

 Open access • Posted Content • DOI:10.1101/2021.01.15.426639

Dopamine enhances model-free credit assignment through boosting of retrospective model-based inference — [Source link](#)

Lorenz Deserno, Rani Moran, Ying Lee, Jochen Michely ...+2 more authors

Institutions: University of Würzburg, University College London, Dresden University of Technology, Charité ...+1 more institutions

Published on: 17 Jan 2021 - bioRxiv (Cold Spring Harbor Laboratory)

Related papers:

- [Dopamine enhances model-free credit assignment through boosting of retrospective model-based inference](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/dopamine-enhances-model-free-credit-assignment-through-1gmoqysap2>

Dopamine enhances model-free credit assignment through boosting of retrospective model-based inference

Lorenz Deserno^{1,2,3*}, Rani Moran^{1,2*}, Jochen Michely^{1,2}, Ying Lee^{1,2}, Peter Dayan^{1,4,5},
Raymond J. Dolan^{1,2}

¹ Max Planck UCL Centre for Computational Psychiatry and Ageing Research, London, United Kingdom;

² Wellcome Centre for Human Neuroimaging, Institute of Neurology, University College London, London, United Kingdom;

³ Department of Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, University of Würzburg, Würzburg, Germany;

⁴ Max Planck Institute for Biological Cybernetics, Max Planck Ring 8, 72076 Tübingen, Germany;

⁵ University of Tübingen, 72074 Tübingen, Germany

* denotes equal contribution to authorship

Contact Information:

Prof. Dr. Lorenz Deserno
Margarete-Höppel-Platz 1
97080 Würzburg, Germany
deserno_l@ukw.de

Dr. Rani Moran
10-12 Russel Square
London WC1B5EH, United Kingdom
rani.moran@gmail.com

Abstract

Dopamine is implicated in signalling model-free (MF) reward prediction errors and various aspects of model-based (MB) credit assignment and choice. Recently, we showed that cooperative interactions between MB and MF systems include guidance of MF credit assignment by MB inference. Here, we used a double-blind, placebo-controlled, within-subjects design to test the hypothesis that enhancing dopamine levels, using levodopa, boosts the guidance of MF credit assignment by MB inference. We found that levodopa enhanced retrospective guidance of MF credit assignment by MB inference, without impacting on MF and MB influences per se. This drug effect positively correlated with working memory, but only in a context where reward needed to be recalled for MF credit assignment. The dopaminergic enhancement in MB-MF interactions correlated negatively with a dopamine-dependent change in MB credit assignment, possibly reflecting a potential trade-off between these two components of behavioural control. Thus, our findings demonstrate that dopamine boosts MB inference during guidance of MF learning, supported in part by working memory, but trading-off with a dopaminergic enhancement of MB credit assignment. The findings highlight a novel role for a DA influence on MB-MF interactions.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27

Introduction

Dual system theories of reinforcement learning (RL) propose behaviour is controlled by competitive and cooperative interactions between a prospective, model-based (MB), planning system and a retrospective, model-free (MF), value-caching system (Daw and Dayan, 2014; Dolan and Dayan, 2013). MF value-caching is driven by reward prediction error (RPE) signalling via phasic dopamine (DA, Montague et al., 1996; Schultz et al., 1997; Steinberg et al., 2013), a finding mirrored in human neuroimaging studies (D'Ardenne et al., 2008; O'Doherty et al., 2004). While DA RPEs are assumed to train MF values (a process we refer to as MF credit assignment or MFCA), there is evidence that DA neuromodulation also impacts MB learning (MB credit assignment or MBCA) and control (Doll et al., 2012; Langdon et al., 2018). For example, the activity of DA neurons reflects MB values (Sadacca et al., 2016), DA RPEs reflect hidden-state inference (Starkweather et al., 2017), and optogenetic activation and silencing of DA neurons impact the efficacy of MB learning (Sharpe et al., 2017). Human studies also show that higher DA levels are linked to enhanced MB influences (Deserno et al., 2015; Doll et al., 2016; Sharp et al., 2016; Wunderlich et al., 2012), which was confirmed in a non-human animal study (Groman et al., 2019), potentially mediated by a modulation in the efficiency of working memory or motivation.

RL theory has proposed cooperative interactions between MB and MF systems, including the idea that a MB controller instructs a MF system about the structure of the environment (Daw and Dayan, 2014; Mattar and Daw, 2018; Sutton, 1991). For instance, inferences made in a MB manner can disambiguate different possible states of the world in cases in which the MF system is otherwise unable to learn properly because it does not know the state. We recently provided empirical evidence for this sort of MB-MF cooperation, showing that retrospective MB inference guides MFCA via provision of knowledge regarding the environment's transition structure (Moran et al., 2019). Given DA's contribution to both MF and MB systems, we set out to examine whether this aspect of MB-MF cooperation is subject to DA influence.

1 To address this question, we used a dual-outcome bandit task (Moran et al., 2019) in
2 a double-blind, placebo-controlled, within-subjects pharmacological study, employing
3 levodopa to boost the brain's overall DA levels. This task allows a separate measurement of
4 MB and MF systems, and specifically probes guidance of MF learning based on MB knowledge
5 of the environmental transition structure. Our hypothesis was that enhancing DA would
6 strengthen the guidance of MFCA by MB inference. Importantly, at the time MB inference is
7 possible, some rewards are no longer perceptually available to participants. Thus, we expected
8 that DA-induced boosting of the MB guidance of MFCA would depend on working memory
9 capacity exclusively for perceptually absent rewards. Finally, in light of previous reports that
10 levodopa enhanced MB influences (Sharp et al., 2016; Wunderlich et al., 2012), we examined
11 whether this is also true for our dual-outcome task, expecting that inter-individual differences
12 in the effect of boosting DA on MB influences and on MB guidance of MFCA would be related.

13 Foreshadowing our results, we found that boosting DA levels via levodopa enhanced
14 guidance of a MFCA by MB inference, an effect moderated by inter-individual differences in
15 working memory but only when reward needed to be recalled. While boosting DA did not alter
16 the overall influence of a MB system on choice per se, the drug effects on guidance of MFCA
17 by MB inference and on MB choice were negatively correlated.

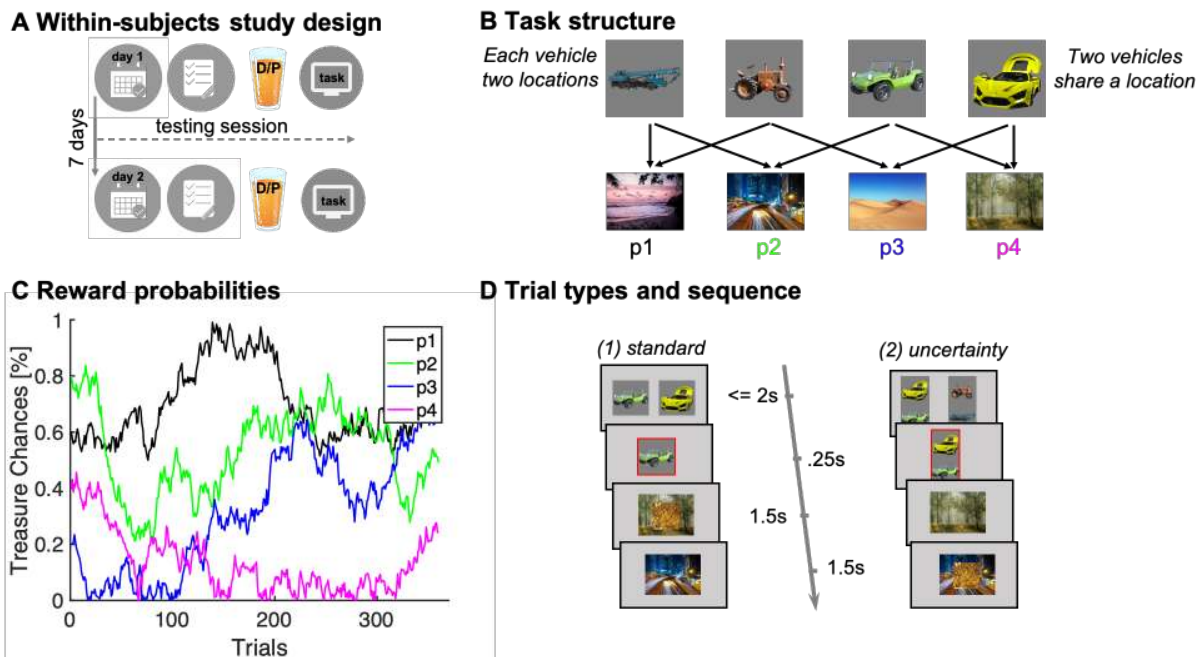
18

19

Results

20 **Study design and task logic.** We conducted a placebo-controlled, double-blind,
21 within-subjects pharmacological study using levodopa to enhance presynaptic DA levels, as
22 in previous studies (Chowdhury et al., 2013; Wunderlich et al., 2012). Participants were tested
23 twice, once under the influence of 150mg levodopa, and once on placebo, where drug order
24 was counterbalanced across individuals (n=62, Figure 1A; cf. Methods). On each lab visit,
25 participants performed a task first introduced previously by Moran et al. (2019). The task was
26 framed as a treasure hunt game called the "Magic Castle". Initially, participants were trained
27 extensively on a transition structure between states, under a cover narrative of four vehicles

1 and four destinations. Subjects learned that each vehicle (state) travelled to two different
2 sequential destinations in a random order (Figure 1B). The mapping of vehicles and
3 destinations remained stationary throughout a session, but the two test sessions featured
4 different vehicles and destinations. At each destination, participants could potentially earn a
5 reward with a probability that drifted across trials according to four independent random walks
6 (Figure 1C).



7
8 **Figure 1. A)** Illustration of within-subjects design. On each of two testing days, approximately
9 7 days apart, participants started with either a medical screening and brief physical exam (day
10 1) or a working memory test (day 2). Subsequently they drank an orange squash containing
11 either levodopa (D) or placebo (P). **B)** Task structure of the Magic Castle Game. Following a
12 choice of vehicle, participants “travelled” to two associated destinations. Each vehicle shared
13 a destination with another vehicle. At each destination, participants could win a reward (10
14 pence) with a probability that drifted slowly as Gaussian random walks, illustrated in **C)**
15 **D)** Depiction of trial types and sequences. (1) On *standard* trials (2/3 of the trials), participants
16 made a choice out of two options in trial-*n* (max. choice 2s). The choice was then highlighted
17 (.25s) and participants subsequently visited each destination (.5s displayed alone). Reward, if
18 obtained, was overlaid to each of the destinations for 1s. (2) On *uncertainty* trials, participants
19 made a choice between two pairs of vehicles. Subsequently, the ghost nominates, unbeknown
20 to the participant, one vehicle out of the chosen pair. Firstly, the participant is presented the
21 destination shared by the chosen pair of vehicles and this destination is therefore *non-*
22 *informative* about the ghost’s nominee. Secondly, the destination unique to the ghost-
23 nominated vehicle is then shown. This second destination is *informative* because it enables
24 inference of the ghost’s nominee with perfect certainty based on a MB inference that relies on
25 task transition structure. Trial timing was identical for standard and uncertainty trials.

26

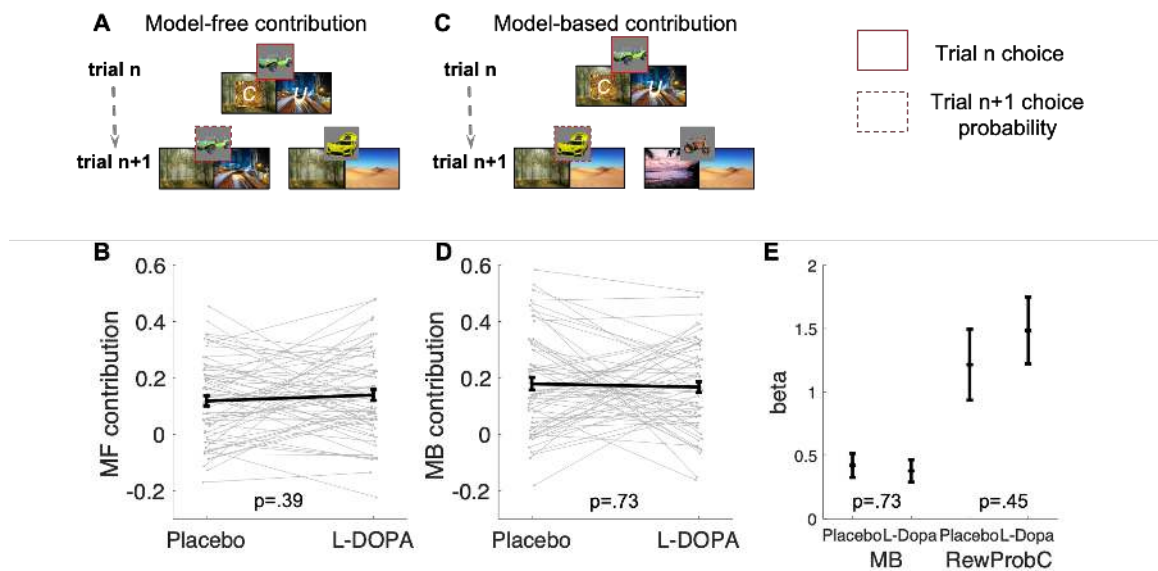
1 The task included two trial types (Figure 1D): (1) standard trials (2/3 of the trials) and
2 (2) uncertainty trials (1/3 of the trials). On standard trials, participants were offered two vehicles
3 and upon choosing one, they visited both its associated destinations where they could earn
4 rewards. On uncertainty trials, participants likewise chose a pair of vehicles (from two offered
5 vehicle-pairs). Next, an unseen ghost randomly nominated a choice of one of the vehicles in
6 the chosen pair, and a visit to its two destinations followed. Critically, participants were not
7 privity to which vehicle was nominated by the ghost. However, they could resolve this
8 uncertainty after seeing both visited destinations based on their knowledge of task transition
9 structure. We refer to this as retrospective MB inference. Such inference can only occur after
10 exposure to the second destination, as only then can subjects know which of the two vehicles
11 the ghost had originally selected.

12 We first present ‘model-agnostic’ analyses focusing on how events on trial n affect
13 choices on trial $n+1$. This allows identification of MF and MB choice signatures, the guidance
14 of MFCA by retrospective MB inference, and, crucially, whether these signatures varied as a
15 function of drug treatment (levodopa vs. placebo). These analyses are supported by validating
16 simulations using computational models as provided in a later section.

17 **Logic of model-free and model-based contributions to choices.** A MF system
18 updates values based on earned rewards only for a chosen vehicle (illustrated in Figure 2A).
19 A MB system does not maintain and update values for the vehicles directly. Instead, the MB
20 system updates the values of destinations and calculates prospectively on-demand values for
21 each offered vehicle (see computational modelling). This enables the MB system to generalize
22 value across vehicles which share a common destination (illustrated in Figure 2B).

23 **No evidence of dopaminergic modulation for MF choice repetition.** Consider a pair
24 of standard trials n and $n+1$ for which the vehicle chosen on the former is also offered on the
25 latter, against another vehicle (Figure 2A). The two vehicles offered on trial $n+1$ reach a
26 common destination, but the vehicle previously chosen on trial n also visits a unique
27 destination. In a logistic mixed effects model, we regressed a choice repetition of this vehicle
28 on whether the common and/or unique destinations were rewarded on trial n (reward/non-

1 reward) and on drug status (levodopa/placebo). Replicating a previous finding (Moran et al.,
 2 2019), we found a main effect for common reward ($b=0.67$, $t(7251)=9.14$, $p<.001$). This effect
 3 constitutes MF choice repetition, as the MB system appraises that the common destination
 4 favours both trial n+1 vehicles (see Figure S1 for validating simulations). As expected on both
 5 MB and MF grounds, there was a main effect for unique reward ($b=1.54$, $t(7251)=17.40$,
 6 $p<.001$). There was no drug x common-reward interaction ($b=0.07$, $t(7251)=.67$, $p=.500$),
 7 providing no evidence for a drug-induced change in MF choice repetition on standard trials
 8 (Figure, 2B). None of the remaining (main or interaction) effects were significant (Table S1).



9

10 **Figure 2. A)** Illustration of MF choice repetition. We consider only standard trials n+1 that offer
 11 for choice the standard trial n chosen vehicle (e.g. green antique car) alongside another vehicle
 12 (e.g. yellow racing car), sharing a common destination. Following choice of a vehicle in trial n
 13 (framed in red), participants visited two destinations of which one can be labelled on trial n+1
 14 as common to both offered vehicles (C, e.g. forest, which was also rewarded in the example)
 15 and the other labelled as unique (U, e.g. city highway, unrewarded in this example) to the trial
 16 n chosen vehicle. The trial n common-destination reward effect on the probability to repeat the
 17 previously chosen vehicle constitutes a MF choice repetition. **B)** The empirical reward effect
 18 at the common destination (i.e., the difference between rewarded and unrewarded on trial n,
 19 see Figure S3 for a more detailed plot) on repetition probability in trial n+1 is plotted for placebo
 20 and levodopa (L-DOPA) conditions. There was a positive common-reward main effect and this
 21 reward effect did not differ significantly between placebo and levodopa conditions. **C)**
 22 Illustration of the MB contribution. We considered only standard trials n+1 that excluded from
 23 the choice set the standard trial n chosen vehicle (e.g. green antique car). One of the vehicles
 24 offered on trial n+1 shared one destination in common with the trial-n chosen vehicle (e.g.,
 25 yellow racing car and we term its choice a generalization). A reward (on trial n) effect for the
 26 common destination on the probability to generalize on trial n+1 constitutes a signature of MB
 27 choice generalization. **D)** The empirical reward effect at the common destination (i.e., the
 28 difference between rewarded and unrewarded, see Figure S3 for a more detailed plot) on
 29 generalization probability is plotted for placebo and levodopa conditions. **E)** In the regression
 30 analysis described in the text, we also include the current (subject- and trial-specific) state of

1 the drifting reward probabilities (at the common destination) because we previously found this
2 was necessary to control for temporal auto correlations in rewards (Moran et al., 2019). For
3 completeness, we plot beta regression weights of reward versus no reward at the common
4 destination (indicated as MB) and for the common reward probability (RewProbC) each for
5 placebo and levodopa conditions. No significant interaction with drug session was observed.
6 Error bars correspond to SEM reflecting variability between participants.

7

8 **No evidence of dopaminergic modulation for MB choice generalization.** Consider a
9 standard trial- $n+1$, which excludes the vehicle chosen on trial n from the choice set. This trial-
10 n chosen vehicle shares a destination with one of the trial- $n+1$ offered vehicles, allowing an
11 analysis of MB choice generalization. Using a logistic mixed effects model, wherein we
12 regressed choice generalization on trial- n rewards at the common destination, on the current
13 reward probability of the common destination and on drug session, replicated our previous
14 finding (Moran et al., 2019) of a positive main effect for the common-reward ($b=0.40$,
15 $t(7177)=6.22$, $p<.001$). This positive common trial- n reward-effect on choice constitutes a MB
16 choice generalization (even after controlling for the drifting reward probability at the common
17 destination, see Figure S1 for validating simulations). The common-reward x drug interaction
18 was not significant ($b=0.05$, $t(7177)=0.39$, $p=.695$), providing no evidence for a drug-induced
19 change in MB choice (Figure 2D & E). Except for the main effect of the drifting reward
20 probability at the common destination, no other effects were significant (Table S1).

21 In summary, we replicate previous findings (Moran et al., 2019) of mutual MF and MB
22 contributions to choices. There was no evidence, however, that these contributions were
23 modulated by levodopa.

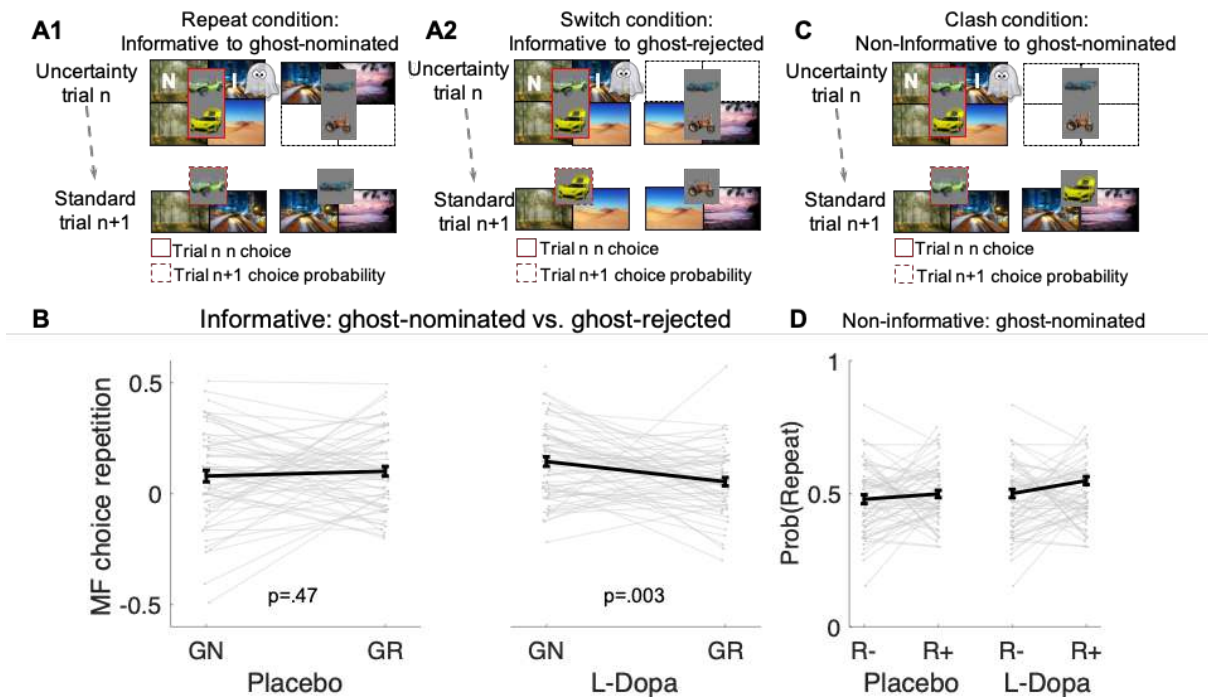
24

25 **Retrospective MB inference guides MFCA.** We next addressed our main question:
26 Does levodopa administration boost a MB guidance of MFCA through a retrospective MB
27 inference? In an uncertainty trial, participants choose one out of the two pairs of vehicles
28 (Figure 1D). Next, a ghost randomly nominates a vehicle from the chosen pair (Figure 3).
29 Participants then observe a destination common to both of the vehicles of the chosen pair,
30 followed by a destination unique to the ghost-nominated vehicle. As participants are
31 uninformed about the ghost nominee, they have a 50-50% belief initially and observing the first

1 destination is *non-informative* with respect to the ghost's nominee (as it is shared between
2 vehicles). Critically, following observation of the second destination, a MB system can infer the
3 ghost-nominated vehicle with absolute certainty based upon knowledge of the task transition
4 structure. Thus, the second destination is retrospectively *informative* with respect to inference
5 of the ghost's nominee. Subsequently, the inferred vehicle information can be shared with a
6 MF system to direct MFCA towards the ghost-nominated vehicle. We predicted guidance of
7 MFCA occurs for both vehicles in the chosen pair, but to a different extent. Specifically,
8 guidance of MFCA for the ghost-nominated, as compared to the ghost-rejected, vehicle would
9 support an hypothesis that retrospective MB inference preferentially guides MFCA (Moran et
10 al., 2019). See Figure S2 for validating model simulations. Our novel hypothesis here is that
11 this effect will be strengthened under levodopa as compared to placebo, which we examine,
12 firstly via the informative and, secondly, via the non-informative destination.

13 **Dopamine enhances preferential guidance of MFCA for the informative**
14 **destination.** MFCA for the ghost-nominated vehicle is tested in a “repeat” standard trial n+1
15 that follows an uncertainty trial n, as depicted in Figure 3 A1. MFCA of the ghost-rejected
16 vehicle is examined in a “switch” standard trial n+1 following an uncertainty trial n, as depicted
17 in Figure 3 A2. For a detailed analysis of repeat and switch trials, see Supplementary
18 Information (SI) and Figure S4. The key metric of interest for our drug analysis is the contrast
19 between MFCA for ghost-nominated *versus* ghost-rejected vehicles, based on the reward
20 effects at the informative destination in repeat and switch trials (repeat or ghost-nominated /
21 switch or ghost-rejected), separately for each nomination trial type (repeat/switch) x drug
22 condition (levodopa /placebo) (Figure 3B). In a mixed effects model (Table S2), we found no
23 main effect either of nomination ($b=.043$ $t(239)=1.60$, $p=.110$) or of drug ($b=.01$, $t(239)=.40$,
24 $p=.690$). Crucially, we found a significant nomination x drug interaction ($b=.11$, $t(239)=2.56$,
25 $p=.011$). A simple effects analysis revealed a preferential MFCA of the ghost-nominated over
26 the ghost-rejected vehicle was significant under levodopa ($b=.09$, $F(243,1)=9.07$, $p=.003$) but
27 not under placebo ($b=-.02$, $F(243,1)=.53$, $p=.472$). This supports our hypothesis that levodopa
28 preferentially enhanced MFCA for the ghost-nominated, compared to ghost-rejected, vehicle

1 under the guidance of retrospective MB inference. The nomination x drug interaction was not
 2 affected by session order (see Table S2).



3
 4 **Figure 3.** In an uncertainty trial n, participants choose a pair of vehicles. The ghost nominates
 5 one vehicle out of this pair (e.g., green antique car). Participants have a chance belief about
 6 the ghost-nominated vehicle. The firstly presented destination holds no information about the
 7 ghost-nominated vehicle, the non-informative ("N") destination. The destination presented
 8 second enables retrospective MB inference about the ghost's nomination and is therefore
 9 informative ("I"). **A1.** Illustration of the repeat condition. The ghost-nominated vehicle (e.g.,
 10 green antique car) is offered for choice in standard trial n+1 alongside a vehicle from the non-
 11 chosen pair (e.g., blue building crane). A higher probability to repeat the ghost-nominated
 12 vehicle in standard trial n+1 after a reward as compared to no reward at the informative
 13 destination constitutes MFCA for the ghost's nomination (GN). **A2.** Illustration of the switch
 14 condition. The ghost-rejected vehicle (e.g., the yellow racing car) is offered for choice in
 15 standard trial n+1 alongside a vehicle from the non-chosen pair (e.g. brown farming tractor). A
 16 higher probability to choose the ghost-rejected vehicle in standard trial n+1 after a reward as
 17 compared to no reward at the informative destination constitutes MFCA for the ghost's
 18 rejection (GR). Both ghost-based assignments depend on retrospective MB inference. **B.**
 19 Preferential effect of retrospective MB inference on MFCA (effects of GN>GR) based on the
 20 informative destination is enhanced under levodopa (L-Dopa) as compared to placebo. This is
 21 indicated by a significant trial type (GN/GR) x drug (placebo/ levodopa) interaction. Under
 22 levodopa, MFCA for GN is significantly higher than of GR, which is not the case under placebo
 23 (see Figure S4 for a more detailed plot). **C.** Illustration of the clash condition. The previously
 24 chosen pair is offered for choice in standard trial n+1. A higher probability to repeat the ghost-
 25 nominated vehicle in standard trial n+1 following reward (relative to non-reward) at the non-
 26 informative destination constitutes a signature of preferential MFCA for GN over GR. **D.** Choice
 27 repetition in clash trial is plotted as a function of reward and drug-group (see Figure S5 for a
 28 more detailed plot). While there was a main effect for drug, there was no interaction of non-
 29 informative reward x drug, providing no evidence that drug modulated MFCA based on the
 30 non-informative outcome. R+: reward; R-: non-reward. Error bars correspond to SEM reflecting
 31 variability between participants.

32

1 **Dopaminergic modulation of preferential MFCA for the non-informative**
2 **destination.** A second means to examine MB influences over MFCA is to consider the non-
3 informative destination. In a standard “clash” trial- $n+1$ following an uncertainty trial- n , the
4 ghost-nominated vehicle is offered for choice alongside the ghost-rejected vehicle as depicted
5 in Figure 3C. We previously showed that a positive effect of reward at the non-informative
6 destination on choice repetition (i.e., a choice of the previously ghost-nominated vehicle)
7 implicates a preferential guidance of MFCA towards the ghost-nominated vehicle guided by
8 retrospective MB inference (Moran et al., 2019). In contrast, a MB system has knowledge that
9 a non-informative destination is common to both standard trial $n+1$ vehicles. Note, this effect
10 of reward at the non-informative destination can only occur when uncertainty about the ghost’s
11 nomination was resolved retrospectively, once the informative destination was encountered.

12 In a logistic mixed effects model, we regressed choice repetition on trial- n rewards at
13 informative and non-informative destinations as well as on drug session. A marginally
14 significant main effect for the reward at the non-informative destination provides some support
15 for preferential MFCA of the ghost-nominated vehicle ($b=0.13$, $t(4861)=1.96$, $p=.051$).
16 Additionally, we found a main effect for reward at the informative destination ($b=1.01$,
17 $t(4861)=9.95$, $p<.001$), as predicted by both the enhanced MFCA for the ghost-nominated
18 vehicle and by an MB contribution. The interaction effect between drug and non-informative
19 reward, however, was not significant ($b=0.05$, $t(4861)=.39$, $p=.696$, Figure 3D), nor were any
20 other interactions in the model (Table S2). This analysis yielded no evidence that levodopa
21 enhanced preferential guidance of MFCA based on reward at a non-informative destination.
22 Unexpectedly, we found a positive main effect of drug ($b=0.15$, $t(4861)=2.31$, $p=.021$, Figure
23 3D), indicating that participants’ tendency to repeat choices of the ghost-nominated vehicle
24 was generally enhanced under levodopa, but this finding that was only seen in this specific
25 subset of trials and could not be corroborated based on computational modelling. We further
26 dissect effects at the non-informative destination, in particular with respect to inter-individual
27 differences in working memory, using computational modelling.

1 **Computational Modelling.** One limitation of the analyses reported above is that they
2 isolate the effects of the immediately preceding trial on a current choice. However, values and
3 actions of RL agents are influenced by an entire task history and, to take account of such
4 extended effects, we formulated a computational model that specified the likelihood of choices
5 (Moran et al., 2019, also see Moran et al., in press, 2021). In brief, at choice, MF values (Q^{MF})
6 of the two presented vehicles feed into a decision module. During learning, the MF system
7 updates Q^{MF} of the *chosen* vehicle based on earned rewards alone. By contrast, the MB
8 system prospectively calculates on-demand Q^{MB} -values for each offered vehicle based on an
9 arithmetic sum of the values of its two destinations:

10 (Eq. 1) $Q^{MB}(\text{vehicle}) = Q^{MB}(\text{corresponding destination 1}) + Q^{MB}(\text{corresponding destination 2})$

11 During learning, the MB system updates the values of the two visited destinations. We
12 refer to these updates as MB credit assignment (MBCA). Unlike MFCA, which does not
13 generalize credit from one vehicle to another, MBCA generalizes across the two vehicles which
14 share a common destination. Thus, when a reward is collected in the forest destination,
15 $Q^{MB}(\text{forest})$ increases. As the forest is a shared destination, both vehicles that lead to this
16 destination benefit during ensuing calculations of the on-demand Q^{MB} -values. Critically, our
17 model included five free “MFCA parameters” of focal interest, quantifying the extent of MFCA
18 on standard trials (one parameter), on uncertainty trials (four parameters) for each of the
19 objects in the chosen pair (nominated/rejected), and for each destination (informative/non-
20 informative). We verified that the inclusion of these parameters was warranted using
21 systematic model comparisons. A description of the sub-models and the model selection
22 procedure is reported in the methods section and in Figure S6. We fitted our full model to each
23 participant’s data in drug and placebo sessions based on Maximum Likelihood Estimation (see
24 methods).

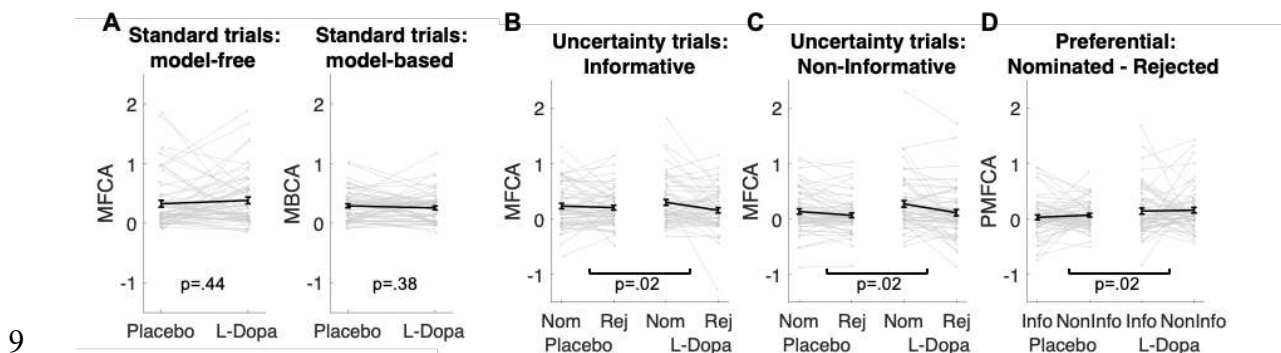
25 **Absence of dopaminergic modulation for MBCA and MFCA on standard trials.** In
26 line with our model-agnostic analyses of standard trials, we found positive contributions of
27 MFCA (parameter c_{standard}^{MF} ; Fig. 4A) for both levodopa ($M = 0.381$, $t(61) = 6.84$, $p < .001$) and

1 placebo ($M = 0.326$, $t(61) = 5.76$, $p < .001$), with no difference between drug conditions ($t(61) = -$
2 0.78 , $p = .442$). Likewise, MBCA (parameter c^{MB} ; Fig. 4B) contributed positively for both
3 levodopa ($M = 0.255$, $t(61) = 7.88$, $p < .001$) and placebo ($M = 0.29$, $t(61) = 8.88$, $p < .001$), with no
4 significant difference between drugs ($t(61) = 0.88$, $p = .3838$). Thus, while both MBCA and MFCA
5 contribute to choice, there was no evidence for a drug-related modulation. Forgetting and
6 perseveration parameters of the model did not differ as a function of drug (see SI).

7 **Levodopa enhances guidance of preferential MFCA by retrospective MB**
8 **inference on uncertainty trials.** To test our key hypothesis, that guidance of preferential
9 MFCA by retrospective MB inference on uncertainty trials is enhanced by levodopa, we
10 focused on the four computational parameters that pertaining to MFCA on uncertainty trials
11 ($c_{\text{nom,info}}^{MF}$, $c_{\text{rej,info}}^{MF}$, $c_{\text{nom,noninfo}}^{MF}$, $c_{\text{rej,noninfo}}^{MF}$, Figure 4B,C). In a mixed effects model, we regressed
12 these MFCA parameters on their underlying features: nomination (nominated / rejected),
13 informativeness (informative / non-informative) and drug session (levodopa / placebo).
14 Crucially, we found a positive nomination x drug interaction ($b = 0.10$, $t(480) = 2.43$, $p = .015$). A
15 simple effects analysis revealed preferential MFCA (the effect of nomination) to be significant
16 under levodopa ($b = .13$, $F(488,1) = 9.71$, $p = .002$), and stronger than in the placebo condition
17 ($b = 0.08$, $F(488,1) = 4.83$, $p = .029$), indicating that preferential MFCA was stronger under
18 levodopa as compared to placebo. Importantly, this interaction was not qualified by a triple
19 interaction ($b = .02$, $t(480) = 0.32$, $p = .738$), providing no evidence that the extent of preferential
20 MFCA differed for informative and non-informative outcomes. No other effect pertaining to drug
21 reached significance (Table S3).

22 To examine in more fine-grained detail whether a MFCA is indeed preferential, we
23 calculated, for each participant, in each session (drug/placebo), and for each level of
24 informativeness (informative/non-informative), the extent to which MFCA was preferential for
25 the ghost-nominated as opposed to the ghost-rejected vehicle (as quantified by $c_{\text{nom,info}}^{MF} -$
26 $c_{\text{rej,info}}^{MF}$, $c_{\text{nom,noninfo}}^{MF} - c_{\text{rej,noninfo}}^{MF}$; Figure 4D). Using a mixed effects model, we regressed
27 preferential MFCA (PMFCA), based on MB guidance on informativeness and drug session.

1 We found a positive main effect for drug ($b=0.10$, $t(240)=2.41$, $p=.017$), but neither the main
 2 effect of informativeness ($b=-0.03$, $t(240)=-0.57$, $p=.568$) nor the informativeness x drug
 3 interaction ($b=.02$, $t(240)=0.33$, $p=.739$) were significant. Using simple effects, MFCA preferred
 4 the ghost-nominated vehicle in the levodopa condition ($b=0.15$, $F(1,244)=15.45$, $p<.001$),
 5 while the same effect was only marginally significant in the placebo condition ($b=0.05$,
 6 $F(1,244)=2.86$, one-sided $p=.046$). Thus, our computational modelling analysis indicates that
 7 preferential MFCA is boosted by levodopa as compared to placebo across informative and
 8 non-informative destinations.



10 **Figure 4.** Analyses based on estimated credit assignment (CA) parameters from
 11 computational modelling. **A)** Model-free and model-based credit assignment parameters
 12 (MFCA; MBCA) did not differ significantly for placebo and levodopa conditions. **B)** MFCA
 13 parameters based on the informative outcome for the ghost-nominated and the ghost-rejected
 14 destinations as a function of drug condition. **D)** Same as C but for the non-informative
 15 destination. **E)** The extent to which MFCA prefers the nominated over the rejected vehicle for
 16 each destination and drug condition. We name this preferential MFCA (PMFCA).

17

18 **Drug effect correlates positively with working memory only for reward at the non-**

19 **informative destination.** We hypothesized that working memory (WM) would moderate the

20 boosting effect of levodopa, but only based on reward at the non-informative destination. When

21 the informative destination is delivered on uncertainty trials, a MB system can infer the hidden

22 choice and guide PMFCA. PMFCA based on reward at the non-informative destination can

23 prefer the ghost-nominated vehicle only if it is at least partially postponed until uncertainty has

24 been resolved by retrospective MB inference, in other words after delivery of the informative

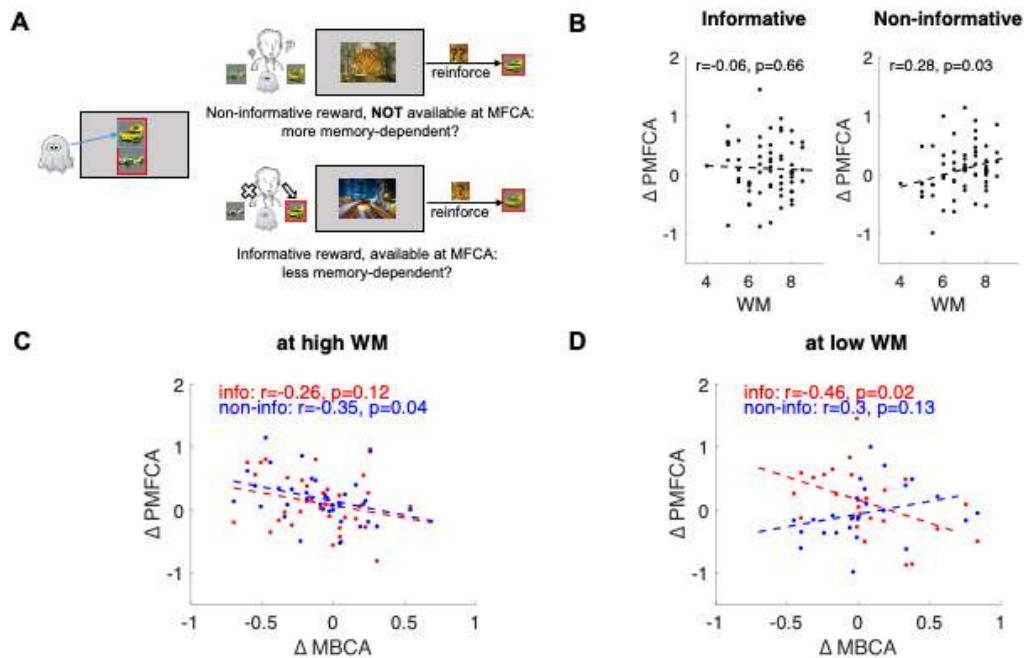
25 destination. At this time, reward received at the non-informative destination is no longer

26 perceptually available and needs to be recalled (as illustrated in Figure 5A). Subjects' WM

1 capacity, as ascertained with the digit span test, showed a positive across-participants
2 Spearman correlation with the drug effect (levodopa vs placebo) on PMFCA in the non-
3 informative ($r = .278$, $p = .029$, Figure 5B), but not for the informative destination ($r = -.057$,
4 $p = .659$, Figure 5B). The difference between these correlations was significant ($p = .044$,
5 permutation test; see methods). There was no significant correlation of WM capacity with drug-
6 induced change in MBCA or with MBCA at levodopa or placebo (see SI).

7 **Inter-individual differences in drug effects.** Previous studies, using a task that
8 cannot dissociate cooperative and competitive interactions between MB and MF systems,
9 reported that boosting DA levels leads to enhanced MB choices (Sharp et al., 2016;
10 Wunderlich et al., 2012), an effect we did not observe at a group level on our measure of
11 MBCA. To explore the possibility that drug effects in different task conditions (guidance of
12 MFCA vs. MBCA) are related, we analyzed inter-individual differences in the effects of boosting
13 DA levels on guidance of MFCA and on MBCA. Because WM capacity correlated positively
14 with drug effects at the non-informative destination as reported above, we included WM in the
15 analysis of inter-individual differences in drug effects. Thus, we regressed DA-dependent
16 differences (levodopa vs placebo) in PMFCA against informativeness, DA-dependent
17 differences in MBCA and WM capacity. This model revealed an informativeness \times MBCA \times
18 WM interaction ($b = 0.16$, $t(116) = 2.16$, $p = .032$). To unpack the interaction, we ran the model
19 separately at high and low WM capacity based on a median split. In individuals with high WM
20 capacity, this revealed a negative main effect of MBCA ($b = -0.13$, $t(48) = -2.45$, $p = .018$, see
21 Figure 5C) which was not qualified by an interaction between informativeness \times MBCA ($b = -$
22 0.07 , $t(48) = -0.86$, $p = .40$). This means that, for high WM individuals, the drug-effects on PMFCA
23 and MBCA are negatively related for informative and non-informative destinations. In contrast,
24 in individuals with low WM capacity, there was a significant negative informativeness \times MBCA
25 interaction ($b = -0.23$, $t(68) = -2.43$, $p = .018$; Figure 5D). A simple effects analysis revealed that
26 the drug-effect on MBCA had a significant negative relation on the drug effect on PMFCA for
27 the informative destination ($b = -.18$, $F(1,68) = 6.13$, $p = .015$; Figure 5D) but not for the non-
28 informative destination ($b = .05$, $F(1,68) = 0.42$, $p = .517$; Figure 5D). Using model-agnostic

1 metrics of DA-dependent change in guidance of MFCA and in MB choice, the negative
2 correlation was also significant (see Figure S7). These inter-individual differences may reflect
3 a trade-off between PMFCA and MBCA under boosted DA levels.



4

5 **Figure 5.** Inter-individual differences. **A)** Illustration of MFCA based on rewards at informative
6 and non-informative destination. The latter is likely to depend more on memory recall because
7 the reward is no longer perceptually available when MFCA can take place (after state
8 uncertainty was resolved). **B)** Scatter plots of the drug effect (levodopa minus placebo) on
9 preferential MFCA (Δ PMFCA) based on the informative destination reward and for the non-
10 informative destination reward against working memory (WM). **C)** Scatter plot of the drug effect
11 (levodopa minus placebo) on preferential MFCA (Δ PMFCA) based on the informative
12 destination reward (info, red) and for the non-informative destination reward (non-info, blue)
13 against drug-induced change in MBCA (Δ MBCA) at high working memory (WM) capacity. **D)**
14 Same scatter plot as in C) but at low working memory (WM) capacity. In panels B, C and D
15 regression lines are dashed. r refers to the Spearman correlation coefficient in panel B and
16 Pearson correlation coefficient in C and D.

17

18

19

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

Discussion

We show that enhancing dopamine boosted the guidance of model-free credit assignment by retrospective model-based inference. This pharmacological effect was associated with higher working memory capacity just for rewards that were no longer perceptually available and had to be recalled for credit assignment to be correct. Whereas both MF and MB influences were unaffected by the drug manipulation at the group level, analysis of inter-individual differences in drug effects showed that enhanced guidance of MFCA by retrospective MB inference was negatively correlated with drug-related change in MBCA. The findings provide, to our knowledge, the first human evidence that DA directly influences cooperative interactions between MB and MF systems, highlighting a novel role for DA in how MB information guides MFCA.

The effect of levodopa on prefrontal DA levels can lead to the enhancement of general aspects of cognition, for example WM (Cools and D'Esposito, 2011), probably depending on DA synthesis capacity in an inverted U-curved manner. The latter is likely to be important for supporting the computationally sophisticated operation of a MB system (Otto et al., 2013). One might therefore expect a primary drug effect on prefrontal DA to result in boosted MB influences (Sharpe et al., 2017; Wunderlich et al., 2012) – but we found no such influence. Equally, a long-standing proposal that phasic DA relates to a MF learning signal might predict that the main effect of the drug would be to speed or bias MF learning (Pessiglione et al., 2006). We observed no such effect, nor has it been seen in two previous studies (Sharp et al., 2016; Wunderlich et al., 2012). Instead, we found levodopa had a more specific influence, impacting the preferential MB guidance of MFCA in a situation where individuals needed to rely on retrospective MB inference to resolve state uncertainty. Thus, MB instruction about *what* (unobserved or inferred) state the MF system might *learn about*, was boosted under levodopa. In other words, DA boosts an exploitation of a model of task structure so as to facilitate retrospective learning about the past. These findings indicate an enhanced integration of MB information in DA signalling (Sadacca et al., 2016). Our results thus may provide a fine-grained view of the various processes involved – with the specificities of our task allowing us

1 to separate out a rather particular component of WM, and an important, but restricted influence
2 of MB information on MFCA.

3 First, preferential MFCA based on reward at the uninformative destination can only take
4 place after seeing the informative destination and inferring the ghost's choice. Thus, the
5 uninformative destination's reward has to be maintained in WM to support preferential MFCA.
6 In other words, an ability to maintain information in working memory is a prerequisite for a DA-
7 dependent boosting of preferential MFCA based on the uninformative destination. In line with
8 this, we found a DA-boosting of MB guidance of MFCA depended on WM for the non-
9 informative destination alone. This underlines the importance of accounting for inter-individual
10 differences in supportive cognitive processes particularly when it comes to providing a detailed
11 understanding of DA drug effects of interest (Cools, 2019; Kroemer et al., 2019).

12 Second, given that the information about the uninformative destination is stored in WM,
13 what might be the neural mechanisms associated with its use in MB guidance of MFCA. Animal
14 and human work points to a crucial role for orbitofrontal cortex in representing the model of a
15 task model, including unobserved and inferred states, and in guiding behaviour accordingly
16 (Howard et al., 2020; Jones et al., 2012; Schuck et al., 2016). This orbitofrontal function has
17 also been related to the degree of sequential offline replay in the hippocampus (Schuck and
18 Niv, 2019). Theoretical treatments of hippocampal offline neural replay proposes it informs
19 credit assignment based on RPE (Mattar and Daw, 2018), a suggestion gaining support in
20 recent empirical evidence in humans (Eldar et al., 2020; Liu et al., 2019, 2020). In our task,
21 offline replay seems especially necessary to support preferential MFCA based on the first,
22 uninformative, destination, because at this stage participants are still uncertain about the
23 ghost's choice. Under this account, we would predict enhanced offline replay (during rest
24 between trials) of the non-informative destination (including its reward) and the inferred ghost's
25 choice under the influence of L-Dopa. Whether this enhanced replay occurs indirectly, via the
26 interaction with WM, or is also a direct consequence of the L-Dopa is a pressing question for
27 future work.

1 Previous studies, using a task not designed to test cooperative interactions between
2 MB and MF systems (Daw et al., 2011), indicated a positive relationship between boosted DA
3 and MB contributions to choice (Deserno et al., 2015; Doll et al., 2012; Sharp et al., 2016;
4 Wunderlich et al., 2012). While MB choice contributions were not elevated at the group level
5 by the drug in our data, we found a negative correlation between drug-related change on these
6 contributions and on MB guidance of MFCA, in keeping with a trade-off between DA influences
7 on these two components of behavioural control. In arbitrating between MB choice and
8 retrospective MB inference to guide MFCA, participants need to weigh their respective
9 cognitive costs vs. instrumental value. In independent recent work, a balance of costs and
10 benefits was recently shown to be modulated by DA (Westbrook et al., 2020). Future studies
11 will be needed to detail how the relative costs of planning vs. retrospective state-inference are
12 influenced by DA, which can also inform DA contributions to trade-offs pertaining to strategy
13 selection.

14 A limitation in our study is that guidance of informative MFCA by MB inference was
15 significant in the levodopa condition alone but not in the placebo condition in model-agnostic
16 measures (which are based on a subset of trials and consider only very recent influences on
17 choice). However, computational modelling, informed by the entire trial-by-trial history of one's
18 experiences is arguably more sensitive, and this consideration enabled us to capture a
19 preferential guidance of MFCA by MB inference also in the placebo condition.

20 In sum, our study provides first evidence that DA enhances cooperative interactions
21 between MB and MF systems. The finding provides a unified perspective on previous
22 research in humans and animals, suggesting a closely integrated architecture of how MF and
23 MB systems interact under the guidance of DA-mediated so as to improve learning. DA-
24 mediated cooperation between MB and MF control is a potentially exciting target for
25 disentangling the precise role played by MB control in the development of impulsive and
26 compulsive psychiatric symptoms.

27

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

Methods

Procedures. A total of 64 participants (32 females) completed a bandit at each of the two sessions with drug or placebo in counterbalanced order in a double-blinded design. One participant failed to reach required performance during training (see below) and task data could not be collected. Out of remaining 63 participants, one participant experienced side effects during task performance and was therefore excluded. Results reported above are based on a sample of n=62. All participants attended on two sessions approximately 1 week apart. Participants were screened to have no psychiatric or somatic condition, no regular intake of medication before invitation and received a short on-site medical screening at the beginning of their day 1 visit. At the beginning of the day 2 visit, they performed a working memory test, the digit span, which was thus only collected once.

Drug protocol. The order of drug and placebo was counterbalanced. The protocol contained two decision-making tasks, which started at least 60min after ingestion of either levodopa (150 mg of levodopa + 37.5 mg of benserazide dispersed in orange squash) or placebo (orange squash alone with ascorbic acid). Benserazide reduces peripheral metabolism of levodopa, thus, leads to higher levels of DA in the brain and minimizes side effects such as nausea and vomiting. To achieve comparable drug absorption across individuals, subjects were instructed not to eat for up to 2h before commencing the study. Repeated physiological measurements (blood pressure and heart rate) and subjective mood rating scales were recorded under placebo and levodopa. A doctor prepared the orange squash such that data collection was double-blinded.

Task Description. Participants were introduced to a minor variant of a task developed by Moran et al. (2019) using pictures of vehicles and destinations rather than objects and coloured rooms, and lasting slightly less time. The was presented as a treasure hunt called the 'Magic Castle'. Before playing the main task, all participants were instructed that they can choose out of four vehicles from the Magic Castle's garage that each vehicle could take them to two destinations (see Figure 1B). The mapping between vehicles and destination was randomly created for each participant and each session (sessions also had different sets of

1 stimuli) but remained fixed for one session. They were then extensively trained on the specific
2 vehicle-destination mapping. In this training, participants first saw a vehicle and had to press
3 the space bar in self-paced time to subsequently visits the two associated destinations in
4 random. The initial training run contained 12 repetitions per vehicle-destination mapping (48
5 trials). This training was followed by two types of each 8 quiz trials which asked to match one
6 destination out of two to a vehicle or to match a vehicle out of two to a destination (time limit
7 of 3sec). Each quiz trial had to be answered correctly and in time otherwise another training
8 session was started with only 4 repetitions per vehicle-destination mapping (16 trials) followed
9 again by the quiz. This procedure was repeated until participants passed all quiz. Participants
10 were then introduced to the general structure of standard trials of bandit task (18 practice
11 trials). This was followed by instructions introducing the ghost trials, which were complemented
12 by another 16 practice trials including standard and ghost trials. Before starting the main
13 experiment, participants performed a shorter refresher training of the vehicle-destination
14 mapping with 4 repetitions per vehicle-destination mapping followed by the same quiz trials to
15 passed as described above. In case of not passing at this stage, the refresher training was
16 repeated with 2 repetitions per vehicle-destination mapping until the quiz was passed.

17 During the subsequent main task, participants should try to maximize their earnings. In
18 each trial, they could probabilistically find a treasure (reward) at each of the two destinations
19 (worth 1 penny). Reward probabilities varied over time independently for each of the four
20 destinations according to Gaussian random walks with boundaries at $p=0$ and $p=1$ and a
21 standard deviation of .025 per trial (Figure 1C). Random walks were generated anew per
22 participant and session. A total of 360 trials split in 5 blocks of each 72 trials were played with
23 short enforced breaks between blocks. Two of three trials were 'standard trials', in which a
24 random pair of objects was offered for choice sharing one common outcome (choice time \leq
25 2s). After making a choice, they visited each destination subsequently in random order. Each
26 destination was presented for 1s and overlaid with treasure or not (indicating a reward or not).
27 The lag between the logged choice and the first destination as well as between first and second
28 destinations was 500ms. Every third trial was an "uncertainty trial" in which two disjoint pairs

1 of vehicles were offered for choice. Crucially, each of the presented pairs of vehicles shared
2 one common outcome. Participants were told before the main task that after their choice of a
3 pair of vehicles, the ghost of the Magic Castle would randomly pick one vehicle out of the
4 chosen pair. Because this ghost was transparent, participants could not see the ghost's choice.
5 However, participants visited the two destinations subsequently and collected treasure reward
6 (or not). Essentially, when the ghost nominated a vehicle, the common destination was
7 presented first and the destination unique to this vehicle was presented second. At this time of
8 presentation of the unique destination, participants could retrospectively infer the choice made
9 by the ghost. Trial timing was identical for standard and ghost trials. The 120 standard trials
10 following a previous trial $n-1$ standard trial included 30 presentations of each of the four eligible
11 pairs of vehicles in a random order. The 120 uncertainty trials included 60 presentations of the
12 two eligible pairings in a random order. The standard trials following uncertainty trials were
13 defined according to the observed transition based on the (ghost's) choice in the preceding
14 (uncertainty) trial. These 120 trials contained 40 presentations of each of the "repeat", "switch"
15 or "clash" trial types in a random order. A repeat trial presented the ghost-nominated object
16 alongside its vertical counterpart, a switch trial presented the ghost-rejected object alongside
17 its vertical counterpart and a clash trial presented the previously selected pair.

18 **Model-agnostic analysis.** Model agnostic analyses were performed with logistic
19 mixed effects models using MATLAB's "fitglme" function with participants serving as random
20 effects with a free covariance matrix. All models included the variable ORDER as regressor
21 (coded as +.5 for the first and -.5 for the second session) to control for unspecific effects and
22 participants (PART) served as random effects. Details of are reported in Table S1.

23 The analysis of MF and MB contributions is restricted to standard trials followed by a
24 standard trial. For MF contributions, we consider only a trial- $n+1$, which offers the trial- n chosen
25 object for choice (against another object). Regressors C (common destination) and U (unique
26 destination) indicated whether rewards were received at trial n (coded as +.5 for reward and -
27 .5 for no reward) and were included to predict the variable REPEAT indicating whether the
28 previously chosen vehicle was repeated or not. The variable DRUG was included as regressor

1 indicating within-subject Levodopa or placebo session (coded as +.5 for levopdopa and -.5 for
2 placebo). The model, in Wilkinson notation, can be found on Table S1. For MB contributions,
3 we specifically examined trials in which the trial-n chosen vehicle was excluded on trial n+1.
4 The regressors C, PART and DRUG were coded as for the analysis of the MF contribution.
5 One additional regressor P was included, which coded the reward probability of the common
6 destination and was centralized by subtracting .5. These regressors were included to predict
7 the variable GENERALIZE indicated whether the choice on trial n+1 was generalized
8 (choosing the vehicle not shown in trial n+1 that shares a destination with the trial-n chosen
9 vehicle). The model, in Wilkinson notation, can be found on Table S1.

10 The analysis of how retrospective MB inference preferentially guides MFCA focused
11 on standard trials following uncertainty trials. The key analysis reported above focuses on MF
12 choice repetition for the ghost-nominated in contrast to the ghost-rejected vehicle. This was
13 achieved by extracting empirical choice proportions from “repeat trials” and from “switch trials”.
14 More specifically, we computed the proportion of repeating or switching after a reward minus
15 no reward at the informative destination averaged across rewards at the non-informative
16 destination (reflecting the main effect of the informative destination, “I”) for each trial type.
17 These two metrics were subjected to a mixed-effects models as dependent variable and with
18 TYPE (nominated / rejected coded as +.5 and -.5) and, as before, DRUG and PART as
19 predictors. The model, in Wilkinson notation, can be found on Table S2. A detailed analysis
20 using separate mixed effects models for repeat and switch conditions is reported in the SI.

21 Another model-agnostic analysis examined learning for the ghost-nominated and -
22 rejected vehicles based on the uncertainty trial n non-informative destination and therefore
23 focused on n+1 “clash” trials, which offer for choice the same pair of objects as chosen on the
24 previous uncertainty trial (the ghost-nominated and ghost-rejected objects). Choice repetition
25 was defined as choice of the ghost-nominated vehicle from uncertainty trial n indicated by the
26 variable REPEAT. Regressors PART, N, I and DRUG are coded as previously. The model, in
27 Wilkinson notation, can be found on Table S2.

1 **Computational Models.** We formulated a hybrid RL model to account for the series of
2 choices for each participant. In the model, choices are contributed by both the MB and MF
3 systems. The MF system caches a Q^{MF} -value for each vehicle, subsequently retrieved when
4 the vehicle is offered for choice. During learning on standard trials, following reward-feedback,
5 rewards from the two visited destinations are used to update the Q^{MF} -value for the chosen
6 vehicle as follows:

$$7 \quad (Eq. 2) \quad Q^{MF}(\text{chosen vehicle}) \leftarrow (1 - f^{MF}) * Q^{MF}(\text{chosen vehicle}) + c_{standard}^{MF} * (r_1 + r_2)$$

8 where $c_{standard}^{MF}$ is a free MFCA parameter on standard trials and the r 's are the rewards for
9 each of the two obtained outcomes (coded as 1 for reward or -1 for non-reward) and
10 f^{MF} (between 0-1) is a free parameter corresponding to forgetting in the MF system.

11 During learning on uncertainty trials, the MF values of the ghost nominated and ghost
12 rejected options were updated according to:

$$13 \quad (Eq. 3) \quad Q^{MF}(\text{nominated vehicle}) \\ 14 \quad \quad \quad \leftarrow (1 - f^{MF}) * Q^{MF}(\text{nominated vehicle}) + c_{nom,info}^{MF} * r_{info} + c_{nom,noninfo}^{MF} \\ 15 \quad \quad \quad * r_{noninfo}$$

$$16 \quad (Eq. 4) \quad Q^{MF}(\text{rejected vehicle}) \\ 17 \quad \quad \quad \leftarrow (1 - f^{MF}) * Q^{MF}(\text{rejected vehicle}) + c_{rej,info}^{MF} * r_{info} + c_{rej,noninfo}^{MF} * r_{noninfo}$$

18 Where the c 's are free MFCA parameters on uncertainty trials for each destination
19 (informative/non-informative) and vehicle type (ghost nominated/rejected) in the chosen pair.
20 The r 's are rewards (once more, coded as 1 or -1) for the informative and non-informative
21 outcomes.

22 The MF values of the remaining vehicles (3 on standard trials; 2 on uncertainty trials)
23 were subject to forgetting:

$$24 \quad (Eq. 5) \quad Q^{MF}(\text{non chosen vehicles}) \leftarrow (1 - f^{MF}) * Q^{MF}(\text{non chosen vehicles})$$

1 Unlike MF, the MB system maintains Q^{MB} -values for the four different destinations.
2 During choices the Q^{MB} - value for each offered vehicle is calculated based on the transition
3 structure (i.e., the two destinations associated with a vehicle):

$$4 \quad (Eq. 6) \quad Q^{MB}(\text{vehicle}) = Q^{MB}(\text{destination 1}) + Q^{MB}(\text{destination 2})$$

5 Following a choice (on both standard and uncertainty trials), the MB system updates the Q^{MB} -
6 values of each of the two observed destination based on its own reward:

$$7 \quad (Eq. 8) \quad Q^{MB}(\text{destination}) \leftarrow (1 - f^{MB}) * Q^{MB}(\text{destination}) + c^{MB} * r$$

8 Where f^{MB} (bet. 0-1) is a free parameter corresponding to forgetting in the MB system, c^{MB} is
9 a free MBCA parameter and r corresponds to the reward (1 or -1) obtained at the destination.

10 Our model additionally included progressive perseveration for vehicles. After each
11 standard trial the perseveration values of each of the 4 vehicles updated according to

$$12 \quad (Eq. 9) \quad PERS(\text{vehicle}) \leftarrow (1 - f^P) * PERS(\text{vehicle}) + pr_{\text{standard}} * 1_{\text{vehicle}=\text{chosen}}$$

13 Where $1_{\text{vehicle}=\text{chosen}}$ is the chosen vehicle indicator, pr_{standard} is a free perseveration
14 parameter for standard trials, and f^P (bet. 0-1) is a free perseveration forgetting parameter.

15 Similarly after each uncertainty trials perseverations values were updated according to:

$$16 \quad (Eq. 10) \quad PERS(\text{vehicle}) \leftarrow (1 - f^P) * PERS(\text{vehicle}) + pr_{\text{uncertainty}} * 1_{\text{vehicle}=\text{nom}}$$

17 where $1_{\text{vehicle}=\text{nom}}$ is the ghost-nominated vehicle indicator, and $pr_{\text{uncertainty}}$ is a free
18 perseveration parameter for uncertainty trials.

19 During a standard trial choice a net Q value was calculated for each offered vehicle:

$$20 \quad (Eq. 11) \quad Q_{\text{net}}(\text{vehicle}) = Q^{MB}(\text{vehicle}) + Q^{MF}(\text{vehicle}) + PERS(\text{vehicle})$$

21 Similarly, during an uncertainty-trial choice the Q_{net} value of each offered vehicle-pair was
22 calculated as a sum of the MB, MF and PERS values of that pair. MF, MB, and PERS values
23 for a vehicle-pair in turn were each calculated as the corresponding average value of the two
24 vehicles in that pair. For example:

1
$$(Eq. 12) Q^{MF}(\text{vehicle pair}) \leftarrow \frac{Q^{MF}(\text{vehicle 1}) + Q^{MF}(\text{vehicle 2})}{2}$$

2 The Q_{net} values for the 2 vehicles offered for choice on standard trials are then injected
3 into a softmax choice rule such that the probability to choose an option is:

4
$$(Eq. 13) Prob(\text{vehicle}) = \frac{e^{Q_{\text{net}}(\text{vehicle})}}{e^{Q_{\text{net}}(\text{vehicle})} + Q_{\text{net}}(\text{other vehicle})}$$

5 Similarly, on uncertainty trials the probability to choose a vehicle pair was based on softmaxing
6 the net Q-values of the two offered pairs. Q^{MF} and *PERS* person-values and Q^{MB} vegetables-
7 values where initialized to 0 at the beginning of the experiment.

8 **Model Comparison and Fitting.** Our full hybrid agents, which allowed for contributions
9 from both an MB and an MF system, served as a super-model in a family of six nested sub-
10 models of interest: 1) a pure MB model, which was obtained by setting the contribution of the
11 MF to 0 (i.e. $c_{\text{standard}}^{MF} = c_{\text{nom,info}}^{MF} = c_{\text{nom,noninfo}}^{MF} = c_{\text{rej,info}}^{MF} = c_{\text{rej,noninfo}}^{MF} = 0$), 2) a pure MF-
12 action model, which was obtained by setting the contribution of the MB system to choices to 0
13 (i.e. $c^{MB} = 0$; Note that in this model, MB inference was still allowed to guide MF inference),
14 3) a ‘no informativeness effect on MFCA’ sub-model obtained by constraining equality between
15 the MFCA for the informative and non-informative destination (i.e., $c_{\text{nom,info}}^{MF} = c_{\text{nom,noninfo}}^{MF}$,
16 $c_{\text{rej,info}}^{MF} = c_{\text{rej,noninfo}}^{MF}$), 4) a ‘no MB guided MFCA’ sub-model obtained by constraining equality
17 between the MFCA parameters, for both the informative and non-informative destination, for
18 the ghost-nominated and rejected objects ($c_{\text{nom,info}}^{MF} = c_{\text{rej,info}}^{MF}$, $c_{\text{nom,noninfo}}^{MF} = c_{\text{rej,noninfo}}^{MF}$), 5) a
19 ‘no MB guidance of MFCA for the informative outcome’ obtained by constraining equality
20 between the MFCA parameters for the ghost-nominated and ghost-rejected objects for the
21 informative outcome ($c_{\text{nom,info}}^{MF} = c_{\text{rej,info}}^{MF}$), and 6) a ‘no MB guidance of MFCA for the non-
22 informative outcome’ which was similar to 5 but for the non-informative outcome ($c_{\text{nom,noninfo}}^{MF} =$
23 $c_{\text{rej,noninfo}}^{MF}$).

24 We conducted a bootstrapped generalized likelihood ratio test, BGLRT (Moran and
25 Goshen-Gottstein, 2015), for the super-model vs. each of the sub-models separately. In a

1 nutshell, this method is based on the classical-statistics hypothesis testing approach and
2 specifically on the generalized-likelihood ratio test (GLRT). However, whereas GLRT assumes
3 asymptotic Chi-squared null distribution for the log-likelihood improvement of a super model
4 over a sub-model, in BGLRT these distributions are derived empirically based on a parametric
5 bootstrap method. In each of our model comparison the sub model serves as the H0 null
6 hypothesis whereas the full model as the alternative H1 hypothesis. For each participant and
7 drug condition, we created 1001 synthetic experimental sessions by simulating the sub-agent
8 with the ML parameters on novel trial sequences which were generated as in the actual data.
9 We next fitted both the super-agent and the sub-agent to each synthetic dataset and calculated
10 the improvement in twice the logarithm of the likelihood for the full model. For each participant
11 and drug condition, these 1001 likelihood-improvement values served as a null distribution to
12 reject the sub-model. The p-value for each participant in each drug condition was calculated
13 based on the proportion of synthetic dataset for which the twice logarithm of the likelihood-
14 improvement was at least as large as the empirical improvement. Additionally, we performed
15 the model comparison at the group level. We repeated the following 10,000 times. For each
16 participant and drug condition we chose randomly, and uniformly, one of his/her 1,000
17 synthetic twice log-likelihood super-model improvements and we summed across participant
18 and drug conditions. These 10,000 obtained values constitute the distribution of group super-
19 model likelihood improvement under the null hypothesis that a sub-model imposes. We then
20 calculated the p-value for rejecting the sub-agent at the group level as the proportion of
21 synthetic datasets for which the super-agent twice logarithm of the likelihood improvement was
22 larger or equal to the empirical improvement in super-model, summed across participants.
23 Results, as display in Figure S6 in detail, fully supported the use of our full model including all
24 effects of interest regarding MFCA in uncertainty trials.

25 We next fit our choice models to the data of each individual, separately for each drug
26 condition (levodopa/placebo) maximizing the likelihood (ML) of their choices (we optimized
27 likelihood using MATLAB's 'fmincon', with 200 random starting points per participant * drug

1 condition; Table S4 for distribution best-fitting parameters). See Table S2 for the distribution
2 of full model's fitted parameters.

3 **Model simulations.** To generate model predictions with respect to choices, we
4 simulated for each participant and each drug condition, 25 synthetic experimental sessions
5 (novel trial sequences were generated as in the actual experiment), based on ML parameters
6 obtained from the corresponding model fits. We then analysed these data in the same way as
7 the original empirical data (but with datasets that were 25 times larger, as compared to the
8 empirical data, per participant). Results are reported in Figures S1 and S2 of the SI. We also
9 tested recoverability of model parameters (see Figure S7).

10 **Analysis of model parameters.** All models included the variable ORDER as regressor
11 (coded as +.5 for the first and -.5 for the second session) to control for unspecific effects and
12 participants (PART) served as random effects. Details of are reported in Table S2.

13 For each participant in each drug condition, we obtained, based on the full model, four
14 MFCA parameter estimates corresponding to destination (informative/non-informative) and
15 vehicle (nominated/rejected) types. We conducted a mixed effects model (again implemented
16 with MATLAB's function "fitglme") with TYPE (nominated/rejected coded as +.5 and -.5), INFO
17 (informative/non-informative coded as +.5 and -.5) and DRUG (drug/placebo coded as +.5 and
18 -.5) as regressors. The model, in Wilkinson notation, can be found on Table S3.

19 After finding significant drug by NOM * DRUG interaction, we followed this up in detail:
20 we calculated for each participant in each drug condition and for each destination type the
21 "preferential MFCA" (denoted PMFCA) effect as the difference between the corresponding
22 nominated and rejected MFCA parameters. We next ran a mixed effects model for PMFCA.
23 Our regressors where the destination type (denoted INFO; coded as before), and DRUG
24 (coded as before). The model, in Wilkinson notation, can be found on Table S3.

25 **Correlations between preferential MFCA and WM.** For each participant and
26 destination type (Informative/non-informative), we contrasted the "preferential MFCA"
27 estimates (as defined in the previous section) for levodopa minus placebo to obtain a drug-

1 induced PMFCA effect. For each destination, we calculated across-participants Spearman
2 correlations between these drug induced effects and WM. We compared the two correlations
3 (for informative and non-informative destinations) using a permutation test. First, we z-scored
4 the PMFCA separately for each destination type. Next we repeated the following steps (1-3),
5 10,000 times: 1) For each participant we randomly reshuffled (independent of other
6 participants) the outcome type labels “informative” and “non-informative”, 2) We calculated the
7 “synthetic” Spearman correlations between drug induced PMFCA effects and WM for each
8 outcome type subject to the relabelling scheme and, 3) We subtracted the two correlations
9 (non-informative minus informative). These 10,000 correlation-differences constituted a null
10 distribution for testing the null hypothesis that the two correlations are equal. Finally, we
11 calculated the p-value for testing the hypothesis of a stronger correlation for the non-
12 informative destination as the percentage (of the 10,000) synthetic correlation-differences that
13 were at least as large (in absolute value) as the empirical correlation-difference.

14 **Relationship between drug effects.** We used the same score for drug-dependent
15 change in PMFCA (levodopa minus placebo) and regressed it against informativeness, drug-
16 dependent change in MBCA and working memory capacity in a mixed effects model.

17

18

Acknowledgements. RJD is supported by a Wellcome Trust Investigator Award (098362/Z/12/Z) under which the above study was carried out. This work was carried out whilst R.J.D. was in receipt of a Lundbeck Visiting Professorship (R290-2018-2804) to the Danish Research Centre for Magnetic Resonance. RM is supported by the Max Planck Society and LD was at the time when the study was performed. The UCL-Max Planck Centre for Computational Psychiatry and Ageing is funded by a joint initiative between UCL and the Max Planck Society. RJD and LD are supported by a grant from the German Research Foundation (DFG TRR 265, project A02) and YL was at the time when the study was performed. For the purpose of Open Access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

1

References

- 2 Chowdhury, R., Guitart-Masip, M., Lambert, C., Dayan, P., Huys, Q., Duzel, E., and Dolan,
3 R.J. (2013). Dopamine restores reward prediction errors in old age. *Nat Neurosci* 16, 648–
4 653.
- 5 Cools, R. (2019). Chemistry of the Adaptive Mind: Lessons from Dopamine. *Neuron* 104,
6 113–131.
- 7 Cools, R., and D’Esposito, M. (2011). Inverted-U-shaped dopamine actions on human
8 working memory and cognitive control. *Biol Psychiatry* 69, e113-25.
- 9 D’Ardenne, K., McClure, S.M., Nystrom, L.E., and Cohen, J.D. (2008). BOLD responses
10 reflecting dopaminergic signals in the human ventral tegmental area. *Science* 319, 1264–
11 1267.
- 12 Daw, N.D., and Dayan, P. (2014). The algorithmic anatomy of model-based evaluation.
13 *Philos Trans R Soc Lond B Biol Sci* 369.
- 14 Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., and Dolan, R.J. (2011). Model-based
15 influences on humans’ choices and striatal prediction errors. *Neuron* 69, 1204–1215.
- 16 Deserno, L., Huys, Q., Boehme, R., Buchert, R., Heinze, H.J., Grace, A.A., Dolan, R.J.,
17 Heinz, A., and Schlagenhauf, F. (2015). Ventral striatal presynaptic dopamine reflects
18 behavioral and neural signatures of model-based control during sequential decision-making.
19 *Proc Natl Acad Sci U S A*.
- 20 Dolan, R.J., and Dayan, P. (2013). Goals and habits in the brain. *Neuron* 80, 312–325.
- 21 Doll, B.B., Simon, D.A., and Daw, N.D. (2012). The ubiquity of model-based reinforcement
22 learning. *Curr Opin Neurobiol* 22, 1075–1081.
- 23 Doll, B.B., Bath, K.G., Daw, N.D., and Frank, M.J. (2016). Variability in Dopamine Genes
24 Dissociates Model-Based and Model-Free Reinforcement Learning. *J Neurosci* 36, 1211–
25 1222.
- 26 Eldar, E., Lièvre, G., Dayan, P., and Dolan, R.J. (2020). The roles of online and offline replay
27 in planning. *ELife* 9, e56911.
- 28 Groman, S.M., Massi, B., Mathias, S.R., Curry, D.W., Lee, D., and Taylor, J.R. (2019).
29 Neurochemical and Behavioral Dissections of Decision-Making in a Rodent Multistage Task.
30 *J Neurosci* 39, 295–306.
- 31 Howard, J.D., Reynolds, R., Smith, D.E., Voss, J.L., Schoenbaum, G., and Kahnt, T. (2020).
32 Targeted Stimulation of Human Orbitofrontal Networks Disrupts Outcome-Guided Behavior.
33 *Current Biology* 30, 490-498.e4.
- 34 Jones, J.L., Esber, G.R., McDannald, M.A., Gruber, A.J., Hernandez, A., Mirenzi, A., and
35 Schoenbaum, G. (2012). Orbitofrontal Cortex Supports Behavior and Learning Using Inferred
36 But Not Cached Values. *Science* 338, 953–956.
- 37 Kroemer, N.B., Lee, Y., Pooseh, S., Eppinger, B., Goschke, T., and Smolka, M.N. (2019). L-
38 DOPA reduces model-free control of behavior by attenuating the transfer of value to action.
39 *NeuroImage* 186, 113–125.

- 1 Langdon, A.J., Sharpe, M.J., Schoenbaum, G., and Niv, Y. (2018). Model-based predictions
2 for dopamine. *Curr Opin Neurobiol* 49, 1–7.
- 3 Liu, Y., Dolan, R.J., Kurth-Nelson, Z., and Behrens, T.E.J. (2019). Human Replay
4 Spontaneously Reorganizes Experience. *Cell* 178, 640-652.e14.
- 5 Liu, Y., Mattar, M.G., Behrens, T.E.J., Daw, N.D., and Dolan, R.J. (2020). Experience replay
6 supports non-local learning (Neuroscience).
- 7 Mattar, M.G., and Daw, N.D. (2018). Prioritized memory access explains planning and
8 hippocampal replay. *Nat Neurosci* 21, 1609–1617.
- 9 Montague, P.R., Dayan, P., and Sejnowski, T.J. (1996). A framework for mesencephalic
10 dopamine systems based on predictive Hebbian learning. *J Neurosci* 16, 1936–1947.
- 11 Moran, R., and Goshen-Gottstein, Y. (2015). Old processes, new perspectives: Familiarity is
12 correlated with (not independent of) recollection and is more (not equally) variable for targets
13 than for lures. *Cognitive Psychology* 79, 40–67.
- 14 Moran, R., Dayan, P., and Dolan, R.J. (in press). Human subjects exploit a cognitive map for
15 credit assignment. *Proc Natl Acad Sci U S A*.
- 16 Moran, R., Keramati, M., Dayan, P., and Dolan, R.J. (2019). Retrospective model-based
17 inference guides model-free credit assignment. *Nat Commun* 10, 750.
- 18 Moran, R., Keramati, M., and Dolan, R.J. (2021). Model based planners reflect on their
19 model-free propensities. *PLoS Comput Biol* 17, e1008552.
- 20 O’Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R.J. (2004).
21 Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304,
22 452–454.
- 23 Otto, A.R., Gershman, S.J., Markman, A.B., and Daw, N.D. (2013). The curse of planning:
24 dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychol*
25 *Sci* 24, 751–761.
- 26 Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., and Frith, C.D. (2006). Dopamine-
27 dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* 442,
28 1042–1045.
- 29 Sadacca, B.F., Jones, J.L., and Schoenbaum, G. (2016). Midbrain dopamine neurons
30 compute inferred and cached value prediction errors in a common framework. *ELife* 5.
- 31 Schuck, N.W., and Niv, Y. (2019). Sequential replay of nonspatial task states in the human
32 hippocampus. *Science* 364, eaaw5181.
- 33 Schuck, N.W., Cai, M.B., Wilson, R.C., and Niv, Y. (2016). Human Orbitofrontal Cortex
34 Represents a Cognitive Map of State Space. *Neuron* 91, 1402–1412.
- 35 Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and
36 reward. *Science* 275, 1593–1599.
- 37 Sharp, M.E., Foerde, K., Daw, N.D., and Shohamy, D. (2016). Dopamine selectively
38 remediates “model-based” reward learning: a computational approach. *Brain* 139, 355–364.

- 1 Sharpe, M.J., Chang, C.Y., Liu, M.A., Batchelor, H.M., Mueller, L.E., Jones, J.L., Niv, Y., and
2 Schoenbaum, G. (2017). Dopamine transients are sufficient and necessary for acquisition of
3 model-based associations. *Nat Neurosci* 20, 735–742.
- 4 Starkweather, C.K., Babayan, B.M., Uchida, N., and Gershman, S.J. (2017). Dopamine
5 reward prediction errors reflect hidden-state inference across time. *Nat Neurosci* 20, 581–
6 589.
- 7 Steinberg, E.E., Keiflin, R., Boivin, J.R., Witten, I.B., Deisseroth, K., and Janak, P.H. (2013).
8 A causal link between prediction errors, dopamine neurons and learning. *Nat Neurosci* 16,
9 966–973.
- 10 Sutton, R.S. (1991). Dyna, an integrated architecture for learning, planning, and reacting.
11 *SIGART Bull.* 2, 160–163.
- 12 Westbrook, A., van den Bosch, R., Määttä, J.I., Hofmans, L., Papadopetraki, D., Cools, R.,
13 and Frank, M.J. (2020). Dopamine promotes cognitive effort by biasing the benefits versus
14 costs of cognitive work. *Science* 367, 1362–1366.
- 15 Wunderlich, K., Smittenaar, P., and Dolan, R.J. (2012). Dopamine enhances model-based
16 over model-free choice behavior. *Neuron* 75, 418–424.
- 17
- 18

Supplementary Information

Repeat and switch standard trials following uncertainty trials. We showed previously on “repeat” trials (Moran et al., 2019), a positive effect of an informative destination reward (on trial n) on choice-repetition implicates MFCA to the ghost-nominated object (while the MB system knows that the value of the informative destination favours both vehicles on trial n+1). We also ran a separate analysis that examined MFCA for the ghost-nominated alone. In trial n+1 “repeat” trials, the ghost-nominated vehicle from trial-n is offered for choice alongside a vehicle from the trial-n non-chosen pair that shares the inference-allowing destination with the ghost-nominated object. Choice repetition was defined as choice of the ghost-nominated vehicle from uncertainty trial n as indicated by the variable REPEAT. Regressor PART is coded as previously. Regressors N (non-informative destination) and I (informative destination) indicate whether a reward was received at the destinations or not in trial n (coded as +.5/-.5). The model is $REPEAT \sim N*I + (N*I | PART)$. This showed a main effect for the informative (I) destination ($b=0.60$, $t(4885)=7.56$, $p=4e-14$), supporting MFCA to the ghost-nominated object. Additionally, we found a main effect for the non-informative (N) destination ($b=1.23$, $t(4885)=10.83$, $p=9e-42$) as predicted by both MF and by MB contributions, and a significant interaction between the Informative and Non-informative destinations ($b=0.31$, $t(4885)=2.09$, $p=.04$). See Figure S4, A & B.

We showed previously on switch trials that a positive main effect of the informative outcome reward on choice-switching implicates MFCA for the ghost-rejected vehicle (because the MB system knows the informative destination is unrelated to both vehicles on trial n+1). A second separate analysis examined MFCA for the ghost-rejected vehicle. In uncertainty trial n+1 “switch” trials, the ghost-rejected vehicle from trial-n is offered for choice alongside a vehicle from the trial-n non-chosen pair that shares a destination with the ghost-rejected object. Choice switching was defined as choice of the ghost-rejected vehicle from uncertainty trial n as indicated by the variable SWITCH. Regressors PART, N and I are coded as previously. The model is $SWITCH \sim N*I + (N*I | PART)$. This showed a main effect for the reward at informative

1 destination ($b=0.38$, $t(4866)=5.60$, $p=2e-8$), supporting MFCA to the ghost-rejected vehicle.
2 While, this challenges any notion of *perfect* MB guidance of MFCA, it is consistent with the
3 possibility that some participants, at least some of the time, do not rely on MB-inference
4 because when MB inference does not occur, or when it fails to guide MF credit-assignment,
5 the MF system has no basis to assign credit unequally to both vehicles in the selected pair.
6 Additionally, we found a main effect for the non-informative destination reward ($b=0.98$,
7 $t(4866)=13.18$, $p=5e-39$), as predicted by an MF credit-assignment to the ghost-rejected
8 vehicle account but also by MB contributions. We found no significant interaction between
9 rewards at the informative and non-informative destinations (Table S1). In Figure S2, we plot
10 empirical choice proportions from both repeat and switch conditions (reflecting the effects
11 reported above) in the manner as in the original paper by Moran et al. (2019) but separately
12 for drug and placebo conditions. See Figure S4 C & D.

13
14 **Absence of drug effects on perseveration and forgetting parameters of the**
15 **computational model.** No difference between drug conditions was observed for
16 perseveration parameter on standard trials ($t(61)= 0.48$, $p=.63$), perseveration parameter on
17 uncertainty trials ($t(61)= 0.51$, $p=.61$), MF forgetting parameter ($t(61)= 1.37$, $p=.17$), MB
18 forgetting parameter ($t(61)= -0.33$, $p=.74$), perseveration forgetting parameter ($t(61)=0.30$,
19 $p=.77$).

20
21 **Correlation between WM and MBCA.** Working memory moderated the boosting drug effect
22 on guidance of MFCA based on retrospective MB inference but only based on non-informative
23 reward (see main text). No moderating effect of working memory on a drug-dependent
24 difference in MBCA was observed ($r=-.07$, $p=.59$). Working memory correlated positively with
25 MBCA separately at placebo and at drug but this was non-significant (placebo: $r=.21$, $p=.08$;
26 drug: $r=.15$, $p=.23$).

Table S1. Mixed-effects models on model-agnostic choice data from standard trials

Name	Estimate	SE	tStat	DF	pValue	LowerCI	UpperCI
MF choice (standard trials)							
REPEAT ~ 1+ C*U*DRUG*ORDER + (C+U+DRUG+ORDER PART)							
(Intercept)	0.34	0.06	5.54	7251	.000	0.22	0.46
C (common)	0.67	0.07	9.14	7251	.000	0.53	0.81
U (unique)	1.54	0.09	17.40	7251	.000	1.36	1.71
DRUG	0.03	0.07	0.46	7251	.643	-0.11	0.18
ORDER	0.07	0.07	0.91	7251	.365	-0.08	0.21
C*U	0.19	0.11	1.72	7251	.085	-0.03	0.40
C*DRUG	0.07	0.11	0.67	7251	.500	-0.14	0.29
U*DRUG	0.06	0.11	0.56	7251	.577	-0.15	0.27
C*ORDER	0.12	0.11	1.09	7251	.276	-0.09	0.33
U*ORDER	-0.11	0.11	-0.99	7251	.321	-0.32	0.11
DRUG*ORDER	-0.25	0.24	-1.02	7251	.309	-0.73	0.23
C*U*DRUG	0.14	0.22	0.64	7251	.524	-0.29	0.57
C*U*ORDER	0.13	0.22	0.59	7251	.554	-0.30	0.56
C*DRUG*ORDER	-0.02	0.29	-0.06	7251	.952	-0.59	0.56
U*DRUG*ORDER	-0.18	0.35	-0.51	7251	.609	-0.87	0.51
C*U*DRUG*ORDER	-0.22	0.44	-0.50	7251	.618	-1.07	0.64
MB choice (standard trials)							
GENERALIZE ~ C*P*DRUG*ORDER + (C+P+DRUG+ORDER PART)							
(Intercept)	0.30	0.04	6.96	7177	.000	0.22	0.38
C (common)	0.40	0.06	6.22	7177	.000	0.27	0.52
P (common reward probability)	1.33	0.21	6.39	7177	.000	0.92	1.74
DRUG	-0.13	0.08	-1.65	7177	.099	-0.29	0.03
ORDER	-0.13	0.08	-1.57	7177	.116	-0.29	0.03
C*P	-0.23	0.23	-1.01	7177	.311	-0.67	0.21
C*DRUG	0.05	0.12	0.39	7177	.695	-0.19	0.28
P*DRUG	-0.34	0.23	-1.48	7177	.140	-0.79	0.11
C*ORDER	-0.06	0.12	-0.52	7177	.606	-0.30	0.17
P*ORDER	0.16	0.23	0.70	7177	.482	-0.29	0.61
DRUG*ORDER	-0.24	0.17	-1.41	7177	.158	-0.58	0.09
C*P*DRUG	-0.08	0.45	-0.18	7177	.856	-0.97	0.80
C*P*ORDER	0.57	0.45	1.26	7177	.207	-0.31	1.45
C*DRUG*ORDER	-0.38	0.25	-1.48	7177	.140	-0.87	0.12
P*DRUG*ORDER	0.46	0.83	0.55	7177	.583	-1.18	2.09
C*P*DRUG*ORDER	1.40	0.91	1.54	7177	.123	-0.38	3.17

Table S2. Mixed-effects models on model-agnostic choice data from uncertainty trials

Name	Estimate	SE	tStat	DF	pValue	LowerCI	UpperCI
Preferential MFCA for the informative destination (Ghost-nominated, “repeat trials” > ghost-rejected, “switch trials”)							
MFCA ~ NOM*DRUG*ORDER + (NOM*DRUG+ORDER PART)							
(Intercept)	0.10	0.01	8.54	239	.000	0.07	0.12
NOM (Nomination)	0.03	0.02	1.60	239	.110	-0.01	0.08
DRUG	0.01	0.02	0.40	239	.690	-0.04	0.05
ORDER	0.00	0.02	0.15	239	.878	-0.04	0.05
NOM*DRUG	0.11	0.04	2.56	239	.011	0.03	0.20
NOM*ORDER	0.02	0.04	0.45	239	.650	-0.07	0.11
DRUG*ORDER	-0.03	0.04	-0.74	239	.463	-0.12	0.06
NOM*DRUG*ORDER	-0.01	0.09	-0.06	239	.951	-0.18	0.16
MFCA for non-informative destination (Ghost-nominated > ghost-rejected, “clash trials”)							
REPEAT ~ N*I*DRUG*ORDER + (N*I*DRUG+ORDER PART)							
(Intercept)	0.05	0.04	1.27	4861	.203	-0.03	0.12
N (non-informative)	0.13	0.07	1.96	4861	.051	0.00	0.26
I (informative)	1.01	0.10	9.95	4861	.000	0.81	1.21
DRUG	0.15	0.07	2.31	4861	.021	0.02	0.29
ORDER	0.03	0.07	0.41	4861	.684	-0.10	0.16
N*U	0.08	0.14	0.57	4861	.568	-0.19	0.35
N*DRUG	0.05	0.13	0.39	4861	.696	-0.21	0.31
I*DRUG	0.03	0.14	0.24	4861	.810	-0.25	0.32
N*ORDER	-0.05	0.13	-0.34	4861	.733	-0.30	0.21
I*ORDER	0.06	0.14	0.43	4861	.664	-0.22	0.35
DRUG*ORDER	-0.20	0.15	-1.37	4861	.171	-0.49	0.09
N*I*DRUG	0.07	0.29	0.26	4861	.798	-0.49	0.64
N*I*ORDER	0.25	0.29	0.86	4861	.388	-0.32	0.81
N*DRUG*ORDER	-0.47	0.26	-1.80	4861	.072	-0.99	0.04
I*DRUG*ORDER	-0.12	0.41	-0.31	4861	.759	-0.92	0.67
N*I*DRUG*ORDER	0.86	0.55	1.56	4861	.118	-0.22	1.94

Table S3. Mixed-effects models on parameters of the computational model.

Name	Estimate	SE	tStat	DF	pValue	LowerCI	UpperCI
MFCA for ghost-nominated vs. ghost-rejected and informative vs non-informative							
MFCA ~ NOM*INFO*DRUG* + (NOM*INFO*DRUG PART)							
(Intercept)	0.18	0.02	7.60	480	.000	0.14	0.23
NOM (nomination)	0.10	0.03	3.72	480	.000	0.05	0.15
INFO (informativeness)	0.08	0.04	2.19	480	.029	0.01	0.15
DRUG	0.05	0.05	1.00	480	.316	-0.05	0.15
ORDER	0.04	0.05	0.78	480	.434	-0.06	0.14
NOM*INFO	-0.03	0.05	-0.57	480	.567	-0.12	0.06
NOM*DRUG	0.10	0.04	2.43	480	.015	0.02	0.18
INFO*DRUG	-0.08	0.07	-1.16	480	.247	-0.22	0.06
NOM*ORDER	0.02	0.04	0.37	480	.715	-0.07	0.10
INFO*ORDER	0.10	0.07	1.42	480	.157	-0.04	0.23
DRUG:ORDER	-0.09	0.10	-0.98	480	.328	-0.28	0.10
NOM*INFO*DRUG	0.02	0.07	0.33	480	.738	-0.12	0.17
NOM*INFO*ORDER	-0.01	0.07	-0.08	480	.934	-0.15	0.14
NOM*DRUG*ORDER	-0.06	0.11	-0.60	480	.551	-0.27	0.15
INFO*DRUG*ORDER	0.16	0.14	1.10	480	.272	-0.12	0.44
NOM*INFO*DRUG*ORDER	0.10	0.19	0.55	480	.585	-0.26	0.47
Preferential MFCA for informative vs. non-informative							
PMFCA ~ INFO*DRUG*ORDER + (INFO*DRUG+ORDER PART)							
(Intercept)	0.10	0.03	3.72	240	.000	0.05	0.15
INFO (informativeness)	-0.03	0.05	-0.57	240	.568	-0.12	0.07
DRUG	0.10	0.04	2.41	240	.017	0.02	0.18
ORDER	0.02	0.04	0.36	240	.717	-0.07	0.10
INFO*DRUG	0.02	0.07	0.33	240	.739	-0.12	0.17
INFO*ORDER	-0.01	0.07	-0.08	240	.934	-0.15	0.14
DRUG*ORDER	-0.06	0.11	-0.60	240	.551	-0.27	0.15
INFO*DRUG*ORDER	0.10	0.19	0.55	240	.585	-0.27	0.47

Table S4. Distribution of parameters from the full computational model.

Cond.	%	MFCA standard	MFCA info-nom	MFCA info-rej	MFCA non-info-nom	MFCA non-info-rej	MBCA	perseveration - standard	perseveration - nominated	forget_MF	forget_MB	forget_Pers
Placebo	25	0.053	-0.056	-0.026	-0.070	-0.074	0.059	-0.197	-0.093	0.002	0.038	0.010
	50	0.147	0.168	0.149	0.048	0.030	0.273	0.042	0.071	0.058	0.148	0.123
	75	0.364	0.479	0.391	0.333	0.204	0.454	0.383	0.353	0.519	0.521	0.428
L-DOPA	25	0.060	-0.025	-0.073	-0.011	-0.098	0.026	-0.086	-0.047	0.019	0.022	0.008
	50	0.272	0.165	0.130	0.178	0.070	0.278	0.098	0.084	0.190	0.127	0.089
	75	0.574	0.517	0.383	0.390	0.291	0.367	0.346	0.374	0.598	0.508	0.492

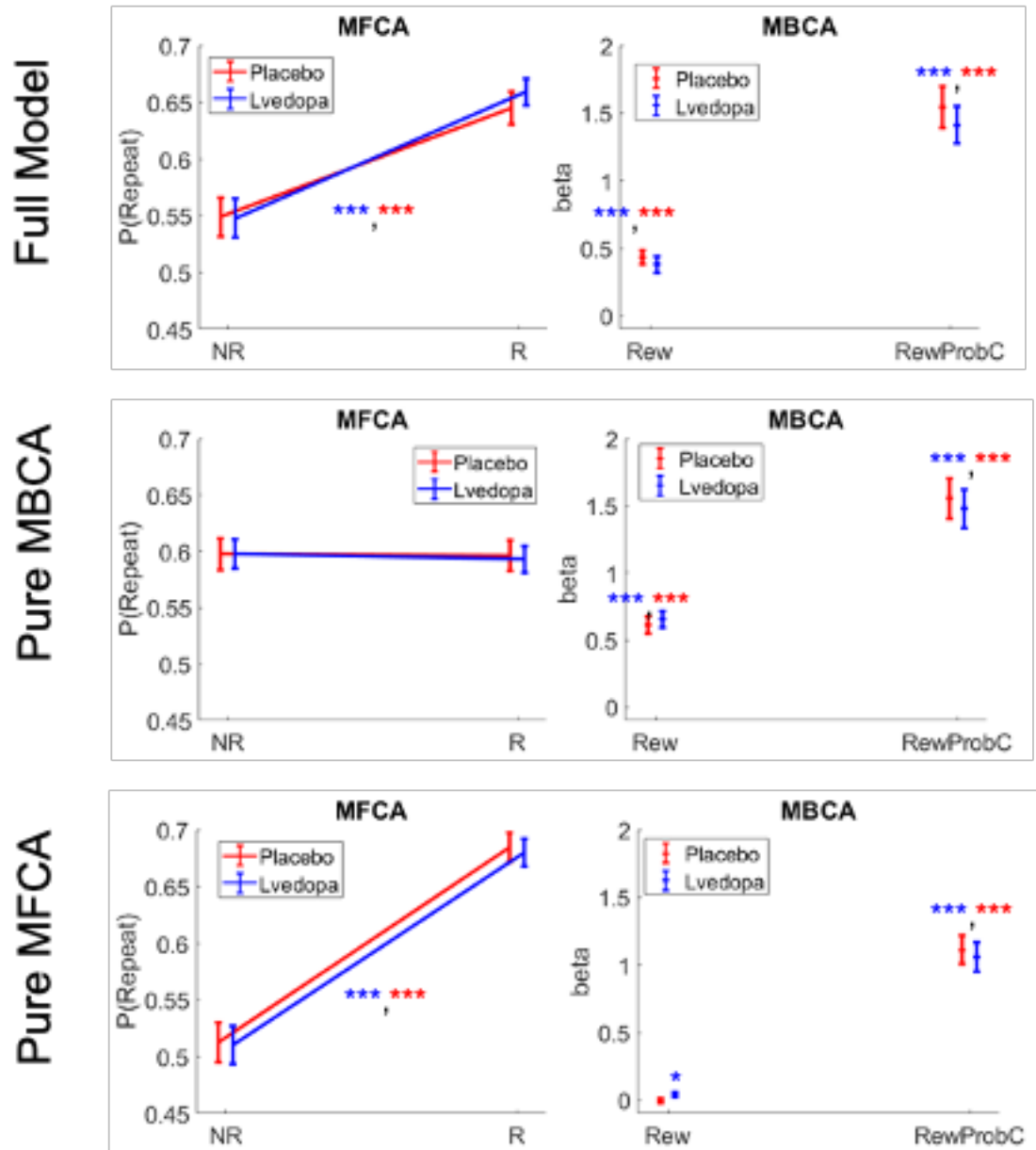


Figure S1. Simulations for standard trials based on the full model and sub-models. NR=no reward, R=reward. Rew=reward at the common destination, RewProbC=Reward Probability at the common destination.

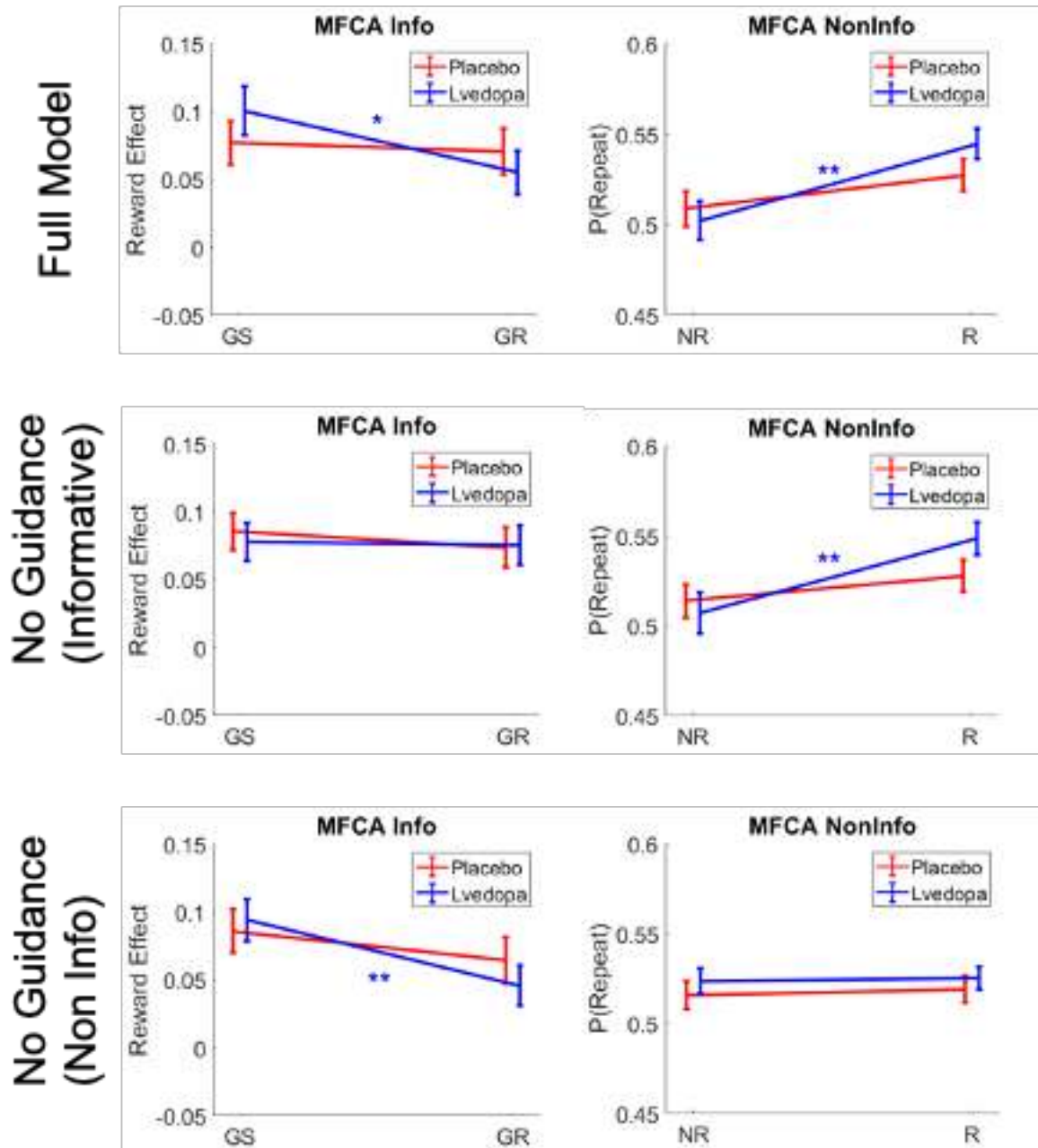


Figure S2. Simulations for uncertainty trials based on the full model and sub-models. GS=Ghost-selected, GR=Ghost-rejected.

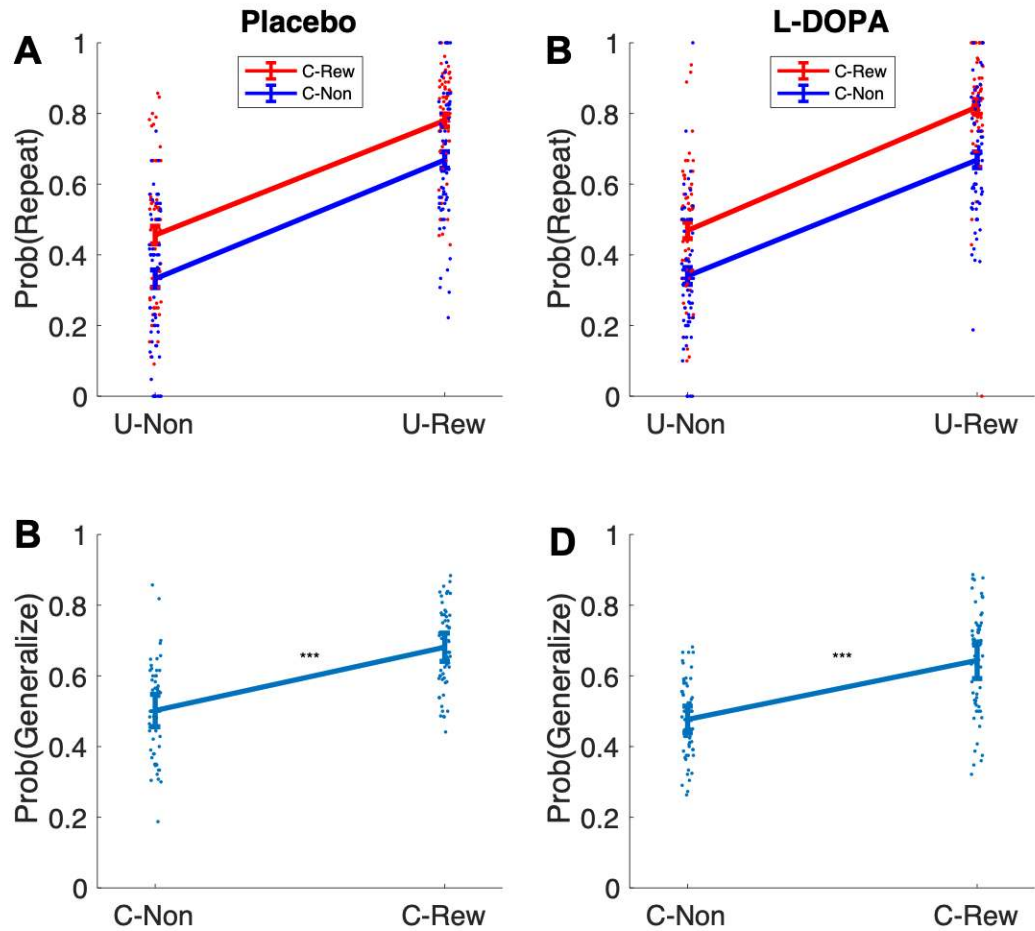


Figure S3. Empirical probabilities of model-agnostic MF (A & B) and MB (C & D) choice contribution under placebo and levodopa (L-DOPA). U-Non=no reward at unique destination, U-Rew= reward at unique destination, C-Non=no reward at common destination, C-Rew= reward at common destination.

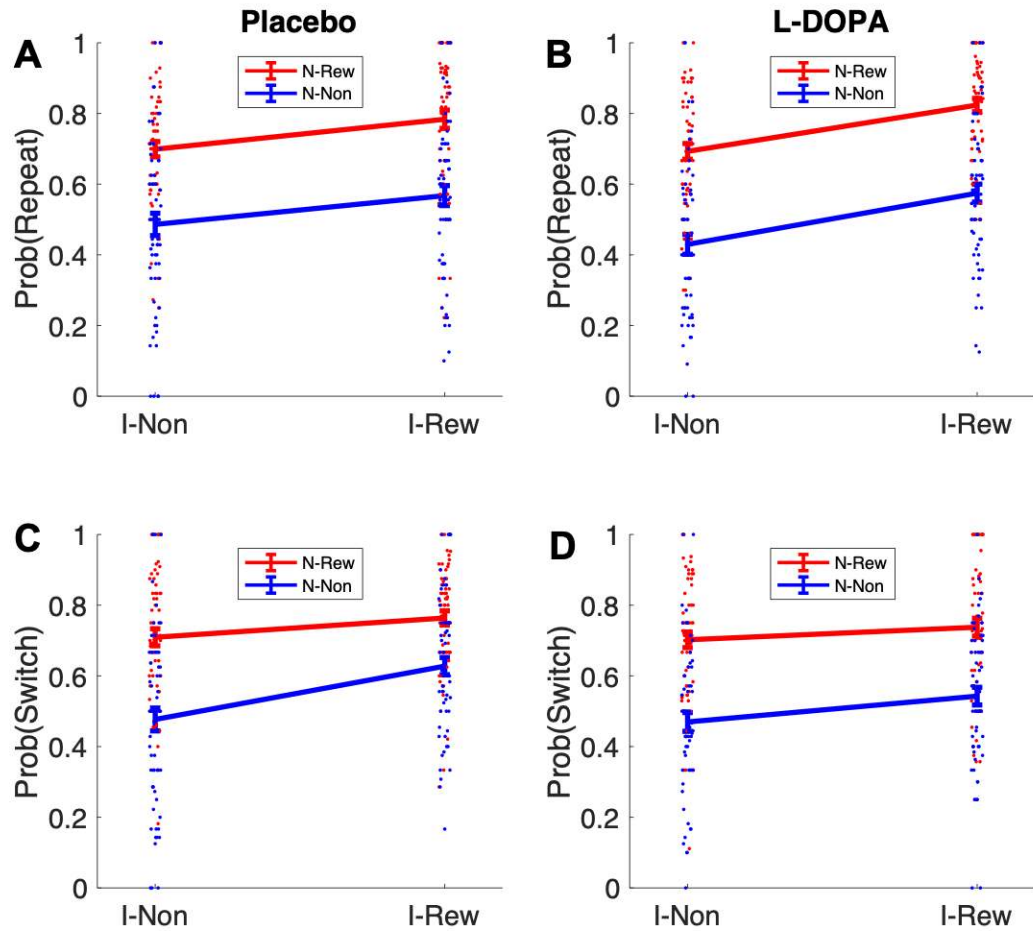


Figure S4. Retrospective MB inference using the informative destination based on repeat and switch signatures after uncertainty trials. I-Non=no reward at informative destination, I-Rew= reward at informative destination, N-Non=no reward at non-informative destination, N-Rew= reward at non-informative destination.

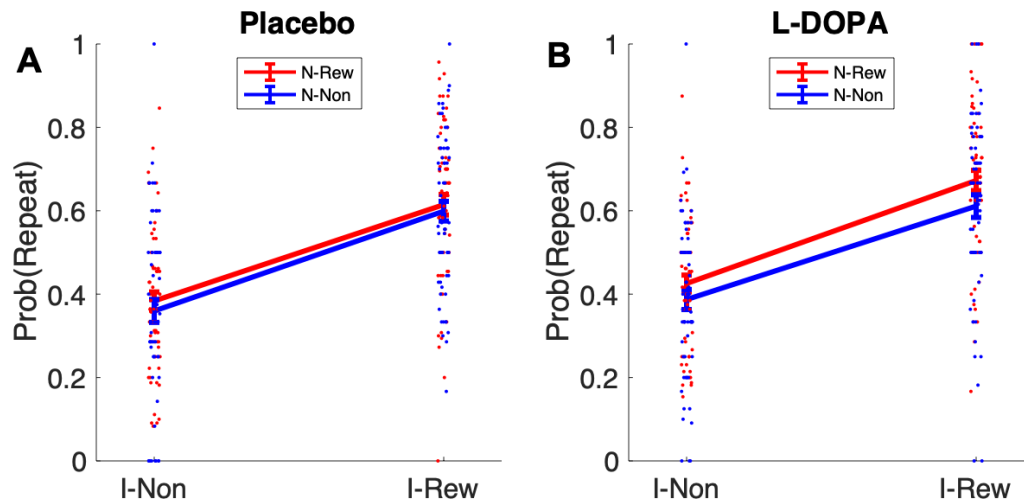


Figure S5. Retrospective MB inference using the non-informative destination based on choice repetition in “clash” trials $n+1$ following an uncertainty trial- n . I-Non=no reward at informative destination, I-Rew=reward at informative destination, N-Non=no reward at non-informative destination, N-Rew=reward at non-informative destination.

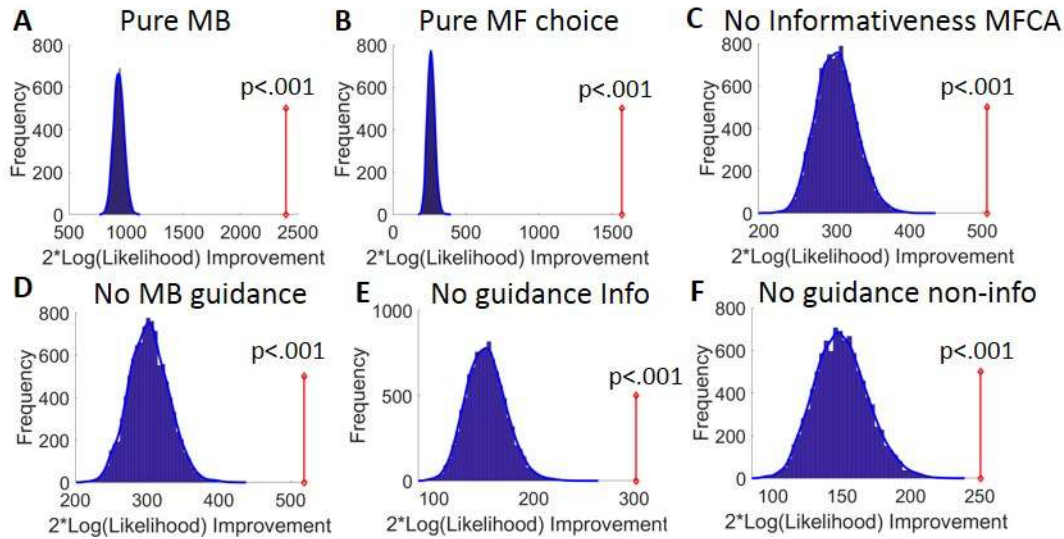


Figure S6. Model-comparison results. A) Results of the bootstrap-GLRT model-comparison for the pure MB sub-model. The blue bars show the histogram of the group twice log-likelihood improvement (model vs. sub-model) for synthetic data simulated using the sub-model (10000 simulations). The blue line displays the smoothed null distribution (using Matlab's "ksdensity"). The red line shows the empirical group twice log-likelihood improvement. p-value reflect the proportion of 10000 simulations that yielded an improvement in likelihood that was at least as large as the empirical improvement. B-E) Same as (A), but for the pure MF choice, the no informativeness effects on MFCA, the no MB-guidance for MFCA, the no MB-guidance for the informative destination and the no-MB guidance for the non-informative destination sub models.

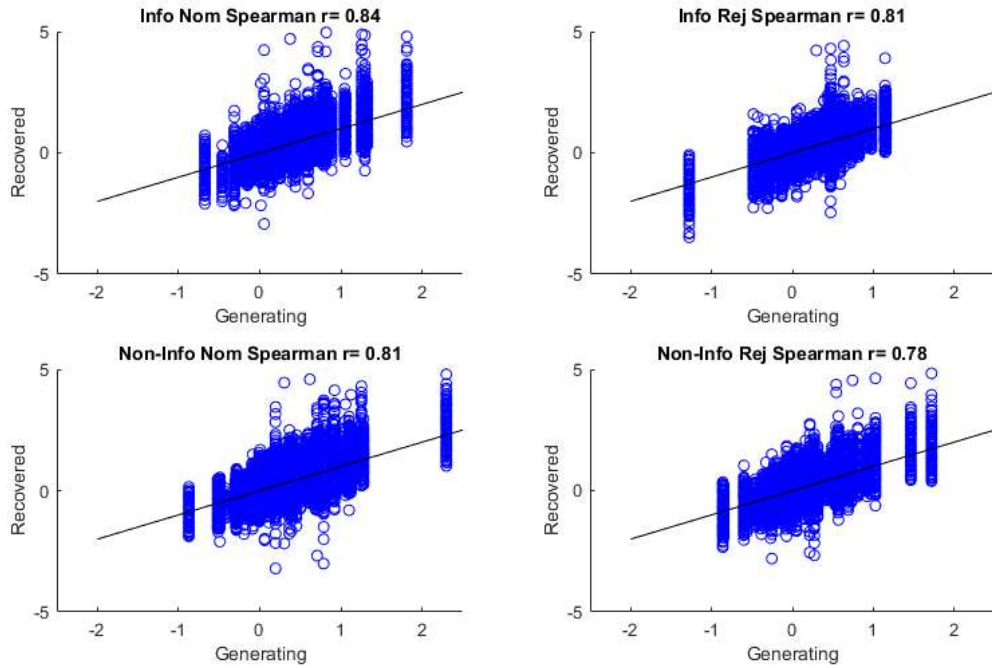


Figure S7. Parameter recoverability. For each of the 2*62 full model parameter-combinations 1000 synthetic (simulated) datasets were created by simulating the full model on experimental sessions as in the true experiment. Then the full model was fit to each of these generated datasets. For each MFC parameter (info/non-info x nom/rej), we plot the recovered against the generating parameters (and impose black diagonals where "recovered = generating").

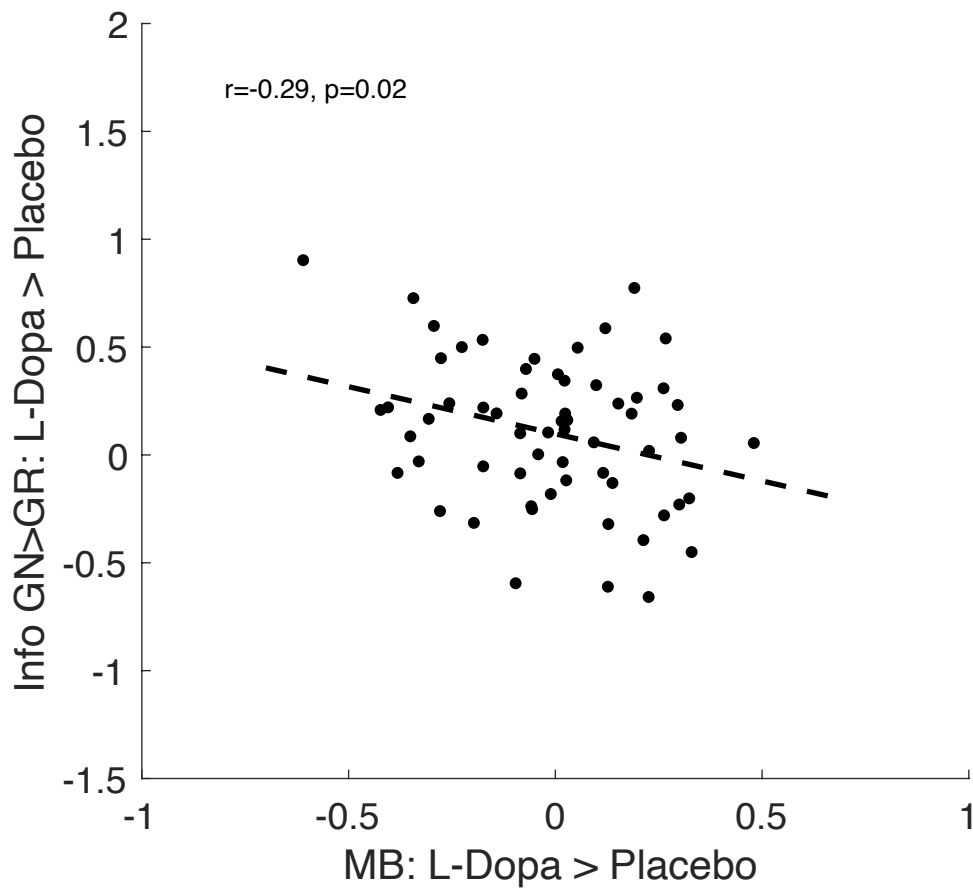


Figure S8. When using a model-agnostic measure of MB choice (probability to generalize after reward minus no-reward) and of preferential MFCA at the informative destination (repeat or ghost-nominate minus switch or ghost-rejected), dopamine dependent differences (levodopa minus placebo) in those measures were correlated negatively ($r=-.29$, $p=.021$) mirroring the finding as reported on parameters from the computational model in the main text.