivity and RPEs

Our analysis identified substantial inter-individual variability among older adults for both reward and expected value representations in the nucleus accumbens at baseline (that is, under placebo; **Supplementary Fig. 5**), whereby the latter was associated with task performance. We hypothesized that this might be associated with the known variability in the age-related decline of dopamine neurons from the SN/VTA, which may, in principle, be indexed through anatomical nigro-striatal connectivity. Using DTI and probabilistic tractography ($n = 30$ older adults), we defined a measure of connection strength between the right SN/VTA and right striatum (**Supplementary Fig. 6** and Online Methods). Nigro-striatal tract connectivity strength measured with DTI correlated with the fMRI parameter estimate under placebo associated with expected value ($Q_{a(t)}(t)$) (Spearman's $\rho = -0.46$, $P = 0.010$), but not with that associated with reward ($R(t)$) (Spearman's $\rho = 0.12$, $P = 0.54$) (**Fig. 4**). These correlations were significantly different from each other, suggesting that individual functional activation differences of the representation of expected value, but not reward, were linked to anatomical connectivity strength between the SN/VTA and striatum (Fisher's r-to-z transformation, $z = -2.32$, $P = 0.002$). This relationship between greater tract connectivity strength and more negative expected value parameter estimates remained significant after controlling for age, gender, total intracranial volume, size of the seed region from which tractography was performed and global white matter integrity indexed by fractional anisotropy (partial Spearman's $\rho = -0.44$, $P = 0.027$). There was no difference in this correlation between subgroups of older adults (win more on L-DOPA group, $n = 14$, $\rho = -0.54$, $P = 0.047$; win less on L-DOPA group, $n = 16$, $\rho = -0.37$, $P = 0.154$; Fisher's r-to-z transformation comparing both groups, $z = 0.53$, $P = 0.596$; **Supplementary Fig. 7**). Neither fractional anisotropy values of SN/VTA nor nucleus accumbens functional ROI correlated with expected value (Pearson's $r = 0.26$ and $r = 0.17$, $P = 0.16$ and $P = 0.38$, respectively), suggesting that this correlation was related to circuit strength rather than to local structural integrity as determined by fractional anisotropy.

Older participants with equivalent baseline performance levels to young adults (win less on L-DOPA group on placebo) had stronger connectivity between SN/VTA and the striatum than older participants with worse baseline performance than young adults (win more on L-DOPA group on placebo; between groups comparison $t_{29} = 2.40$, $P = 0.023$). This suggests that these older individuals had higher baseline integrity of the nigro-striatal dopamine circuit than older adults with lower baseline levels of performance.

DISCUSSION

We used a probabilistic reinforcement learning task in combination with a pharmacological manipulation of dopamine, as well as structural

and functional imaging, to probe reward-based decision-making in old age. Overall, older adults had an incomplete RPE signal in the nucleus accumbens consequent on a lack of a neuronal response to expected reward value. Baseline inter-individual differences of the expression of expected value were linked to performance and tightly coupled to nigro-striatal structural connectivity strength, determined using DTI. L-DOPA increased the task-based learning rate and modified the BOLD representation of expected value in the nucleus accumbens. Notably, this effect was only observed for those participants that showed a substantial drug-induced improvement on task performance.

Previous studies have shown that older adults perform worse on probabilistic learning tasks than their younger counterparts[2,22,23]. As it is widely held that dopamine neurons encode an RPE signal, it is conceivable that a dopamine decline, occurring as part of the normal aging process, could account for these behavioral deficits. Indeed, this was a prime motivation for our use of L-DOPA. Although there was no significant difference in task performance in older adults as a group on placebo versus L-DOPA, we found that older adults with low baseline levels of performance improved following L-DOPA treatment. Using a reinforcement-learning model, we found that those older adults who performed better under L-DOPA had a higher learning rate on L-DOPA than on placebo. This is consistent with findings in patients with Parkinson's disease (a dopamine deficit disorder) whose learning rates when on dopaminergic medication are higher than when off their medication, albeit, in this instance, without any significant difference in overall performance[11]. As in that study[11], it is impossible to make a definitive distinction between learning rate, the magnitude of the prediction error that arises from learning and the stochastic way that learning leads to choice.

There are two important points in each trial at which a temporal difference error type signal can be anticipated, namely at choice, when the temporal difference error is the expected value of the chosen option, and at the time of outcome, when the temporal difference error is the difference between the reward actually provided and the expected value. Decomposing the outcome signal into these separate positive and negative components is important because the response to reward is highly correlated with the full prediction error, potentially readily confusing the two[19,21,24]. Overall, in our experiment, under placebo, although the representation of the actual reward appeared to be normal, neither of the components of the expected value signal at choice or outcome was present in nucleus accumbens BOLD signal. This absence is consistent with the few behavioral[23] and neuroimaging studies[16,17] that have suggested that older adults, on average, have abnormal expected value representations, although it is important to note that we did not find a substantial behavioral impairment. Notably, we found that, under L-DOPA, both components of the expected value signal were restored. However, a closer inspection, taking individual differences in drug-induced effects on performance into account, revealed that this was only the case for those older adults whose performance improved under L-DOPA.

There are at least two possible explanations for the absence of the expected value signal. One is that a putative model-free decision-making system, closely associated with neuromodulatory effects[3,25], is impaired. This would render reward-based behavior subject to the operation of a model-based system, which is thought to be less dependent on dopaminergic transmission[26]. This possibility is supported by evidence that older adults perform better than younger adults in tasks requiring a model of the environment (for example, where future outcomes are dependent on previous choices)[27]. Reconciling it with the observation that suppressing[28] or boosting[29]

dopamine in healthy young volunteers suppresses or boosts, respectively, model-based over model-free control is more of a challenge. In relation to this point, we identified two parietal clusters where expected value representations correlated with task performance in the L-DOPA condition. Notably, these clusters overlap with regions purported to signal state prediction errors[30]. One possibility is that these regions may be a neural signature of model-based calculations, which have also recently been shown to be enhanced by L-DOPA in young participants[29]. Although previous studies have shown dopaminergic modulation of value representations in the prefrontal cortex[31], we did not find strong evidence for the involvement of any other extra-striatal regions implicated in the effects of L-DOPA on reward processing in our sample of older adults. However, L-DOPA may have also influenced other extra-striatal learning mechanisms in our task. For example, episodic learning mediated by the hippocampus has also been linked to the dopaminergic system[32] and could support aspects of rapid learning when it occurs.

Another possibility for the absence of a model-free expected value signal is that it is still calculated normally, but that when dopamine levels are low, it is not manifest in nucleus accumbens BOLD signal. One can reasonably expect that dopamine levels will affect the state of striatal neurons[33]. However, dopamine effects on local activity in the striatum as well as on the BOLD signal of cortical and dopaminergic inputs to the striatum remain unclear. In the future, it would be interesting to use procedures based on recent studies (for example, see refs. 34,35) in older participants with and without L-DOPA to investigate the balance of model-free and model-based control.

Enriching the above picture are recent studies in healthy young participants showing that at least some aspects of the representation in striatal BOLD of the expected value component of the temporal difference error are conditional on a requirement for action. In one such study, the representation of expected value in young adults was not modulated by L-DOPA[36]. However, it is not clear whether this is an effect of the more extensive training provided in that study (which can render behaviors insensitive to dopamine manipulations[37]) or that the expected value did not fluctuate in a way that was relevant for choice. Together with recent findings[38], these results raise the possibility that dopamine might only modulate the neural representation of expected value when it is behaviorally relevant for the task at hand.

Our DTI connectivity analysis supports the notion that neuronal representations of expected value, and thus appropriate RPE signaling, rely on the integrity of the dopaminergic system. The connectivity strength of tracts is one DTI metric that has been reported to predict age-related performance differences[39,40]. Notably, older adults who performed the task under placebo as well as young adults had higher nigro-striatal connectivity strength than older adults with lower baseline levels of task performance. Furthermore, older individuals with stronger connectivity between SN/VTA and striatum had more robust value representations in the nucleus accumbens. Although our findings can be interpreted in the context of a well-defined decline of nigro-striatal dopamine neurons with increasing age[12,13], we acknowledge that DTI measures of connectivity are not a direct mapping of dopamine neurons, but instead reflect white matter tract strength between the SN/VTA and striatum. In addition, the direction of information flow cannot be inferred from DTI-based tractography[41]. We did not observe a relationship between fractional anisotropy of either the SN/VTA or striatum with functional activity in the accumbens. Fractional anisotropy values characterize the extent of water diffusion, thereby providing an indirect measure of myelin, axons and the structural organization of both gray and white matter[15,42]. Our results are therefore an indication that inter-individual anatomical

differences at the level of nigro-striatal circuit strength, rather than local gray-matter integrity in SN/VTA or striatum, determine the success of prediction error signaling in healthy older adults.

In summary, our findings suggest that a subgroup of older adults who underperform at baseline can show a drug-induced improvement in task performance. For these older adults, L-DOPA increased a task-based learning rate and led to a canonical RPE signal by restoring the representation of expected value in the nucleus accumbens. On the other hand, participants that performed better on the task under placebo (that is, on a par with young controls) had a greater representation of expected value in the striatum and stronger nigro-striatal connectivity, suggesting higher baseline dopamine status. After receiving L-DOPA, their performance decreased, perhaps because of increased noise in the representations of RPEs. One possibility is that, in these participants, the administration of L-DOPA overdoses the system, an interpretation that is consistent with both a previously described inverted U-shape (that is, nonlinear dose dependent) effect of dopamine on cognition[20,43] and a variable dopamine decline among older adults. By establishing a link between dopaminergic signaling in the nucleus accumbens and the representations of expected value in the brain, our results provide a potential therapeutic route for tackling age-related impairments in decision-making.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

AUTHOR CONTRIBUTIONS
R.C. and M.G.-M. conducted the experiment, analyzed the data and prepared the manuscript. C.L., P.D., Q.H. and E.D. contributed to data analysis and manuscript preparation. R.J.D. contributed to data analysis and manuscript preparation and supervised the project.

1. Eppinger, B., Hämmerer, D. & Li, S.-C. Neuromodulation of reward-based learning and decision making in human aging. *Ann. NY Acad. Sci.* **1235**, 1–17 (2011).
2. Mell, T. *et al.* Effect of aging on stimulus-reward association learning. *Neuropsychologia* **43**, 554–563 (2005).
3. Schultz, W., Dayan, P. & Montague, P.R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
4. Salamone, J.D., Correa, M., Mingote, S.M. & Weber, S.M. Beyond the reward hypothesis: alternative functions of nucleus accumbens dopamine. *Curr. Opin. Pharmacol.* **5**, 34–41 (2005).
5. Haber, S.N., Fudge, J.L. & McFarland, N.R. Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *J. Neurosci.* **20**, 2369–2382 (2000).
6. O'Doherty, J. *et al.* Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* **304**, 452–454 (2004).
7. O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H. & Dolan, R.J. Temporal difference models and reward-related learning in the human brain. *Neuron* **38**, 329–337 (2003).
8. Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B. & Dolan, R.J. Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876–879 (2006).
9. Knutson, B. & Gibbs, S. Linking nucleus accumbens dopamine and blood oxygenation. *Psychopharmacology (Berl.)* **191**, 813–822 (2007).
10. Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J. & Frith, C.D. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* **442**, 1042–1045 (2006).
11. Rutledge, R.B. *et al.* Dopaminergic drugs modulate learning rates and perseveration in parkinson's patients in a dynamic foraging task. *J. Neurosci.* **29**, 15104–15114 (2009).
12. Bäckman, L., Nyberg, L., Lindenberger, U., Li, S.-C. & Farde, L. The correlative triad among aging, dopamine and cognition: current status and future prospects. *Neurosci. Biobehav. Rev.* **30**, 791–807 (2006).
13. Düzel, E., Bunzeck, N., Guitart-Masip, M. & Duzel, S. Novelty-related motivation of anticipation and exploration by dopamine (NOMAD): implications for healthy aging. *Neurosci. Biobehav. Rev.* **34**, 660–669 (2010).
14. Fearnley, J.M. & Lees, A.J. Ageing and Parkinson's disease: substantia nigra regional selectivity. *Brain* **114**, 2283–2301 (1991).
15. Vaillancourt, D.E., Spraker, M.B., Prodoehl, J., Zhou, X.J. & Little, D.M. Effects of aging on the ventral and dorsal substantia nigra using diffusion tensor imaging. *Neurobiol. Aging* **33**, 35–42 (2012).
16. Samanez-Larkin, G.R., Kuhnen, C.M., Yoo, D.J. & Knutson, B. Variability in nucleus accumbens activity mediates age-related suboptimal financial risk taking. *J. Neurosci.* **30**, 1426–1434 (2010).
17. Schott, B.H. *et al.* Ageing and early-stage Parkinson's disease affect separable neural mechanisms of mesolimbic reward processing. *Brain* **130**, 2412–2424 (2007).
18. Cox, K.M., Aizenstein, H.J. & Fiez, J.A. Striatal outcome processing in healthy aging. *Cogn. Affect. Behav. Neurosci.* **8**, 304–317 (2008).
19. Behrens, T.E.J., Hunt, L.T., Woolrich, M.W. & Rushworth, M.F.S. Associative learning of social value. *Nature* **456**, 245–249 (2008).
20. Cools, R. & D'Esposito, M. Inverted U-shaped dopamine actions on human working memory and cognitive control. *Biol. Psychiatry* **69**, e113–e125 (2011).
21. Li, J. & Daw, N.D. Signals in human striatum are appropriate for policy update rather than value prediction. *J. Neurosci.* **31**, 5504–5511 (2011).
22. Eppinger, B., Kray, J., Mock, B. & Mecklinger, A. Better or worse than expected? Aging, learning and the ERN. *Neuropsychologia* **46**, 521–539 (2008).
23. Samanez-Larkin, G.R., Wagner, A.D. & Knutson, B. Expected value information improves financial risk taking across the adult life span. *Soc. Cogn. Affect. Neurosci.* **6**, 207–217 (2011).
24. Guitart-Masip, M. *et al.* Go and no-go learning in reward and punishment: interactions between affect and effect. *Neuroimage* **62**, 154–166 (2012).
25. Dickinson, A. & Balleine, B. *The Role of Learning in the Operation of Motivational Systems* (John Wiley & Sons, New York, 2002).
26. Dickinson, A., Smith, J. & Mirenowicz, J. Dissociation of Pavlovian and instrumental incentive learning under dopamine antagonists. *Behav. Neurosci.* **114**, 468–483 (2000).
27. Worthy, D.A., Gorlick, M.A., Pacheco, J.L., Schnyer, D.M. & Maddox, W.T. With age comes wisdom. *Psychol. Sci.* **22**, 1375–1380 (2011).
28. de Wit, S. *et al.* Reliance on habits at the expense of goal-directed control following dopamine precursor depletion. *Psychopharmacology (Berl.)* **219**, 621–631 (2012).
29. Wunderlich, K., Smittenaar, P. & Dolan, R.J. Dopamine enhances model-based over model-free choice behavior. *Neuron* **75**, 418–424 (2012).
30. Gläscher, J., Daw, N., Dayan, P. & O'Doherty, J.P. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* **66**, 585–595 (2010).
31. Jocham, G., Klein, T.A. & Ullsperger, M. Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. *J. Neurosci.* **31**, 1606–1613 (2011).
32. Shohamy, D. & Adcock, R.A. Dopamine and adaptive memory. *Trends Cogn. Sci.* **14**, 464–472 (2010).
33. Nicola, S.M., Surmeier, D.J. & Malenka, R.C. Dopaminergic modulation of neuronal excitability in the striatum and nucleus accumbens. *Annu. Rev. Neurosci.* **23**, 185–215 (2000).
34. Daw, N.D. & Gershman, S.J. Seymour, B., Dayan, P. & Dolan, R.J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69**, 1204–1215 (2011).
35. Simon, D.A. & Daw, N.D. Neural correlates of forward planning in a spatial decision task in humans. *J. Neurosci.* **31**, 5526–5539 (2011).
36. Guitart-Masip, M. *et al.* Action controls dopaminergic enhancement of reward representations. *Proc. Natl. Acad. Sci. USA* **109**, 7511–7516 (2012).
37. Choi, W.Y., Balsam, P.D. & Horvitz, J.C. Extended habit training reduces dopamine mediation of appetitive response expression. *J. Neurosci.* **25**, 6729–6733 (2005).
38. Klein-Flügge, M.C., Hunt, L.T., Bach, D.R., Dolan, R.J. & Behrens, T.E.J. Dissociable reward and timing signals in human midbrain and ventral striatum. *Neuron* **72**, 654–664 (2011).
39. Coxon, J.P., Van Impe, A., Wenderoth, N. & Swinnen, S.P. Aging and inhibitory control of action: cortico-subthalamic connection strength predicts stopping performance. *J. Neurosci.* **32**, 8401–8412 (2012).
40. Forstmann, B.U. *et al.* Cortico-subthalamic white matter tract strength predicts interindividual efficacy in stopping a motor response. *Neuroimage* **60**, 370–375 (2012).
41. Le Bihan, D. & Johansen-Berg, H. Diffusion MRI at 25: exploring brain tissue structure and function. *Neuroimage* **61**, 324–341 (2012).
42. Draganski, B. *et al.* Regional specificity of MRI contrast parameter changes in normal ageing revealed by voxel-based quantification (VBQ). *Neuroimage* **55**, 1423–1434 (2011).
43. Li, S.-C. & Sikström, S. Integrative neurocomputational perspectives on cognitive aging, neuromodulation, and representation. *Neurosci. Biobehav. Rev.* **26**, 795–808 (2002).

ONLINE METHODS

Older subjects. 32 healthy adults aged 65–75 years participated in the study (**Supplementary Table 1**). Written informed consent was obtained from all participants. The study received ethical approval from the North West London Research Ethics Committee 2. Four participants experienced side effects (emesis) from L-DOPA administration. These participants remained in all analyses as they vomited more than 2.5 h after L-DOPA ingestion, well after completion of the task, and they did not feel unwell when performing the task in the scanner.

Young subjects. 22 healthy young adults (25.18 ± 3.85 years, mean ± s.d., 12 females) were recruited via the University College London subject pool and word of mouth. Participants were screened to ensure they were healthy, with no history of neurological or psychiatric disorders, no medications, no recent illicit drug use and no recent participation in other research studies involving medication. These subjects performed the task on a laptop and did not undergo MRI scanning or pharmacological manipulation.

Study procedure. This was a double-blind, within-subject, placebo-controlled study. Older participants attended on two occasions, 1 week apart, and performed the same task on both days, 60 min after ingestion of either levodopa (150 mg of levodopa + 37.5 mg of benserazide mixed in orange juice, L-DOPA) or placebo (orange juice alone), the order of which was randomized and counterbalanced. Benserazide promotes higher levels of dopamine in the brain while minimizing peripheral side effects such as nausea and vomiting. To achieve comparable drug absorption across individuals, subjects were instructed not to eat for up to 2 h before commencing the study. Repeated physiological measurements (blood pressure and heart rate) and subjective mood rating scales were recorded under placebo and L-DOPA (**Supplementary Table 5**). After completing the task, participants performed an unrelated episodic memory task on both days and had DTI scanning on just one of the days.

Two-armed bandit task. All participants performed a two-armed bandit task (**Fig. 1a**). Participants were given written and verbal instructions and undertook five practice trials before pharmacological manipulation. The probabilities of obtaining a reward for each stimulus were independent of each other and varied on a trial-to-trial basis according to a Gaussian random walk, generated using a previously described procedure[8]. Different pairs of fractal images were used on the 2 d of testing and randomly assigned among participants.

Reinforcement learning models. We fitted choice behavior to a standard reinforcement learning model on a trial-by-trial basis. This involves $Q_a(t)$ values for each action $a \in \{0,1\}$ on trial t, which are updated if the subject chooses action $a(t)$ as

$$
\begin{aligned}
Q_{a(t)}(t+1) &= Q_{a(t)}(t) + \alpha\delta(t) \\
\delta(t) &= R(t) - Q_{a(t)}(t)
\end{aligned}
$$

Here, $Q_{a(t)}(t)$ is the expected value of the chosen option, which was set to zero at the beginning of the experiment. $\delta(t)$ is the RPE, which represents the difference between the actual outcome $R(t)$ and the expected outcome $Q_{a(t)}(t)$, where $R(t)$ was one (win) or zero (no win). The free parameter α defined subjects' learning rate, with higher values reflecting greater weight being given to more recent outcomes and leading to a more rapid updating of expected value.

As standard, we used a softmax rule to determine the probability of choosing between the two stimuli on trial t. If $m_a(t)$ are the propensities for doing action a on trial t, this uses

$$
P(a(t) = a) = \frac{\exp[\beta m_a(t)]}{\exp[\beta m_0(t)] + \exp[\beta m_1(t)]}
$$

in which the inverse temperature parameter β indexes how deterministic choices were.

We consider two cases for $m_a(t)$. The simplest makes $m_a(t) = Q_a(t)$. However, it is often found that subjects have a tendency either to repeat or avoid doing the same action twice[44]. To account for this, we also considered a model in which $m_a(t) = Q_a(t) + b\chi_{a=a(t-1)}$, allowing an extra boost or suppression b associated with the action performed on the previous trial. We fit all sessions (L-DOPA and placebo) for each participant using expectation maximization in a hierarchical random effects model.

It has previously been noted that it can be hard to infer both α and β independently of each other[11,44], as it is their product that dominates behavior in certain regimes of learning. We therefore adopted the strategy of first fitting a full random effects model as if they were independent, and then clamping β to the mean of its posterior distribution and re-inferring α using the random effects model. Among other things, this limits any strong claims about having inferred differences in true learning rates. Thus, when we describe differences in the learning rate, we acknowledge the possible contribution of both α and β.

In a second step, we used the mean posterior β parameter at the group level obtained on the preceding step (single fixed $\beta = 1.27$ for older adults; single fixed $\beta = 1.13$ for young adults; note that data for young and older adults were analyzed separately) as a fixed parameter in two, nested, reinforcement learning models reflecting the two possibilities for $m_a(t)$. The first has one parameter, the learning rate α. The second has the learning rate α and the perseveration/alternation parameter b.

For older adults only, we then repeated the two steps described above, but instead estimated two separate β terms for the L-DOPA and placebo conditions. We then fixed each β at their respective posterior group means ($\beta = 1.43$ and 1.10 for L-DOPA and placebo, respectively) and proceeded as before to test the two models outlined above.

Model fitting procedure and comparison. For older adults, we first compared the two full random effects models to choose between a model with one or with two separate β terms, and we compared the two nested reinforcement learning models described above with a single fixed β. For completeness, we also compared the same two reinforcement learning models with two fixed β values (**Supplementary Table 2**). For young adults, we compared the two reinforcement learning models with a single fixed β, as young adults did not undergo pharmacological manipulation. Procedures for fitting the models were identical to those used previously[24,45].

Behavioral analysis. We analyzed task performance (amount of money won) using two-tailed paired t tests (L-DOPA versus placebo in older adults) and independent t tests (young versus old). Reinforcement learning model parameters (learning rate and perseveration) were not normally distributed (Shapiro-Wilk and Kolmogorov-Smirnov, $P < 0.05$). Thus, we used two-tailed Wilcoxon signed ranks tests to compare these parameters between the L-DOPA and placebo conditions. Two-tailed Pearson's and Spearman's correlations were used to analyze normally distributed and non-normally distributed data, respectively.

We performed a median split according to difference in performance (total won on L-DOPA minus total won on placebo) among older adults (**Supplementary Fig. 1**). This resulted in a group who won less on L-DOPA (total won L-DOPA < placebo, $n = 17$) and a group who won more on L-DOPA (total won L-DOPA > placebo, $n = 15$).

Image acquisition. All MRI images were acquired using a 3.0T Trio MRI scanner (Siemens) using a 32-channel head coil. Functional data using echo-planar imaging was acquired on 2 d. On each day, scanning consisted of two runs, each containing 194 volumes (matrix 64×74, 48 slices per volume, image resolution = $3 \times 3 \times 3$ mm, repetition time = 70 ms, echo time = 30 ms). Six additional volumes at the beginning of each series were acquired to allow for steady state magnetization and were subsequently discarded. Individual field maps were recorded using a double-echo fast low-angle shot (FLASH) sequence (matrix size = 64×64, 64 slices, spatial resolution = $3 \times 3 \times 2$ mm, gap = 1 mm, short echo time = 10 ms, long echo time = 12.46 ms, repetition time = 1,020 ms) and estimated using the FieldMap toolbox for distortion correction of the acquired EPI images. A structural multi-parameter map protocol employing a three-dimensional multi-echo FLASH sequence at 1-mm isotropic resolution was used to acquire magnetization transfer–weighted (echo time = 2.2–14.7 ms, repetition time = 23.7 ms, flip angle = 6 degrees) and T1-weighted (echo time = 2.2–14.7 ms, repetition time = 18.7 ms, flip angle = 20 degrees) images. A double-echo FLASH sequence (echo time 1 = 10 ms, echo time 2 = 12.46 ms, $3 \times 3 \times 2$-mm resolution and 1-mm gap) and B1 mapping (echo time = 37.06 and 55.59 ms, repetition time = 500 ms, 4-mm³ isotropic resolution) were acquired[46]. Diffusion-weighted images were acquired using spin-echo echoplanar imaging, with twice refocused diffusion encoding to reduce eddy current–induced distortions[47]. We acquired 75 axial slices (whole brain to mid-pons) in an interleaved order

(1.7-mm isotropic resolution, image matrix = 96 × 96, field of view = 220 × 220 mm², slice thickness = 1.7 mm with no gap between slices, repetition time = 170 ms, echo time = 103 ms, asymmetric echo shifted forward by 24 phase-encoding lines, readout bandwidth = 2,003 Hz per pixel) for 61 images with unique diffusion encoding directions. The first seven reference images were acquired with a b value of 100 s mm^{-2} (low b images) and the remaining 61 images with a b value of 1,000 s mm^{-2} (ref. 48). Two DTI sets were acquired with identical parameters, except that the second was acquired with a reversed k-space readout direction allowing removal of susceptibility artifacts post-processing[49]. Given that the SN/VTA was a major region of interest, we optimized the quality of our images by using pulse gating to minimize pulsation artifact in the brainstem.

fMRI data analysis. Data were analyzed using SPM8 (Wellcome Trust Centre for Neuroimaging). Pre-processing included bias correction, realignment, unwarping using individual field maps, co-registration and spatial normalization to the MNI space using diffeomorphic registration algorithm (DARTEL)[50] with spatial resolution after normalization of 2 × 2 × 2 mm. Data were smoothed with a 6-mm FWHM Gaussian kernel. The fMRI time series data were high-pass filtered (cutoff = 128 s) and whitened using an AR(1) model. For each subject, a statistical model was computed by applying a canonical hemodynamic response function combined with time and dispersion derivatives.

Our study builds on a wealth of pre-existing literature describing RPEs in healthy young adults performing probabilistic learning tasks, whereby the striatum signals the difference between actual and expected rewards (for example, see refs. 7,51). A region showing a canonical RPE should show a positive correlation with reward ($R(t)$) and a negative correlation with expected value ($Q_{a(t)}(t)$). This is different from a region showing a 'putative' prediction error, which shows a correlation with [$R(t) - Q_{a(t)}(t)$]. Studies in young participants have shown a canonical RPE signal with both a positive response to reward and negative response to expected value in value learning tasks[19]. Separating the RPE into its components has often not been done, although recent studies have shown that, as $R(t)$ is highly correlated with the RPE, a correlation between BOLD signal and $R(t) - Q_{a(t)}(t)$ may lead to false positive results suggesting that areas whose BOLD signal only correlates with $R(t)$ may be thought of as representing an RPE. This is an important distinction that allows us to determine whether, as a group, healthy older adults do not represent all components of a canonical RPE signal, and the extent to which the two components ($R(t)$ and $Q_{a(t)}(t)$) may be related to task performance, the integrity of anatomical connectivity between the dopaminergic midbrain and the striatum, and modulation by L-DOPA.

The general linear model for each subject at the first level consisted of regressors at the time of stimulus display separately for when a choice was made, when no choice was made and at the time of stimulus outcome. All trials in which participants made a response and obtained an outcome were included in the analysis. BOLD responses to outcomes were parametrically modeled with two separate parametric regressors ($R(t)$ and $Q_{a(t)}(t)$); $R(t)$ was a binary regressor with value 0 when no reward was obtained and value 1 when a reward was obtained, and $Q_{a(t)}(t)$ included the expected value on that trial and was built using the group posterior mean α distribution from the winning model. Thus, a brain region that correlates with $R(t)$ indicates that BOLD responses are higher after a reward is obtained when compared to no reward. Separate design matrices were calculated for the L-DOPA and placebo conditions. To capture residual movement-related artifacts, six covariates were included (the three rigid-body translation and three rotations resulting from realignment) as regressors of no interest. Finally, we also included 18 regressors for cardiac and respiratory phases to correct for physiological noise.

At the first level, we implemented the contrasts $R(t)$ and $Q_{a(t)}(t)$, which are the individual components of the RPE, and the putative RPE $R(t) > Q_{a(t)}(t)$. At the second level, we first defined a functional ROI with the contrast $R(t) > Q_{a(t)}(t)$ collapsed across L-DOPA and placebo conditions. We used an uncorrected threshold of $P < 0.001$ to produce a whole-brain statistical parametric map of regions encoding putative RPEs from which we identified a region in the right nucleus accumbens. We used a subject-derived anatomical mask (**Supplementary Fig. 2**) to constrain this functional ROI.

We examined the effects of placebo and L-DOPA on the reward ($R(t)$) and expected value ($Q_{a(t)}(t)$) components of the canonical RPE signal in the functional ROI. Here we used the Marsbar toolbox[52] to extract the parameter estimates

from the functional ROI to enter into a two ($R/Q_{a(t)}$) by two (L-DOPA/placebo) repeated-measures ANOVA. We conducted *post hoc* tests to characterize the impairment in expected value representation (one-tailed one-sample t tests for each condition to test the null hypothesis that they are not different from zero, and two-tailed paired t tests to compare the effect of L-DOPA to placebo). Furthermore, for each participant, we measured the standard error of the parameter estimates for reward ($R(t)$) and expected value ($Q_{a(t)}(t)$) on L-DOPA and placebo and calculated the drug-induced change (L-DOPA minus placebo). We performed two-tailed Spearman correlations between these measures and the drug-induced change in task performance (total won L-DOPA minus placebo; **Supplementary Fig. 3**).

For completeness, we also performed a whole-brain voxel-based analysis of drug effects (L-DOPA > placebo) across participants using the contrasts defined at the first level for reward, expected value and the putative RPE (no regions survived whole-brain family-wise error correction; **Supplementary Table 6**). We also performed separate multiple regression analyses for the same contrasts on the placebo and L-DOPA conditions separately using task performance (total won) as a between-subjects regressor.

Time course extraction. The main aim of this analysis was to visualize the effect of reward and expected value on the BOLD signal, at the time of the choice and at the time of the outcome, from the nucleus accumbens functional ROI over the course of a trial. In the fMRI SPM analysis, it was not possible to simultaneously test for the effects of value expectation on the choice and the outcome phases. This is because the time of the choice and the time of the outcome were very close together in time (3 s apart) and including the same parametric modulator on both time points would have resulted in highly correlated regressors. Thus, although the SPM model included regressors at the time of the choice and time of the outcome, we only included parametric modulators at the time of the outcome, thereby focusing on just outcome prediction errors.

Time courses were extracted from preprocessed data in MNI space. We upsampled the extracted BOLD signal to 100 ms. The signal was divided into trials and resampled to a duration of 15 s with the onset (presentation of the stimuli) occurring at 0 s, the time of the choice occurring between 0–2 s and the time of the outcome at 3 s. We then estimated a general linear model across trials at every time point in each subject independently, where reward and expected value were the regressors of interest. These regressors were not orthogonalized and therefore competed for variance, which is a particularly stringent test[19]. We calculated group mean effect sizes at each time point and their standard errors, plotted separately for the placebo and L-DOPA conditions.

DTI connectivity strength analysis. One participant was unable to tolerate scanning; thus, DTI data was collected from 31 older individuals. DTI tractography and generation of relative connectivity strength maps was performed as described previously[53] (**Supplementary Fig. 6**). The aim of our tractography analysis was to determine whether inter-individual differences in nigro-striatal connectivity influenced the observed baseline variability in functional prediction error signaling (**Supplementary Fig. 5**). We used Spearman's correlations to relate connectivity strength to prediction error signaling ($R(t)$ and $Q_{a(t)}(t)$ parameter estimates from the functional nucleus accumbens ROI)[54]. To identify outliers, we converted connectivity strength to z scores and used the conventional definition of $z < -3$ or $z > 3$. Although none of our participants were outside this range, one participant had a z score of 2.83 (equivalent to connectivity strength = 0.006) and was therefore excluded from the reported results (therefore reported DTI results are for $n = 30$). Even so, including this potential outlier in the analysis did not change the results. We performed partial Spearman's correlations with the following covariates: age, gender, total intracranial volume, size of the manually defined seed (right SN/VTA) region and global white matter integrity. Note that the size of the target region was not included, as this was the same for all participants. Global white matter integrity was measured by segmenting fractional anisotropy maps and calculating mean fractional anisotropy values of the white matter fractional anisotropy maps. As a control, we performed additional Pearson's correlations between fractional anisotropy values of the SN/VTA seed or nucleus accumbens functional ROI (fractional anisotropy values were normally distributed) and $Q_{a(t)}(t)$ on placebo.

44. Schönberg, T., Daw, N.D., Joel, D. & O'Doherty, J.P. Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *J. Neurosci.* **27**, 12860–12867 (2007).

45. Huys, Q.J.M. *et al.* Disentangling the roles of approach, activation and valence in instrumental and Pavlovian responding. *PLoS Comput. Biol.* **7**, e1002028 (2011).

46. Lutti, A., Hutton, C., Finsterbusch, J., Helms, G. & Weiskopf, N. Optimization and validation of methods for mapping of the radiofrequency transmit field at 3T. *Magn. Reson. Med.* **64**, 229–238 (2010).

47. Reese, T.G., Heid, O., Weisskoff, R.M. & Wedeen, V.J. Reduction of eddy-current-induced distortion in diffusion MRI using a twice-refocused spin echo. *Magn. Reson. Med.* **49**, 177–182 (2003).

48. Nagy, Z., Weiskopf, N., Alexander, D.C. & Deichmann, R. A method for improving the performance of gradient systems for diffusion-weighted MRI. *Magn. Reson. Med.* **58**, 763–768 (2007).

49. Andersson, J.L.R., Skare, S. & Ashburner, J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage* **20**, 870–888 (2003).

50. Ashburner, J. A fast diffeomorphic image registration algorithm. *Neuroimage* **38**, 95–113 (2007).

51. Knutson, B. & Cooper, J.C. Functional magnetic resonance imaging of reward prediction. *Curr. Opin. Neurol.* **18**, 411–417 (2005).

52. Brett, M., Anton, J.-L., Valabregue, R. & Poline, J.-B. Region of interest analysis using an SPM toolbox. *Neuroimage* **16**, 1140–1141 (2002).

53. Forstmann, B.U. *et al.* The speed-accuracy tradeoff in the elderly brain: a structural model-based approach. *J. Neurosci.* **31**, 17242–17249 (2011).

54. Schwarzkopf, D.S., de Haas, B. & Rees, G. Better ways to improve standards in brain-behavior correlation analysis. *Front. Hum. Neurosci.* **6**, 200 (2012).