LETTER

doi:10.1038/nature19845

A basal ganglia circuit for evaluating action outcomes

Marcus Stephenson-Jones[1], Kai Yu[1], Sandra Ahrens[1], Jason M. Tucciarone[1,2], Aile N. van Huijstee[1], Luis A. Mejia[1], Mario A. Penzo[1], Lung-Hao Tai[3], Linda Wilbrecht[3,4] & Bo Li[1]

The basal ganglia, a group of subcortical nuclei, play a crucial role in decision-making by selecting actions and evaluating their outcomes[1,2]. While much is known about the function of the basal ganglia circuitry in selection[1,3,4], how these nuclei contribute to outcome evaluation is less clear. Here we show that neurons in the habenula-projecting globus pallidus (GPh) in mice are essential for evaluating action outcomes and are regulated by a specific set of inputs from the basal ganglia. We find in a classical conditioning task that individual mouse GPh neurons bidirectionally encode whether an outcome is better or worse than expected. Mimicking these evaluation signals with optogenetic inhibition or excitation is sufficient to reinforce or discourage actions in a decision-making task. Moreover, cell-type-specific synaptic manipulations reveal that the inhibitory and excitatory inputs to the GPh are necessary for mice to appropriately evaluate positive and negative feedback, respectively. Finally, using rabies-virus-assisted monosynaptic tracing[5], we show that the GPh is embedded in a basal ganglia circuit wherein it receives inhibitory input from both striosomal and matrix compartments of the striatum, and excitatory input from the 'limbic' regions of the subthalamic nucleus. Our results provide evidence that information about the selection and evaluation of actions is channelled through distinct sets of basal ganglia circuits, with the GPh representing a key locus in which information of opposing valence is integrated to determine whether action outcomes are better or worse than expected.

The GPh, a phylogenetically conserved non-motor output of the basal ganglia[6–8], excites the lateral habenula (LHb) that, in turn, drives inhibition onto dopamine neurons when an outcome is worse than expected[8–11]. GPh neurons may thus play a key role in evaluating action outcomes by providing a source of 'prediction error' (PE) to the reward system, to drive reinforcement learning. In order to test this hypothesis, we first verified that we could selectively target GPh neurons in the entopeduncular nucleus (EP), the rodent homologue of the primate globus pallidus interna (GPi) where GPh neurons are located[7,8,12,13], on the basis of their expression of vesicular glutamate transporter 2 (Vglut2) and the neuropeptide somatostatin (Som)[6,8,14], and that GPh neurons project exclusively to the LHb[13] (Extended Data Fig. 1).

To examine the function of GPh neurons in relation to outcome evaluation, we modified a classical conditioning task designed for studying value coding in dopamine neurons[15]. Here, a unique auditory conditioned stimulus (CS) predicted the delivery of one of five unconditioned stimuli (US): water rewards (1 and 5 μl), nothing, or air puffs to the face (100 and 500 ms). As training progressed, mice began licking or blinking in response to the reward- or punishment-predicting cues, respectively. The lick rate and blinking occurrence were significantly higher for cues that predicted large rewards and punishments than for cues predicting small rewards and punishments (Fig. 1a, b), indicating that mice had learnt the CS–US associations.

We recorded the activity of EP neurons in *Vglut2-Cre;Ai35* mice, in which Vglut2$^+$ GPh neurons could be optogenetically tagged with the light-sensitive proton pump archaerhodopsin (Arch) (see Methods) (Extended Data Fig. 2a–c), while they performed this task. Hierarchical clustering revealed that all the tagged neurons belonged to one class of neurons, which we classified as putative GPh neurons (Extended Data Fig. 2d–g). By contrast, neurons in two other functional clusters were never optogenetically tagged, and showed an activity profile resembling that of the classic GABAergic movement-related neurons found in the GPi[16] (Extended Data Fig. 2d–j).

Putative GPh neurons were phasically excited by both punishment-predicting tones and punishments (Fig. 1c; Extended Data Fig. 2h).

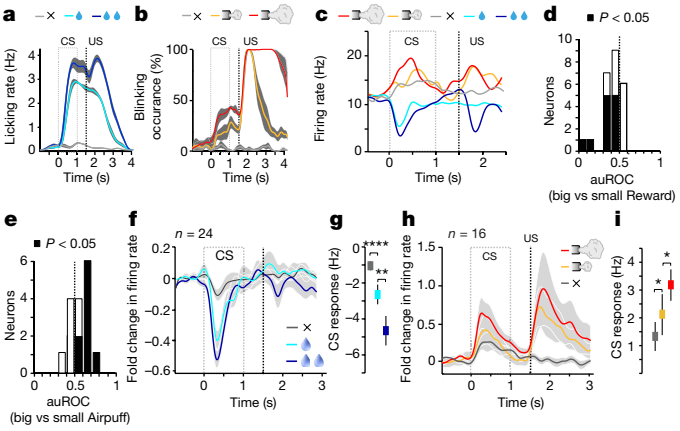


Figure 1 | GPh neurons bidirectionally integrate reward and punishment related information. a, **b**, Licking (**a**) and blinking (**b**) behaviour from a representative experimental session. The dashed boxed area and dashed line indicate the time of CS and US delivery, respectively. Licking rate ($n = 30$ sessions from 7 mice, $F_{(2,18)} = 41.59$, $P < 0.0001$, $P < 0.05$ for all comparisons) and blinking occurrence ($n = 32$ sessions from 4 mice, $F_{(2,9)} = 33.13$, $P < 0.001$, $P < 0.05$ for all comparisons) during the delay between CS and US in recording sessions were compared across different US magnitudes with one-way ANOVA followed by Tukey's test. **c**, Responses of an example putative GPh neuron, shown as spike density functions. **d**, **e**, auROC (area under the receiver operating characteristic) analysis of differences in firing rate between big and small reward trials (**d**), or between big and small punishment trials (**e**), during the peak response to the CS presentation (180–480 ms). Filled bars, $P < 0.05$, *t*-test. **f**, Average response of putative GPh neurons to reward. **g**, Firing rate change during CS predicting reward of different amplitudes (big versus small reward, $z = -3.2$, $^{**}P < 0.01$; small versus no reward, $z = -4.11$, $^{****}P < 0.0001$; Wilcoxon signed-rank test). **h**, Average response of putative GPh neurons to punishment. **i**, Firing rate change during CS predicting punishment of different durations (big versus small punishment, $z = 2.27$, $^{*}P < 0.05$; small versus no punishment, $z = 2.06$, $^{*}P < 0.05$, Wilcoxon signed-rank test). Data are represented as mean ± s.e.m. in **a**, **b**, **f**–**i**.

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. [2]Medical Scientist Training Program & Program in Neuroscience, Stony Brook University, Stony Brook, New York 11790, USA. [3]Department of Psychology, University of California Berkeley, Berkeley, California 94720, USA. [4]Helen Wills Neuroscience Institute, University of California Berkeley, Berkeley, California 94720, USA.

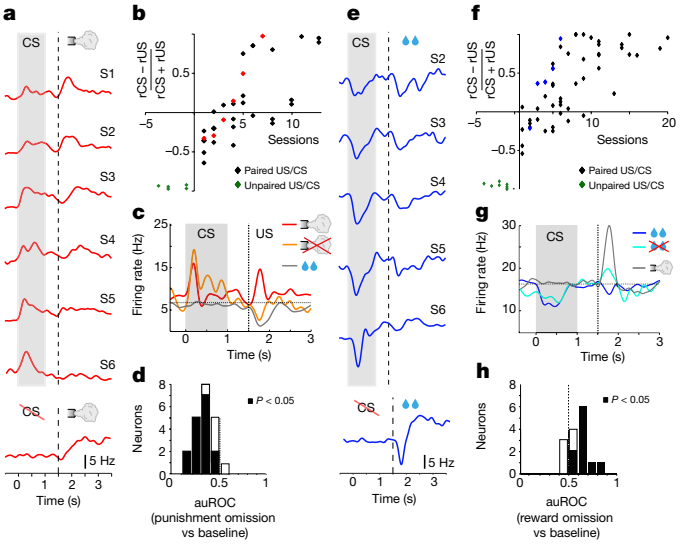


Figure 2 | GPh responses to unconditioned stimuli are modulated by expectation. a, CS and US (air puff) responses of an example putative GPh neuron tracked over multiple sessions (S1–S6). Session-by-session waveform correlations for this individual unit were >0.96. **b**, CS–US (air puff) response index for 36 putative GPh neurons (6 mice) across different stages of training (red dots represent values of the sample neuron in **a**) ($r^2 = 0.56$, $P < 0.0001$ by a linear regression). rCS and rUS represent the CS and US responses respectively. **c**, Responses of an example GPh neuron to an expected air puff (red), an unexpectedly omitted air puff (orange), or an unsignalled reward (grey). **d**, auROC analysis of differences in firing rate between baseline and US presentation time (1.7–1.9 s) in air puff omission trials ($n = 21$ neurons from 6 mice). Filled bars, $P < 0.05$, *t*-test. **e**, CS and US (reward) responses of an example putative GPh neuron tracked over multiple sessions. Session-by-session waveform correlations for this individual unit were >0.97. **f**, CS–US (reward) response index for 60 putative GPh neurons (9 mice) across different stages of training (the blue dots represent values of the sample neuron in **e**) ($r^2 = 0.48$, $P < 0.0001$ by a linear regression). **g**, Responses of an example GPh neuron to an expected reward (blue), an unexpectedly omitted reward (light blue), or an unsignalled air puff (grey). **h**, auROC analysis of differences in firing rate between baseline and US presentation time (1.7–1.9 s) in reward omission trials ($n = 15$ neurons from 4 mice). Filled bars, $P < 0.05$, *t*-test.

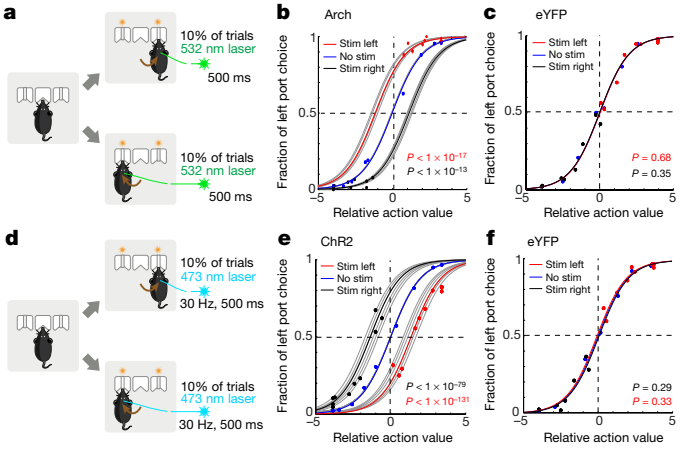


Figure 3 | Optogenetic inhibition or activation of the GPh–LHb pathway bidirectionally influences reinforcement. a, Schematic of the optogenetic inhibition. **b**, Probability of left port choice as a function of action value, for trials immediately following the trials in which photo-inhibition was delivered when mice entered the left (stim left) or right (stim right) port, or was not delivered (no stim). Coloured lines indicate the fit by the logistic regression model on the pooled data for each of the three conditions; grey lines indicate the pooled data for each individual mouse ($n = 5$ mice, 34,627 trials, 6,943 ± 1,330 trials per mouse). **c**, Similar to **b**, except that control mice with eYFP-expressing GPh neurons were used ($n = 6$ mice, 79,589 trials, 13,265 ± 596 trials per mouse). **d**, **e**, Similar to **a** and **b**, except that optogenetic activation of the GPh–LHb projection was applied ($n = 6$ mice, 42,292 trials, 4,424 ± 1,806 trials per mouse). **f**, Similar to **e**, except that control mice with eYFP-expressing GPh neurons were used ($n = 6$ mice, 45,389 trials, 7,564 ± 2,120 trials per mouse). In **b**, **c**, **e** and **f**, *P* values are reported for *t*-tests where H_0: $\beta_{stim} = 0$.

In a portion of these neurons, the magnitude of single-neuron tone responses was greater when the tone predicted a larger punishment (Fig. 1c, e). In addition, as in primates[7,17], GPh neurons were phasically inhibited by both reward-predicting tones and rewards (Fig. 1c; Extended Data Fig. 2d–f); the inhibition was greater when the tone predicted a larger reward (Fig. 1c, d). The average magnitude of the CS responses in these neurons was graded, reflecting the expected magnitude of reward or punishment (Fig. 1f–i). We conclude that individual GPh neurons bidirectionally encode the expected value of an action as well as the value of its outcome.

To determine how these expectation and outcome signals develop, we recorded GPh neurons over the course of the behavioural training (Extended Data Fig. 3a). Prior to training, GPh neurons showed no response to the CS (Extended Data Fig. 3b, c). During training, these neurons rapidly acquired a CS response, which was initially smaller but became gradually larger than the US response as training proceeded (Fig. 2a, b, e, f; Extended Data Fig. 3d, e). Indeed, once the animals had fully learnt the task, and thus could readily predict the US, the response to the US (reward or punishment) was markedly suppressed or even absent in the majority of neurons (Fig. 2a, b, e, f; Extended Data Fig. 3d, e). However, unexpected delivery of the US still evoked a response, even in neurons that no longer responded to the US when it was predicted (Fig. 2a, e; Extended Data Fig. 3d, e). This reduction in US response in GPh neurons is consistent with encoding of PE, a function well-described for dopamine neurons[15,18].

Another hallmark of PE coding is that neurons respond when an expected outcome is omitted[15,19]. To test whether GPh neurons have such a property, we omitted an expected US in 10% of the large reward or punishment trials. When an expected punishment was omitted, putative GPh neurons displayed a decrease in firing either compared to when the punishment was delivered (Fig. 2c; Extended Data Fig. 3f, g), or compared with baseline (Fig. 2d). In contrast, upon omission of an expected reward, putative GPh neurons showed an increase in firing rate relative to delivery of the reward (Fig. 2g; Extended Data Fig. 3h, i), and relative to baseline (Fig. 2h). Together, these results demonstrate that GPh neurons encode reward and punishment PEs, bidirectionally signalling when an outcome is better or worse than expected.

The observed bidirectional responses to reward and punishment suggest that inhibition or excitation of the GPh may influence behaviour. Whereas excitation of the GPh is aversive[8], inhibition of it may be rewarding. To test this, we introduced Arch selectively into GPh neurons of *Vglut2-Cre* mice and optogenetically inhibited these neurons when the mice performed a real-time place preference task. Inhibition of either the somata of GPh neurons (Extended Data Fig. 4a–c) or their axon terminals in the LHb (Extended Data Fig. 4d–f) induced a preference in these mice for the inhibition-paired chamber (Extended Data Fig. 5a–i). Mice would also actively work to receive inhibition of the GPh (Extended Data Fig. 5j). These effects were not specific to the *Vglut2-Cre* mice, as optogenetic inhibition or excitation (using channelrhodopsin (ChR2), the light-gated cation channel) of the GPh, targeted in wild-type mice with a retrograde canine adenovirus (CAV2–Cre)[20], also induced real-time place preference or aversion, respectively (Extended Data Figs 4g–l, 5k–m). None of our optogenetic manipulations had an effect on the velocity or distance of movement (Extended Data Fig. 5n, o). These data show that excitation and inhibition of the GPh have opposing motivational valence, with the former being aversive and the latter rewarding.

GPh neurons' response properties and their influence on behaviour point to the possibility that they could play a fundamental role in evaluating action outcomes. To test this possibility, we trained *Vglut2-Cre* mice in a probabilistic switching task, where animals had to rely on the evaluation of previous choice outcomes to update their future decisions (Extended Data Fig. 6a–f). Previous studies have shown that mice adopt a 'win-stay, lose-switch' strategy in this task that can be best described by a multivariate logistic regression model[21] (Extended Data Fig. 6a–f; also see Methods).

We used optogenetics to specifically inhibit or activate GPh neurons (Extended Data Fig. 4a–c, m–o) at the moment of outcome evaluation in 10% of the trials in this task. Inhibition of GPh neurons at the time that a mouse poked its nose in a water port significantly biased the mouse to return to the same port on the subsequent trial (Fig. 3a, b; Supplementary Video 1). Conversely, activation of the GPh–LHb pathway when a mouse poked its nose in a water port significantly promoted the mouse to switch to the alternative port on the subsequent trial (Fig. 3d, e; Supplementary Video 2). In both cases, the probability of repeating the same choice was dependent on both the optogenetic manipulation and the previous reward history. Optogenetic inhibition or activation of GPh neurons shifted the sigmoidal decision curve along the *x* axis (Fig. 3b, e), indicating that these manipulations mimicked a fixed increase or decrease, respectively, in the value of a chosen action (Extended Data Fig. 6g–i; also see Methods). Alternative models did not provide a better explanation for our data (Supplementary Table 1). In control experiments, light illumination of GPh neurons expressing eYFP produced no effect on choice behaviour (Fig. 3c, f; Extended Data Fig. 6g, h). These results indicate that bidirectional changes in GPh activity are sufficient to bias outcome evaluation, thereby reinforcing or discouraging particular actions.

To examine whether the GPh also contributes to action selection, we optogenetically inhibited or activated the GPh in the same mice, but at the time when they nose-poked the central port to initiate a trial, a time frame that may coincide or overlap with the moment of action selection[21] (Extended Data Fig. 7). Neither of these manipulations produced an effect on choice (Extended Data Fig. 7a–h). In addition, these manipulations did not appear to influence ongoing behaviour (Extended Data Fig. 7i–l). Thus, activation or inhibition of the GPh influences the evaluation, but not selection, of actions.

Although these optogenetic experiments demonstrate that bidirectional changes in GPh activity are sufficient to reinforce or discourage actions, they do not determine whether endogenous GPh activity is required for such function. We reasoned that if the excitatory and inhibitory responses of GPh neurons to action outcomes are essential in providing negative and positive feedback, respectively, then reducing the excitatory or inhibitory inputs onto these neurons should accordingly

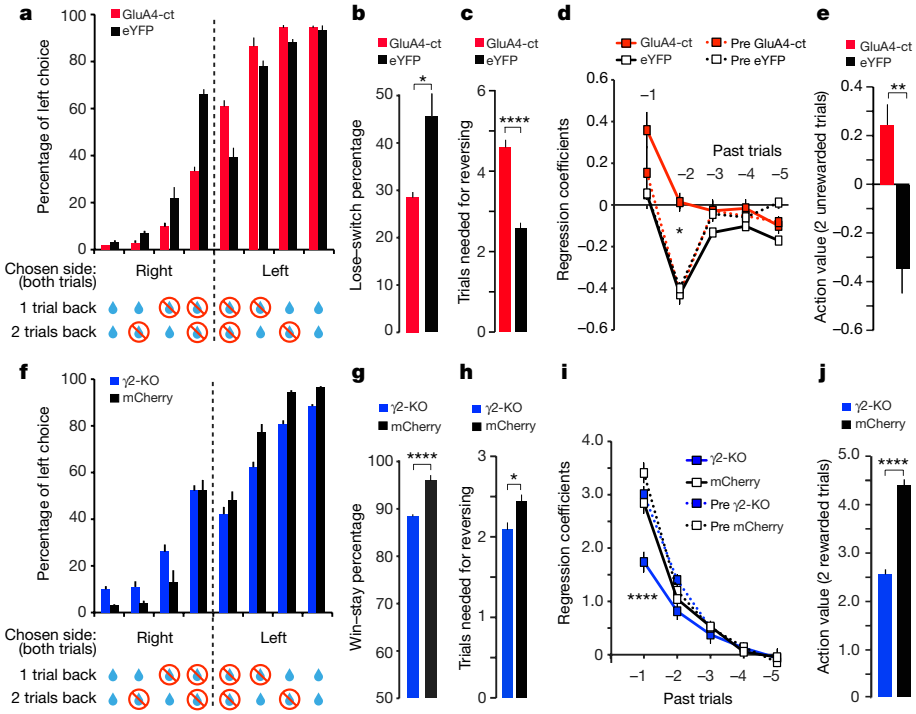


Figure 4 | Reducing glutamatergic or GABAergic drive onto GPh neurons decreases sensitivity to negative or positive feedback, respectively. **a**, Bar graphs showing the increased perseverance of $GPh^{GluA4\text{-}ct}$ mice ($n = 10$) compared to GPh^{eYFP} controls (in which eYFP was introduced into GPh neurons by a Cre-dependent virus; $n = 7$). For clarity, only choices for which mice had previously made two consecutive responses at the same port are shown. **b**, The lose–switch percentage in these mice ($GPh^{GluA4\text{-}ct}$, 31.13 ± 1.7%; GPh^{eYFP}, 45.67 ± 5.1%; $t_{(13)} = 2.58$, $^{*}P < 0.05$, *t*-test). **c**, The number of trials mice took before reversing choice after reward contingencies were switched ($GPh^{GluA4\text{-}ct}$, 4.59 ± 0.18 trials; GPh^{eYFP}, 2.56 ± 0.15 trials; $t_{(13)} = 6.02$, $^{****}P < 0.0001$, *t*-test). **d**, The negative regression coefficients associated with the past five trials for $GPh^{GluA4\text{-}ct}$ mice and GPh^{eYFP} mice before and after surgery (first two trials back × groups, $F_{(3,33)} = 6.566$, $P < 0.01$; $^{*}P < 0.05$ for $GPh^{GluA4\text{-}ct}$ compared to all other groups on the second trial back; two-way ANOVA followed by Bonferroni's test). **e**, The action value of two sequentially unrewarded trials, derived from the sum of their regression coefficients ($t_{(16)} = 3.46$, $^{**}P < 0.01$, *t*-test). **f**, Bar graphs showing decreased perseverance in $GPh^{\gamma2\text{-}KO}$ mice ($n = 8$) compared to $GPh^{mCherry}$ controls (in which mCherry was introduced into GPh neurons by a Flp-dependent virus; $n = 8$). Only choices where mice previously made two consecutive responses at the same port are shown. **g**, The win–stay percentage in these mice ($GPh^{\gamma2\text{-}KO}$, 89.0 ± 0.7%; $GPh^{mCherry}$, 95.6 ± 0.5%; $t_{(16)} = -6.61$, $^{****}P < 0.0001$, *t*-test). **h**, The number of trials mice took before reversing choice after reward contingencies were switched ($GPh^{\gamma2\text{-}KO}$, 2.08 ± 0.26 trials; $GPh^{mCherry}$, 2.45 ± 0.27 trials; $t_{(16)} = -2.74$, $^{*}P < 0.05$, *t*-test). **i**, The positive regression coefficients associated with the past five trials for $GPh^{\gamma2\text{-}KO}$ mice and $GPh^{mCherry}$ mice before and after surgery (first two trials back × groups, $F_{(3,31)} = 42.10$, $P < 0.0001$; $^{****}P < 0.0001$ for $GPh^{\gamma2\text{-}KO}$ compared to all other groups on the first trial back; two-way ANOVA followed by Bonferroni's test). **j**, The action value of two sequentially rewarded trials, derived from the sum of their regression coefficients ($t_{(16)} = -7.49$, $^{****}P < 0.0001$, *t*-test). All data are represented as mean ± s.e.m.

impair choice behaviour. To test this hypothesis, we first selectively weakened glutamatergic synapses onto GPh neurons by expressing in these neurons the C-terminal tail of AMPA receptor (AMPAR) subunit GluA4 (GluA4-ct) (Extended Data Fig. 8a, e). The GluA4-ct inhibits excitatory synaptic transmission by blocking AMPAR synaptic trafficking[22] (Extended Data Fig. 8b–d). Notably, mice expressing the GluA4-ct in the GPh ($GPh^{GluA4\text{-}ct}$) were significantly less likely to switch their choice following an unrewarded outcome (lose–switch) (Fig. 4a, b), and were slower to reverse their choice when the reward contingencies were switched (Fig. 4c). On the other hand, $GPh^{GluA4\text{-}ct}$ mice had no change in the ability to repeat the same choice following a rewarded outcome (win–stay) (Extended Data Fig. 8f). Logistic regression analysis revealed that $GPh^{GluA4\text{-}ct}$ mice had a substantial reduction in their negative regression coefficients, the weighted contribution that past unrewarded trials had on current choice (Fig. 4d, e; Extended Data Fig. 8g), but had no change in their positive regression coefficients (Extended Data Fig. 8h, i). These results indicate that reducing the glutamatergic transmission onto the GPh impairs negative feedback by selectively diminishing the impact of unrewarded outcomes on future decisions.

Next, we selectively weakened GABAergic synapses onto GPh neurons. To this end, we injected the EP of *Som-Flp;Gabrg2^{flox}* mice[23] with a virus expressing Cre in a Flp-dependent manner[22] (Extended Data Fig. 8j, n). In these mice, only GPh neurons within the EP could express Cre and thus have the γ2 subunit of $GABA_A$ receptor ablated ($GPh^{\gamma2\text{-}KO}$). As expected, this approach led to a significant reduction of $GABA_A$-mediated synaptic transmission onto GPh neurons (Extended Data Fig. 8k–m). Compared with control mice, $GPh^{\gamma2\text{-}KO}$ mice were less persistent in response to positive feedback (Fig. 4f), showed a substantial reduction in win–stay (Fig. 4g) but no change in lose–switch behaviour (Extended Data Fig. 8o), and were faster to reverse their choice when the reward contingencies were switched (Fig. 4h), presumably owing to decreased sensitivity to reward. Logistic regression analysis revealed that the positive regression coefficients—the weighted contributions that past rewarded trials had on current choice—were reduced in $GPh^{\gamma2\text{-}KO}$ mice (Fig. 4i, j; Extended Data Fig. 8p). By contrast, there was no overall change in the negative regression coefficients in $GPh^{\gamma2\text{-}KO}$ mice (Extended Data Fig. 8q, r). These results indicate that reducing the inhibitory input onto the GPh impairs positive feedback by selectively reducing the impact past rewarded outcomes had on future choice.

To identify the circuits upstream of the GPh that may provide the reward and punishment information, we used a modified rabies virus system to trace the monosynaptic inputs onto GPh neurons[5] (Extended Data Fig. 9a–c). We found that, like the canonical basal ganglia output nuclei GPi/SNr (substantia nigra pars reticulate)[4], the GPh received inputs directly from the striatum (Fig. 5a; Extended Data Fig. 9d). However, unlike the GPi/SNr[12,24], a large proportion of the inputs to the GPh arose from the striosomal compartment of the striatum (Fig. 5b, c). These striatal inputs could drive monosynaptic GABAergic responses in the GPh (Fig. 5d, e). These data indicate that reward-related evaluation signals in the GPh may at least in part arise from subsets of neurons in both the striosomal and matrix compartments of the striatum.

In addition to this direct projection from the striatum, we found that the GPh was also regulated by distinct nuclei associated with the 'indirect pathway'. In contrast to the GPi, which receives excitatory input from the core of the subthalamic nucleus (STN)[25] and input from parvalbumin-positive GPe neurons[26], the GPh received input from the subthalamic cells located in the 'limbic' region of the STN[27], on the medial border of this nucleus and in the surrounding parasubthalamic nucleus (pSTN), and GPe input from mainly parvalbumin-negative neurons (Fig. 5a, f; Extended Data Fig. 9e–i). This medial limbic STN could drive monosynaptic excitation in GPh neurons (Fig. 5g, h), and may thus provide these neurons with negative valence information[28]. These results indicate that the GPh is embedded in a basal ganglia circuit that is intermingled with, but distinct from, the circuitry that regulates the GPi/SNr (Extended Data Fig. 9j).

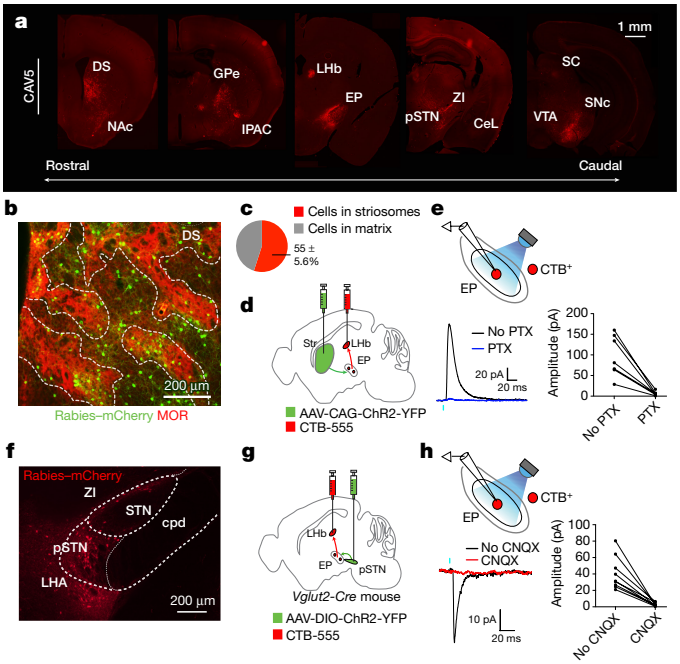


Figure 5 | Identification of monosynaptic inputs to the GPh. **a**, Series of coronal sections, ipsilateral to site of injection, from a representative mouse (that is, CAV5) showing the major monosynaptic inputs to the GPh. **b**, Confocal image of the dorsal striatum (DS) with monosynaptically labelled neurons (green) and μ opioid receptor (MOR) immunostaining that labels the striosomes (red). **c**, Quantification of monosynaptically labelled cells in striatal subcompartments. **d**, The injection strategy. **e**, Schematic of recording configuration (upper) and sample inhibitory postsynaptic currents (IPSCs) induced by optogenetic activation of the striatal input to the GPh (lower left). These IPSCs were blocked by picrotoxin (PTX) (lower right). **f**, Image of the pSTN with monosynaptically labelled neurons (red). **g**, The injection strategy. **h**, Schematic of recording configuration (upper) and sample excitatory postsynaptic currents (EPSCs) induced by optogenetic activation of pSTN input to the GPh (lower left). These EPSCs were blocked by CNQX (lower right). cpd, cerebal peduncle, Cel, central amygdalar nucleus, lateral part, DS, dorsal striatum, IPAC, interstitial nucleus of the posterior limb of the anterior commissure, LHA, lateral hypothalamic area, NAc, nucleus accumbens, SC, superior colliculus, SNc, substantia nigra pars compacta, VTA, ventral tegmental area, ZI, zona incerta. Diagrams in **d** and **g** are modified from the Allen Mouse Brain Atlas, Allen Institute for Brain Science; available from http://mouse.brain-map.org/.

Together, our results demonstrate that the GPh is a key locus where information of opposing valence is integrated, from a subset of basal ganglia circuits, to determine if an action is better or worse than expected. The outcome evaluation function of the GPh is probably mediated through bidirectional control of dopamine neurons, in which PE coding is critical for reinforcement learning. The GPh is well placed to bidirectionally influence dopaminergic activity as it provides tonic excitatory input to the LHb[6,8], which in turn regulates dopamine neurons[7,9,11] disynaptically via the GABAergic rostromedial tegmental nucleus[7,29]. Indeed, bidirectional changes in LHb firing have opposing effects on dopamine cell firing[7,9,11], and lesions of the LHb disrupt negative- and impair positive-reward PE coding in dopaminergic neurons[30]. We propose that an increase in GPh activity when an outcome is worse than expected increases the excitatory drive onto the LHb to inhibit dopamine neurons and discourage actions, whereas decreases in GPh activity when an outcome is better than expected remove the tonic excitation of the LHb to increase dopaminergic activity and reinforce actions (Extended Data Fig. 10).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 October 2015; accepted 14 September 2016.
Published online 21 September 2016.

1. Nelson, A. B. & Kreitzer, A. C. Reassessing models of basal ganglia function and dysfunction. *Annu. Rev. Neurosci.* **37,** 117–135 (2014).
2. Amemori, K., Gibb, L. G. & Graybiel, A. M. Shifting responsibly: the importance of striatal modularity to reinforcement learning in uncertain environments. *Front. Hum. Neurosci.* **5,** 47 (2011).
3. Hikosaka, O. Basal ganglia mechanisms of reward-oriented eye movement. *Ann. NY Acad. Sci.* **1104,** 229–249 (2007).
4. Alexander, G. E. & Crutcher, M. D. Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends Neurosci.* **13,** 266–271 (1990).
5. Callaway, E. M. & Luo, L. Monosynaptic circuit tracing with glycoprotein-deleted rabies viruses. *J. Neurosci.* **35,** 8979–8985 (2015).
6. Stephenson-Jones, M., Kardamakis, A. A., Robertson, B. & Grillner, S. Independent circuits in the basal ganglia for the evaluation and selection of actions. *Proc. Natl Acad. Sci. USA* **110,** E3670–E3679 (2013).
7. Hong, S. & Hikosaka, O. The globus pallidus sends reward-related signals to the lateral habenula. *Neuron* **60,** 720–729 (2008).
8. Shabel, S. J., Proulx, C. D., Trias, A., Murphy, R. T. & Malinow, R. Input to the lateral habenula from the basal ganglia is excitatory, aversive, and suppressed by serotonin. *Neuron* **74,** 475–481 (2012).
9. Matsumoto, M. & Hikosaka, O. Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature* **447,** 1111–1115 (2007).
10. Hong, S., Jhou, T. C., Smith, M., Saleem, K. S. & Hikosaka, O. Negative reward signals from the lateral habenula to dopamine neurons are mediated by rostromedial tegmental nucleus in primates. *J. Neurosci.* **31,** 11457–11471 (2011).
11. Stamatakis, A. M. & Stuber, G. D. Activation of lateral habenula inputs to the ventral midbrain promotes behavioral avoidance. *Nat. Neurosci.* **15,** 1105–1107 (2012).
12. Rajakumar, N., Elisevich, K. & Flumerfelt, B. A. Compartmental origin of the striato-entopeduncular projection in the rat. *J. Comp. Neurol.* **331,** 286–296 (1993).
13. Parent, M., Lévesque, M. & Parent, A. Two types of projection neurons in the internal pallidum of primates: single-axon tracing and three-dimensional reconstruction. *J. Comp. Neurol.* **439,** 162–175 (2001).
14. Vincent, S. R. & Brown, J. C. Somatostatin immunoreactivity in the entopeduncular projection to the lateral habenula in the rat. *Neurosci. Lett.* **68,** 160–164 (1986).
15. Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B. & Uchida, N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* **482,** 85–88 (2012).
16. DeLong, M. R., Crutcher, M. D. & Georgopoulos, A. P. Primate globus pallidus and subthalamic nucleus: functional organization. *J. Neurophysiol.* **53,** 530–543 (1985).
17. Bromberg-Martin, E. S., Matsumoto, M., Hong, S. & Hikosaka, O. A pallidus-habenula-dopamine pathway signals inferred stimulus values. *J. Neurophysiol.* **104,** 1068–1076 (2010).
18. Pan, W. X., Schmidt, R., Wickens, J. R. & Hyland, B. I. Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. *J. Neurosci.* **25,** 6235–6242 (2005).
19. Schultz, W. Dopamine reward prediction-error signalling: a two-component response. *Nat. Rev. Neurosci.* **17,** 183–195 (2016).
20. Bru, T., Salinas, S. & Kremer, E. J. An update on canine adenovirus type 2 and its vectors. *Viruses* **2,** 2134–2153 (2010).
21. Tai, L. H., Lee, A. M., Benavidez, N., Bonci, A. & Wilbrecht, L. Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. *Nat. Neurosci.* **15,** 1281–1289 (2012).
22. Ahrens, S. *et al.* ErbB4 regulation of a thalamic reticular nucleus circuit for sensory selection. *Nat. Neurosci.* **18,** 104–111 (2015).
23. Wulff, P. *et al.* From synapse to behavior: rapid modulation of defined neuronal types with engineered GABAA receptors. *Nat. Neurosci.* **10,** 923–929 (2007).
24. Fujiyama, F. *et al.* Exclusive and common targets of neostriatofugal projections of rat striosome neurons: a single neuron-tracing study using a viral vector. *Eur. J. Neurosci.* **33,** 668–677 (2011).
25. Kita, H. & Kitai, S. T. Efferent projections of the subthalamic nucleus in the rat: light and electron microscopic analysis with the PHA-L method. *J. Comp. Neurol.* **260,** 435–452 (1987).
26. Mastro, K. J., Bouchard, R. S., Holt, H. A. & Gittis, A. H. Transgenic mouse lines subdivide external segment of the globus pallidus (GPe) neurons and reveal distinct GPe output pathways. *J. Neurosci.* **34,** 2087–2099 (2014).
27. Hamani, C., Saint-Cyr, J. A., Fraser, J., Kaplitt, M. & Lozano, A. M. The subthalamic nucleus in the context of movement disorders. *Brain* **127,** 4–20 (2004).
28. Breysse, E., Pelloux, Y. & Baunez, C. The good and bad differentially encoded within the subthalamic nucleus in rats(1,2,3). *eNeuro* **2,** ENEURO.0014-15.2015 (2015).
29. Jhou, T. C., Fields, H. L., Baxter, M. G., Saper, C. B. & Holland, P. C. The rostromedial tegmental nucleus (RMTg), a GABAergic afferent to midbrain dopamine neurons, encodes aversive stimuli and inhibits motor responses. *Neuron* **61,** 786–800 (2009).
30. Tian, J. & Uchida, N. Habenula lesions reveal that multiple mechanisms underlie dopamine prediction errors. *Neuron* **87,** 1304–1316 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Cutrone, D. Li, and G.-R. Hwang for technical assistance, V. Rao and N. Uchida for sharing the Matlab code for the ROC and clustering analysis, S. D. Shea and S. H. Ebbesen for critical reading of the manuscript, Z. J. Huang for providing mouse strains, and members of the Li laboratory for helpful discussions. This work was supported by grants from the National Institutes of Health (NIH) (R01MH108924 to B.L.), the Dana Foundation (to B.L.), NARSAD (to B.L., M.S. and S.A.), Louis Feil Trust (to B.L.), the Stanley Family Foundation (to B.L.), Simons Foundation (to B.L.), Wodecroft Foundation (to B.L.), and an EMBO Long-Term Fellowship Award (to M.S.).

Author Contributions M.S. and B.L. designed the study. M.S. conducted experiments and analysed the data. K.Y. assisted with analysis and implementation of the *in vivo* recording experiments. S.A. and M.P. performed the patch clamp recording experiments. A.N.H. assisted with behavioural training and tracing experiments. J.T. designed and produced the starter virus for the rabies tracing and assisted all the rabies tracing experiments. L.M. assisted with the *in vitro* electrophysiology experiments. L.H.T. and L.W. assisted with the analysis of behaviour in the probabilistic switching task. M.S. and B.L. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.S. (mstephen@cshl.edu) or B.L. (bli@cshl.edu).

Reviewer Information *Nature* thanks G. Stuber and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Animals. All procedures were approved by the Institutional Animal Care and Use Committee of Cold Spring Harbour Laboratory (CSHL) and conducted in accordance to the United States National Institutes of Health guidelines. Mice were housed under a 12-h light–dark cycle (8 a.m. to 8 p.m. light). All behavioural experiments were performed during the light cycle. All mice had free access to food, but water was restricted for mice used in certain behavioural experiments. Free water was provided on days with no experimental sessions. Male and female mice 2–4 months of age were used in all experiments. No differences were observed in the behaviour of male or female mice during the switching task or in our optogenetic or synaptic manipulations for this behaviour (see below). All animals were randomly allocated to the different experimental conditions used in this study. The *Vglut2-Cre* (*Slc17a6*$^{tm2(Cre)Lowl}$/J, stock #016963 from Jackson Laboratory), *Ai35* (*Gt(ROSA)26Sor*$^{tm35.1(CAG-aop3/GFP)Hze}$/J, stock #012735 from Jackson Laboratory), *Rosa26-stop*^{flox}*-H2b-GFP* (from Z. J. Huang)[31], *Som-Flp* (from Z. J. Huang)[32], *Gabrg2*^{flox} (*Gabrg2*^{tm1Wul}/J, stock #021197 from Jackson Laboratory)[23], *Rosa26-stop*^{flox}*-tTA* (stock #012266 from Jackson Laboratory)[32,33] mouse strains have all been previously characterized. All mice were bred onto a C57BL/6J background.

Viral vectors. All adeno-associated viruses (AAV) were produced by the University of North Carolina vector core facility or the University of Pennsylvania vector core and have previously been described: AAV9-Ef1a-DIO-hChR2(H134R)-eYFP, AAV9-CAG-FLEX-ArchT-GFP, AAV9-Ef1a-DIO-eYFP, AAV8-hSyn-DIO-mCherry, AAV1-Syn-GCAMP6f.WPRE.SV40 (used for non-Cre dependant viral tracing), AAV9-CAG-ChR2-GFP, AAV9-DIO-GluA4-ct-GFP[22], AAV9-CAG-FSF-GFP-T2A-nCre (which expresses Cre in a Flp-dependent manner[22]), and AAV8-Ef1a-fDIO-2A-mCherry (which expresses mCherry in a Flp-dependent manner[22]). CAV2-Cre was purchased from Montpellier vector platform (Plateforme de Vectorologie de Montpellier (PVM), Biocampus Montpellier, Montpellier, France). All viral vectors were aliquoted and stored at −80 °C until use.

Stereotaxic surgery. Mice were anaesthetized with 100 mg kg^{-1}/0.4 mg kg^{-1} ketamine/dexmedetomidine hydrochloride and head-fixed in a stereotaxic injection frame (myNeuroLab, Leica Microsystems Inc., Buffalo Grove, Illinois, USA). Lidocaine (20 µl) was injected subcutaneously into the head and neck area as a local anaesthetic. For *in vivo* recordings, mice were implanted with a head-bar and a microdrive containing the recording electrodes and an optical fibre. Viral injections were performed using previously described procedures[32] at the following stereotaxic coordinates: ventral medial nucleus of the thalamus (VM), −1.4 to −1.5 mm from bregma, 1.3 mm lateral from midline, and 4.10 mm ventral from cortical surface; GPh, −1.22 mm from bregma, 1.77 mm lateral from midline, and 4.64 mm ventral from cortical surface; and LHb, −1.7 mm from bregma, 0.53 mm lateral from midline, and 2.8 mm ventral from cortical surface. During the surgical procedure, mice were kept on a heating pad and were brought back to their home-cage for post-surgery recovery and monitoring. Post-operative care included intraperitoneal injection with 0.3–0.5 ml of Lactated Ringer's solution and Metacam (1–2 mg kg^{-1} meloxicam; Boehringer Ingelheim Vetmedica, Inc., St. Joseph, Missouri, USA) for analgesia and anti-inflammatory purposes. All AAVs were injected at a total volume of approximately 0.6 µl, and were allowed at least 4 weeks for maximal expression. For retrograde tracing of projection cells in the EP, CTB-555 or CTB-488 (0.3 µl, 0.5% in phosphate-buffered saline (PBS); Invitrogen, Thermo Fisher Scientific, Waltham, Massachusetts, USA) was injected into the VM or the LHb and allowed 3–5 days for sufficient retrograde transport.

Immunohistochemistry. Immunohistochemistry experiments were performed following standard procedures. In brief, mice were anaesthetized with Euthasol (0.4 ml; Virbac, Fort Worth, Texas, USA) and transcardially perfused with 40 ml of PBS, followed by 40 ml of 4% paraformaldehyde in PBS. Coronal sections (40–50 µm) were cut using a freezing microtome (Leica SM 2010R, Leica). Sections were first washed in PBS (3 × 5 min), incubated in PBST (0.3% Triton X-100 in PBS) for 30 min at room temperature (RT) and then washed with PBS (3 × 5 min). Next, sections were blocked in 5% normal goat serum in PBST for 30 min at RT and then incubated with primary antibodies overnight at 4 °C. Sections were washed with PBS (5 × 15 min) and incubated with fluorescent secondary antibodies at RT for 1 h. After washing with PBS (5 × 15 min), sections were mounted onto slides with Fluoromount-G (eBioscience, San Diego, California, USA). Images were taken using a LSM 710 laser-scanning confocal microscope (Carl Zeiss, Oberkochen, Germany). The primary antibodies used were: rabbit anti-Somatostatin-14 (Peninsula Laboratories Inc., San Carlos, California, USA; catalogue number T-4103), mouse anti-Parvalbumin (Swant, Switzerland; PV 235), chicken anti-GFP (Aves Labs, catalogue number GFP1020, lot number GFP697986), rabbit anti-RFP (Rockland, catalogue number 600-401-379, lot number 34135). Primary antibodies were incubated with appropriate fluorophore-conjugated secondary antibodies (Life Technologies, Carlsbad, California, USA) depending on the desired fluorescence colour.

Monosynaptic tracing with pseudotyped rabies virus. Retrograde tracing of mono-synaptic inputs onto genetically defined cell populations of the GPh was accomplished using a previously described method[5,32]. In brief, *Vglut2-Cre;Rosa26-stop*^{flox}*-tTA* mice that express tTA in Vglut2^{+} cells were injected into the GPh with AAV-TRE-hGFP-TVA-G (0.2–0.3 µl) that expresses the following components in a tTA-dependent manner: a fluorescent reporter histone GFP (hGFP); TVA (which is a receptor for the avian virus envelope protein EnvA); and the rabies envelope glycoprotein (G). Alternatively, CAV2-Cre virus was injected into the LHb of *Rosa26-stop*^{flox}*-tTA* mice, so that any input to the LHb will express tTA. As above, these mice were also injected into the GPh with AAV-TRE-hGFP-TVA-G (0.2–0.3 µl). Two weeks later, mice were injected in the same GPh location with the rabies-EnvA-SAD-DG-mCherry (1.2 µl), a rabies virus that is pseudotyped with EnvA, lacks the envelope glycoprotein, and expresses mCherry. This method ensures that the rabies virus exclusively infects cells expressing TVA. Furthermore, complementation of the modified rabies virus with envelope glycoprotein in the TVA-expressing cells allows the generation of infectious particles, which then can trans-synaptically infect presynaptic neurons.

Center of mass analysis. To compare average location of rabies-infected neurons in the GPe and STN, the centre of mass of a brain section was obtained by averaging positions of neurons. In order to standardize the results from individual animals onto a standard atlas, each neuron's position was normalized by anatomical landmarks: for Extended Data Fig. 9g, we used the midline and the most ventral part of the GPe; for Extended Data Fig. 9i, we used the midline and the most ventral medial portion of the STN.

Classical conditioning task. Nine *Vglut2-Cre;Ai35* mice were trained on an auditory classical conditioning task. One week after surgery, mice were water-deprived in their home-cage. During training, mice were head restrained using custom-made clamps and metal head-bars. Each mouse was habituated to head restraint for one day before training. There were five possible outcomes (unconditioned stimuli, US), each associated with a different auditory cue (conditioned stimulus, CS): a large water reward (5 µl), a small water reward (1 µl), nothing, a small air puff (100 ms) or a large air puff (500 ms). The air puff was delivered to the animal's face. Each trial began with a CS (1 s sound), followed by a 0.5 s delay and then a US (the outcome). In each session, reward and punishment trials were presented in two sequential blocks, with each cue chosen pseudorandomly. Each block contained the neutral stimulus.

Eye blinking was tracked using a CMOS camera (QSICC2). Offline video analysis was conducted using EthoVision XT software (Noldus; Wageningen, The Netherlands). Oval regions of interest (ROI) surrounding the eye were manually drawn. Pixels corresponding to the eye were detected as those that were darker than the background within the ROI. As each blink reduced the observable area of the eye, a threshold number of pixels corresponding to the eye was used to define a blink, and thus to determine the time and duration of each blink.

***In vivo* electrophysiology.** Custom-built screw-driven microdrives with four implantable tetrodes and a 50 µm fibre-optic were used to record simultaneously from multiple neurons. Each tetrode was glued to the fibre optic with epoxy, such that the end of each tetrode was 200–400 µm from the end of the fibre optic. Neural recordings and time stamps for behavioural variables were acquired with a Tucker-Davis Technologies RZ recording system (with a 32-channel preamp PZ2-32 and a RZ5D Bioamp processor; Alachua, Florida, USA).

Broadband signals from each wire were filtered between 0.2 and 8,500 Hz and recorded continuously at 25 kHz. To extract the timing of spikes, signals were band-pass-filtered between 300–5,000 Hz. Data analyses were carried out using software in Matlab (The Mathworks, Inc., Natick, Massachusetts, USA). Spike waveforms were manually sorted offline based on amplitude and waveform energy features using MClust-3.5 (from A. D. Redish). Individual neurons were only included in the dataset if they were well isolated based on their isolation distance (>20) and L-ratio (<0.1)[34]. Prior to implantation, tetrodes were dipped in DiI to aid the post hoc visualization of the recording locations.

In order to convert raster plots of firing rate into continuous spike density functions, spike times were first binned into 1-ms time windows and then convolved with a Gaussian kernel ($\sigma = 15$ ms). To determine the response to the CS or US presentation, the average firing rates were calculated using a 300 ms window defined as 180–480 ms following the stimulus. These time windows were chosen to cover the time of the peak neuronal response. Average baseline firing was calculated using a 300 ms window immediately preceding the delivery of the CS.

To identify putative GPh neurons—the Vglut2^{+} EP neurons—we used Arch-mediated optic tagging[35], whereby 200-ms light pulses ($\lambda = 532$ nm; OEM Laser Systems Inc., Bluffdale, Utah, USA) were delivered every 3 s for 100 trials following

each behavioural recording session. In early sessions we also used 500-ms ($n=3$) or 1-s ($n=1$) light pulses, which tagged Vglut2$^+$ EP neurons in a similar way to that of the 200-ms light pulses.

In addition to their response to light, putative GPh neurons were identified based on their firing pattern through a previously described unsupervised clustering approach[15]. In brief, to calculate receiver operating characteristic (ROC) curves, the distribution of firing rates within 100-ms bins were compared (from 1 s before the CS presentation to 1.5 s after delivery of the US) to the baseline firing rate 900 ms before CS presentation. The first three principal components (PCs) of the auROC (area under the receiver operating characteristic) curves were then calculated using principal component analysis (PCA), with the singular value decomposition algorithm. Hierarchical clustering of the auROC curves' first three PCs was then performed using a Euclidean distance metric and a complete agglomeration method.

Cross-correlations between spike waveforms across sessions were used to determine whether the same unit was recorded over multiple sessions. The cross-correlations were calculated after aligning the negative peak of each waveform, averaging separately, and aligning the peaks of the averages. A conservative session-to-session cross-correlation coefficient of >0.95 was used to positively classify two sets of waveforms as belonging to the same unit. The correlation was calculated using the full duration of the spike in a window 10 ms before and 40 ms after the peak negative response.

CS–US indices were calculated as (CS − US)/(CS + US), where CS is the difference between the peak firing rate (maximum value of the PSTH) in the 500 ms after CS onset and the baseline firing rate, and US is the difference between the peak firing rate in the 500 ms after US onset and the baseline firing rate. The baseline firing rate was calculated as the mean of the PSTH in the 0.5 s before CS onset.

Probabilistic switching task. Mice were trained in a two alternative choice probabilistic switching task. The initiation port was located in the centre between two reward ports. Infrared photodiode/phototransistor pairs placed on the inside of each port detected each nose poke (Island Motion Corporation, Tappan, New York, USA). Water valves (Neptune Research & Development, Inc., West Caldwell, New Jersey, USA) were calibrated to deliver a volume of water (2 μl) for rewarded choices. Water-deprived mice initiated a trial by a nose poke into the central port, which triggered a 'Go' light cue over the two peripheral ports. Mice then chose to enter either the left or right peripheral port where they received water rewards. On each trial, reward was delivered only at one port, and only for 75% of correct choices. The rewarded port was switched across blocks—the lengths of which were randomly distributed between 7–23 rewards—with no external instruction.

To characterize how reward and lack of reward influenced choice on a trial-by-trial basis in this task, we used a previously established logistic regression model[21,36]

$$\log\left(\frac{P_L(i)}{1-P_L(i)}\right) = \sum_{j=1}^{n} \beta_j^{\text{Reward}}(Y_L(i-j) - Y_R(i-j)) + \sum_{j=1}^{n} \beta_j^{\text{No Reward}}(N_L(i-j) - N_R(i-j)) + \beta_0 \quad (1)$$

where $P_L(i)$ indicates the probability of choosing the left port on the ith trial; $Y_L(i)$ and $Y_R(i)$ indicate a reward was delivered when choosing the left or right port, respectively, on the ith trial (1 for chosen and 0 for non-chosen port); $N_L(i)$ and $N_R(i)$ specify the lack of reward when choosing the left or right port, respectively, on the ith trial (1 for chosen and 0 for non-chosen port); n represents the number of past trials that were included in the regression model ($n=5$ was used, except for the optogenetic experiments where $n=2$ was used); the regression coefficients β^{Reward} and $\beta^{\text{No Reward}}$ represent the weighted contributions past rewards or lack of rewards have to the current choice; β_0 indicates the intrinsic bias a mouse may have for choosing the left or right port.

For *in vivo* optogenetic manipulations in the probabilistic switching task, *Vglut2-Cre* mice were bilaterally implanted with optical fibre cannulae (Thorlabs, Inc., Newton, New Jersey, USA), before behavioural training and following the surgery procedure for viral injection (described above). Optical fibres (200 μm) were implanted with the tips placed 0.4 mm dorsal to the site of virus injection and were secured to the skull with C&B Metabond quick adhesive cement (Parkell Inc., Edgewood, New York, USA) followed by dental cement (Lang Dental Manufacturing Co., Inc., Wheeling, Illinois, USA). Viruses were allowed to express for 3–4 weeks. The optic fibres were connected to a laser source ($\lambda=532$ nm or 473 nm; OEM Laser Systems) via a dual fibre rotary joint (FRJ_1x2i_FC-2FC; Doric Lenses, Inc., Québec, Canada) using an optic fibre sleeve (Thorlabs). Following training and habituation, optical stimulation was delivered at two time points during the task, at the time of action selection when the mouse nose pokes in the centre port to initiate a trial or during the evaluation phase when the mouse nose pokes in the peripheral choice ports. For ChR2-mediated stimulation, 5-ms optical light pulses were delivered at 30 Hz for 500 ms. For Arch-mediated inhibition, 500 ms of continuous illumination was delivered. In each session stimulation occurred randomly at either choice port in 10% of the trials. Stimulation sessions were interspersed by training sessions.

The effect of optogenetic manipulation on outcome evaluation was model by the following logistic regression equation

$$\log\left(\frac{P_L(i)}{1-P_L(i)}\right) = \sum_{j=1}^{n} \beta_j^{\text{Reward}}(Y_L(i-j) - Y_R(i-j)) + \sum_{j=1}^{n} \beta_j^{\text{NoReward}}(N_L(i-j) - N_R(i-j)) + \beta_0 + \beta_{\text{stim}} \times X_{\text{stim}}(i-1) \quad (2)$$

where β_{stim} is added to equation (1) to represent the effects of photostimulation on the current choice, and X_{stim} represents whether in the previous trial the stimulation was delivered when mice nose poked the left (1) or right (−1) reward port to collect reward, or was not delivered (0). When analysing the longevity of the effect optogenetic stimulation had on upcoming choices, additional β_{stim} terms were added to equation (2) to account for the stimulation on the $(i-n)$ trials.

The effect of optogenetic manipulations on action selection was model by the following logistic regression equation

$$\log\left(\frac{P_L(i)}{1-P_L(i)}\right) = \sum_{j=1}^{n} \beta_j^{\text{Reward}}(Y_L(i-j) - Y_R(i-j)) + \sum_{j=1}^{n} \beta_j^{\text{NoReward}}(N_L(i-j) - N_R(i-j)) + \beta_0 + \beta_{\text{stim}} \times X_{\text{stim}}(i) \quad (3)$$

where β_{stim} is added to equation (1) to represent the effects of photostimulation on the current choice, and X_{stim} represents whether the stimulation was delivered in the current trial when mice nose poked the centre port to initiate the trial (1 for stimulated trials and 0 otherwise).

For cell-type specific synaptic manipulations and their controls, mice were trained on the task before surgery and then tested again 4 weeks after surgery. The first five sessions before surgery and after surgery were used as the comparative sessions.

***In vitro* electrophysiology.** Patch clamp recording was performed as previously described[32]. In brief, mice were anaesthetized with isoflurane before they were decapitated; their brains were then dissected out and placed in ice-chilled dissection buffer (110 mM choline chloride, 25 mM $NaHCO_3$, 1.25 mM NaH_2PO_4, 2.5 mM KCl, 0.5 mM $CaCl_2$, 7.0 mM $MgCl_2$, 25.0 mM glucose, 11.6 mM ascorbic acid and 3.1 mM pyruvic acid, gassed with 95% O_2 and 5% CO_2). An HM650 Vibrating-blade Microtome (Thermo Fisher Scientific) was then used to cut 300 μm thick coronal sections that contained the EP. These slices were subsequently transferred to a storage chamber that contained oxygenated artificial cerebrospinal fluid (ACSF) (118 mM NaCl, 2.5 mM KCl, 26.2 mM $NaHCO_3$, 1 mM NaH_2PO_4, 20 mM glucose, 2 mM $MgCl_2$ and 2 mM $CaCl_2$, at 34 °C, pH 7.4, gassed with 95% O_2 and 5% CO_2). Following 40 min of recovery time, slices were transferred to RT (20–24 °C), where they were continuously bathed in the ACSF.

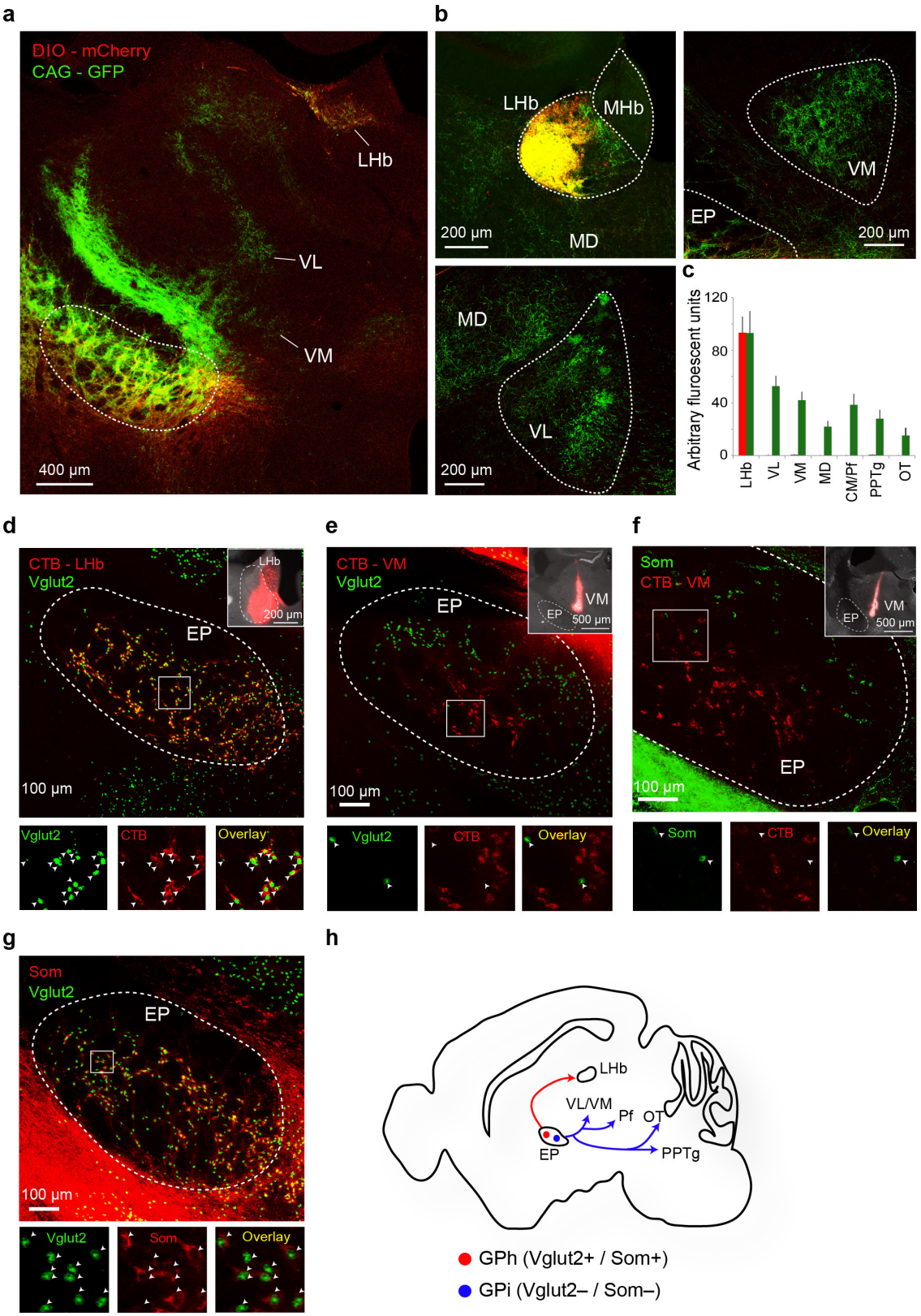
Visually guided whole-cell patch clamp recording from GPh neurons was obtained with Multiclamp 700B amplifiers and pCLAMP 10 software (Molecular Devices, Sunnyvale, California, USA), and was guided using an Olympus BX51 microscope equipped with both transmitted and epifluorescence light sources (Olympus Corporation, Shinjuku, Tokyo, Japan). GPh neurons with fluorescence of different colours were patched. To evoke excitatory postsynaptic currents (EPSCs), a bipolar stimulating electrode was placed on the medial dorsal border of the EP. Electrical stimulation was delivered every 10 s and synaptic responses were low-pass filtered at 1 KHz and recorded at holding potentials of −70 mV (for AMPA-receptor-mediated responses) and +40 mV (for NMDA-receptor-mediated responses). The NMDA-receptor-mediated component of the response was quantified as the mean current amplitude between 50–60 ms after electrical stimulation. Recordings were made in ACSF. The internal solution contained 115 mM caesium methanesulphonate, 20 mM CsCl, 10 mM HEPES, 2.5 mM $MgCl_2$, 4 mM Na_2ATP, 0.4 mM Na_3GTP, 10 mM sodium phosphocreatine and 0.6 mM EGTA (pH 7.2). The evoked EPSCs were recorded with picrotoxin (100 μM) added to the ACSF, and were analysed using pCLAMP 10 software. Miniature inhibitory postsynaptic currents (mIPSCs) were recorded at 0 mV holding potential with tetrodotoxin (TTX; 1 μM), APV (100 μM), and CNQX (5 μM) added to the ACSF, and were analysed using Mini Analysis software (Synaptosoft, Inc., Decatur, Georgia, USA).

To evoke striatal or pSTN synaptic transmission onto GPh neurons, AAV-ChR2-YFP or AAV-DIO-ChR2-YFP was injected into the striatum of wild-type C57BL/6 mice or the pSTN of *Vglut2-Cre* mice, respectively, and allowed to

express for 3 weeks. Acute brain slices were prepared and a blue light was used to stimulate ChR2-expressing axons. The light source was a single-wavelength LED system ($\lambda = 470$ nm; http://www.coolled.com/) connected to the epifluorescence port of the Olympus BX51 microscope. Single light pulses of 1 ms, triggered by a TTL signal from the Clampex software, were delivered to drive synaptic responses.

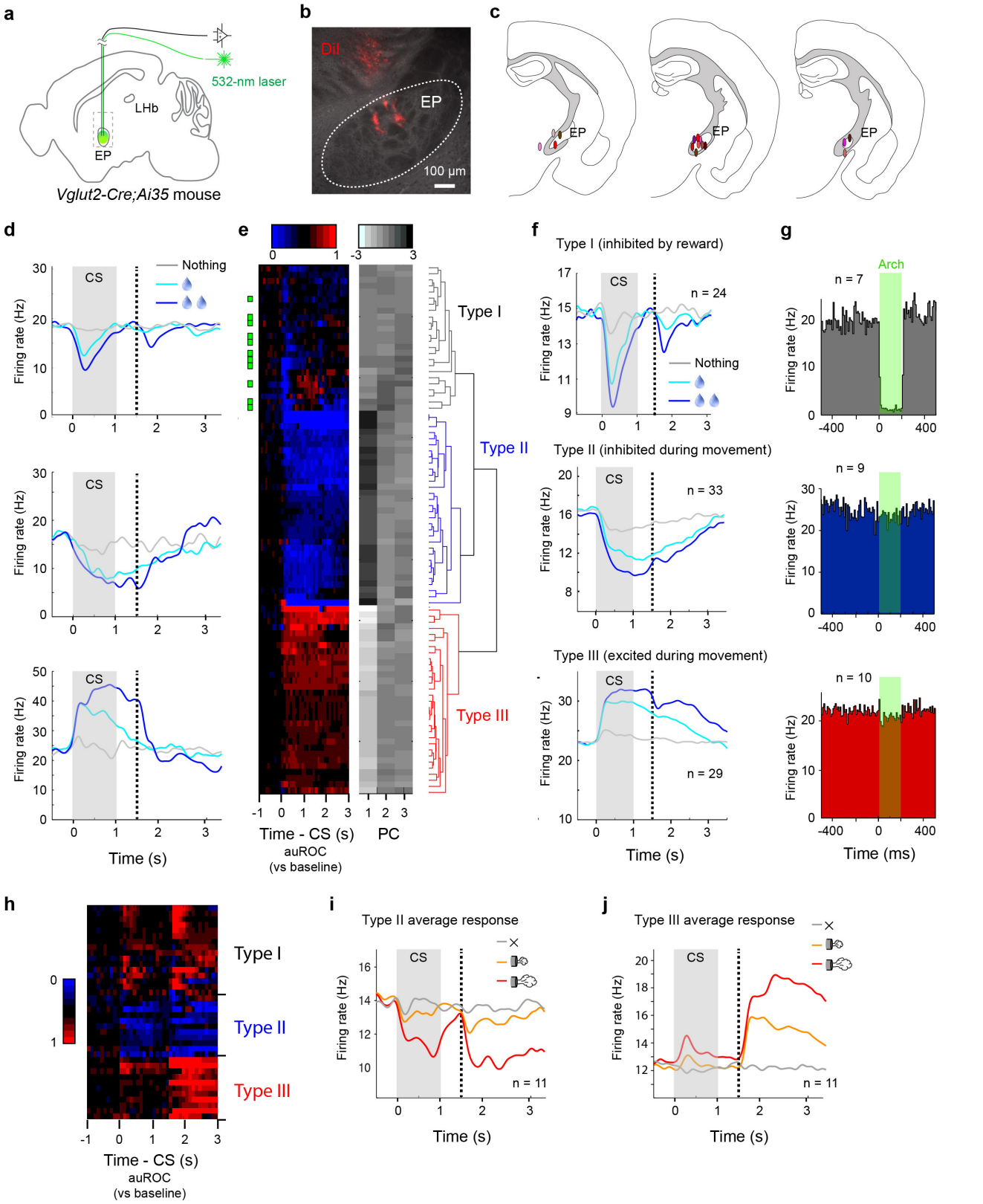
Statistics and data presentation. To determine whether parametric tests could be used, the Shapiro–Wilk Test was performed on all data as a test for normality. The statistical test used for each comparison is indicated when used. The sample sizes used in this study were based on estimations by a power analysis. Behavioural tests and electrophysiological data acquisition were performed by investigators with knowledge of the identities of experimental groups. All these experiments were controlled by computer systems, with data collected and analysed in an automated and unbiased way. For *in vivo* recordings, the data from a mouse were excluded if the tetrode tracts and tips were outside of the EP. No other mice or data points were excluded.

31. He, M. *et al.* Cell-type-based analysis of microRNA profiles in the mouse brain. *Neuron* **73,** 35–48 (2012).
32. Penzo, M. A. *et al.* The paraventricular thalamus controls a central amygdala fear circuit. *Nature* **519,** 455–459 (2015).
33. Li, L. *et al.* Visualizing the distribution of synapses from individual neurons in the mouse brain. *PLoS One* **5,** e11503 (2010).
34. Schmitzer-Torbert, N., Jackson, J., Henze, D., Harris, K. & Redish, A. D. Quantitative measures of cluster quality for use in extracellular recordings. *Neuroscience* **131,** 1–11 (2005).
35. Courtin, J. *et al.* Prefrontal parvalbumin interneurons shape neuronal activity to drive fear expression. *Nature* **505,** 92–96 (2014).
36. Lau, B. & Glimcher, P. W. Dynamic response-by-response models of matching behavior in rhesus monkeys. *J. Exp. Anal. Behav.* **84,** 555–579 (2005).



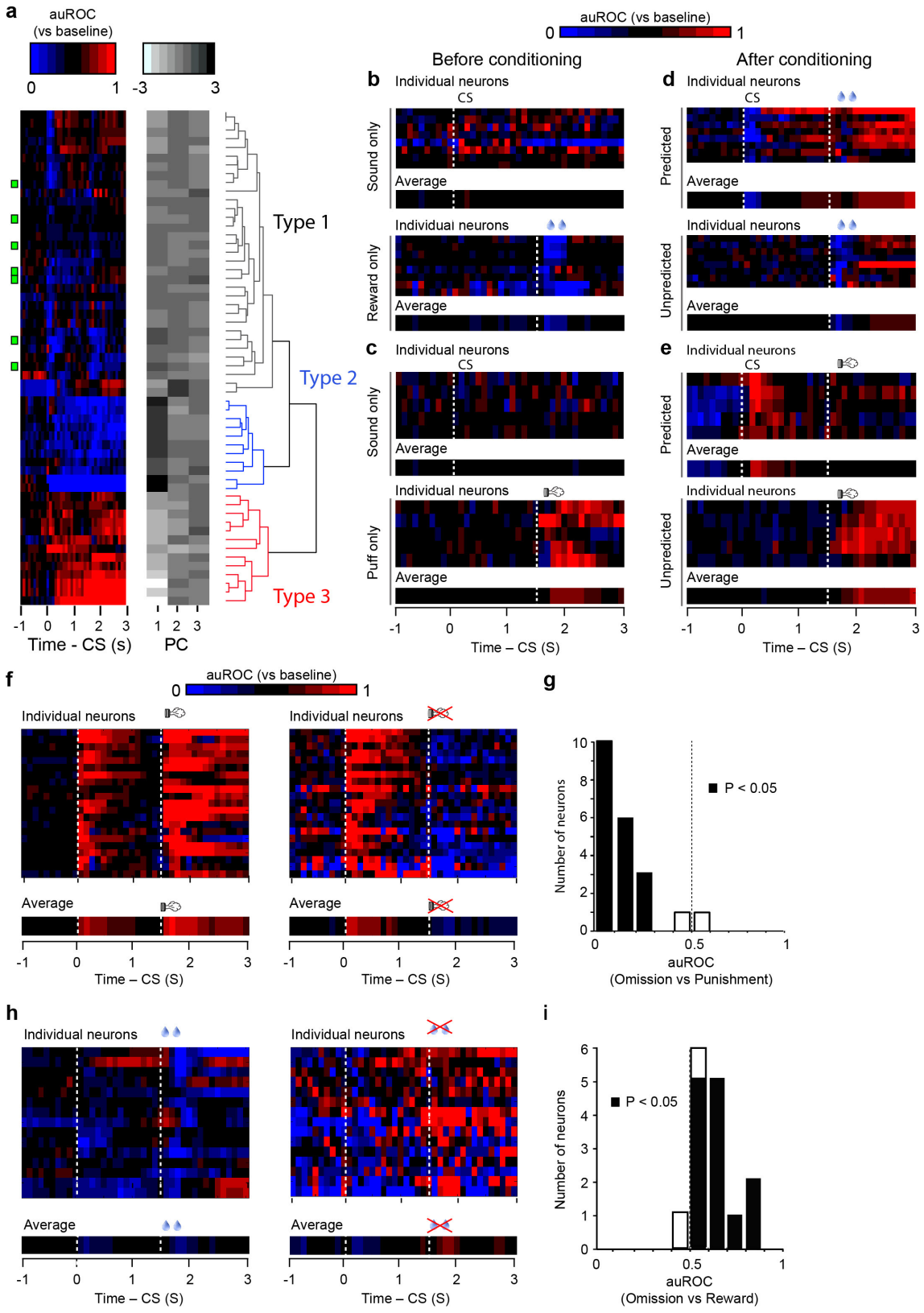
Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | Vglut2 and Somatostatin are markers for GPh neurons. a, Image showing the projection patterns of non-specifically labelled neurons (green, infected with adeno-associated virus (AAV) expressing GCaMP6 (AAV1-Syn-GCAMP6f.WPRE.SV40); signal was enhanced by anti-GFP antibody; see Methods) and Vglut2$^+$ neurons (red, infected with AAV expressing mCherry in a Cre-dependent manner (AAV8-hSyn-DIO-mCherry); signal was enhanced by anti-mCherry antibody; see Methods) in the EP of a *Vglut2-Cre* mouse. **b**, Confocal images of the LHb, ventrolateral thalamus (VL) and ventromedial thalamus (VM), showing fibres originating from the non-specifically labelled neurons (green) and Vglut2$^+$ neurons (red) in the EP. **c**, Quantification of the GFP and mCherry fluorescence intensity in the projection targets of the EP neurons. **d**, Upper panel, representative image showing retrograde labelling of GPh neurons by injection of the cholera toxin subunit B conjugated to Alexa Fluor 594 (CTB-594) into the LHb (inset) of *Vglut2-Cre;Rosa26-stop^{flox}-H2b-GFP* mice, in which Vglut2$^+$ cells can be identified based on their expression of nuclear GFP. Lower panels, high-magnification pictures of the boxed area in the EP in the upper panel, showing the co-labelling of GPh neurons by CTB-594 and Vglut2 (arrowheads). The vast majority of CTB-labelled neurons expressed Vglut2 (95.45 ± 1.2% (mean ± s.e.m.), $n = 6$ mice). **e**, Upper panel, a representative image showing retrograde labelling of VM-projecting EP neurons by injection of CTB-594 into the VM (inset) of *Vglut2-Cre;Rosa26-stop^{flox}-H2b-GFP* mice. Lower panels, high-magnification pictures of the boxed area in the EP in the upper panel, showing the segregation of the EP neurons labelled by CTB-594 and those labelled by Vglut2 (arrowheads). Very few CTB-labelled neurons expressed Vglut2 (0.51 ± 0.45%, $n = 6$ mice). **f**, Upper panel, a representative image showing retrograde labelling of VM-projecting EP neurons by injection of CTB-594 into the VM (inset). Lower panels, high-magnification pictures of the boxed area in the EP in the upper panel, showing the segregation of the EP neurons labelled by CTB-594 and those labelled by anti-Som antibody (arrowheads). Very few CTB-labelled cells expressed Som (0.88 ± 0.72%, $n = 5$ mice). **g**, Upper panel, a representative image showing antibody labelling of Som in the EP of *Vglut2-Cre;Rosa26-stop^{flox}-H2b-GFP* mice. Lower panels, high-magnification pictures of the boxed area in the upper panel, showing the co-labelling of EP neurons by Som and Vglut2 (arrowheads). The vast majority of Vglut2 neurons expressed Som (90.87 ± 0.79%, $n = 6$ mice). **h**, A cartoon showing the only projection target of GPh neurons (red) and the different projection targets of classic GPi neurons (blue). Diagram in **h** was modified from the Allen Mouse Brain Atlas, Allen Institute for Brain Science; available from http://mouse.brain-map.org/.



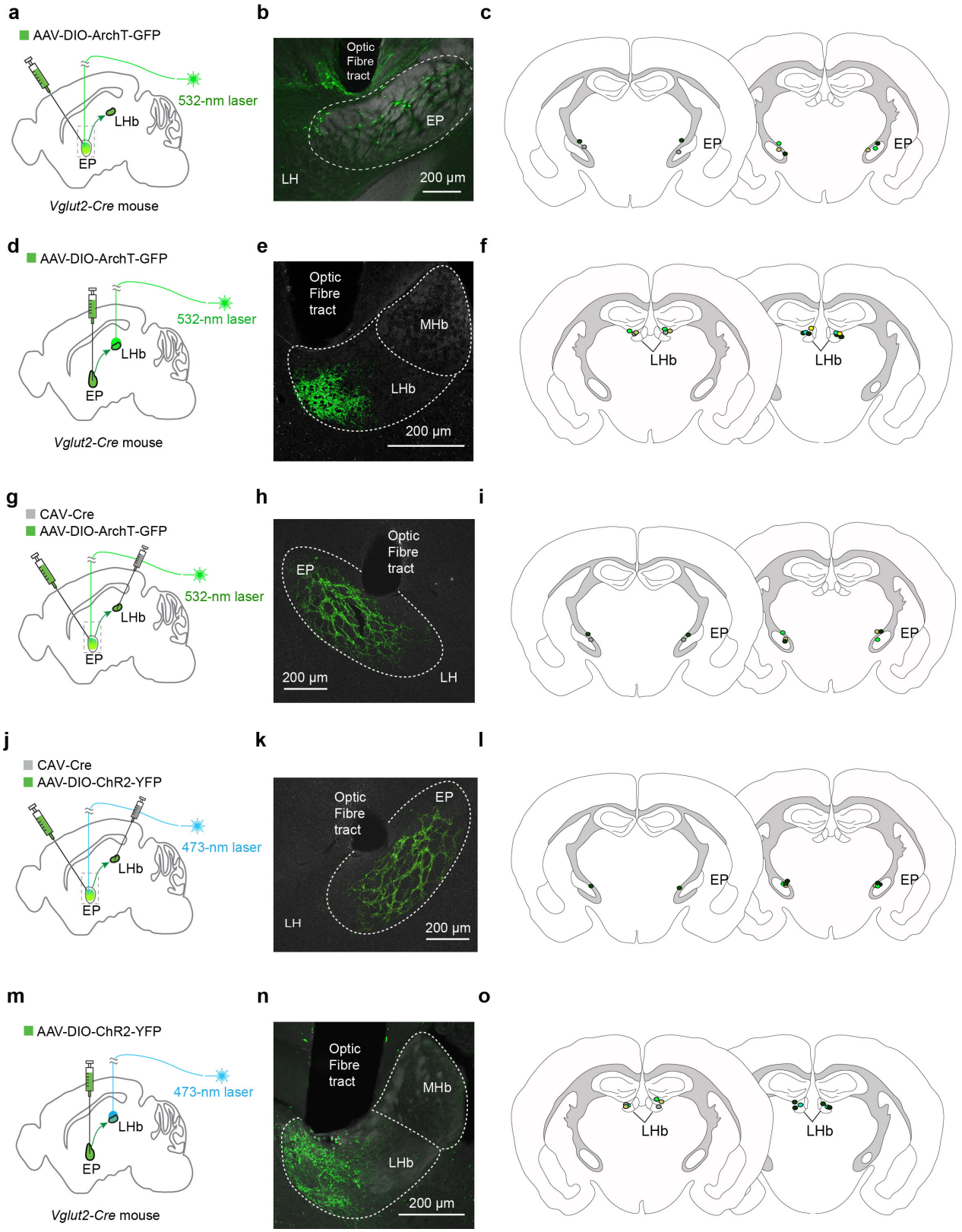
Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Classification of EP neurons on the basis of their distinct response profiles. a, Schematic of the experimental approach used for *in vivo* recording and optogenetic tagging. **b**, Photomicrograph showing a DiI-labelled recording site. **c**, Schematics showing the locations of the recording sites ($n = 15$ mice). **d**, Responses of three example neurons in the classic conditioning task. **e**, Left, auROC plots of the responses of all neurons during large reward trials. Red, increase from baseline; blue, decrease from baseline; each row represents one neuron. Green bars indicate the neurons that were optogenetically tagged ($n = 11$ neurons). The three main clusters are arranged in order to match the neurons presented in **d**. Right, first three principle components and hierarchical clustering dendrogram showing the relationship of each neuron within the three clusters. **f**, Average firing rates of the three types of neurons ($n = 86$ neurons from 9 mice). **g**, Plots of peristimulus time histogram (PSTH) showing inhibition for type I (top, $n = 7$ neurons from 4 mice), but no change for type II (middle, $n = 9$ neurons from 4 mice) or type III (bottom, $n = 10$ neurons from 4 mice) neurons in response to green light pulses (green bars, 200 ms; 100 trials per neuron, 0.3 Hz). Only type II and type III neurons that were recorded in the same sessions and animals as those of the light-responsive type I neurons represented in **g** are shown. **h**, auROC plots of the responses of all 38 neurons ($n = 9$ mice) recorded during large punishment trials. **i**, **j**, Average firing rates of type II ($n = 11$ neurons from 9 mice) (**i**) and type III ($n = 11$ neurons from 9 mice) (**j**) neurons during punishment trials. Diagrams in **a** and **c** were modified from the Allen Mouse Brain Atlas, Allen Institute for Brain Science; available from http://mouse.brain-map.org/.



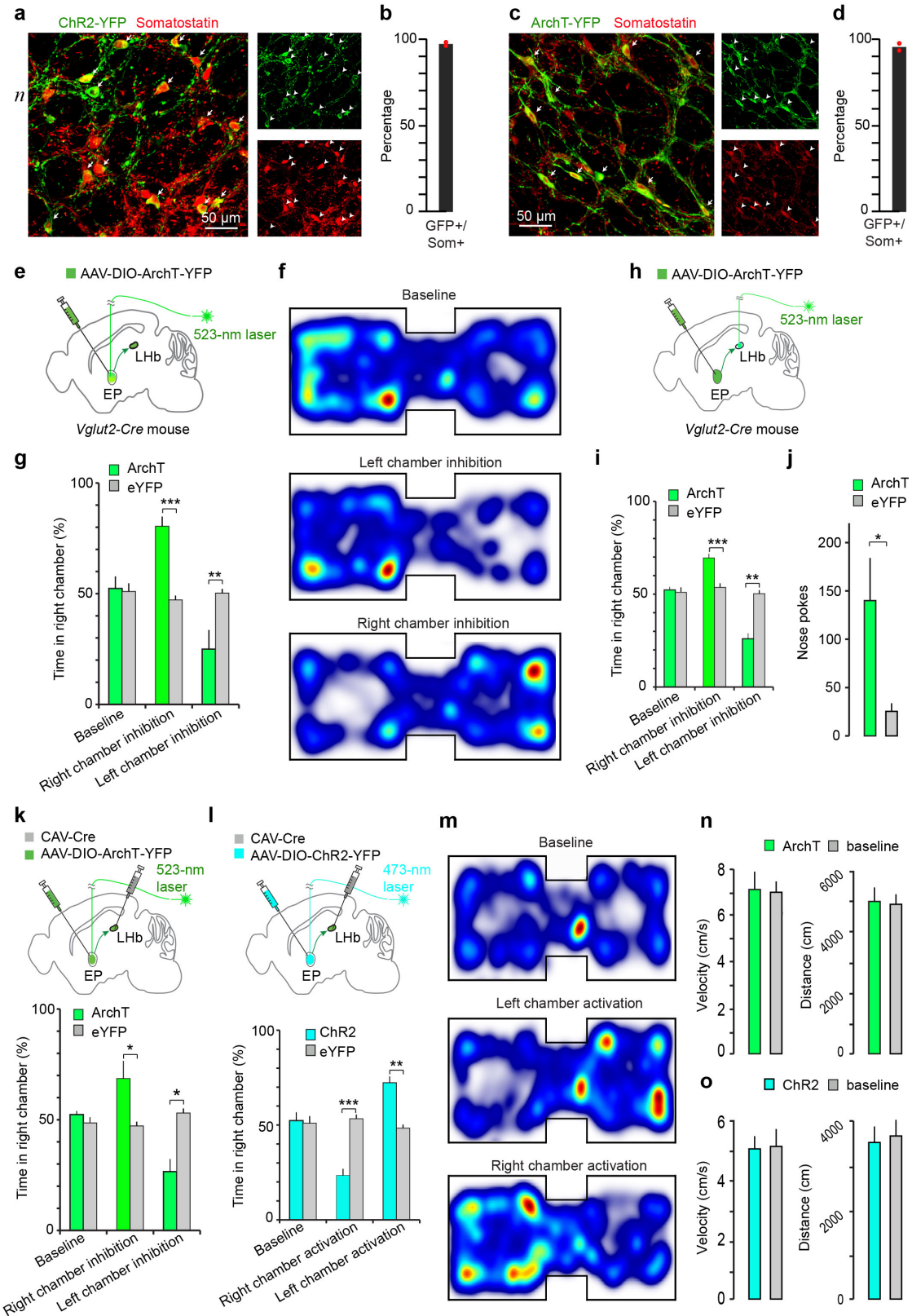
Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | Response profiles of putative GPh neurons during different CS–US contingencies. a, Graphs showing hierarchical clustering used to identify additional putative GPh neurons used in the analysis for this figure. All data shown (**b**–**i**) are from type I neurons only. Left, auROC plots of the responses of all additional neurons recorded. Red, increase from baseline; blue, decrease from baseline. Each row represents one neuron. Green bars indicate the neurons that were optogenetically tagged. Right, first three principle components and hierarchical clustering dendrogram showing the relationship of each neuron within the three clusters. **b**, auROC plots showing the firing rate changes in response to CS (top) and reward (bottom) before behavioural training. **c**, auROC plots showing the firing rate changes in response to CS (top) and air puff (bottom) before behavioural training. **d**, auROC plots showing the firing rate changes in response to an expected (top) or unexpected (bottom) reward. **e**, auROC plots showing the firing rate changes in response to an expected (top) or unexpected (bottom) air puff. **f**, auROC plots showing the firing rate changes in response to receiving an expected air puff (left) or having an expected air puff omitted (right). **g**, Histogram of difference in firing rate between air puff omission and air puff (filled bars, $P < 0.05$, *t*-test). Values are represented using auROC. **h**, auROC plots showing the firing rate changes in response to receiving an expected reward (left) or having an expected reward omitted (right). **i**, Histogram of difference in firing rate between reward omission and reward (filled bars, $P < 0.05$, *t*-test). Values are represented using auROC.



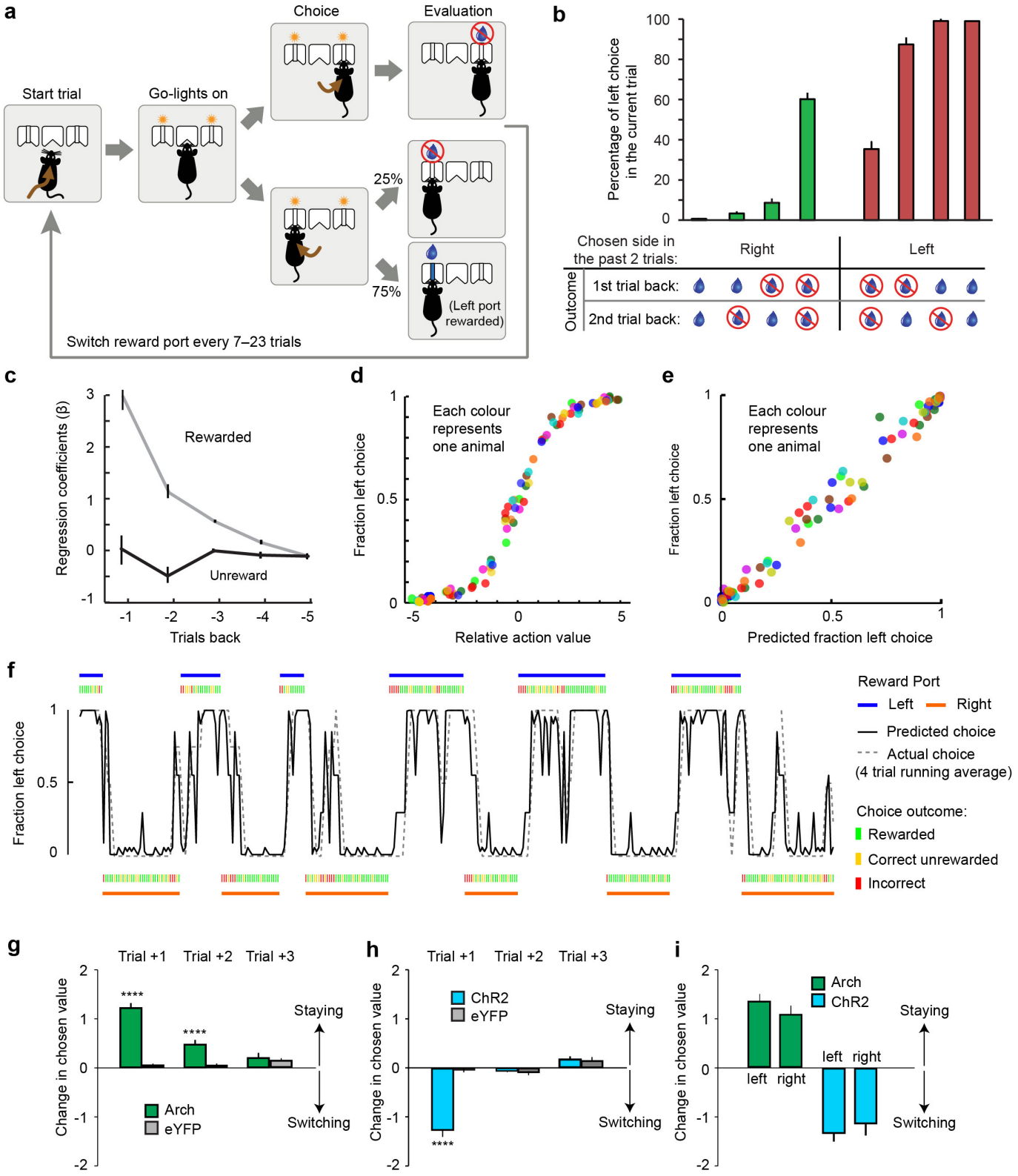
Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | Optic fibre implantation locations. **a**, A schematic of the experimental approach used for Arch-mediated inhibition of GPh neurons. **b**, A photomicrograph showing the location of optic fibre placement and ArchT–GFP$^+$ GPh neurons within the EP. **c**, Schematics showing the location of the optic fibre placements ($n = 5$). **d**, A schematic of the experimental approach used for Arch-mediated inhibition of the GPh–LHb projection. **e**, A photomicrograph showing the location of optic fibre placement and ArchT–GFP$^+$ axon fibres within the LHb. **f**, Schematics showing the location of the optic fibre placements ($n = 7$). **g**, A schematic of the experimental approach used for Arch-mediated inhibition of the GPh, which was targeted retrogradely by injection of the LHb with CAV2–Cre. **h**, A photomicrograph showing the location of the optic fibre placement and ArchT–GFP$^+$ neurons in the EP. **i**, Schematics showing the location of the optic fibre placements ($n = 5$). **j**, A schematic of the experimental approach used for ChR2-mediated excitation of the GPh, which was targeted retrogradely by injection of the LHb with CAV2–Cre. **k**, A photomicrograph showing the location of the optic fibre placement and ChR2–GFP$^+$ neurons in the EP. **l**, Schematics showing the location of the optic fibre placements ($n = 5$). **m**, Schematic of the experimental approach used for ChR2-mediated activation of the GPh–LHb projection. **n**, A photomicrograph showing the optic fibre placement and ChR2–YFP$^+$ axon fibres in the LHb. **o**, Schematics showing the location of the optic fibre placements ($n = 6$). Diagrams in **a**, **c**, **d**, **f**, **g**, **i**, **j**, **l**, **m** and **o** were modified from the Allen Mouse Brain Atlas, Allen Institute for Brain Science; available from http://mouse.brain-map.org/.



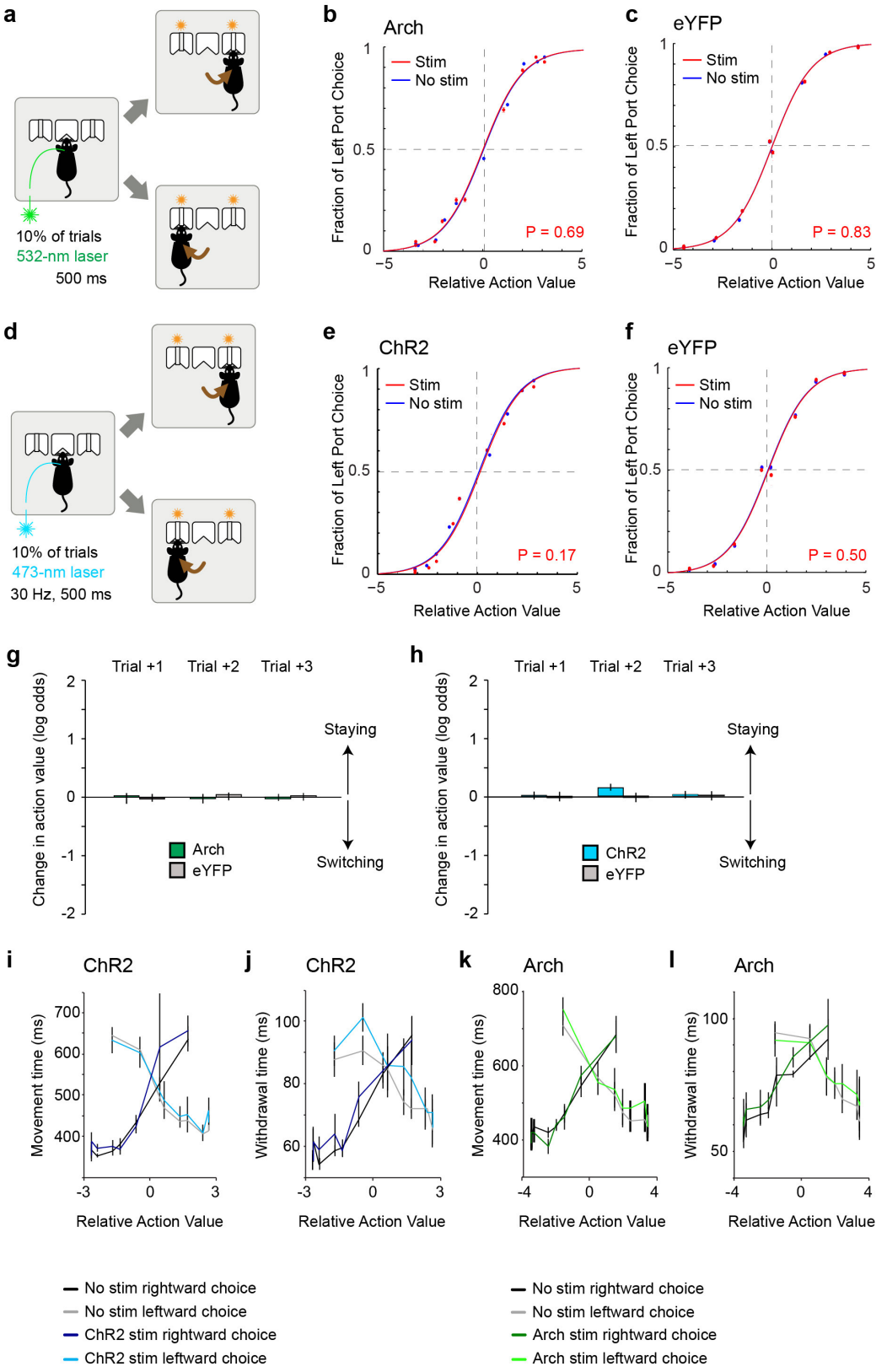
Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Optogenetic inhibition of the GPh drives reward-related behaviours. a, Confocal images from a *Som-Cre;Ai14* mouse, showing the overlap in expression of ChR2–YFP and tdTomato (indicating Som+ neurons) in GPh neurons. **b**, Quantification of the percentage of ChR2–YFP^{+} neurons that expressed tdTomato ($n = 2$). **c**, Confocal images from a *Som-Cre;Ai14* mouse, showing the overlap in expression of ArchT–YFP and tdTomato in GPh neurons. **d**, Quantification of the percentage of ArchT–YFP^{+} neurons that expressed tdTomato ($n = 2$). **e**, Schematic of the experimental approach used for ArchT-mediated inhibition of GPh neurons. **f**, Heat maps for the activity of a representative mouse at baseline (top), or during optogenetic inhibition of the GPh in either the left (middle) or right (bottom) chamber. **g**, GPh^{Arch} mice ($n = 5$), but not GPh^{eYFP} mice ($n = 5$), showed a significant place preference for the chamber paired with laser stimulation in the GPh ($F_{(5,29)} = 14.95$, $P < 0.0001$, $^{***}P < 0.001$, $^{**}P < 0.01$, two-way ANOVA followed by Tukey's test). **h**, Schematic of the experimental approach used for ArchT-mediated inhibition of GPh axon terminals in the LHb. **i**, GPh^{ArchT} mice ($n = 7$), but not GPh^{eYFP} mice ($n = 5$), showed a significant place preference for the chamber paired with laser stimulation in the GPh ($F_{(5,35)} = 52.22$, $P < 0.0001$, $^{***}P < 0.001$, $^{**}P < 0.01$, two-way ANOVA followed by Tukey's test). **j**, GPh^{Arch} mice ($n = 5$) made significantly more nose pokes than GPh^{eYFP} mice ($n = 5$) to obtain laser stimulation in the GPh ($t_{(8)} = 2.61$, $^{*}P < 0.05$, *t*-test). **k**, Schematic of the retrograde labelling approach used to target the GPh for ArchT-mediated optical inhibition (top). GPh$^{CAV\text{-}Cre/Arch}$ mice ($n = 5$), but not GPh^{eYFP} mice ($n = 5$), showed a significant place preference for the chamber paired with laser stimulation in the GPh (bottom) ($F_{(5,29)} = 5.98$, $P < 0.01$, $^{*}P < 0.05$, two-way ANOVA followed by Tukey's test). **l**, Schematic of the retrograde labelling approach used to target the GPh for ChR2-mediated optical excitation (top). GPh$^{CAV\text{-}Cre/ChR2}$ mice ($n = 5$), but not GPh^{eYFP} mice ($n = 5$), showed a significant place aversion for the chamber paired with laser stimulation in the GPh (bottom) ($F_{(5,29)} = 26.50$, $P < 0.0001$; $^{***}P < 0.001$, $^{**}P < 0.01$, two-way ANOVA followed by Tukey's test). **m**, Heat maps for the activity of a representative mouse at baseline (top), or during optogenetic excitation of the GPh in either the left (middle) or right (bottom) chamber. **n**, Mice did not move faster (left) or further (right) during the Arch stimulation sessions when compared to their baseline activity ($t_{(32)} = 0.15$, $P > 0.05$; $t_{(32)} = 0.16$, $P > 0.05$; *t*-test, $n = 17$). **o**, Mice did not move faster (left) or further (right) during the ChR2 stimulation sessions when compared to their baseline activity ($t_{(8)} = 0.12$, $P > 0.05$; $t_{(8)} = 0.26$, $P > 0.05$; *t*-test, $n = 5$). All data are presented as mean ± s.e.m. Diagrams in **e**, **h**, **k** and **i** were modified from the Allen Mouse Brain Atlas, Allen Institute for Brain Science; available from http://mouse.brain-map.org/.



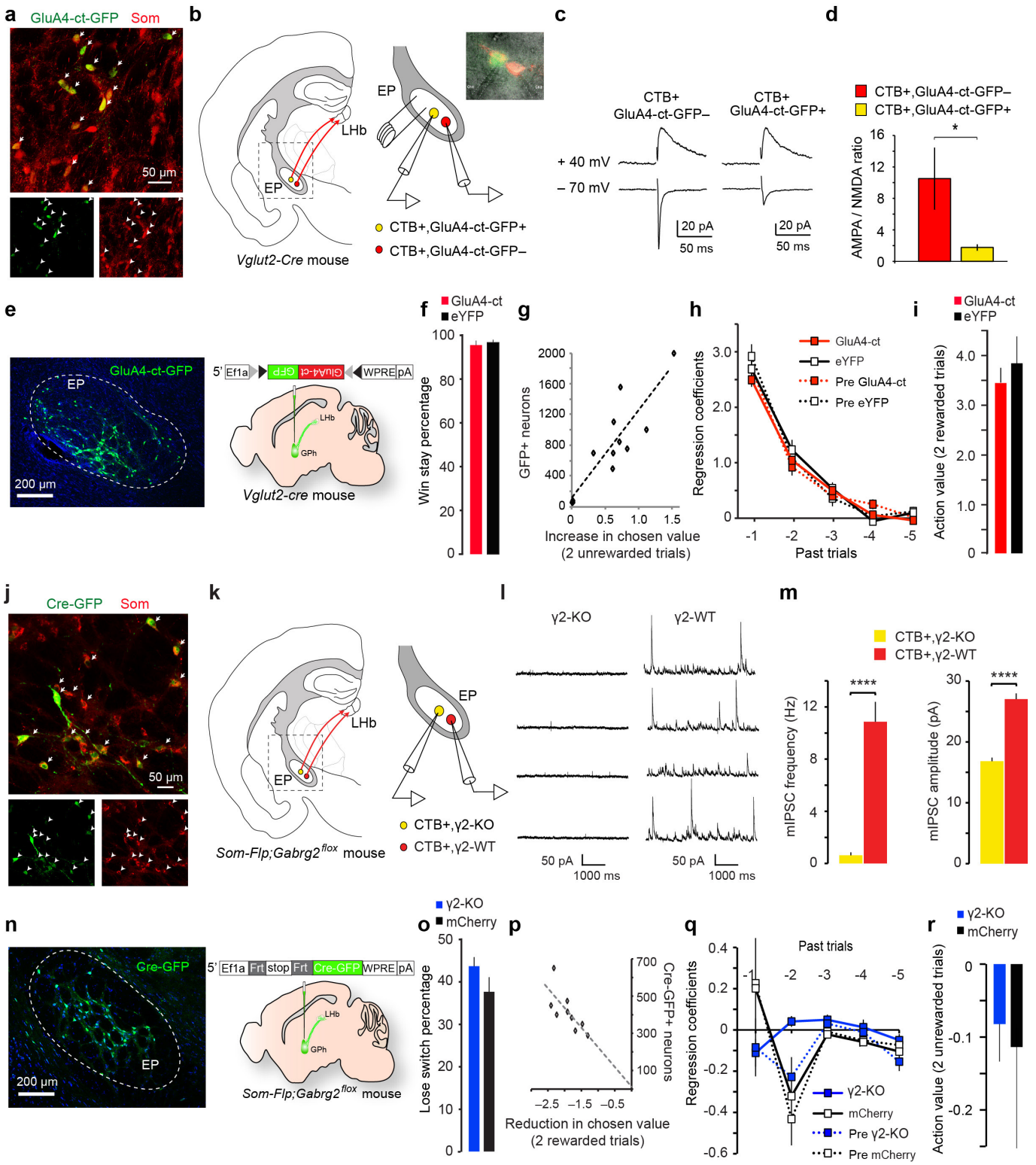
Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | A probabilistic switching task for studying action evaluation. a, Schematic of the task. **b**, The probability of choosing the left port by one mouse for reward history in which consecutive choices to either the right or the left port were made during the previous two trials. **c**, The contribution of rewarded and unrewarded outcomes in the previous five trials—represented by regression coefficients β^{Reward} and $\beta^{\text{No Reward}}$, respectively—to choices in the current trial ($n = 10$ mice, $4{,}685 \pm 786$ trials per mouse). **d**, The fraction of left port choice for ten mice plotted against the relative action value (sum of the regression coefficients from the previous two trials). Data from each mouse was grouped into ten bins and represented by a distinct colour. **e**, The actual probability of choosing the left port plotted against the probability of choosing the left port predicted by the logistic regression model. **f**, Example data from one session showing 12 trial blocks. Blue bars represent left reward blocks (top); orange bars indicate right reward blocks (bottom). Green, orange, and red ticks represent whether a particular trial was a correct rewarded trial, a correct unrewarded trial, or an incorrect trial, respectively. The grey dashed line represents a four-trial running average of the mouse's probability of choosing the left port, and the black line indicates the probability of choosing the left port predicted by the logistic regression model. **g**, **h**, Change in chosen value one to three trials after optogenetic inhibition of the GPh (**g**), or activation of the GPh–LHb pathway (**h**). **i**, Changes in chosen value one trial after optogenetic activation or inhibition at the left or right reward port. **g**, **h**, $^{****}P < 0.0001$, *t*-test. **b**, **c**, **g**–**i**, Data are represented as mean ± s.e.m.



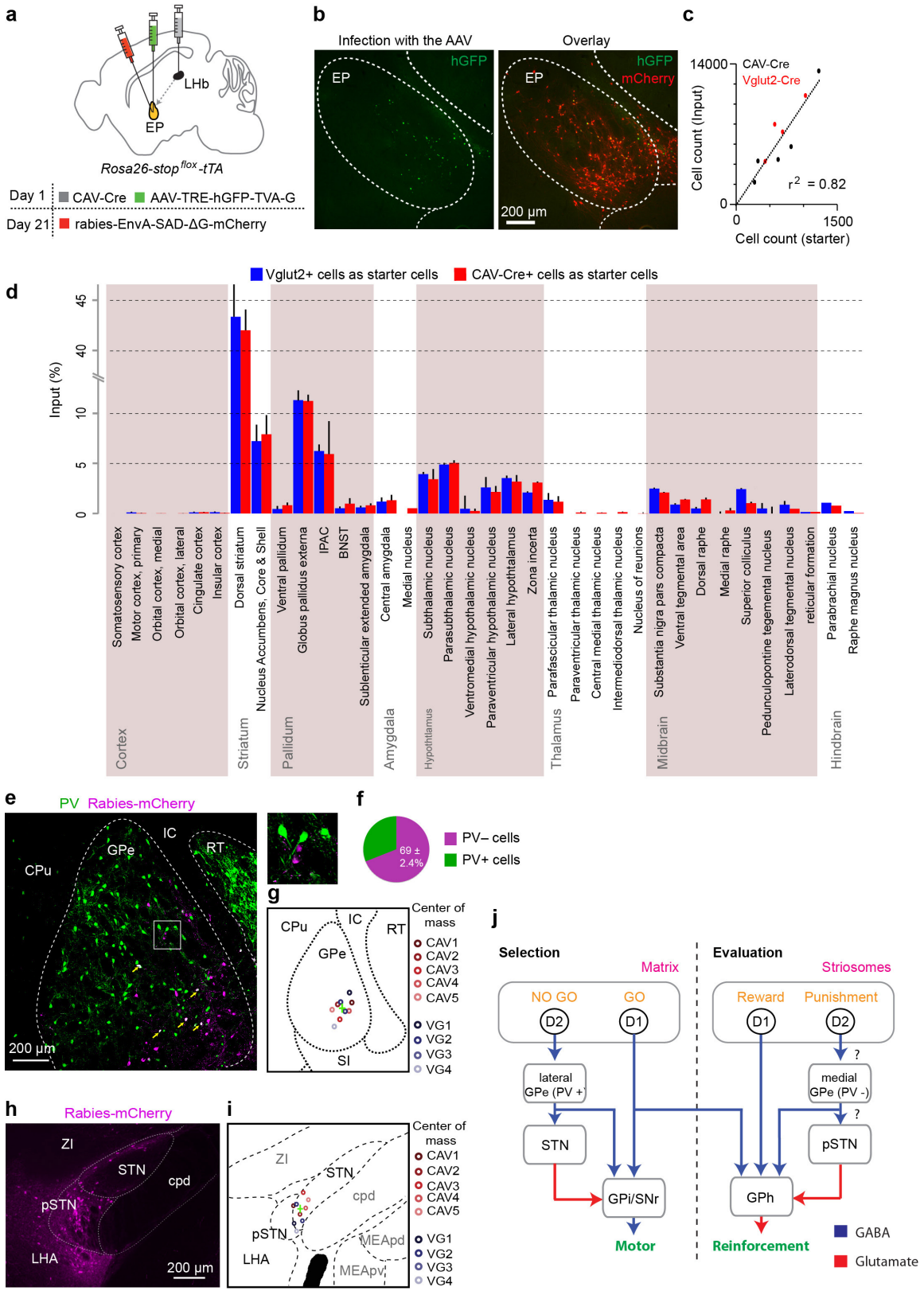
Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Optogenetic inhibition or activation of the GPh–LHb pathway does not influence action selection. a, A schematic of optogenetic inhibition of the GPh at the point of action selection. **b**, Data points indicate the probability of left port choice as a function of action value for the trials in which the photo-stimulation was delivered at the centre port (stim) or was not delivered (no stim). Lines indicate the fit by the logistic regression model on the pooled data for each of the two conditions ($n = 5$ mice, 15,411 trials, 3,082 ± 1,063 trials per mouse). **c**, Similar to **b**, except that control mice with eYFP-expressing GPh neurons were used ($n = 6$ mice, 56,241 trials, 9,373 ± 596 trials per mouse). **d**, **e**, Similar to **a** and **b**, except that optogenetic activation of the GPh–LHb projection was applied at the point of action selection ($n = 6$ mice, 41,557 trials, 8,311 ± 2,565 trials per mouse). **f**, Similar to **e**, except that control mice with eYFP-expressing GPh neurons were used ($n = 6$ mice, 72,423 trials, 12,070 ± 1,673 trials per mouse). **g**, **h**, The changes in action value in response to optogenetic stimulation of the GPh–LHb pathway one to three trials after the photostimulation, for mice in which GPh neurons expressed Arch ($n = 5$) or eYFP ($n = 6$) (**g**), or ChR2 ($n = 6$) or eYFP ($n = 6$) (**h**). In **b**, **c**, **e** and **f**, P values reported for t tests: H_0: $\beta_{stim} = 0$. **i**–**l**, Graphs showing the average withdrawal (calculated as the time from centre port entry to exit) and movement (calculated as the time from centre port exit to the poke at the chosen port) time for trials with or without light stimulation. Both withdrawal time and movement time were shorter when the action value associated with the chosen action was higher. Neither activation of GPh neurons with ChR2 ($n = 6$ mice) (**i**, **j**) (movement time for leftward choices, ChR2 stimulated trials (ChR2) versus unstimulated trials (no stim), $F_{(1,8)} = 0.174$, $P > 0.05$; rightward choices, ChR2 versus no stim, $F_{(1,8)} = 1.352$, $P > 0.05$; withdrawal time preceding leftward choices, ChR2 versus no stim, $F_{(1,8)} = 0.667$, $P > 0.05$; preceding rightward choices, ChR2 versus no stim, $F_{(1,8)} = 0.599$, $P > 0.05$; two-way ANOVA), nor inhibition of these neurons with Arch ($n = 5$ mice) (**k**, **l**) (movement time for leftward choices, Arch stimulated trails (Arch) versus unstimulated trials (no stim), $F_{(1,8)} = 0.105$, $P > 0.05$; rightward choices, Arch versus no stim, $F_{(1,8)} = 0.023$, $P > 0.05$; withdrawal time preceding leftward choices, Arch versus no stim, $F_{(1,8)} = 0.821$, $P > 0.05$; preceding rightward choices, Arch versus no stim, $F_{(1,8)} = 0.459$, $P > 0.05$; two-way ANOVA) had any significant effect on the ongoing behaviour. Data in **g**–**l** are presented as mean ± s.e.m.



Extended Data Figure 8 | See next page for caption.

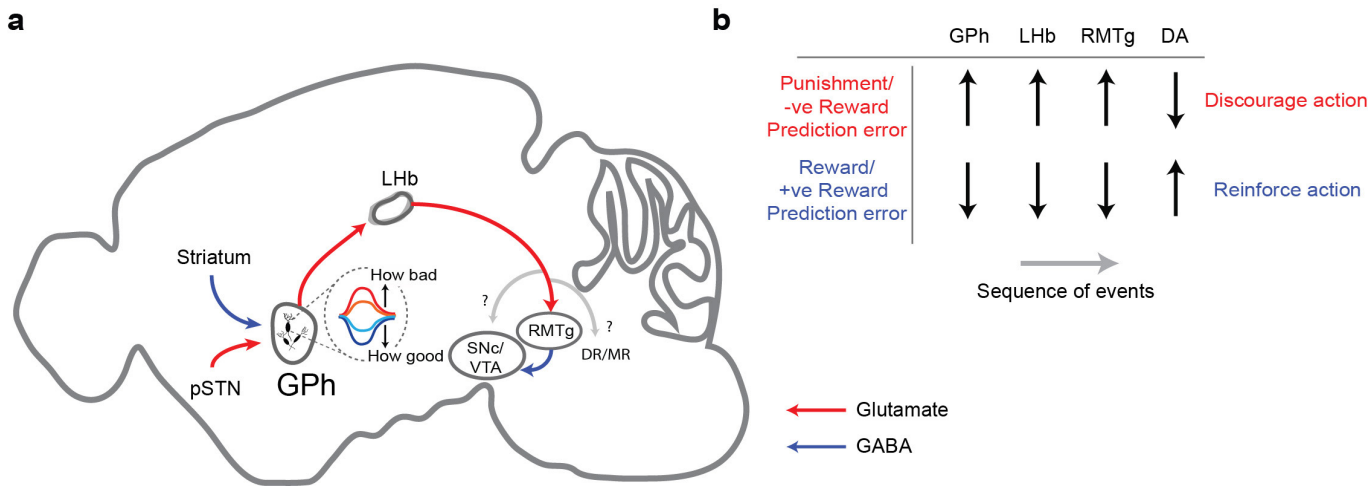
Extended Data Figure 8 | Weakening of excitatory or inhibitory synapses onto GPh neurons and its effects on the sensitivity to negative or positive feedback. a, Confocal images from a *Som-Cre;Ai14* mouse, showing the overlap in expression of GluA4-ct–GFP (delivered by injecting the EP with the AAV-DIO-GluA4-ct-GFP) and tdTomato (indicating the expression of Som) in GPh neurons. 97.86 ± 2.9% of GluA4-ct–GFP$^+$ neurons expressed tdTomato ($n = 2$ mice). **b**, Schematics of the experimental approach. CTB-594 was injected into the LHb to label GPh neurons in the EP. On the right is an enlarged graph of the boxed area in the cartoon on the left. Inset is a photomicrograph showing simultaneous recording of a CTB$^+$/GluA4-ct$^+$ GPh neuron and a nearby CTB$^+$/GluA4-ct$^-$ GPh neuron. **c**, EPSC traces recorded from the two neurons shown in **b**. **d**, Quantification of the ratio between AMPA receptor-mediated EPSC amplitude and NMDA receptor-mediated EPSC amplitude (AMPA/NMDA ratio) for the two populations of GPh neurons (CTB$^+$/GluA4-ct$^+$, $n = 6$ cells; CTB$^+$/GluA4-ct$^-$, $n = 8$ cells; $n = 3$ mice; $t_{(12)} = -1.89$, $^*P < 0.05$, *t*-test). **e**, A representative image showing the expression of GluA4-ct–GFP (delivered by injecting the EP of a *Vglut2-Cre* mouse with the AAV-DIO-GluA4-ct-GFP) in GPh neurons (left) and a schematic of the approach (right). **f**, The win–stay percentage in these mice (GPh$^{GluA4\text{-}ct}$, 94.17 ± 1.02%; GPh^{eYFP}, 95.82 ± 0.51%; $P > 0.05$, *t* test). **g**, For animals ($n = 10$ mice) used in Fig. 4a–e, the number of GPh neurons that were infected with the GluA4-ct–GFP virus correlated with the change in animal behaviour in the switching task, measured as an increase in action value following two consecutive unrewarded trials ($R^2 = 0.72$, $P < 0.05$ by a linear regression). **h**, Contributions of rewarded outcomes over the past five trials, as reflected by their regression coefficients, to the current choice. GPh$^{GluA4\text{-}ct}$ mice were not significantly different from control mice or their pre-surgery condition (first two trials back × groups, $F_{(3,33)} = 0.5412$, $P > 0.05$; two-way ANOVA, $n = 10$ GPh$^{GluA4\text{-}ct}$ mice and $n = 7$ control mice). **i**, The action value following two sequentially rewarded trials was not significantly different between GPh$^{GluA4\text{-}ct}$ mice and GPh^{eYFP} mice ($P > 0.05$, *t*-test). **j**, Confocal images from a Som-flp mouse, showing the overlap in expression of Cre–GFP (delivered by injecting the EP with the AAV-FSF-GFP-Cre) and somatostatin, recognized through antibody labelling. 96.25 ± 2.3% of Cre–GFP$^+$ neurons expressed somatostatin ($n = 2$ mice). **k**, Schematics of the experimental approach. CTB-594 was injected into the LHb to label GPh neurons in the EP. On the right is an enlarged graph of the boxed area in the cartoon on the left. **l**, Sample miniature IPSC (mIPSC) traces recorded from a GPh neuron that expressed Cre–GFP (and thus had γ2 ablated (γ2-KO)) and a control GPh neuron that did not express the Cre–GFP (γ2-WT). **m**, Quantification of the frequency (left) and amplitude (right) of mIPSCs recorded from the two groups of GPh neurons (γ2-KO, $n = 7$ cells; γ2-WT, $n = 10$ cells; $n = 3$ mice; frequency, $t_{(15)} = 5.51$, $^{****}P < 0.0001$; amplitude, $t_{(15)} = 8.19$, $^{****}P < 0.0001$; *t*-test). **n**, A representative image showing the expression of Cre–GFP (delivered by injecting the EP of a *Som-Flp;Gabrg2^{flox}* mouse with the AAV-FSF-GFP-Cre) in GPh neurons (left) and a schematic of the approach (right). **o**, The lose–switch percentage in these mice ($P > 0.05$, *t*-test). **p**, For animals ($n = 9$) used in Fig. 4f–j, the number of GPh neurons that were infected with the Cre–GFP virus correlated with the change in animal behaviour in the switching task, measured as a reduction in action value following two consecutive rewarded trials ($R^2 = 0.53$, $P < 0.05$ by a linear regression). **q**, The negative regression coefficients associated with the past five trials were not significantly different between GPh$^{\gamma2\text{-}KO}$ mice and control mice either before or after surgery (first two trials back × groups, $F_{(3,35)} = 0.9072$, $P > 0.05$, $n = 9$ GPh$^{\gamma2\text{-}KO}$ mice and $n = 9$ control mice). **r**, The action value following two sequentially unrewarded trials was not significantly different between GPh$^{\gamma2\text{-}KO}$ mice and GPh^{mCherry} mice ($P > 0.05$, *t*-test). All data are represented as mean ± s.e.m. Diagrams in **b**, **e**, **k** and **n** were modified from the Allen Mouse Brain Atlas, Allen Institute for Brain Science; available from http://mouse.brain-map.org/.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Monosynaptic inputs onto the GPh and a schematic of the circuitry for reinforcement learning. a, Schematics of experimental design. The GPh neurons in the EP were targeted using either *Vglut2-Cre;Rosa26-stop^{flox}-tTA* mice or by injecting the LHb of *Rosa26-stop^{flox}-tTA* mice with the retrograde CAV2–Cre. **b**, Images showing the starter cell location in the EP. **c**, Relationship between the number of starter and input neurons. **d**, Graph showing the fraction of monosynaptically labelled neurons in each brain region that projects to the GPh ($n = 9$ mice) **e**, Confocal images of the rabies virus and parvalbumin (PV) labelled neurons in the GPe. Only a small fraction of the virally labelled GPe cells expressed PV (arrows). On the right is a high-magnification image of the boxed area in the GPe. **f**, Quantification of the fraction of rabies virus-labelled GPe neurons that expressed PV ($n = 3$ mice). **g**, Center of mass analysis for all GPe labelled neurons ($n = 9$ mice). **h**, A confocal image of the parasubthalamic nucleus (pSTN) showing monosynaptically labelled neurons. **i**, Center of mass analysis for all pSTN labelled neurons ($n = 9$ mice). **j**, A schematic showing the proposed selection and evaluation circuits within the basal ganglia. Question marks indicate elements of the proposed circuit that remain to be tested experimentally. Diagrams in **a**, **g** and **i** were modified from the Allen Mouse Brain Atlas, Allen Institute for Brain Science; available from http://mouse.brain-map.org/.

LETTER RESEARCH



Extended Data Figure 10 | The proposed function of the basal ganglia and midbrain evaluation circuits. a, Schematic showing the activity of GPh neurons and the downstream circuitry controlling the midbrain dopaminergic system. **b**, Proposed sequence of events by which GPh activity may influence the firing rate in downstream structures. Upward arrows indicate an increase in firing; downward arrows indicate a decrease in firing. RMTg, Rostromedial tegmental nucleus; SNc, Substantia nigra pars compacta; VTA, ventral tegmental area; DA, dopamine. DR, dorsal raphe; MR, median raphe. A question mark (?) indicates that alternative circuits downstream of the LHb, including the serotonergic raphe nuclei, may constitute other key pathways that also process the GPh–LHb prediction error signals that we demonstrate in this study. Diagram in **a** was modified from the Allen Mouse Brain Atlas, Allen Institute for Brain Science; available from http://mouse.brain-map.org/.