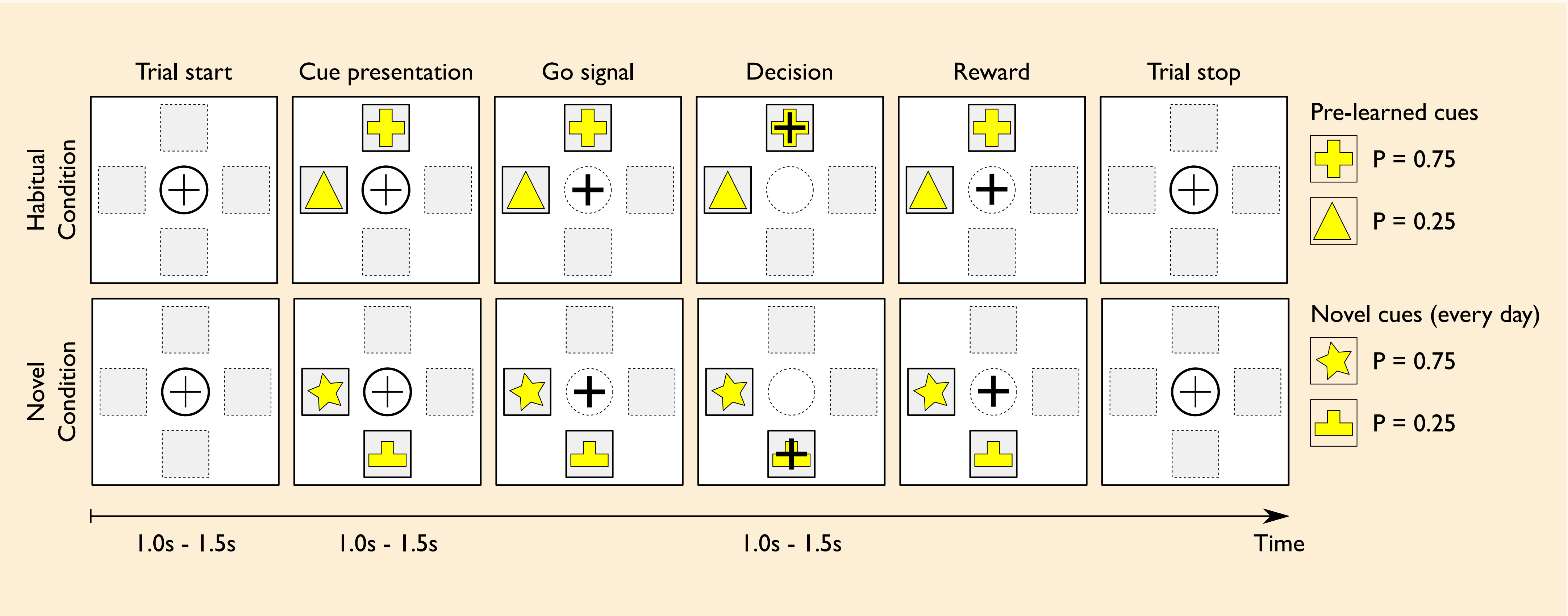
If basal ganglia are widely accepted to participate in the high-level cognitive function of decision-making, their role is less clear regarding the formation of habits. One of the biggest problem is to understand how goal-directed actions are transformed into habitual responses, or, said differently, how an animal can shift from an action-outcome (A-O) system to a stimulus-response (S-R) one while keeping a consistent behaviour?

We introduce a computational model that can solve a simple **two armed bandit task** using reinforcement learning and explicit valuation of the outcome. Hebbian learning has been added at the cortical level such that the model learns each time a move is issued, rewarded or not. Then, by inhibiting the output nuclei of the model (GPi), we show how learning has been transferred from the basal ganglia to the cortex, simply as a consequence of the statistics of the choice. Because best (in the sense of most rewarded) actions are chosen more often, this directly impacts the amount of Hebbian learning and lead to the formation of habits within the cortex. These results have been confirmed in monkeys doing the same tasks where the BG has been inactivated using muscimol. This tends to show that the basal ganglia implicitely teach the cortex in order for it to learn the values of new options.

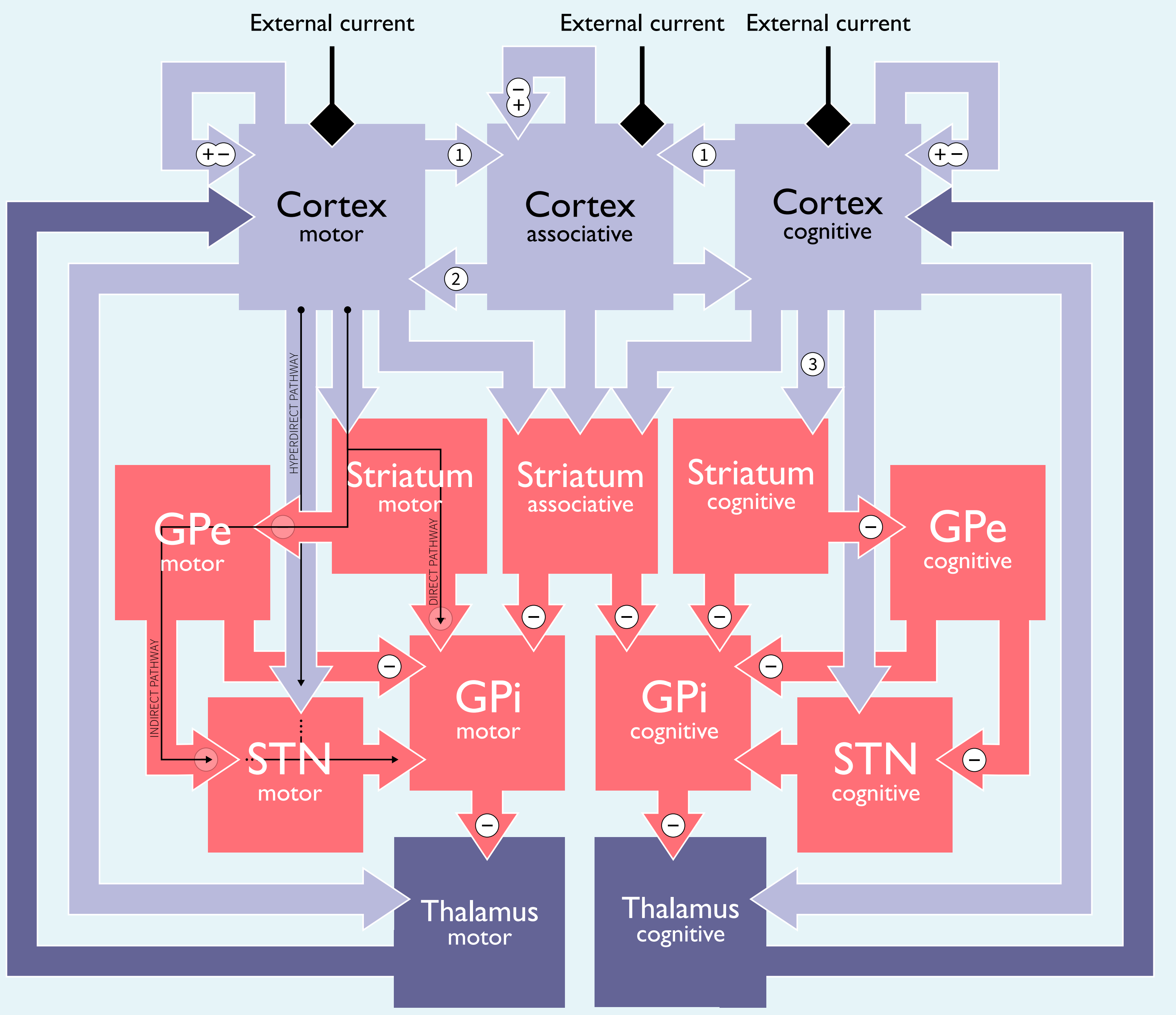


THE FORMATION OF HABITS

The implicit supervision of the basal ganglia

MEROPI TOPALIDOU - CAMILLE PIRON - DAISUKE KASE - THOMAS BORAUD - NICOLAS ROUGIER

institut des maladies neurodégénératives



The model is based on the model presented in [1] which itself derives from the competition principles introduced in [2]. This former model introduces an action selection mechanism that is based upon the competition between a positive feedback through the direct pathway and a negative feedback through the hyperdirect pathway. The model has been further extended in [2] and exploits the parallel organization of circuits between the basal ganglia and the cortex [3] using segregated loops: one for making the selection between the two presented cue shapes, and the other for making the selection between the two possible movement directions. However, to solve the task described previously, it is necessary for the model to first choose the cue shape and then (and only then) to select the right movement direction which depends upon the chosen cue. The model has been further refined in this study such as to have a competition mechanisms within each cortical group. Using short range excitation and long range inhibitions, this competition ensures that a unique cognitive and motor decision eventually emerges, even if these decisions might be unrelated at this stage. Learning occurs between the cognitive cortex and the cognitive striatum using a simple reinforcement learning where the value of the different cues are updated after each decision (see [2] for details). We added Hebbian learning (LTP) at the cortical level between the cognitive/motor cortical groups and the associative cortical group. This learning is enforced once per trial, at the time a move is made and independently of the actual reward. The model has been trained on cues 1 and 2 (cognitive cortex) that are presented simultaneously at random positions in the motor cortex. Cue 1 is associated with a reward probability of 75% while cue 2 is associated with a reward probability of 25%. The model is trained until it achieves a performance of 0.95, meaning it chooses cue 1 most of the time. This takes between 10 and 20 trials depending on the initial conditions (noise) and whether first cues are rewarded or not. In the meantime, this training impacts significantly Hebbian learning at the cortical level because cue 1 is chosen most of the time and consequently, the associative link relative to cue 1 is reinforced compared to associative link relative to cue 2 (while cues 3 and 4 are never reinforced since they are never presented).

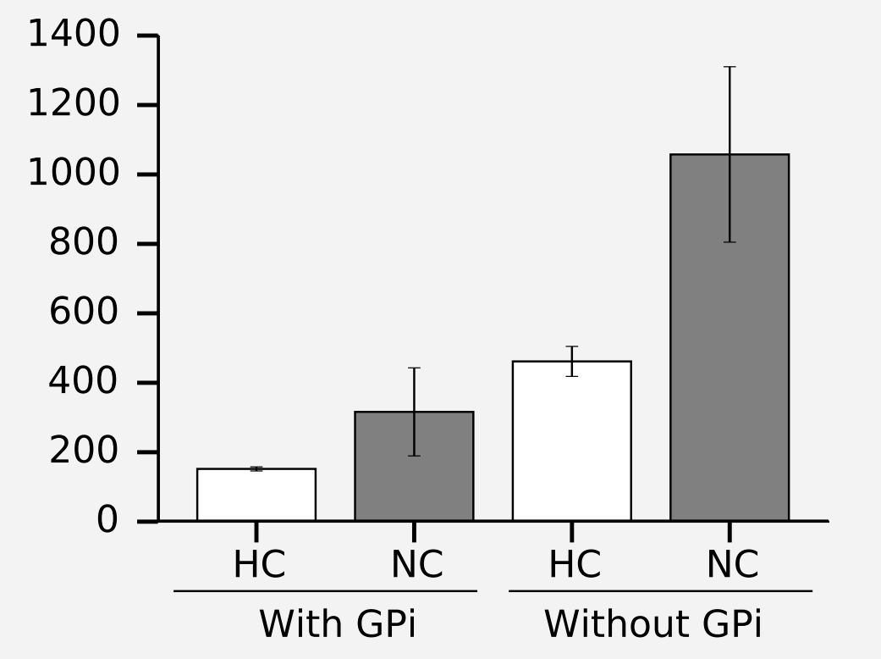
[1] M. Guthrie, A. Leblois, A. Garenne, and T. Boraud. Interaction between cognitive and motor cortico-basal ganglia loops during decision making: a computational study. Journal of Neurophysiology, 109:3025–3040, 2013.
[2] Leblois A, Boraud T, Meissner W, Bergmann H, Hansel D. Competition betweenfeedback loops underlies normal and pathological dynamics in the basal ganglia. J Neurosci 26: 3567–3583, 2006.
[3] Alexander G, Crutcher M, De Long M. Basal ganglia-thalamocortical circuits: parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. Prog Brain Res 85: 119–146, 1991.

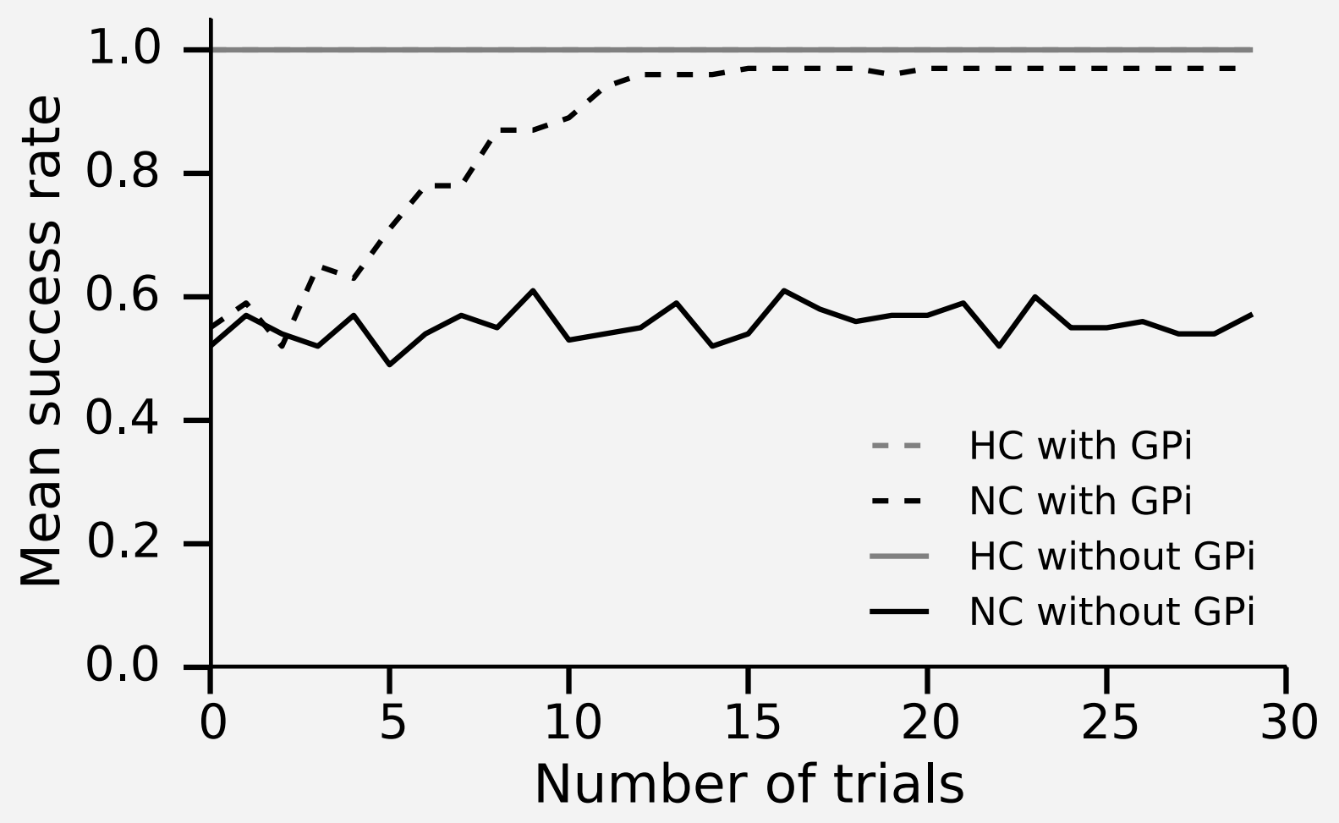
After the initial training phase, we tested the model using four different paradigms that corresponds to the experiments:

- HC/GPi (saline): Habitual Condition using cues 1 & 2 with intact GPi
- NC/GPi (saline): Novel Condition using cues 3 & 4 with intact GPi
- HC/NoGPi (muscimol): Habitual Condition using cues 1 & 2 with lesioned GPi
- NC/NoGPi (muscimol): Novel Condition using cues 3 & 4 with lesioned GPi

An experiment is made of 120 consecutive trials. Each trial starts with a settling period that last for 500ms until two cues are presented to the model at random position. Once a motor decision is made, the reward is computed according to the chosen cue, that is, the one that corresponds to the actual motor choice and not the cognitive one. Response time has been recorded as the time of the motor decision, that is, when the difference between two greatest motor activation is greater than the decision threshold (40Hz). This time is relative to the stimulus onset. Each of the four conditions has been averaged over 250 experiments

Model results are in accordance with the experiments in monkeys. In the habitual condition (HC), performances are optimal with or without lesion, indicating the cortex is able to make the optimal decision without the help of the basal ganglia if it has been learned previously. In novel condition, performances of the intact model (NC/Gpi) are initially at chance level but after a few trials (around 15), it reaches a near-optimal performance, indicating the model has learned the respective reward probability associated with each novel cues. However, for the lesioned model (NC/NoGPi), performances stay at chance level, indicating the cortex is unable to learn the new task without the help of the basal ganglia.

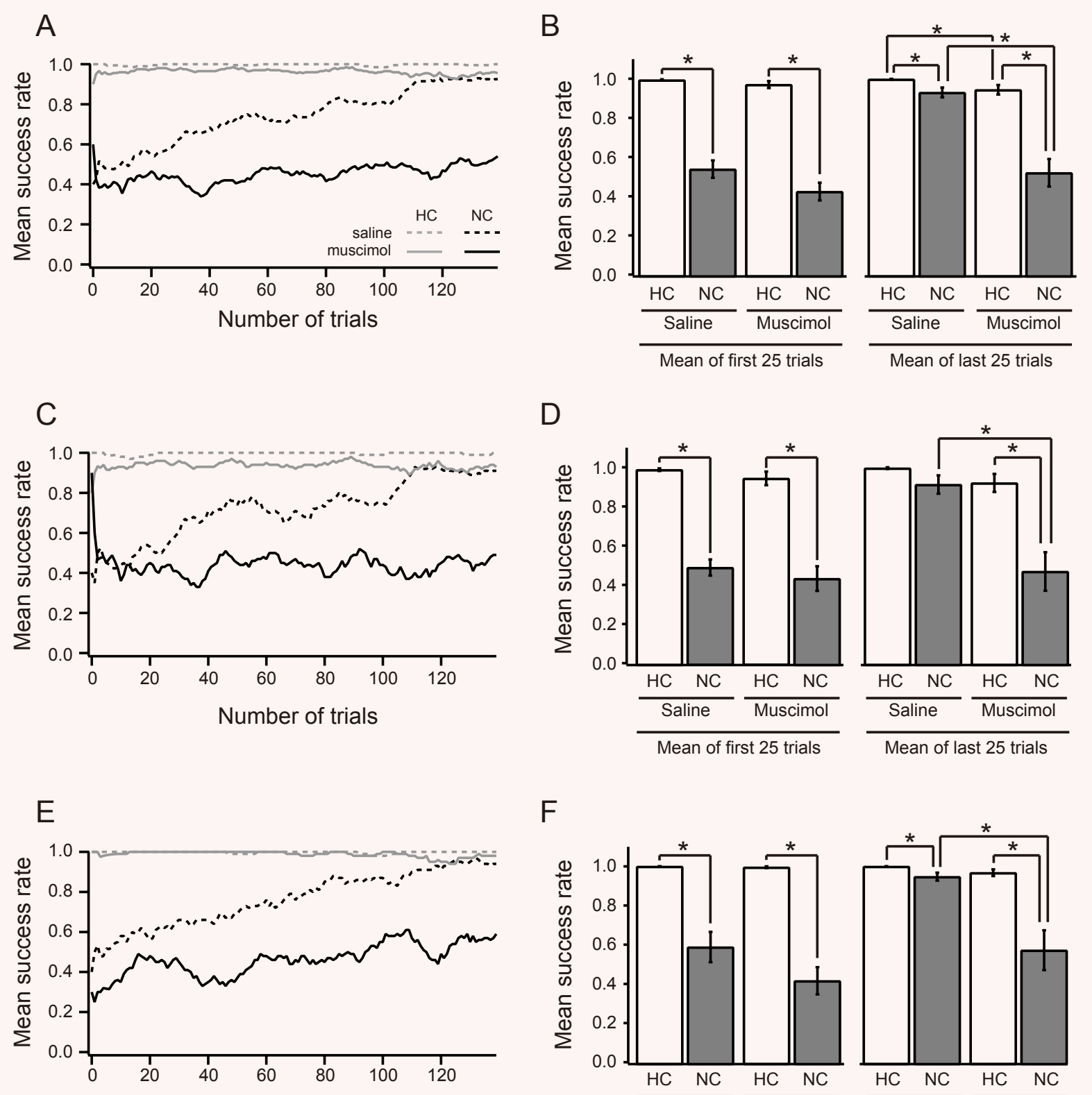




Model performances in the four conditions. Each trial has been averaged over 250 experiments. In habitual conditions, performances are optimal with (GPi) or without GPi (NoGPi). In novel conditions, only the intact model (GPi) is able to learn the new cues while lesioned model (NoGPi) performances stay at chance level.

Monkeys were tested using 20 sessions in control conditions and 20 in muscimol conditions (10 for each monkey in each condition). We defined as success rate, the number of trials in which the animals chose the optimal target (i.e. HC1 or NC1 in Habitual or Novelty conditions respectively), normalized by the number of trials in which a choice has been achieved. When a trial is interrupted before a choice has been achieved and validated, it is considered as an error trial.

In control conditions, the **animals** maximize their choice in the Habitual Condition. The mean success rate is 99.4±3.3% (Figure A), respectively 98.8 ± 0.6% for Monkey F (Figure C) and 100.0 ± 0.0% for monkey Z (Figure E), P>0.05 between the two animal (unpaired t-test). In the Novelty Condition, they learn progressively the difference between the two cues (Figure A,C & E). They choose randomly at the beginning of training to finally display a preference for NC1, the target associated to the best utility (mean 53.8±4.4% to 93.0±2.5%, Figure B). The two monkeys displayed the same behavior (48.8±4.1% to 91.2±4.7% for Monkey F and 58.8±7.8% to 94.8 ± 2.0% for monkey Z, P>0.05 between the two animal and P<0.01 between the start and the end of sessions, Figure D &F).


A
B
C
D
E
F
Mean success rate
Number of trials
HC
NC
saline
muscimol
HC
NC
Saline
Muscimol
Mean of first 25 trials
Mean of last 25 trials
*