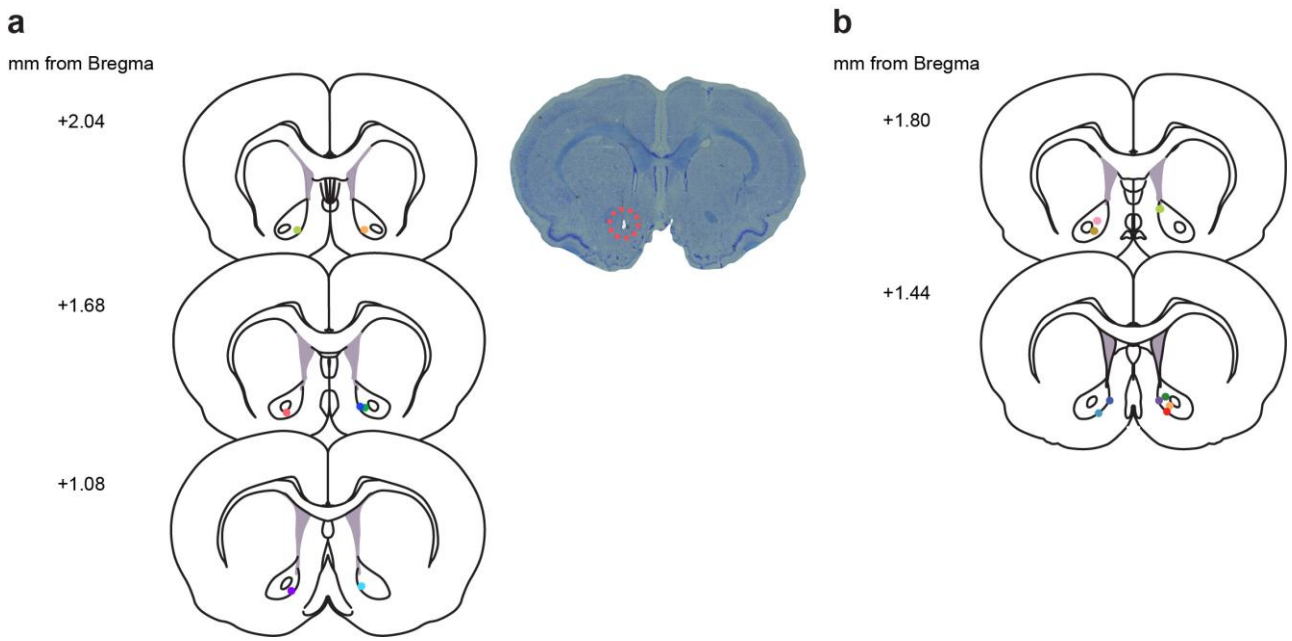
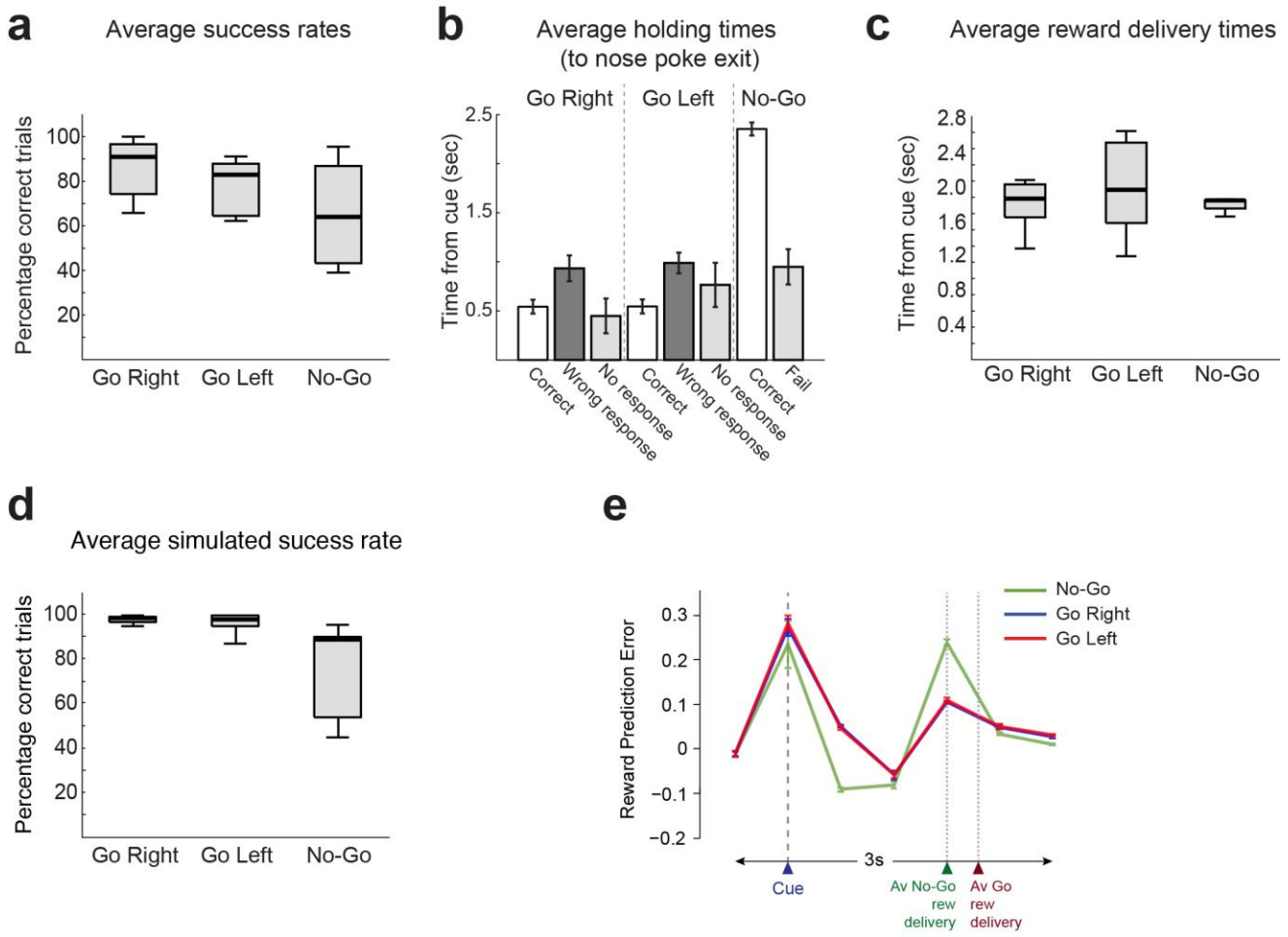


Supplementary Figure 1

Representation of recording sites.

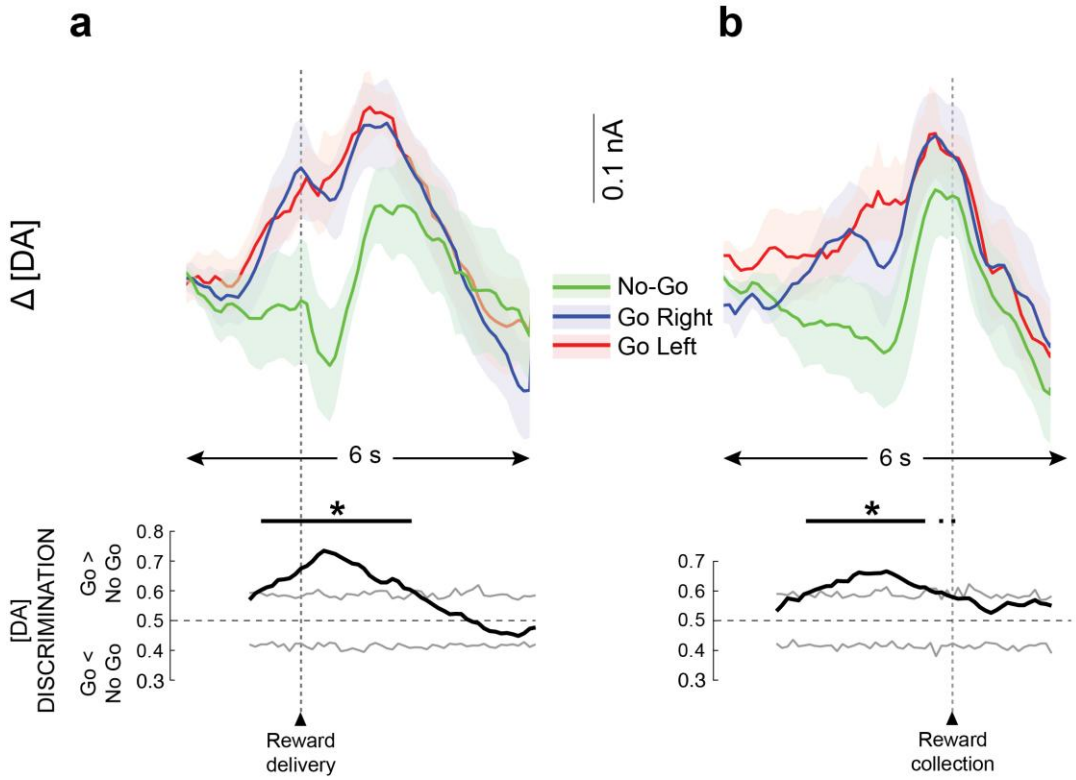
(a) Schematic, along with an example photomicrograph, of the recording locations in the nucleus accumbens core in experiment 1 (n = 7 electrodes in 7 rats). **(b)** Schematic of the recording locations in the nucleus accumbens core in experiment 2 (n = 9 electrodes in 6 rats). The numbers next to each section indicate distance in mm anterior to bregma. Adapted from the atlas of Paxinos and Watson (2005).



Supplementary Figure 2

Behavioral performance and simulations of the actor-critic model in experiment 1.

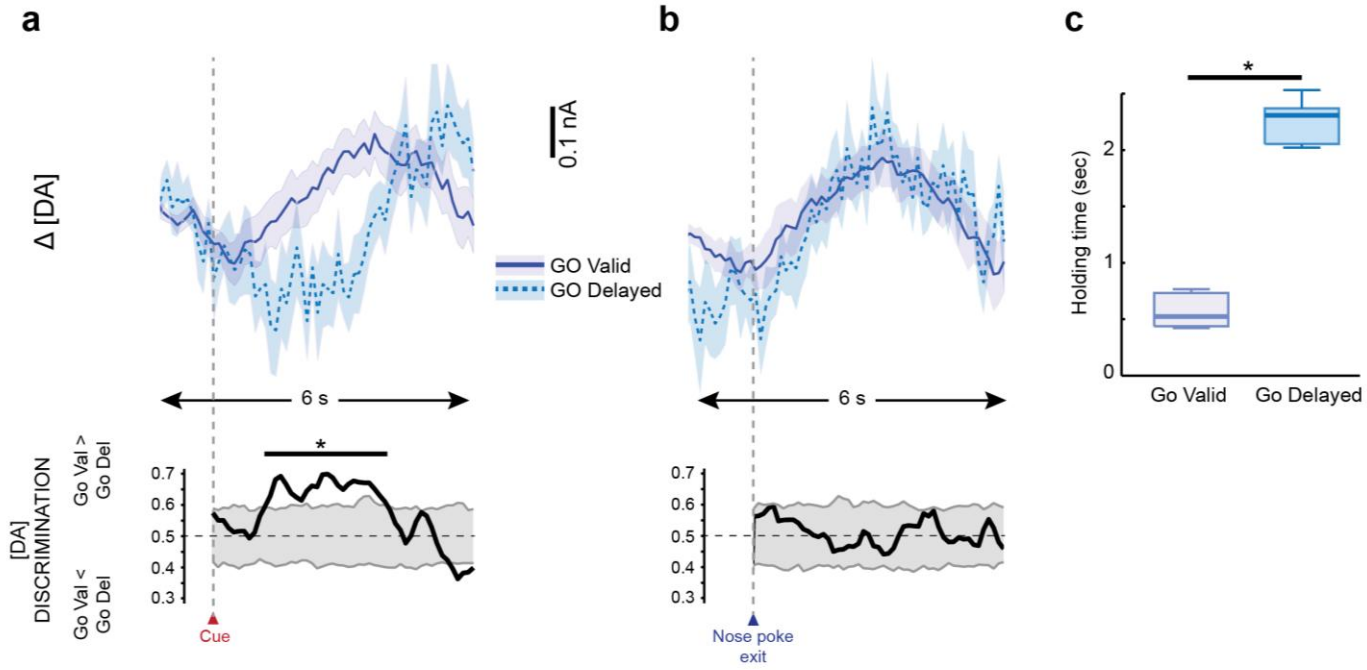
(a) Average success rates, **(b)** holding times from cue onset to head exit (mean ± S.D.), and **(c)** reward delivery times after cue onset during experiment 1. For all box plots, the central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points. **(d)** Box plots of success rates derived from simulations. Note that the model also found the No-Go trials more difficult to perform correctly as there was only a single sequence of actions that resulted in the delivery of the pellet, while all other sequences of actions resulted in the light being turned on signaling an error. By contrast, on Go trials there existed multiple potential sequences that were not immediately incorrect. **(e)** Average RPEs (mean ± S.E.M) recorded in 0.5 s time steps for the different trial types for 7 simulated "rats". Though the model was trained over a long period (that parallels the extended training received by the rats in experiment 1) to achieve qualitatively similar discrepancies in success rates as the real rats, there is nonetheless a positive RPE at cue presentation in all conditions. This occurred because the simulated animal was (i) unable to estimate time precisely from past events, and (ii) its accuracy was not at 100%. In particular, there was a large increase in the RPE at cue onset, because after reward delivery, the next trial could only begin after the ITI had ended, 5 s after reward delivery, so the simulated animal could not fully predict if entering and staying in the nose-poke would initiate a trial. The real rats in experiment 1 in fact made nose-poke responses during the ITI (< 4.5 s since previous reward delivery) on ~20% of trials, which could indicate that they also did not estimate time precisely and hence could not fully predict if staying in the nose poke would initiate a trial. The RPE at cue presentation was numerically smaller on No-Go than on Go trial, because, just like the real rats, the simulated animals had lower accuracy on No-Go trials; thus, they estimated a lower value for the state associated with No-Go cue. There was also an increase in the RPE at the time of reward delivery. This occurred because the same action had to be executed multiple times to result in the pellet delivery, and the state of the simulated animals did not include any information on its past actions, so the simulated animal could not fully predict if its action would result in pellet delivery in the next time step.



Supplementary Figure 3

Dopamine signals in experiment 1 to reward delivery and collection.

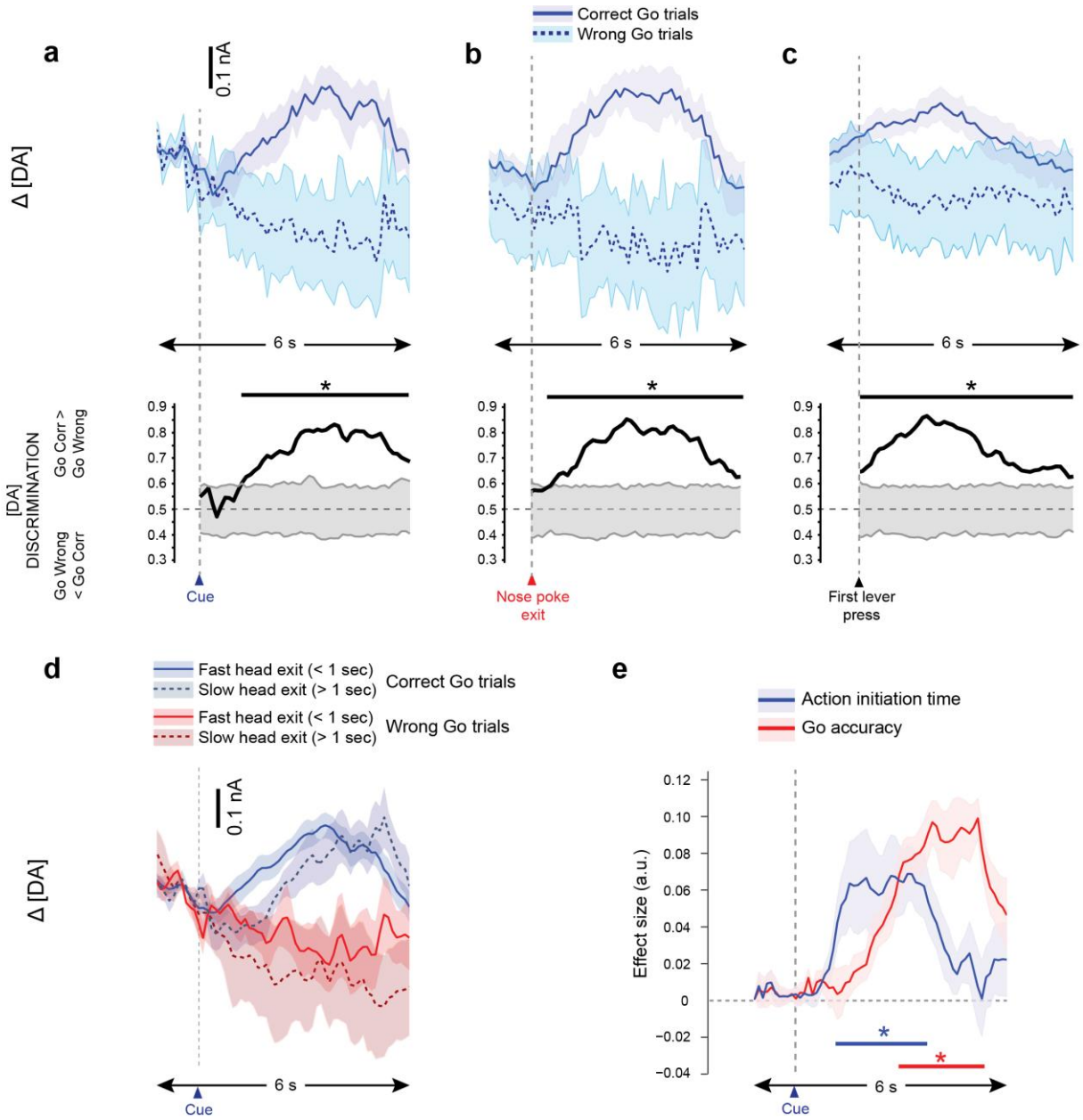
Dopamine data was aligned to reward delivery **(a)** and reward collection from the food magazine **(b)**. Reward delivery occurred after the 2nd lever press on Go trials, and at cue offset on No-Go trials. The food magazine was situated on the opposite wall of the chamber to the nosepoke and levers. It took rats on average 1.85 s or 2.04 s for Go or No-Go trials, respectively, to reach the food magazine after reward delivery.



Supplementary Figure 4

Dopamine signals in delayed Go trials.

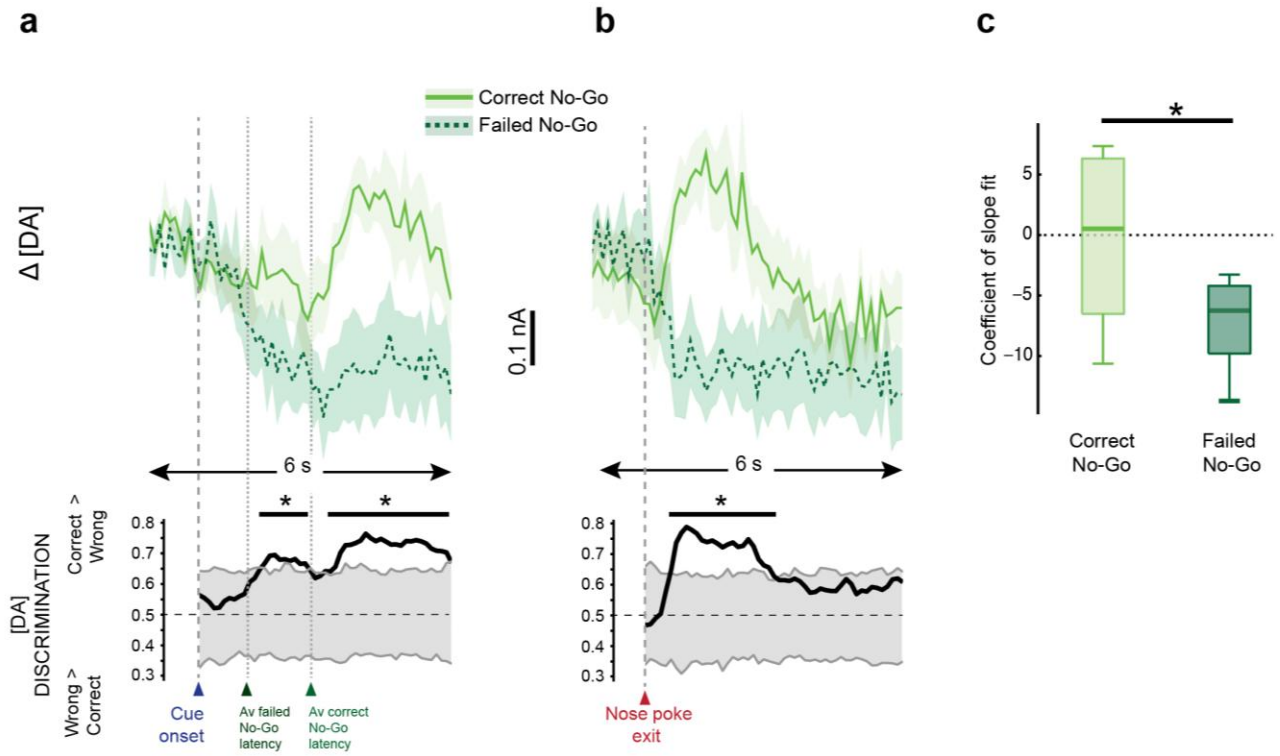
Upper panels display the unsmoothed average dopamine signals from experiment 1 (mean ± S.E.M.) recorded during 'valid' Go trials (post-cue RT < 1.7 s; blue filled line) or 'delayed' Go trials (post-cue RT > 1.7 s; dotted cyan line) (NB: delayed Go trials were not included in any other analyses) aligned to either **(a)** cue onset or **(b)** time of head exit from nose-poke. Lower panels show average discriminability between the Valid and Delayed Go trials for each timepoint (shaded area = population of 1000 permuted sessions; line: *, $p < 0.05$ permutation tests, corrected for multiple comparisons). **(c)** Boxplot of the nose-poke holding times (central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points) for the valid or delayed Go trials (*, $p = 0.02$, $W_7 = 0$, Wilcoxon Signed Ranks Test)



Supplementary Figure 5

Dopamine predicts Wrong Go selections prior to the error being signaled.

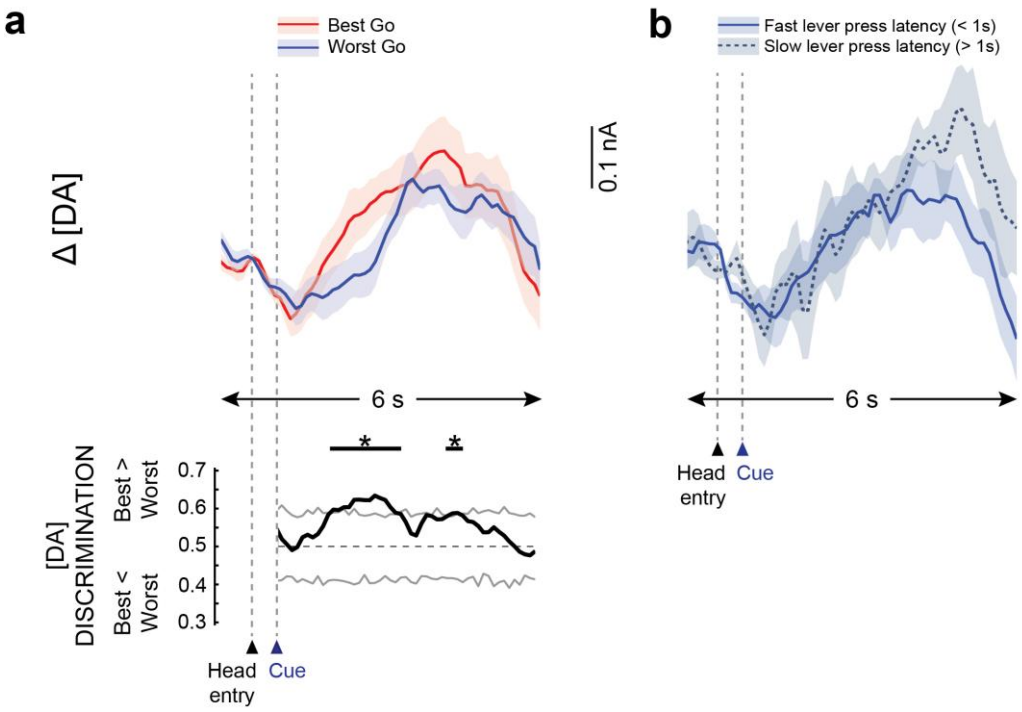
Average unsmoothed dopamine release from experiment 1 (mean ± S.E.M.) recorded during successful (filled line) and wrong (dotted line) Go trials (trials where the animal correctly initiated an action but selected the wrong lever) aligned to the time of **(a, d)** cue onset, **(b)** head exit from nose-poke or **(c)** the first lever press (when the error is signaled). Lower panels show average discriminability between the Correct and Wrong Go trials for each timepoint (shaded area = population of 1000 permuted sessions; line: *, $p < 0.05$ permutation tests, corrected for multiple comparisons). **(d-e)** Comparison of dopamine release on successful and wrong Go trials as a function of speed of action initiation. **(d)** Successful (blue lines) or Wrong (red lines) Go trials sorted by "fast" (< 1 s, filled line) or "slow" (> 1 s, dotted line) nose-poke exit latency. **(e)** Regression weights for trial accuracy and action initiation latency for Go trials (a.u., arbitrary units). On average, rats were both significantly slower to exit the nose poke ($p = 0.03$, $W_7 = 1$, Wilcoxon Signed Rank Test) and to move from the nose poke to make a lever press on Wrong trials ($p = 0.02$, $W_7 = 0$, Wilcoxon Signed Rank Test). Nonetheless, these differences could not fully explain the different patterns of dopamine release on the Correct and Wrong trials.



Supplementary Figure 6

Dopamine on failed No-Go trials decreases when the error is signaled.

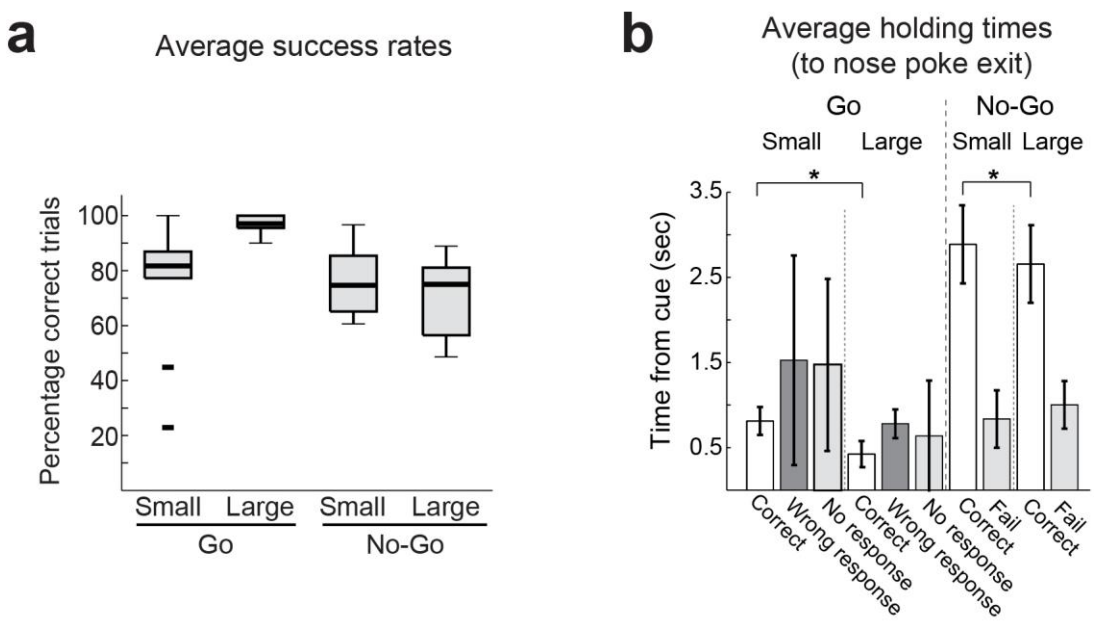
Average unsmoothed dopamine signals from experiment 1 (mean ± S.E.M.) recorded during successful (filled line) and failed (dotted line) No-Go trials (trials where the animal exited the nose poke before the end of the No-Go hold period) aligned to time of **(a)** cue onset or **(b)** head exit from nose-poke (when the error is signaled). Lower panels show average discriminability between the Correct and Failed No-Go trials for each timepoint (shaded area = population of 1000 permuted sessions; line: *, $p < 0.05$ permutation tests, corrected for multiple comparisons). **(c)** Boxplot of the average coefficients for the slopes of the linear fit of the dopamine signal in a 2.5 s window following cue onset for correct and failed No-Go trials (central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points). The decrease in DA following cue onset is significantly greater in failed no-go trials than in successful no-go trials (*, $p = 0.02$, $W_7 = 0$, Wilcoxon Signed Rank Test).



Supplementary Figure 7

Dopamine release scales with success rates on Go trials in experiment 1.

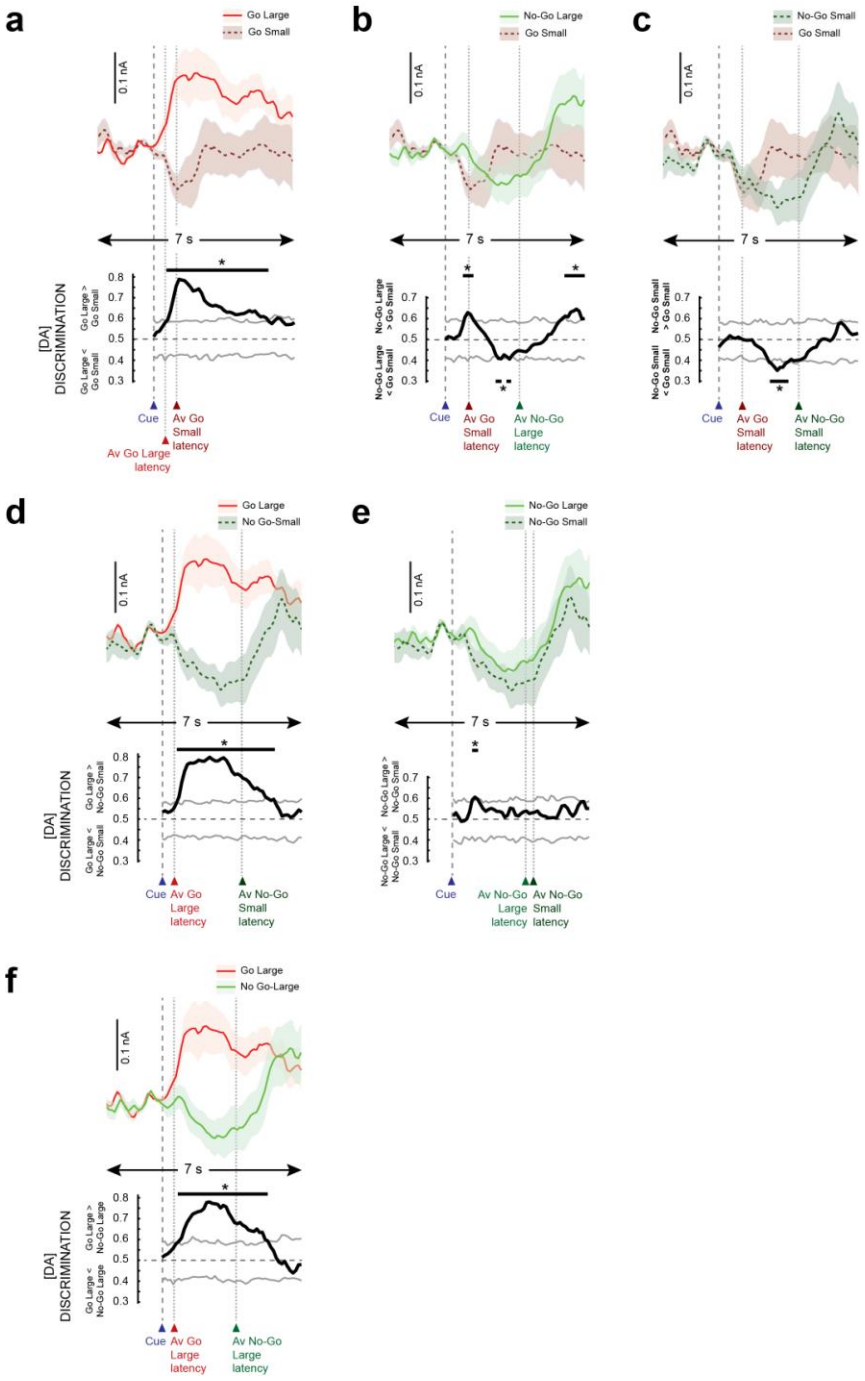
(a) Average dopamine release (mean ± S.E.M.) recorded during successful Go Left and Right trials in experiment 1, arranged in each animal by performance (best / highest success rate = red line, worst / lowest success rate = blue line). Lower panels show average discriminability (black line) between the Best and Worst Go trials for each time point (grey lines = population of 1000 permuted sessions; black bars with *, $p < 0.05$ permutation tests, corrected for multiple comparisons). Note the mean head exit latency was the same between the two conditions (Best Go = 0.59 s, Worst Go = 0.59 s). **(b)** Patterns of dopamine release on Go trials with equivalent nose-poke exit times, divided into trials with fast (< 1 s) or slow (> 1 s) response times to make the 1st lever press (filled and dashed blue lines, respectively), were qualitatively very similar.



Supplementary Figure 8

Behavioral performance in experiment 2.

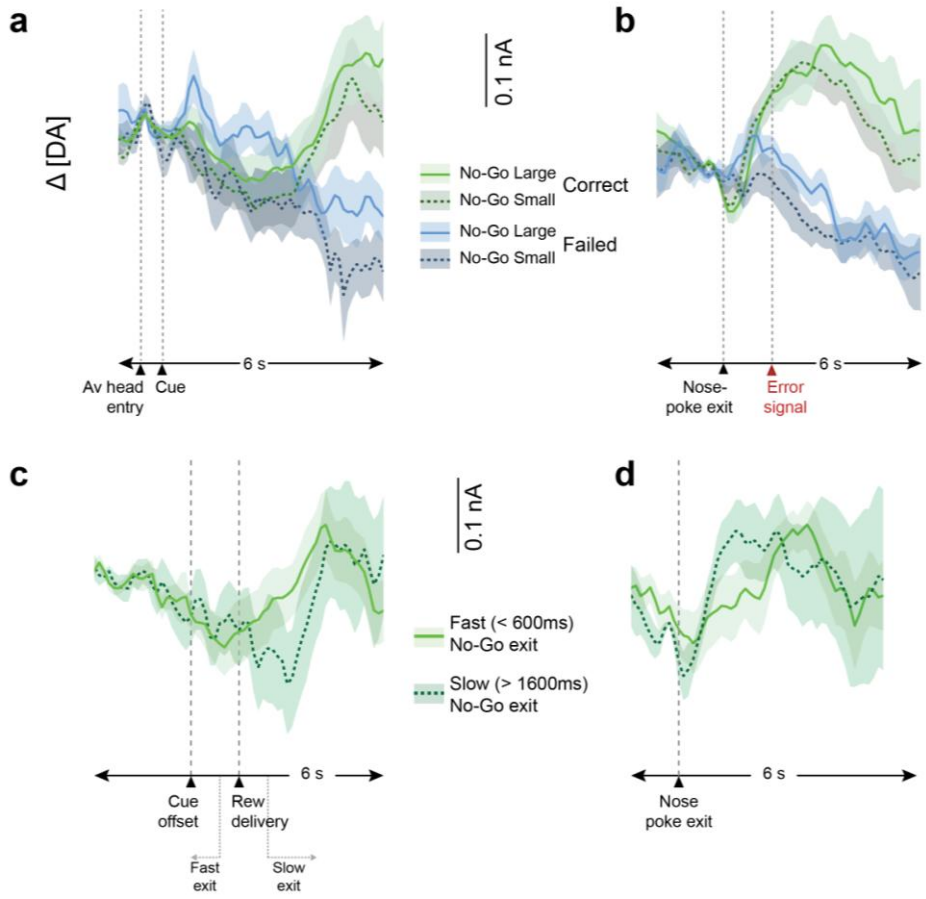
(a) Average success rates in experiment 2. Central mark is the median, box edges are 25th and 75th percentiles, whiskers extend to the most extreme data points not considered outliers (points 1.67 x interquartile range away from the 25th or 75th percentile), and outliers are plotted individually. The two outlier points in the Go Small condition come from the same animal. **(b)** Holding times from cue onset to head exit (mean ± S.D.) on Correct, Wrong Response (wrong lever pressed) or No Response (no lever pressed within 5 s of cue onset) Go trials, or Correct or Failed (premature exit from nose poke) No-Go trials. (*, both $p = 0.03$, $W_6 = 0$, Wilcoxon Signed Rank Test).



Supplementary Figure 9

Comparison between trial types in experiment 2.

Average discriminability between the different trial types in experiment 2 for each time point (grey lines = max and min from population of 1000 permuted sessions; line: *, $p < 0.05$ permutation tests, corrected for multiple comparisons). **(a)** Go Large v Go Small, **(b)** No-Go Large v Go Small, **(c)** No-Go Small v Go Small, **(d)** Go Large v No-Go Small, **(e)** No-Go Large v No-Go Small, **(f)** Go Large v No-Go Small.



Supplementary Figure 10

Dopamine release on No-Go trials in experiment 2 reflects correct action initiation.

(**a, b**) Average dopamine release (mean ± S.E.M.) on correct No-Go trials (green lines) compared to incorrect No-Go trials when the animals left the nose-poke prematurely (blue lines). Data is aligned to cue onset (**a**) or nose-poke exit (**b**). Large Reward trials are plotted with filled lines, Small Reward with dashed lines. Owing to the smaller proportion of Go trials in each session and increased success rate on Go Large trials, there were too few Incorrect Go trials to analyse in experiment 2 (< 2%). (**c, d**) Average No-Go dopamine release (mean ± SEM) on a subset of correctly performed No-Go trials where the rats' exit from the nose-poke was either "fast" (< 600 ms after cue offset, filled green line) or "slow" (> 1,600 ms after cue offset, dashed green line). Data is aligned to cue offset (which occurred 1 s before reward delivery in experiment 2) (**c**) or nose-poke exit (**d**). Therefore, the fast head exit times occur > 400 ms before reward delivery and the slow ones > 600 ms after reward delivery.

Current state			Action	Additional condition	Reward	New state	
Position	Sound	Enviro.				Sound	Enviro.
N	-	-	-	5s passed since reward consumed	0	L or R or N	-
L (or R)	L (or R)	-	P	lever pressed before within the trial	0	-	Pellet
L (or R)	R (or L)	-	P		0	-	Light
N	N	-	-	animal for 1.5s in the nose poke hole	0	-	Pellet
N	N	-	L (or R or F)		0	-	Light
Any	-	Pellet	F		1	-	-
Any	-	Light	Any	5s passed since light turned on	0	-	-

Supplementary Table 1. State transitions in the simulations of the actor-critic model.

Nature Neuroscience: doi:10.1038/nn.4187