Dorsal striatal dopamine D1 receptor availability predicts an instrumental bias in action learning

Lieke de Boer[a,1], Jan Axelsson[b,c], Rumana Chowdhury[d], Katrine Riklund[b,c], Raymond J. Dolan[e], Lars Nyberg[b,c,f], Lars Bäckman[a], and Marc Guitart-Masip[a,e]

[a]Aging Research Center, Karolinska Institutet, SE-17165, Stockholm, Sweden; [b]Department of Radiation Sciences, Diagnostic Radiology, University Hospital, Umeå University, SE-901 87 Umeå, Sweden; [c]Umeå Center for Functional Brain Imaging, Umeå University, SE-907 36 Umeå, Sweden; [d]Department of Neurology, Leeds Teaching Hospitals National Health Service Trust, Leeds LS1 3EX, United Kingdom; [e]Max Planck University College London Centre for Computational Psychiatry and Ageing Research, University College London, London WC1B 5EH, United Kingdom; and [f]Department of Integrative Medical Biology, Physiology, Umeå University, SE-901 87 Umeå, Sweden

Edited by Marcus E. Raichle, Washington University in St. Louis, St. Louis, MO, and approved November 19, 2018 (received for review October 5, 2018)

Learning to act to obtain reward and inhibit to avoid punishment is easier compared with learning the opposite contingencies. This coupling of action and valence is often thought of as a Pavlovian bias, although recent research has shown it may also emerge through instrumental mechanisms. We measured this learning bias with a rewarded go/no-go task in 60 adults of different ages. Using computational modeling, we characterized the bias as being instrumental. To assess the role of endogenous dopamine (DA) in the expression of this bias, we quantified DA D1 receptor availability using positron emission tomography (PET) with the radioligand [^{11}C]SCH23390. Using principal-component analysis on the binding potentials in a number of cortical and striatal regions of interest, we demonstrated that cortical, dorsal striatal, and ventral striatal areas provide independent sources of variance in DA D1 receptor availability. Interindividual variation in the dorsal striatal component was related to the strength of the instrumental bias during learning. These data suggest at least three anatomical sources of variance in DA D1 receptor availability separable using PET in humans, and we provide evidence that human dorsal striatal DA D1 receptors are involved in the modulation of instrumental learning biases.

decision making | dopamine | Pavlovian bias | instrumental learning | positron emission tomography

Instrumental learning occurs through responding to the environment in ways that lead to rewards, and avoiding responding in ways that lead to punishments (1). Theories of instrumental learning generally do not take into account whether such a response is active or passive and assume that action (go and no-go) and valence (win or lose) are independent. However, empirical studies of human learning have systematically shown that instrumental responding is biased: Learning to act to reap a reward and not to act to avoid punishment ["go to win" (GW) and "no-go to avoid losing" (NGL)] is easier than learning inaction to gain a reward and action to avoid punishment ["no-go to win" (NGW) and "go to avoid losing" (GL)] (2–8).

Such biases (often conceived of as Pavlovian biases in nature) facilitate learning in many real-world contexts but can be detrimental in situations where action and valence are not coupled congruently (2, 4, 5, 9, 10). Given the robustness of the findings demonstrating the existence of this bias, any account of instrumental learning that does not consider it will necessarily be incomplete. In turn, studying the mechanisms leading to this bias is crucial to enrich our understanding of instrumental learning.

One obvious source of these biases may be sought in the functional architecture of the basal ganglia and its dopaminergic modulation (11). A widely accepted computational framework posits that dopamine (DA) conveys reward prediction error signals (12, 13) with phasic bursts signaling better than expected events (also called positive prediction errors) and dips below baseline signaling worse than expected events (also called negative prediction errors) (14). In the striatum, increases in DA are thought to reinforce the direct pathway (expressing DA D1 receptors) and promote those actions associated with dopaminergic bursts, while dips in DA are thought to reinforce the indirect pathway (expressing DA D2 receptors) and discourage actions associated with the dopaminergic dips (11, 15). Hence, within this framework, rewards promote action and punishment promotes inhibition, coupling action and valence and naturally leading to biases in action learning.

Although this framework is robustly supported by results from animal experiments (16–19), corresponding evidence in humans is mainly limited to studies of genetic polymorphisms and pharmacological manipulations (7, 20–24). Only one previous study has assessed the relationship between DA receptor availability and predictions derived from this framework using a simple learning paradigm in a small sample (25). If this theoretical framework is correct, one would predict that the extent to which individuals are biased in coupling action and valence can be predicted based on measures of DA receptor availability in striatum. In support of this line of reasoning, Richter et al. (7) found that this kind of bias was stronger in those individuals with genetic variants linked to DA D2 receptor expression in striatum (17). However, it remains unknown whether the behavioral biases coupling action and valence are related to direct measures of endogenous DA function such as receptor availability.

The biases that arise because of a coupling of action and valence have generally been conceived as Pavlovian in nature (2, 4, 5), but a recent study has shown that instrumental mechanisms

Significance

The brain's dopaminergic pathways are crucially important for adaptive behavior. They are thought to enable us to approach rewards and stay away from punishments. During learning, dopaminergic reward prediction errors are thought to reinforce previously rewarded actions, so they become easier to repeat. This dopaminergic activity could lead to a systematic bias by which rewarded actions are more readily learned than rewarded inactions. We present two findings. First, dopamine receptors in cortex, dorsal striatum, and nucleus accumbens provide distinct sources of variance in the human brain. Second, the boost in an individual's learning rate from previously rewarded actions is dependent on the dopamine receptor density in dorsal striatum, a central structure in the dopaminergic circuit.

Author contributions: R.J.D., L.N., L.B., and M.G.-M. designed research; L.d.B., R.C., and M.G.-M. performed research; L.d.B., J.A., K.R., and M.G.-M. contributed new reagents/analytic tools; L.d.B., J.A., and M.G.-M. analyzed data; and L.d.B. and M.G.-M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]To whom correspondence should be addressed. Email: liekelotte@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1816704116/-/DCSupplemental.

Published online December 18, 2018.

Downloaded at Palestinian Territory, occupied on December 1, 2021

NEUROSCIENCE

PSYCHOLOGICAL AND COGNITIVE SCIENCES

www.manaraa.com

also play a role (8). A Pavlovian mechanism reflects an effect of anticipated valence on action selection and is likely to be associated with the anticipatory phase of reward prediction errors. Instrumental mechanisms, on the other hand, reflect an asymmetric impact of positive and negative prediction errors on learning the value of action and inhibition during outcome processing. Both Pavlovian and instrumental mechanisms could in principle arise through the interaction of the functional architecture of the striatum and its dopaminergic modulation discussed above. However, it is currently unknown whether this is the case, and therefore it is of interest to decompose the bias in its Pavlovian and instrumental components when studying its relation to direct measures of endogenous DA function.

Aside from this striatal modulation, DA also plays a role in facilitating working memory and attentional processes in prefrontal cortex (PFC) and parietal cortex (26–28). Although the striatal workings of DA would predict an increase in the bias that couples action with valence as highlighted above, boosting the DA system with L-DOPA has previously resulted in a decrease of such a bias (6). This effect is consistent with the thesis that DA in PFC facilitates attentional mechanisms that can help in overcoming the effects of action/valence-dependent biases on learning (27). In line with this assertion, Cavanagh et al. (2) observed increased frontal theta with EEG in those individuals who performed better under Pavlovian conflicts (i.e., situations where action and valence are not congruently coupled). Therefore, it is unclear whether the effects of DA on these biases are regionally specific.

To investigate how DA availability modulates the behavioral expression of this bias and the anatomical locus of this modulation, we collected behavioral data from 30 younger and 30 older participants on a probabilistic monetary go/no-go (GNG) task (4), where participants were required to learn the correct action–valence contingencies. We used computational modeling to test different parameterizations (i.e., Pavlovian and instrumental) of the effect of this bias on each subject's behavior. We also collected positron emission tomography (PET) data with [^{11}C]SCH23390, quantifying D1 DA binding potentials (BP_{ND}) as a measure of DA D1 receptor availability in cortical and striatal regions of interest (ROIs). We performed principal-component analysis (PCA) on the BP_{ND} values in these ROIs to extract independent sources of variance in DA D1 receptor availability across the brain.

Although previous research does not show any age differences in the strength of action/valence-dependent biases, the inclusion of older participants increases the variance in DA D1 availability and the power to detect a relationship between DA D1 availability and action/valence-dependent learning bias. The functional anatomy of the dopaminergic system allows for two independent, not mutually exclusive predictions: (*i*) In accordance with the striatal direct/indirect pathway model of instrumental learning, the interindividual variation in striatal DA D1 receptor availability would be positively related to the extent to which individuals express a action/valence-dependent bias during learning; and (*ii*) in accordance with the role of cortical DA functions in overcoming such a bias, cortical DA would be negatively related to the extent to which individuals express this bias.

Results

Behavior. A 2 × 2 × 2 ANOVA with action (go/no-go), valence (win/avoid losing), and age group (younger/older) on the 56 participants who completed the task (*Materials and Methods*) detected a main effect of action [$F_{(1,54)} = 5.84$, $P = 0.02$] and a main effect of age group [$F_{(1,54)} = 21.7$, $P < 0.001$], but no action by valence interaction [$F_{(1,54)} = 2.1$, $P = 0.15$] and no significant main effect of valence [$F_{(1,54)} = 0.01$, $P = 0.92$]. We also found an action by valence by age group effect [$F_{(1,54)} = 7.25$, $P = 0.01$], which was driven by a lack of significant action by valence interaction in the older participants [$F_{(1,27)} = 0.42$, $P = 0.52$], but a significant action by valence interaction in the young [$F_{(1,27)} = 15.55$, $P < 0.001$]. This is a surprising result, because a previous study comparing younger and older adults on this task demonstrated an action by valence interaction in both groups of participants (3). Furthermore, we observed lower performance across conditions in our sample compared with the performance typically observed in this task (3, 7).

We reasoned that these findings could be related to some participants performing the task at chance and only contributing noise to the data. To obtain a blind exclusion criteria, we performed a 2-means clustering analysis on the performance during the last 15 trials of the GW condition. GW has previously been shown to be the easiest condition to learn in this task, with participants of all ages consistently performing well in the last 15 trials (3, 7). This analysis detected two groups within the sample. A total of 41 participants (17 old, 24 young), who we will refer to as the "good performers," showed high levels of performance on the last 15 trials for the GW conditions [mean (M) (GW) = 89%]. A total of 15 participants (11 old, 4 young), who we will refer to as the "bad performers," showed low levels of performance on the last 15 trials for the GW condition, as well as low levels of accuracy for all other task conditions [M(GW) = 54%, M(GL) = 63%, M(NGW) = 47%, M(NGL) = 49%] that did not differ from chance level (50%; $P > 0.16$), with the exception of the GL condition ($P < 0.001$).

An overview of the task conditions and the proportion of correct trials on each condition for all good performers are displayed in Fig. 1 *A* and *B*. The 2 × 2 × 2 ANOVA with action (go/no-go), valence (win/avoid losing), and group (younger/older) on the 41 good performers only detected a main effect of action [$F_{(1,40)} = 43.3$, $P < 0.001$] and an action by valence interaction [$F_{(1,40)} = 20.9$, $P < 0.001$] without a significant main effect of valence [$F_{(1,40)} = 0.18$, $P = 0.67$]. We again found an action by valence by group effect [$F_{(1,40)} = 6.59$, $P = 0.01$], which was driven by a lack of significant action by valence interaction in the older participants [$F_{(1,16)} = 1.73$, $P = 0.21$]. Despite the lack of significant interaction in the older participants, we observed a positive interaction term in 14 of the 17 older participants (range, −0.36–0.56), showing that the effect of interest was present in most of the participants. For completeness, 20 of 24 younger participants showed a positive interaction (range, −0.27–0.96).

Computational Modeling of Task Performance. We assessed a collection of models in their ability to account for the observed behavioral data. Based on previous studies (2, 6), the base model included a Rescorla–Wagner Q-learning rule to update the values of go and no-go choices, separate model parameters for sensitivity to rewards and punishments, as well as a learning rate, an irreducible noise parameter, and a constant go bias parameter (*Materials and Methods*). An additional set of parameters were added to the models, and models were compared using the integrated Bayesian information criterion (iBIC), where small iBIC values indicate a model that fits the data better after penalizing for the number of parameters. Comparing iBIC values is akin to a likelihood ratio test (29). The different models were compared, and their iBIC scores are presented in Table 1. The winning model included an instrumental learning bonus κ, which modulated participants' learning rates depending on the outcome of the trial. Contrary to the previous report by Swart et al. (8), model evidence suggested that learning rate was boosted by κ on rewarded go trials but was not decreased by κ on punished no-go trials. Adding a Pavlovian bias parameter to the winning model did not improve model fit further. To our surprise, model comparison does not provide strong evidence for the addition of a Pavlovian bias to a model with separate sensitivities to rewards and punishments (model 1 versus model 2). This model has previously provided the most parsimonious account of data on this task (2, 6).

The descriptive statistics for the parameters of the winning model are summarized in Table 2. When comparing the model

Downloaded at Palestinian Territory, occupied on December 1, 2021
www.manaraa.com

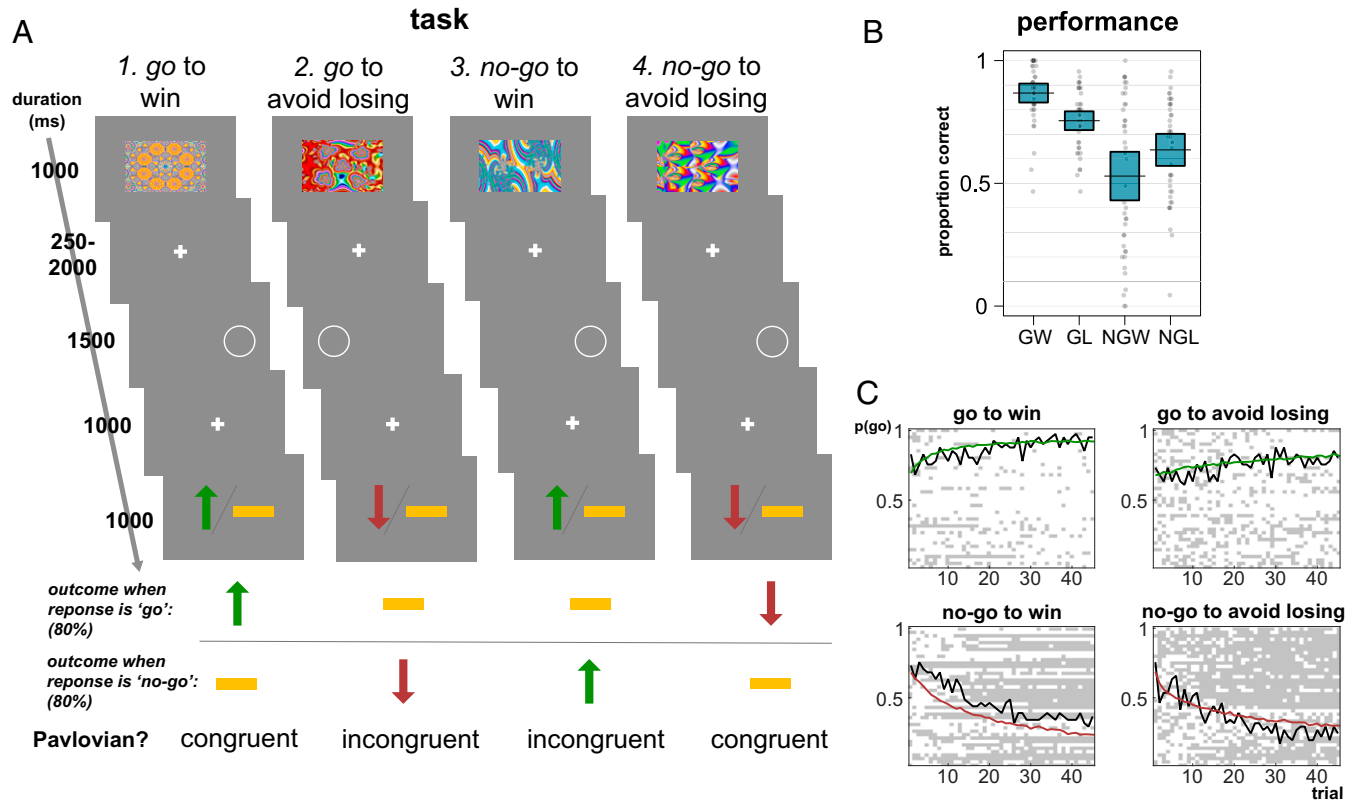

Fig. 1. (*A*) Schematic representation of the rewarded go/no-go task. On each trial, participants were presented with one of four fractals. After a variable delay of 250–2,000 ms, they were presented with a target circle. A "go" was counted as a button press on the same side as the target within 1,500 ms of target presentation. After another delay of 1,500 ms, participants were presented with 80/20 probabilistic feedback. (*B*) Performance on each trial type. All participants are presented as individual data points in gray. The 95% confidence interval around the mean on each condition is presented in blue color. (*C*) Model parameters of the winning model were used to generate simulated choice data. The simulated group mean probability of performing a go on each trial is plotted in colored lines (green for go conditions, where go is the correct response; red for no-go conditions, where no-go is the correct response). The group mean for participants' actual performance is plotted in black lines, reflecting the proportion of actual go responses on each trial. In the plot area, each row represents one participant's choice behavior. Forty-five pixels, one per trial, make up each row. A white pixel reflects that a participant chose go on that trial; a gray pixel represents no-go.

parameters between age groups with independent-sample *t* tests or Mann–Whitney *U* tests, differences between three out of six parameters could be observed: ρ_{lose}, ε, and ξ (Table 2). Simulations for the winning model are presented in Fig. 1*C*.

To assess the models further, we calculated pseudo-R^2 for each model. The pseudo-R^2 is a value that describes how much more variability a model can account for in each participant compared with a baseline model reflecting chance decisions, that is, assuming an equal probability of choosing go and no-go for each choice. Inspection of pseudo-R^2 values for the 17 bad performers further confirmed that our computational modeling analysis for these participants was not meaningful—the mean pseudo-R^2 value for this group did not differ significantly from zero (M = 0.05, SD = 0.11; one-sample *t* test against zero, $P = 0.10$). In contrast, the mean individual pseudo-R^2 values for participants included in our sample was 0.33 (one-sample *t* test against zero, $P < 0.001$; Table 1), suggesting that our blind 2-means clustering separated those participants who performed at chance level from those participants whose performance was meaningfully described by the modeling analysis.

Because we have not used the current model in combination with this task before, we wanted to assess whether behavior on

Table 1. Model comparison for the six models that were used to account for the behavioral data

Model no.	Model parameters	No. of parameters	Likelihood	Pseudo-R^2	iBIC
1	ε, ρ_{win}, ρ_{lose}, ξ, b	5	−3,484	0.32	7,057
2	ε, ρ_{win}, ρ_{lose}, ξ, b, π	6	−3,472	0.32	7,051
3	**ε, ρ_{win}, ρ_{lose}, ξ, b, $\kappa_{\text{rewarded go}}$**	**6**	**−3,444**	**0.33**	**6,996**
4	ε, ρ_{win}, ρ_{lose}, ξ, b, $\kappa_{\text{rewarded go/punished no-go}}$	6	−3,464	0.32	7,042
5	ε, ρ_{win}, ρ_{lose}, ξ, b,π, $\kappa_{\text{rewarded go}}$	7	−3,446	0.33	7,016
6	ε, ρ_{win}, ρ_{lose}, ξ, b, π, $\kappa_{\text{rewarded go/punished no-go}}$	7	−3,482	0.32	7,089

The winning model statistics are presented in boldface type. Parameters: ε, learning rate; ρ_{win}, weighting of reward on win trials; ρ_{lose}, weighting of punishments on lose trials; b, go bias; π, Pavlovian bias; ξ, irreducible noise; κ, instrumental learning bias. iBIC, integrated Bayesian information criterion.

Downloaded at Palestinian Territory, occupied on December 1, 2021
www.manaraa.com

Table 2. Summary statistics for the parameters in the winning model

Parameter	Mean	SD	Min	Q0.25	Median	Q0.75	Max	Mean old	Mean young	*P* value difference old/young
ρ_{win}	16.77	10.34	3.21	9.42	13.98	21.45	56.91	18.13	15.81	0.610
ρ_{lose}	7.17	5.42	1.51	2.59	6.22	10.18	23.58	3.95	9.46	<0.001
ε	0.09	0.08	0.01	0.04	0.06	0.11	0.31	0.06	0.12	0.001
ξ	0.89	0.10	0.59	0.88	0.92	0.94	0.97	0.85	0.91	0.006
κ	0.15	0.98	−1.55	−0.60	0.18	1.11	1.97	−0.08	0.32	0.214
b	0.76	0.71	−0.52	0.27	0.77	1.12	3.01	0.90	0.66	0.610

Parameters: ρ_{win}, weighting of reward on win trials; ρ_{lose}, weighting of punishments on lose trials; «, learning rate; ξ, irreducible noise; *b*, go bias; κ, instrumental learning bonus.

this task can be ascribed to an instrumental, rather than a Pavlovian bias, on all instances of the task. We hypothesized that the reason we have not found a Pavlovian parameter in the winning computational model could be related to our task version containing 15 fewer trials per condition than the version reported in previous publications (3, 6). This hypothesis is supported by the logic that, as opposed to an instrumental learning bias, a Pavlovian bias promoting approach or avoidance responses to a cue can only emerge when a cue carries some anticipated value for the agent (30). This anticipated value arises through learning, which will invariably take time. To test this hypothesis, we revisited an available previously published dataset that included 47 younger [18–30 y old (4)] and 42 older participants [64–75 y old (3)] who performed 60 trials in each condition instead of 45 trials. The task was otherwise identical to the one described here. To investigate whether the length of the task affected the manifestation of a Pavlovian bias, we used the same computational modeling routine as the one described above for the current dataset.

This analysis demonstrated that, in this longer version of the task, the model with both instrumental bias parameter κ and Pavlovian bias parameter π described the data best (model 6, *SI Appendix*, Fig. S1 and Table S1). To further test the hypothesis that the Pavlovian bias emerges over time, we built two linear regression models where we predicted performance on the NGW condition in the first 15 and last 15 trials, respectively, using both κ and π as predictors. Performance on the last 15 trials was significantly better than the first 15 NGW trials of the task ($M_{first} = 0.42$, SD = 0.27; $M_{last} = 0.67$, SD = 0.40; paired *t* test, $t = 5.00$; $P < 0.001$). The linear regression analysis showed that κ (but not π) was a significant predictor of performance in the first 15 NGW trials ($\beta_\kappa = -0.084$, $P = 0.011$; $\beta_\pi = -0.090$, $P = 0.224$, model $P = 0.028$) and π (but not κ) was a significant predictor of performance in the last 15 NGW trials of the task ($\beta_\pi = 0.545$, $P < 0.001$; $\beta_\kappa = -0.073$, $P = 0.094$, model $P < 0.001$).

Relationship Between DA D1 Receptor Availability and Behavioral Bias. Note that [^{11}C]SCH23390 binds to all D1-like receptors, which includes D1 and D5 receptor subtypes. We will refer to these as D1 receptors throughout this paper. We calculated BP_{ND} values, a measure of DA D1 receptor availability, in the ROIs we selected for analysis. An overview of selected ROIs overlaid on a typical participant's T1-weighted image and PET image is presented in *SI Appendix*, Fig. S2. The time–activity curves (TACs) and BP_{ND} values for young and old participants are presented in *SI Appendix*, Fig. S3. We first investigated the relationship between the adjusted BP_{ND} values (age-corrected; see *Materials and Methods* for details) and four measures of the behavioral bias coupling action and valence: the model parameter κ, performance on the NGW condition, general bias (interaction score), and the effect of the bias on win trials only (Table 3). As shown in the table, the direction of the correlations between our measures of bias and BP_{ND} values are positive (and its strength similar; $r = 0.26$–0.52) across all considered ROIs, but it only reaches significance in caudate, putamen, dorsolateral PFC (dlPFC)/ventrolateral PFC (vlPFC), and medial orbitofrontal cortex (mOFC)/lateral orbitofrontal cortex (lOFC). Because the BP_{ND} values in different ROIs that we considered are highly correlated (Fig. 2), it is difficult to make inference about regional specificity of the contributions of different sources of DA receptor availability to the behavioral bias.

To circumvent this problem, we next decomposed the variance in BP_{ND} values across different ROIs into separate components using PCA. We entered the adjusted BP_{ND} values from all participants for whom we had PET data available into a PCA to extract separate sources of variance, followed by a varimax rotation of the retained components. BP_{ND} values in all ROIs were highly correlated, although cortical ROIs correlated more strongly with other cortical ROIs than dorsal striatal ROIs and vice versa (Fig. 2). One eigenvalue above 1 was obtained, but the scree plot showed two elbows: one after the first, and one after the third eigenvalue. To determine the greatest change in slope between the eigenvalues in the scree plot and decide which components should be retained, we used the Catell–Nelson–Gorsuch (Cng) test (31–33). This test showed that the maximum change in slope between eigenvalues happened after the third eigenvalue. Therefore, we retained three components for further analysis. The total variance explained in this three-component solution was 92.6%. The same PCA solution was obtained when no age correction was performed on the BP_{ND} values (*SI Appendix*, Table S2).

In the three-component solution, all cortical ROIs loaded strongly on the first component. The dorsal striatum (caudate and putamen) loaded strongly on the second component, and the nucleus accumbens (NAcc) loaded exclusively on the third component (Table 4). Importantly, we confirmed this PCA solution in another, independently collected, dataset with the same radiotracer and including BP_{ND} values for several ROIs in 20 younger and 20 older participants (34). Because the BP_{ND} values of the exact same ROIs were not available, we selected equivalent ROIs and performed the same PCA on this independent sample. The general separation of variance into cortical, striatal, and ventral striatal components remained despite the fact that the exact same ROIs were not used in this analysis (*SI Appendix*, Fig. S4 and Table S3).

To investigate the relationship between variability in DA D1 receptor availability within these components and distinct measures of behavioral bias on the go/no-go task, we calculated individual's PCA component scores for each identified component. Based on previous findings, we hypothesized that (*i*) component scores reflecting striatal DA D1 receptor availability would be related to an increased bias coupling action and valence in learning and/or that (*ii*) such a bias would be less pronounced in individuals with component scores reflecting greater cortical DA D1 receptor availability.

As predicted, we found positive relationships between all measures of the bias in learning and component scores on the dorsal striatal component (component 2, Fig. 3). We did not observe any relationship between measures of behavioral bias and the cortical component (component 1) or the ventral striatal

Downloaded at Palestinian Territory, occupied on December 1, 2021
www.manaraa.com

Table 3. Correlation coefficients for bivariate correlations between age-corrected BP_{ND} values in different ROIs and measures of the behavioral bias coupling action and valence

Measure of behavioral bias	Caudate	Putamen	NAcc	BAs 44, 45, 46, 9	lOFC/vmPFC	BAs 4, 6	IPL
Instrumental parameter κ	0.48**	0.52***	0.26	0.43**	0.39*	0.28′	0.34*
No-go to win, % correct	−0.40*	−0.46**	−0.26	−0.40*	−0.27′	−0.19	−0.24
Behavioral effect on win trials	0.37*	0.43**	0.20	0.34*	0.18	0.13	0.19
Interaction score	0.30′	0.44**	0.09	0.32*	0.15	0.14	0.23

′$P < 0.1$, *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$.

component (component 3). Statistics for the correlations between all components and measures of behavioral bias are displayed in Table 5. These results were not dependent on the exclusion of the participants who performed the task at chance level (*SI Appendix*, Table S4). Significant correlations survived age correction at P(adjusted) < 0.05.

Statistics for the old and young sample separately are presented in *SI Appendix*, Table S5. These results demonstrate that the same correlations can be observed when considering only young or only old individuals in this sample.

Discussion

We investigated the relationship between DA D1 receptor availability and well-established behavioral biases during learning. Consistent with previous literature, healthy participants were better at learning to emit a behavioral response in anticipation of reward, and better at withholding a response in anticipation of punishment. Using computational modeling, we characterized this tendency as an instrumental learning bias. Despite strong associations in DA D1 receptor availability among the targeted ROIs, we could distinguish distinct sources of variance in cortical, dorsal striatal, and ventral striatal DA D1 receptor availability using PCA. Critically, we showed that variability in dorsal striatal DA D1 receptor availability, independent of age, was related to a range of behavioral measures reflecting the strength of learning biases coupling action with valence, including a quantification of the specific instrumental learning bias.

The functional architecture of the striatum and its dopaminergic modulation has been hypothesized to underlie the behavioral biases studied here. This modulation can occur at two different time points. On the one hand, DA can influence learning at the time of reward receipt, biasing instrumental learning by reinforcing those actions that lead to reward, and inhibiting those that lead to punishment, through the direct and indirect pathways expressing D1 and D2 receptors, respectively (11, 16, 35). Therefore, increased DA D1 receptor availability in striatum may reflect a more sensitive reinforcement mechanism in the direct pathway resulting in a stronger link between the most recently performed active choices and received reward. On the other hand, DA can modulate action selection at the time of choice when anticipation of reward or punishment elicited by cues promote approach and invigoration or withdrawal and inhibition, respectively (4, 36, 37). Such a cue-induced bias is commonly referred to as Pavlovian and is widely studied in animals (38, 39) but has also been demonstrated in humans using Pavlovian-instrumental transfer (PIT) paradigms (40–42). In PIT experiments, stimuli that are associated with rewards or punishments enhance the motivational response to those stimuli, even when that enhancement does not directly benefit the outcome of the action, because the testing is conducted during extinction (39). DA is an important modulator of appetitive PIT (43–45). Therefore, increased D1 receptor availability in the striatum may imply a lower threshold for active choices in response to rewarding cues.

Although our task was not designed for this purpose, we could disentangle these two independent sources of biases (instrumental and Pavlovian) by fitting a range of behavioral models to the data (8). This analysis demonstrated that results were best explained by a model that included a parameter influencing the learning rate of the participants depending on the action/outcome contingency on each trial. Such a mechanism reflects an instrumental learning bias, boosting positive reward prediction errors after go choices. The instrumental learning bias described the data best when it only modulated the learning rate on rewarded go trials and not on punished no-go trials. This asymmetry did not cause the model to overshoot in its predictions of the proportion of go responses to reward-predicting stimuli (Fig. 1*C*). This may appear as contradictory to the study by Swart et al. (8) showing boosted reinforcement of rewarded go choices and dampened reinforcement of punished no-go choices. However, the behavioral biases observed with the present task is historically stronger for appetitive trials compared with aversive trials (2–4, 6, 7, 40, 46). This is interesting in light of a recent study on aversive learning showing a bias on the aversive domain depending on whether the choice involved escape or avoidance (46). Escape choices in response to aversive environments are more easily learned to be active (rather than inactive). The difference in the ease with which a response is learned is magnified for such escape choices compared with choices that lead to the avoidance of an aversive outcome. Thus, to observe similarly strong behavioral bias effects on aversive trials, escape trials may need to be added to the experimental paradigm.

Consistent with the role of DA in reinforcement learning, DA D1 receptor availability in dorsal striatum positively correlated with the strength of learning rate modulation on rewarded go

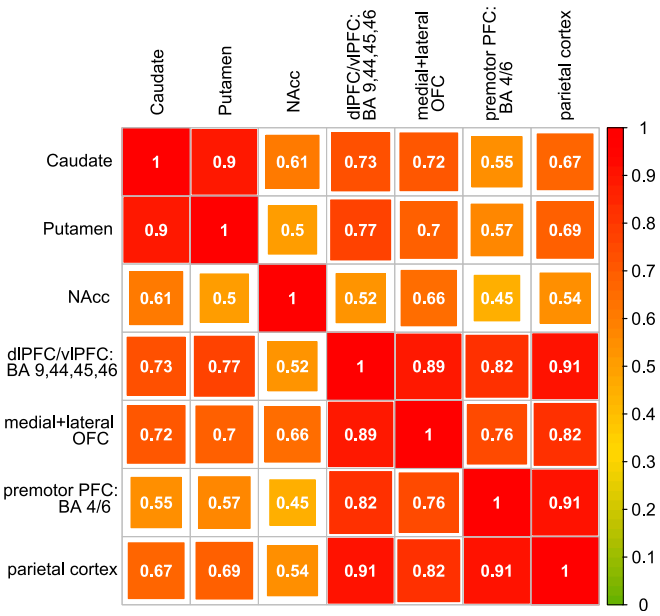

Fig. 2. Correlation matrix that shows the correlations between age-corrected BP_{ND} values in the ROIs selected for analysis.

Downloaded at Palestinian Territory, occupied on December 1, 2021
www.manaraa.com

Table 4. Component loadings for each ROI

Region of interest	Component 1	Component 2	Component 3
Caudate	0.39	**0.88**	0.18
Putamen	0.32	**0.85**	0.34
NAcc	0.24	0.26	**0.93**
dlPFC/vlPFC: BAs 9, 44, 45, 46	**0.80**	0.48	0.22
Limbic PFC: lateral/ medial OFC	**0.72**	0.39	0.45
Premotor PFC: BAs 4, 6	**0.92**	0.19	0.17
Parietal cortex: IPL	**0.79**	0.46	0.20

The component that each ROI loaded on most strongly is displayed in boldface type.

trials. The modeling analysis also demonstrated that no additional variance in choice behavior could be explained with the inclusion of a Pavlovian bias parameter, which would promote an approach/avoidance reaction to the cue at onset depending on the expected value of the condition on a given trial. Thus, the bias studied in the current experiment is more accurately described as an instrumental bias than a Pavlovian bias, and the strength of this instrumental bias is predicted by DA D1 receptor availability in dorsal striatum. The dorsal striatum has previously been shown to be an integral part of the decision-making circuit (47, 48) and a major target for substantia nigra DA neurons. Specifically, the dorsal striatum is involved in learning about actions and their reward consequences (49), whereas the ventral striatum (VS) is thought to be more involved in passive forms of appetitive learning [refs. 50 and 51, but see also Guitart-Masip et al. (52)]. This involvement of the dorsal striatum in action-dependent learning is in line with the findings presented here and suggests that dopaminergic modulation during learning biases the learning from rewarded actions as opposed to rewarded inactions.

Despite the winning model in this experiment only including an instrumental learning bias, we would not argue that choice behavior in this task is not affected by a Pavlovian bias for several reasons. First, Swart et al. (8) performed an experiment designed to computationally disentangle a cue-evoked Pavlovian bias from an instrumental bias in learning from reinforcements at feedback and found that both instrumental and Pavlovian mechanisms played a role. Second, the task in this study was shorter than reported elsewhere (2–4, 6), and reanalysis of one of these datasets (3) demonstrates that both an instrumental learning bias and a Pavlovian bias emerge in the longer version of the task, in agreement with findings reported by Swart et al. (8). Furthermore, in this longer dataset, the individually fitted instrumental learning parameter was a predictor of "no-go to win" performance early on during the task, but not toward the end. Conversely, the Pavlovian parameter was a predictor of "no-go to win" performance toward the end of the task, but not early on. This provides additional evidence for the view that, in our task, instrumental learning biases affect performance mostly during the learning phase, and Pavlovian biases affect performance mostly after learning has occurred.

Although our computational models accurately described behavior observed in this task, we need to acknowledge that our modeling approach does not exhaustively test every possible mechanism by which reward-induced behavioral activation may result in behavioral biases. Whereas we considered and parameterized a Pavlovian mechanism by which valence is coupled to action, alternative non-Pavlovian mechanisms are conceivable. In animal studies, the presentation of rewards results in increased general behavioral output including those behaviors that may eventually lead to reward obtainment (53). Theoretical models demonstrate that learning mechanisms by which stimuli and actions become associated with rewards are dissociable from the general arousing effects of reinforcers leading to undirected increases in behavioral output (54). Because of their independence, it is beyond the scope and hypothesis of this paper to consider general behavioral activation mechanisms.

Measures of instrumental bias correlated strongly with DA D1 receptor availability in the different ROIs we selected. Because BPs are highly correlated across ROIs, it was initially unclear whether these different correlations were masking specific relationships between local DA D1 receptor availability and behavior. With PCA, we could disentangle distinct sources of variance in DA D1 receptor availability. A priori, different PCA solutions were conceivable. We could have found a solution implying a single source of variance across the brain for D1-type receptors. Alternatively, based on the connectivity of the different regions, one could hypothesize a topographical organization whereby motor, associative, and limbic areas of the brain (26, 55) would provide independent source of variance. Finally, based on the ontology of brain development (56), one could hypothesize an anatomical organization with distinct cortical and subcortical sources of variance as the one we found. This is in line with a previous study looking at D2 receptors (57). In the case of D1 receptors, this distinction could also stem from the types of neurons on which these receptors are typically expressed in different brain areas. In the cortex, both D1 and D5 receptors are expressed on pyramidal cells, whereas in the striatum D1 receptors are expressed on medium spiny neurons and D5 receptors are expressed on tonically active neurons (58). The dorsal versus ventral striatum division was less expected. However, there is abundant evidence suggesting that the shell of the NAcc and related ventral striatal regions may have different neurochemical properties compared with the rest of the striatum (15, 26, 59).

In contrast with the suggestion that cortical DA may help overcome learning biases that couple action with valence (6), we did not find any evidence that endogenous DA D1 receptor availability in the cortex is negatively related to the strength of the Pavlovian or instrumental bias. One explanation for the

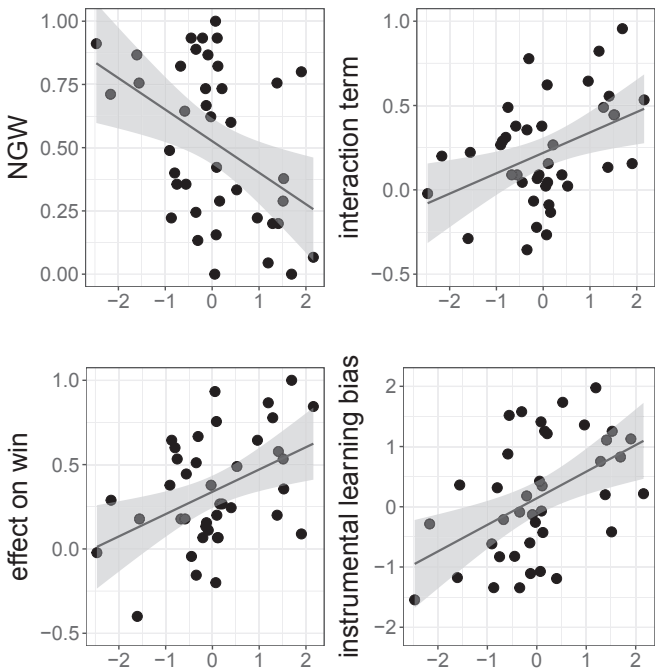

Fig. 3. Correlations between measures of behavioral bias coupling action and valence and component scores for component 2 (dorsal striatal DA D1 receptor availability). Statistics are displayed in Table 5.

Downloaded at Palestinian Territory, occupied on December 1, 2021
www.manaraa.com

Table 5. Correlations coefficients and *P* values for correlations between all component scores and different indicators of a behavioral bias that couples action with valence

	Component 1 scores correlations		Component 2 scores correlations		Component 3 scores correlations	
Measure of behavioral bias	Correlation coefficient	Adjusted *P* value	Correlation coefficient	Adjusted *P* value	Correlation coefficient	Adjusted *P* value
Instrumental parameter κ	0.167	0.485	**0.477**	**0.005**	0.083	0.767
No-go to win performance	−0.082	0.769	**−0.428**	**0.011**	−0.097	0.726
Behavioral effect on win trials	0.121	0.918	**0.432**	**0.010**	0.011	0.867
Interaction score	0.065	0.813	**0.418**	**0.014**	−0.091	0.994

Significant correlations are displayed in boldface type.

absence of a relationship in our data would be that the effects of L-DOPA on behavioral bias were mediated by stimulation of D2 receptors. Alternatively, the effects of L-DOPA on behavioral bias could have been mediated by increased noradrenaline levels. L-DOPA is the precursor of both DA and noradrenaline, and, beyond dopaminergic neurons, noradrenergic neurons have been shown to reuptake L-DOPA (60). Furthermore, previous studies have demonstrated increased DA and noradrenaline release in the rat brain after administration of L-DOPA (61, 62).

We corrected for age before performing the PCA on the D1 receptor BP_{ND} values because the number of DA receptors decreases over the life span (34, 63, 64). Our aim was to investigate how the variation in different sources of receptor availability contributed to the bias coupling action with valence, rather than the overall absolute number of DA receptors, which is strongly associated with age. By regressing out age, we assumed that the way in which different parts of the brain contribute to the variability in receptor availability does not change over the life span. Additionally, it ensures that component scores are not correlated with age. Note that the same PCA on the uncorrected BP_{ND} data yielded a similar solution, supporting the view that this solution is not affected by the age correction.

Despite this, the inclusion of participants from two different age groups may be a limitation of this study. Although these age groups provided increased variability in both behavior and DA D1 receptor availability, many other factors likely differ between older and younger individuals. To address this potential limitation, we regressed out the effects of age in all of our analyses. Furthermore, to make sure that the correlations between DA D1 receptor availability and our measures of behavioral bias do not depend on these uncontrolled factors that covary with age, we assessed these correlations in each age group separately. We found evidence that the correlations between the dorsal striatal component of DA D1 receptor availability are observed when the age groups are considered individually. This control analysis rules out the possibility that our results depend on potentially uncontrolled factors related to the inclusion of participants differing in age.

Another limitation of this study is that performance was low among older participants. Contrary to previous studies (3), the action by valence interaction could not be found in the older group. It is possible that this is caused by a difficulty in understanding the task instructions for older participants in the current study. To address this, we blindly clustered the data and found one cluster of participants performing the task at chance level. By excluding the participants in this cluster, we isolated those participants whose behavioral data are meaningful in a data-driven way. Although we did not find an action by valence interaction in the older participants at the group level, the majority showed a positive action by valence interaction.

In conclusion, our study shows that it is possible to disentangle cortical and striatal sources of variance in DA D1 receptor availability in humans using PET. We provide evidence that higher levels of endogenous DA D1 receptor availability in the human dorsal striatum are related to biases during learning, namely an instrumental learning bias boosting learning from rewarded go trials. This finding suggests that a pervasive bias in instrumental learning stems from the functional architecture of the striatum and its dopaminergic modulation.

Materials and Methods

Participants. Thirty healthy older adults, aged 66–75 y, and 30 younger adults, aged 19–32 y, were recruited through local newspaper advertisements in Umeå, Sweden. The health of all potential participants was assessed before inclusion through a questionnaire administered by research nurses. The questionnaire inquired about past and current neurologic or psychiatric conditions, head trauma, diabetes mellitus, arterial hypertension that required more than two medications, addiction to alcohol or other drugs, and bad eyesight. All participants were right-handed and provided written informed consent before commencing the study. Ethical approval was obtained from the Umeå University Regional Ethical Review Board. Participants were paid 2,000 Swedish crowns (SEK) (~$225) for participation and earned up to 71 additional SEK (~$8.70) in the GNG task.

We excluded four participants (two older, two younger) who emitted >30% incorrect responses (i.e., pressing the button on the opposite side of the target) in any condition. The instructions explicitly stated that this response would never be correct, and therefore these participants are believed to have misinterpreted or failed to understand the instructions. One older participant did not complete the full PET scan, but this participant's behavioral data are still included in the analysis where possible. Additionally, a problem with the injection of the radiotracer in another older participant resulted in a lack of PET signal. Thus, the data from a total sample of 28 younger and 28 older participants for the behavioral analysis, and 30 younger and 28 older participants for the PET analysis were included.

Procedure. Before recruitment, participants completed a health survey questionnaire. On site, all participants performed the Mini Mental State Examination. Scores ranged from 26 to 30 in the young (M = 29.3, SD = 0.88) and from 27 to 30 in the older sample (M = 29.5, SD = 0.85), with no significant difference between age groups ($P = 0.46$). PET scanning and behavioral testing were planned 2 d apart. However, due to a technical problem with the PET scanner, 12 participants were tested with a longer delay (range, 4–44 d apart). On the behavioral testing day, participants completed the GNG learning task and two other tasks inside an MRI scanner. They also completed a battery of tasks outside the scanner. Only results from the GNG learning task will be presented here.

Valence Go/No-Go Task. The task was the learning version of a probabilistic monetary go/no-go paradigm in Swedish, similar to the task first described by Guitart-Masip et al. (4) (Fig. 1*A*). The version described here included only 45 trials per condition instead of 60 as reported previously (2–4). Participants were presented with one of four fractal images on each trial. After a variable delay, they saw a target in the form of a white ring on the left or right side of the screen. Participants were told that for each fractal image, the correct response could be a "go" (press a button corresponding to the same side as the target) or a "no-go" (do not press at all). Participants were instructed that pressing right when the target was on the left or pressing left when the target was on the right would always be counted as an incorrect response.

The four fractals reflected the four conditions of the task. "Win" conditions resulted in a green arrow pointing upward 80% of the trials after correct responses, and in a yellow horizontal bar after 80% of incorrect responses. Win trials either belonged to a Pavlovian congruent condition

Downloaded at Palestinian Territory, occupied on December 1, 2021
NEUROSCIENCE
PSYCHOLOGICAL AND COGNITIVE SCIENCES
www.manaraa.com

(GW, where participants had to respond with a "go" to get a reward) or a Pavlovian incongruent condition (NGW, where participants had to omit a response to get a reward). The same was true for "Lose" trials, which resulted in a neutral outcome after 80% of trials where participants emitted a correct response, and a red arrow pointing down after 80% of incorrect responses. In the Pavlovian congruent condition, participants had to inhibit their response to avoid losing (NGL). In the Pavlovian-incongruent condition, they had to perform a "go" to avoid losing (GL). Responses were counted as a valid go if participants responded within 1,000 ms of being presented with the target. However, participants were able to respond up to 1,500 ms after the target was presented. If they responded between 1,000 and 1,500 ms, the response was counted as a go, but participants saw the words "your response was too slow" on the screen in Swedish. If the participants did not press the button for 1,500 ms after the target was presented, it was counted as a no-go. Participants were presented with feedback after the presentation of a fixation cross for an additional 1,000 ms. Green arrows indicated that participants added 1 SEK (~$0.11) to their running total, horizontal yellow bars meant no loss and no gain, and red arrows pointing down indicated a loss of 1 SEK. The running total was paid to the participants in addition to a participation allowance for the research project.

Participants were explicitly informed about the probabilistic nature of the task but were not informed about the action contingencies. As stated above, there were four trial types, go and no-go to win or avoid losing. Each of these trial types was represented by a different fractal. Participants had to discover, by trial and error, which fractal indicated which condition they were in. Fractal–condition contingencies were randomized between participants. Before performing the learning task, participants performed 20 trials of the target detection task before starting the learning task, to familiarize them with the timing and manner of responding during target detection. After practicing the target detection task, participants performed 45 trials in each of the four task conditions, totaling 180 trials. Trials types were randomly shuffled throughout the duration of the tasks. Participants took a self-paced break every 60 trials, which resulted in their performing three blocks of ~7 min with breaks in between.

Behavioral Analysis. Behavioral data were analyzed using R, 3.4.3. The number of total correct choices per condition was analyzed, defined as any response in line with the task contingencies, even if the go response was performed late (between 1,000 and 1,500 ms).

We present our behavioral data as we observed it in the 56 participants who successfully completed the task. However, a close examination of performance in the older group revealed an overall low performance in all conditions. The low performance in the GW condition (68% correct) is especially surprising considering that this is the easiest condition to learn in this task with performance typically around or over 80% and higher in both younger and older adults (2–4, 6, 7). Because performance was so low, we performed the same analyses that we present in *Results* on a sample of good performers, in which performance levels were more comparable to those previously reported. We selected good performers according to the following procedure: We performed a 2-means clustering analysis on the performance during the last 15 trials of the GW condition.

To investigate the bias coupling action with valence, the proportion of correct responses was entered into a 2×2 ANOVA with action and valence as factors with two levels (go/no-go and win/lose, respectively). To quantify the bias and study its relationship with D1 receptor availability measures, we calculated the overall interaction, reflecting the bias effect on the four conditions (GW + NGL – GL – NGW) and the bias effect on "win" conditions (GW – NGW). Because NGW is the most difficult condition to learn in this task, we took performance on this condition as another indicator of the strength of the bias that couples action with valence.

Computational Modeling. Behavior al data were modeled in MATLAB (Mathworks), similar to Guitart-Masip et al. (4) and Cavanagh et al. (2). We built six parametrized reinforcement learning models to fit participant's behavior (Table 1). These models assigned an action probability for each available action a on each trial t.

Action probabilities depended on an action weight $W(a_t, s_t)$, which was tracked for each action (a), go or no go, and each state (s) determined by the stimulus on that trial (t). The action weights were constructed differently in different models. We added different parameters in a stepwise way. For all models, action weights were passed through a squashed softmax (1):

$$P(a_t,\ s_t) = \left[\frac{\exp[W(a_t, s_t)]}{\sum_{a'} \exp[W(a', s_t)]}\right](1-\xi) + \frac{\xi}{2},$$

where ξ reflected the irreducible noise in the decision rule.

Action weights differed depending on which model was used. All models included the value Q of each action as determined by a Rescorla–Wagner updating rule:

$$Q_t(a_t, s_t) = Q_{t-1}(a_t, s_t) + \varepsilon(\rho r_t - Q_{t-1}(a_t, s_t)).$$

All models included a learning rate ε. Rewards, neutral outcomes, and punishments were entered in the model through r E {−1,0,1}. ρ reflected weighting of reward and punishment, determining the effective size of the reward or punishment. In all models, ρ could take on separate values for rewards and punishments, assuming that forgoing a reward could be more or less aversive than obtaining a punishment. Adding separate sensitivities for rewards and punishments has previously been shown to consistently improve model fit (2, 6).

All models also included an individually fitted static bias parameter b that was added to the value of go:

$$W_t(a_t, s_t) = \begin{cases} Q_t(a_t, s_t) + b & \text{if } a = \text{go}, \\ Q_t(a_t, s_t) & \text{else} \end{cases}$$

In model 2, expected value on the current state $[V_t(s_t)]$ was weighted by another individually fitted free parameter π and added to the value of go choices. Models 2, 5, and 6 therefore included the following action weights for go and no-go:

$$W_t(a_t, s_t) = \begin{cases} Q_t(a_t, s_t) + b + \pi V_t(s) & \text{if } a = \text{go}, \\ Q_t(a_t, s_t) & \text{else} \end{cases}$$

where $\pi \geq 0$.

For models that included a Pavlovian factor, V was computed as follows:

$$V_t(s_t) = V_{t-1}(s_t) + \varepsilon(\rho r_t - V_{t-1}(s_t)).$$

The Pavlovian parameter coupled action and valence and devalued the value of go choices in the punishment conditions in proportion to the value of the stimulus $[V(s)]$, which was negative in these instances. Conversely, the Pavlovian parameter boosted the value of go choices in proportion to the positive value of the stimuli signaling rewarding conditions.

Finally, in line with previous work by Swart et al. (8), we added an instrumental learning bonus κ to some models. κ modulated the participants' learning rate ε depending on choice and feedback on a trial. In the study by Swart et al. (8), the value of this parameter was added to ε on trials that resulted in a rewarded go response, and subtracted from ε on trials that resulted in punished no-go choices. In other trials, ε remained unmodulated.

$$\varepsilon = \begin{cases} \varepsilon_{\text{RewardedGo}} = \varepsilon_0 + \kappa \\ \varepsilon_{\text{PunishedNoGo}} = \varepsilon_0 - \kappa \\ \varepsilon_{\text{other}} = \varepsilon_0 \end{cases}$$

We decided to let κ modulate ε in a stepwise manner, because the difference in learning between rewarded go trials and rewarded no-go trials was larger than the difference in learning between punished go and punished no-go trials. Thus, in model 3 (which included ρ_{win}, ρ_{lose}, ε, ξ, and κ), κ modulated ε only on rewarded go trials. In model 4 (which included ρ_{win}, ρ_{lose}, ε, ξ, and κ), κ modulated ε on rewarded go trials as well as punished no-go trials. Model 5 included ρ_{win}, ρ_{lose}, ε, ξ, π, and κ for rewarded go trials, and model 6 included ρ_{win}, ρ_{lose}, ε, ξ, π, and κ on rewarded go trials as well as punished no-go trials (Table 1).

Model Fitting. Model parameters were fitted using an expectation-maximization approach (4, 37). We used a Laplacian approximation to obtain maximum a posteriori estimates for the parameters for each participant iteratively, starting with flat priors. After an iteration, the resulting group mean posterior and variance for each parameter were used as priors in the next iteration. This method prevents the individuals' parameters from taking on extreme values.

Models were compared using the iBIC calculated as in past work (4, 37). Small iBIC values indicate a model that fits the data better after penalizing for the number of parameters. Comparing iBIC values is akin to a likelihood ratio test. We also calculated a pseudo-R^2 value for each participant. The pseudo-R^2 value compares the difference between the likelihood under a chance model and the likelihood of a given model, and divides the difference by the likelihood of a chance model. The pseudo-R^2 is a measure of how much additional variance can be explained under a given model compared with chance.

Model parameters were assessed for normality with Shapiro–Wilk tests and subsequently compared with between-group tests using independent-sample t tests or Mann–Whitney U tests, depending on the normality of the parameter's distribution.

Downloaded at Palestinian Territory, occupied on December 1, 2021
www.manaraa.com

PET Image Acquisition. PET images were acquired in 3D mode using a Discovery 690 PET/computed tomography (CT) scanner (General Electric), at the Department of Nuclear Medicine, Norrland's University Hospital in Umeå, Sweden. A low-dose helical CT scan (20 mA, 120 kV, 0.8 s/revolution), provided data for PET attenuation correction. Participants were injected with a bolus of 200 MBq of [^{11}C]SCH23390. A 55-min dynamic acquisition commenced at time of injection (9 frames × 2 min, 3 frames × 3 min, 3 frames × 4, 20 min, 3 frames × 5 min). Attenuation- and decay-corrected 256 × 256-pixel transaxial PET images were reconstructed to a 25-cm field-of-view employing the Sharp IR algorithm (six iterations, 24 subsets, 3.0-mm Gaussian post filter). Sharp IR is an advanced version of the OSEM method for improving spatial resolution, in which detector system responses are included (65). The FWHM resolution is below 3 mm. The protocol resulted in 47 tomographic slices per time frame, yielding 0.977 × 0.977 × 3.27-mm^3 voxels. Images were decay-corrected to the start of the scan. Images were deidentified using dicom2usb (https://dicom-port.se). To minimize head movement during the imaging session, the patient's head was fixated with an individually fitted thermoplastic mask (Positocasts Thermoplastic; CIVCO Medical Solutions).

PET Analysis. PET data were analyzed in a ROI-based protocol. This type of analysis requires a priori hypotheses about the regional specificity of dopaminergic modulation of observed behavioral or neuronal effects. All analyses were done using of in-house developed software (imlook4d, version 3.5; https://dicom-port.se/product/imlook4d/).

ROIs for the ROI analysis were based on the division between limbic, associative, and motor areas in the striatum, and their corresponding targets in PFC. Hence, we divided the striatum into NAcc, caudate, and putamen, largely corresponding to limbic, associative, and motor areas of the striatum. In the cortex, a limbic ROI comprised ventromedial PFC (vmPFC) and lOFC. Associative areas including one ROI that comprised the dlPFC and vlPFC, and one ROI comprising the inferior parietal lobule (IPL). Brodmann areas (BAs) 4 and 6 were chosen as representative of motor targets in PFC.

These regions were selected based on their relevance to our task: dlPFC has previously been demonstrated to be involved in executive processes and working memory and cognitive flexibility (66–68), whereas vlPFC is thought to be important for goal-directed action and attention (69). vmPFC has been shown to be responsive to reward magnitude and reward probability in a large number of studies (70). In addition, vmPFC and OFC are active during anticipation of rewards (71, 72). Many connections exist between these regions and VS, an important node in the mesolimbic DA system (26, 30, 73). VS consists of NAcc and parts of the medial caudate nucleus and rostral putamen. The cerebellum was segmented to be used as reference tissue because it is devoid of DA D1 receptors (74). FSL's FIRST algorithm (75) was used to segment subcortical structures.

ROI BP_{ND} values were calculated and presented in a recent publication using this dataset (71). In brief, to obtain ROI BP_{ND} values, the PET time series were first coregistered to the individual T1-weighted images and ROI images. The average TACs were extracted across all voxels within each ROI, and binding potential (BP_{ND}) was calculated using applying the Logan method (76) as implemented in imlook4d. This method was applied to each ROI using the cerebellum as reference tissue. BP values for all ROIs were averaged across hemispheres.

Statistical Analysis of PET Data. Age was not a direct factor of interest in this study, but the age variation ensured both variability in DA D1 receptor availability (63) and performance on the GNG task (3). At the same time, the combined samples provided enough power to perform PCA on PET BP_{ND}. All analyses were performed in such a way that ensured age was regressed out or controlled for. To reduce the collinearity between the DA D1 BP_{ND} values and age, which are highly correlated in a wealth of literature (34, 63, 77, 78), we initially regressed out the effect of age on each BP_{ND} value. We used a similar approach to Raz et al. (79), first computing the β coefficient that reflected the correlation between age and BP_{ND} in each ROI. Then we regressed out the effect of age by calculating the effect of age on BP_{ND} and correcting for this effect:

$$BP_{ND(adj)}(\text{participant, ROI}) = BP_{ND}(\text{participant, ROI}) + \beta_{age(ROI)} * \text{age}(\text{participant}).$$

This procedure is largely similar to performing an analysis on the residuals of BP after regressing out age. We performed simple bivariate Pearson's correlations between these age-corrected BP_{ND} values and four different measures of a behavioral bias that couples action with valence: (*i*) the instrumental learning bias parameter, (*ii*) performance on NGW, (*iii*) the interaction (GW + NGL − GL − NGW), and (*iv*) the effect on no-go trials (NGL − NGW).

The BP_{ND} values in different ROIs were highly correlated ($r > 0.5$; $P < 0.001$ in all ROIs), but an examination of the correlation matrix suggests that BP_{ND} within cortical ROIs are highly correlated to each other but less so to subcortical ROIs (Fig. 2). A similar correlation matrix has previously been observed using the same radioligand to measure DA D1 receptor availability (34). To obtain hypothetically independent sources of variance in DA D1 receptor availability, we performed a PCA on the age-adjusted BP_{ND} data, by first using PCA to extract principal components and subsequently maximizing the variance each component accounted for with an orthogonal varimax rotation. These analyses were performed in R, with the function *principal* (*psych* package). The number of components to retain was determined by performing a Cng test on the eigenvalues (80), done with the R package *nFactors* (function nCng). Cng involves computing the slopes between the eigenvalues in the scree plot. The point at which the greatest change in slope is observed is the cutoff point for the number of components (80).

We then performed Pearson's correlations between participants' component scores on each component obtained in the PCA and behavioral measures described above. To correct for multiple comparisons for correlations between measures of behavioral biases and component loadings, we performed 10,000 permutations where we shuffled the values of the four measures of behavioral biases within participants and correlated these shuffled columns with the DA component loadings. The maximum *t* statistic from the four correlations in each iteration (four columns with shuffled values) was saved and added to the null distribution. This created null distributions that take into account the correlations between the measures of behavioral bias. Correlations in the data with an absolute *t* statistic that exceeded the absolute *t* statistic at the 95th percentile of this new null distribution were considered significant. Adjusted *P* values were calculated by counting the number of *t* values in the new null distribution that exceeded the observed *t* value, divided by 10,000.

ACKNOWLEDGMENTS. We thank Mats Erikson and Kajsa Burström for collecting the data, Dr. Anna Rieckmann for agreeing to sharing her D1 data, and Dr. Rita Almeida for helpful suggestions with regard to the statistical analysis. This research was supported by a research grant from the Swedish Research Council (VR521-2013-2589) (to M.G.-M.), the Humboldt Research Award (to L.B.), and a donation from the af Jochnick Foundation (L.B.). The study was accomplished while L.N. was holding the Söderberg's Professorship in Medicine from Torsten and Ragnar Söderberg's Foundation.

1. Sutton RS, Barto AG (1998) *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, MA).
2. Cavanagh JF, Eisenberg I, Guitart-Masip M, Huys Q, Frank MJ (2013) Frontal theta overrides Pavlovian learning biases. *J Neurosci* 33:8541–8548.
3. Chowdhury R, Guitart-Masip M, Lambert C, Dolan RJ, Düzel E (2013) Structural integrity of the substantia nigra and subthalamic nucleus predicts flexibility of instrumental learning in older-age individuals. *Neurobiol Aging* 34:2261–2270.
4. Guitart-Masip M, et al. (2012) Go and no-go learning in reward and punishment: Interactions between affect and effect. *Neuroimage* 62:154–166.
5. Guitart-Masip M, Duzel E, Dolan R, Dayan P (2014) Action versus valence in decision making. *Trends Cogn Sci* 18:194–202.
6. Guitart-Masip M, et al. (2014) Differential, but not opponent, effects of L-DOPA and citalopram on action learning with reward and punishment. *Psychopharmacology (Berl)* 231:955–966.
7. Richter A, et al. (2014) Valenced action/inhibition learning in humans is modulated by a genetic variant linked to dopamine D2 receptor expression. *Front Syst Neurosci* 8:140.
8. Swart JC, et al. (2017) Catecholaminergic challenge uncovers distinct Pavlovian and instrumental mechanisms of motivated (in)action. *eLife* 6:e22169.
9. Dayan P, Niv Y, Seymour B, Daw ND (2006) The misbehavior of value and the discipline of the will. *Neural Netw* 19:1153–1160.
10. Rutledge RB, Skandali N, Dayan P, Dolan RJ (2015) Dopaminergic modulation of decision making and subjective well-being. *J Neurosci* 35:9811–9822.
11. Frank MJ, Seeberger LC, O'reilly RC (2004) By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science* 306:1940–1943.
12. Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* 47:129–141.
13. Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275:1593–1599.
14. Bayer HM, Lau B, Glimcher PW (2007) Statistics of midbrain dopamine neuron spike trains in the awake primate. *J Neurophysiol* 98:1428–1439.
15. Wickens JR, Budd CS, Hyland BI, Arbuthnott GW (2007) Striatal contributions to reward and decision making: Making sense of regional variations in a reiterated processing matrix. *Ann N Y Acad Sci* 1104:192–212.
16. Collins AGE, Frank MJ (2014) Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychol Rev* 121:337–366.
17. Frank MJ, Fossella JA (2011) Neurogenetics and pharmacology of learning, motivation, and cognition. *Neuropsychopharmacology* 36:133–152.

Downloaded at Palestinian Territory, occupied on December 1, 2021
NEUROSCIENCE
PSYCHOLOGICAL AND COGNITIVE SCIENCES
www.manaraa.com

18. Hikida T, Kimura K, Wada N, Funabiki K, Nakanishi S (2010) Distinct roles of synaptic transmission in direct and indirect striatal pathways to reward and aversive behavior. *Neuron* 66:896–907.
19. Shen W, Flajolet M, Greengard P, Surmeier DJ (2008) Dichotomous dopaminergic control of striatal synaptic plasticity. *Science* 321:848–851.
20. Frank MJ, O'Reilly RC (2006) A mechanistic account of striatal dopamine function in human cognition: Psychopharmacological studies with cabergoline and haloperidol. *Behav Neurosci* 120:497–517.
21. Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD (2006) Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* 442: 1042–1045.
22. Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE (2007) Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc Natl Acad Sci USA* 104:16311–16316.
23. Jocham G, et al. (2009) Dopamine DRD2 polymorphism alters reversal learning and associated neural activity. *J Neurosci* 29:3695–3704.
24. Voon V, et al. (2010) Mechanisms underlying dopamine-mediated reward bias in compulsive behaviors. *Neuron* 65:135–142.
25. Cox SML, et al. (2015) Striatal D1 and D2 signaling differentially predict learning from positive and negative outcomes. *Neuroimage* 109:95–101.
26. Haber SN, Knutson B (2010) The reward circuit: Linking primate anatomy and human imaging. *Neuropsychopharmacology* 35:4–26.
27. Hitchcott PK, Quinn JJ, Taylor JR (2007) Bidirectional modulation of goal-directed actions by prefrontal cortical dopamine. *Cereb Cortex* 17:2820–2827.
28. Seamans JK, Yang CR (2004) The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Prog Neurobiol* 74:1–58.
29. Vrieze SI (2012) Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol Methods* 17:228–243.
30. Salamone JD, Correa M (2012) The mysterious motivational functions of mesolimbic dopamine. *Neuron* 76:470–485.
31. Cattell RB (1966) The scree test for the number of factors. *Multivariate Behav Res* 1: 245–276.
32. Comrey AL, Lee HB (2013) *A First Course in Factor Analysis* (Psychology Press, New York).
33. Raîche G, Walls TA, Magis D, Riopel M, Blais J-G (2013) Non-graphical solutions for Cattell's scree test. *Methodology* 9:23–29.
34. Rieckmann A, et al. (2011) Dopamine D1 receptor associations within and between dopaminergic pathways in younger and elderly adults: Links to cognitive performance. *Cereb Cortex* 21:2023–2032.
35. Clark D, White FJ (1987) D1 dopamine receptor–the search for a function: A critical evaluation of the D1/D2 dopamine receptor classification and its functional implications. *Synapse* 1:347–388.
36. Geurts DEM, Huys QJM, den Ouden HEM, Cools R (2013) Serotonin and aversive Pavlovian control of instrumental behavior in humans. *J Neurosci* 33:18932–18939.
37. Huys QJ, et al. (2011) Disentangling the roles of approach, activation and valence in instrumental and Pavlovian responding. *PLoS Comput Biol* 7:e1002028.
38. Campese VD, et al. (2014) Lesions of lateral or central amygdala abolish aversive Pavlovian-to-instrumental transfer in rats. *Front Behav Neurosci* 8:161.
39. Cartoni E, Balleine B, Baldassarre G (2016) Appetitive Pavlovian-instrumental transfer: A review. *Neurosci Biobehav Rev* 71:829–848.
40. Geurts DEM, Huys QJM, den Ouden HEM, Cools R (2013) Aversive Pavlovian control of instrumental behavior in humans. *J Cogn Neurosci* 25:1428–1441.
41. Prévost C, Liljeholm M, Tyszka JM, O'Doherty JP (2012) Neural correlates of specific and general Pavlovian-to-instrumental transfer within human amygdalar subregions: A high-resolution fMRI study. *J Neurosci* 32:8383–8390.
42. Talmi D, Seymour B, Dayan P, Dolan RJ (2008) Human Pavlovian-instrumental transfer. *J Neurosci* 28:360–368.
43. Corbit LH, Janak PH, Balleine BW (2007) General and outcome-specific forms of Pavlovian-instrumental transfer: The effect of shifts in motivational state and inactivation of the ventral tegmental area. *Eur J Neurosci* 26:3141–3149.
44. Dickinson A, Smith J, Mirenowicz J (2000) Dissociation of Pavlovian and instrumental incentive learning under dopamine antagonists. *Behav Neurosci* 114:468–483.
45. Wassum KM, Ostlund SB, Balleine BW, Maidment NT (2011) Differential dependence of Pavlovian incentive motivation and instrumental incentive learning processes on dopamine signaling. *Learn Mem* 18:475–483.
46. Millner AJ, Gershman SJ, Nock MK, den Ouden HEM (2017) Pavlovian control of escape and avoidance. *J Cogn Neurosci* 30:1379–1390.
47. Balleine BW, Delgado MR, Hikosaka O (2007) The role of the dorsal striatum in reward and decision-making. *J Neurosci* 27:8161–8165.
48. Marche K, Martel A-C, Apicella P (2017) Differences between dorsal and ventral striatum in the sensitivity of tonically active neurons to rewarding events. *Front Syst Neurosci* 11:52.
49. Wickens JR, Reynolds JNJ, Hyland BI (2003) Neural mechanisms of reward-related motor learning. *Curr Opin Neurobiol* 13:685–690.
50. Niv Y, Montague PR (2009) Theoretical and empirical studies of learning. *Neuroeconomics: Decision Making and the Brain* (Academic, New York), Chap 22.
51. O'Doherty J, et al. (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304:452–454.
52. Guitart-Masip M, et al. (2011) Action dominates valence in anticipatory representations in the human striatum and dopaminergic midbrain. *J Neurosci* 31:7867–7875.
53. Killeen PR, Hanson SJ, Osborne SR (1978) Arousal: Its genesis and manifestation as response rate. *Psychol Rev* 85:571–581.
54. Killeen PR, Sitomer MT (2003) MPR. *Behav Processes* 62:49–64.
55. Haber SN, Behrens TEJ (2014) The neural network underlying incentive-based learning: Implications for interpreting circuit disruptions in psychiatric disorders. *Neuron* 83:1019–1039.
56. Lenroot RK, Giedd JN (2006) Brain development in children and adolescents: Insights from anatomical magnetic resonance imaging. *Neurosci Biobehav Rev* 30:718–729.
57. Zald DH, et al. (2010) The interrelationship of dopamine D2-like receptor availability in striatal and extrastriatal brain regions in healthy humans: A principal component analysis of [^{18}F]fallypride binding. *Neuroimage* 51:53–62.
58. Bergson C, et al. (1995) Regional, cellular, and subcellular variations in the distribution of D1 and D5 dopamine receptors in primate brain. *J Neurosci* 15:7821–7836.
59. Zahm DS (1999) Functional-anatomical implications of the nucleus accumbens core and shell subterritories. *Ann N Y Acad Sci* 877:113–128.
60. Navailles S, et al. (2014) Noradrenergic terminals regulate L-DOPA-derived dopamine extracellular levels in a region-dependent manner in Parkinsonian rats. *CNS Neurosci Ther* 20:671–678.
61. Dolphin A, Jenner P, Marsden CD (1976) Noradrenaline synthesis from L-DOPA in rodents and its relationship to motor activity. *Pharmacol Biochem Behav* 5:431–439.
62. Goshima Y, Kubo T, Misu Y (1986) Biphasic actions of L-DOPA on the release of endogenous noradrenaline and dopamine from rat hypothalamic slices. *Br J Pharmacol* 89:229–234.
63. Suhara T, et al. (1991) Age-related changes in human D1 dopamine receptors measured by positron emission tomography. *Psychopharmacology (Berl)* 103:41–45.
64. Volkow ND, et al. (1998) Parallel loss of presynaptic and postsynaptic dopamine markers in normal aging. *Ann Neurol* 44:143–147.
65. Ross S, Stearns C (2010) SharpIR: White paper (GE Healthcare, Waukesha, WI).
66. Barch DM, Sheline YI, Csernansky JG, Snyder AZ (2003) Working memory and prefrontal cortex dysfunction: Specificity to schizophrenia compared with major depression. *Biol Psychiatry* 53:376–384.
67. D'Esposito M, et al. (1995) The neural basis of the central executive system of working memory. *Nature* 378:279–281.
68. Petrides M (2000) The role of the mid-dorsolateral prefrontal cortex in working memory. *Exp Brain Res* 133:44–54.
69. Levy BJ, Wagner AD (2011) Cognitive control and right ventrolateral prefrontal cortex: Reflexive reorienting, motor inhibition, and action updating. *Ann N Y Acad Sci* 1224:40–62.
70. Rushworth MFS, Behrens TEJ (2008) Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat Neurosci* 11:389–397.
71. de Boer L, et al. (2017) Attenuation of dopamine-modulated prefrontal value signals underlies probabilistic reward learning deficits in old age. *eLife* 6:e26424.
72. Kim H, Shimojo S, O'Doherty JP (2011) Overlapping responses for the expectation of juice and money rewards in human ventromedial prefrontal cortex. *Cereb Cortex* 21: 769–776.
73. Rushworth MFS, Noonan MP, Boorman ED, Walton ME, Behrens TE (2011) Frontal cortex and reward-guided learning and decision-making. *Neuron* 70:1054–1069.
74. Hall H, et al. (1994) Distribution of D1- and D2-dopamine receptors, and dopamine and its metabolites in the human brain. *Neuropsychopharmacology* 11:245–256.
75. Patenaude B, Smith SM, Kennedy DN, Jenkinson M (2011) A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* 56:907–922.
76. Logan J, et al. (1996) Distribution volume ratios without blood sampling from graphical analysis of PET data. *J Cereb Blood Flow Metab* 16:834–840.
77. Bäckman L, Nyberg L, Lindenberger U, Li S-C, Farde L (2006) The correlative triad among aging, dopamine, and cognition: Current status and future prospects. *Neurosci Biobehav Rev* 30:791–807.
78. Wang Y, et al. (1998) Age-dependent decline of dopamine D1 receptors in human brain: A PET study. *Synapse* 30:56–61.
79. Raz N, et al. (2004) Aging, sexual dimorphism, and hemispheric asymmetry of the cerebral cortex: Replicability of regional differences in volume. *Neurobiol Aging* 25: 377–396.
80. Gorsuch RL (2014) *Factor Analysis: Classic Edition* (Routledge, New York).

Downloaded at Palestinian Territory, occupied on December 1, 2021

www.manaraa.com