Striatal dopamine encodes the relationship between actions and reward

G. Hart[1*], T.J. Burton[1*] & B.W. Balleine[1#]

1. Decision Neuroscience Lab, School of Psychology, UNSW Sydney

* These authors contributed equally to this work.

Corresponding Author: bernard.balleine@unsw.edu.au

Abstract:

Despite the well described role of dopamine in the striatum for motor learning and movement, its function in encoding actions and their outcomes for goal-directed action is less clear. Here, rats acquired two actions for distinct outcomes while we simultaneously recorded dopamine release in the dorsomedial striatum as these action-outcome associations were incremented and selectively degraded. Goal-directed actions generated a lateralized dopamine signal that reflected the strength of the action-outcome association and tracked increments and decrements in the action-outcome contingency. In addition to this lateralized signal, dopamine signaled reward predictions, which were broadcast bilaterally during the action and updated by a reward prediction error following exposure to the outcome. Therefore, striatal dopamine release directly modulates and updates encoding of the instrumental action-outcome association for goal-directed action.

The capacity for goal-directed action is a core function that, in a changing environment, allows animals to make flexible choices to maintain adaptive behavior (Dickinson and Balleine, 1994; Dolan and Dayan, 2013). Core to this capacity is the ability to encode the relationship between actions and their consequences or outcomes, a process that depends on a prefronto-striatal circuit focused on the posterior dorsomedial striatum (pDMS) (Balleine, 2019; Balleine and O'Doherty, 2010; Hart et al., 2018b, 2018a). Recent evidence suggests that, within the pDMS, the encoding and updating of specific action-outcome associations relies on key changes in plasticity at both types of principal neuron (Balleine et al., 2021): the striato-nigral direct spiny projection neurons (dSPNs), which express dopamine D1 receptors, and striato-pallidal indirect SPNs (iSPNs) expressing the D2 receptor (Gerfen and Surmeier, 2011). Inhibition of dSPNs prevents goal-directed learning and alters its expression (Peak et al., 2020), whereas inhibition of iSPNs prevents the updating of this learning (Matamales et al., 2020; Peak et al., 2020).

The cellular plasticity in the pDMS necessary for goal-directed action is thought to reflect the integration of glutamatergic inputs from cortical, thalamic and limbic regions with the input from midbrain dopaminergic neurons (Bradfield et al., 2013; Fisher et al., 2020; Holly et al., 2019; Lee et al., 2020; Nonomura et al., 2018; Peters et al., 2021; Reynolds et al., 2001; Shiflett and Balleine, 2011). Nevertheless, although much has been learned about the role of striatal dopamine in the cellular and circuit level plasticity necessary for reflexive motor movements and skills (Klaus et al., 2017; Park et al., 2020; Santos et al., 2015) relatively little is known about the specific role that dopamine activity plays in the acquisition and performance of goal-directed actions. Recent studies have shown that dopaminergic terminals in the pDMS show activity consistent with both action value and selection (Lee et al., 2019; Parker et al., 2016); however, it remains unclear whether these signals are specifically related to goal-directed action, and so reflect the development of action-outcome learning, or are a function of instrumental performance generally.

Dopamine release in the pDMS becomes lateralized as goal-directed training progresses

To address these questions, we utilized a genetically-encoded dopamine indicator, dLight1.1 (Patriarchi et al., 2018) (**Fig 1A, Fig 2A**) to measure rapid changes in dopamine release in the pDMS during the acquisition and performance of goal-directed actions. This indicator inserts circularly permutated GFP into dopamine D1 receptors such that the binding of dopamine causes an increase in fluorescence, which we measured using fiber photometry. We examined the release profile during initial learning and subsequently during performance and then assessed effects specific to changes in the strength of the action-outcome association when this was updated during a contingency degradation treatment.

Rats were trained on a goal-directed instrumental conditioning protocol, in which two levers, located to the left and right of a central food port or magazine, each delivered distinct rewards (grain pellets and sucrose solution) on increasing random interval (RI) schedules of reinforcement across sessions that could be collected by entry to the food magazine (**Fig 1B**). Response rates on each lever and mean entry rates across days of training are presented in **Fig 2B**. Levers were allocated as to-be-degraded or to-be-non-degraded, with side of lever (left and right) counterbalanced within each condition. Rats increased responding on both levers across days, and there were no differences in the response rates of each action ($p<0.05$).

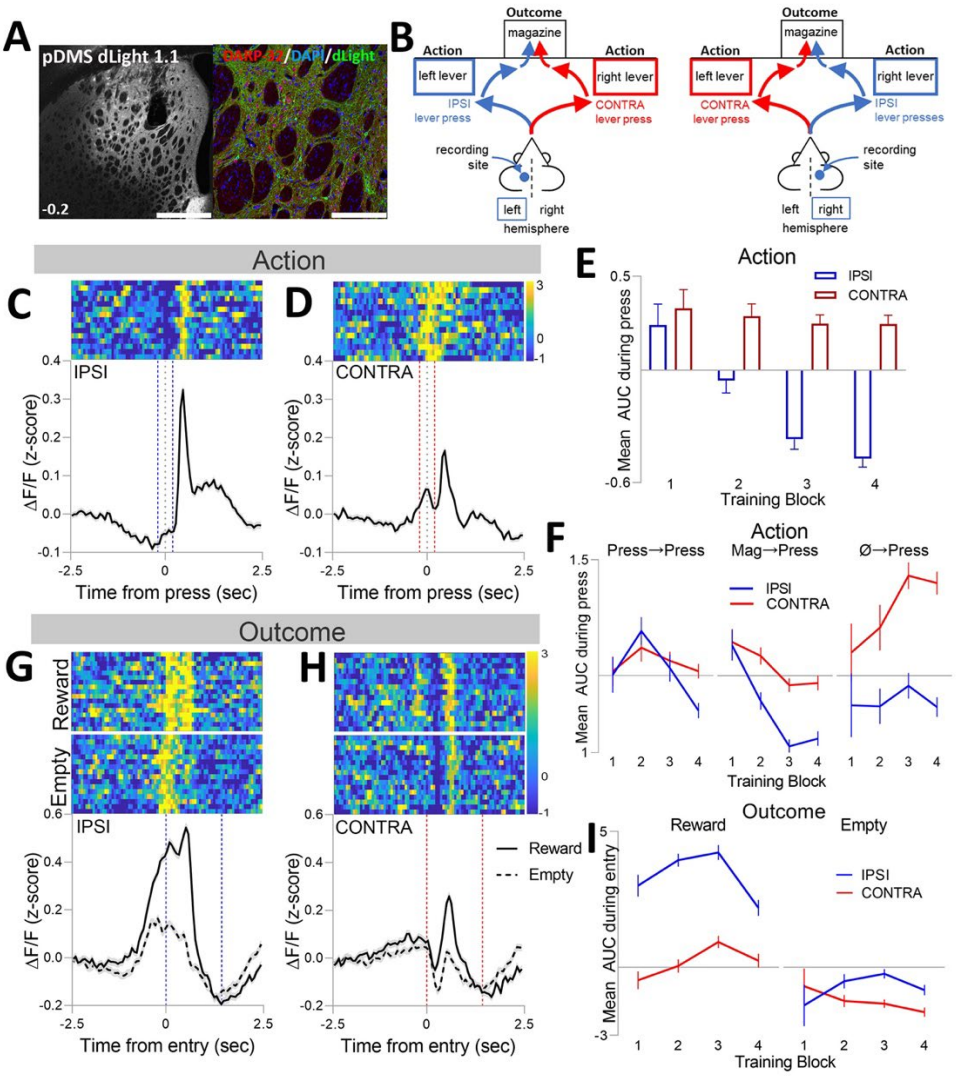

Fig. 1. Dopamine release in the pDMS is lateralized during goal-directed actions and their outcomes. (**A**) Expression of fiber cannula (left) and dLight1.1 in the pDMS (0.2mm posterior to bregma), scale bar =1000um. Right: dLight1.1 (green), DARP-32 (red) and DAPI (blue), scale bar=200um. (**B**) Rats were trained with alternating levers left and right of the magazine, designated as IPSI and CONTRA, relative to the recording hemisphere. (**C**) ΔF/F dLight signals over time aligned to IPSI lever presses. Top: Example trials showing color-coded ΔF/F signals. Bottom: the mean ΔF/F averaged across all IPSI trials during instrumental training, Action Window indicated by blue lines. (**D**) Same as (C), for CONTRA lever presses, Action Window indicated by red lines. (**E**) The mean (+ SEM) area under the curve (AUC) of the ΔF/F signal during the Action Window for all IPSI and CONTRA lever presses, averaged within each training block. (**F**) The mean (± SEM) AUC of the ΔF/F signal for IPSI and CONTRA presses, subdivided according to preceding event, averaged within each training block. (**G**) ΔF/F dLight signals over time aligned to rewarded and non-rewarded (empty) magazine entries after IPSI presses. Top: Example trials showing color-coded ΔF/F signals. Bottom: the mean ΔF/F averaged across all rewarded and empty magazine entries after IPSI presses during instrumental training, Outcome Window indicated by blue lines. (**H**) Same as for (G) after CONTRA presses, Outcome Window indicated by red lines. (**I**) The mean (± SEM) AUC of the ΔF/F signal during the Outcome Window for all rewarded (left) and empty (right) magazine entries after IPSI and CONTRA presses, averaged within each training block.

To identify epochs of interest for photometry analysis, we compiled the temporal dynamics of all press-entry sequences (**Fig 2C**) and sought to analyze non-overlapping windows around the Action (lever press) and Outcome (i.e., when rats either obtained or failed to obtain a food reward). The Action window extended 0.2s before to 0.2s after a lever press, whereas **the** Outcome window captured entry to the reward magazine area and began on entry after a press and continued for 1.5s (see **Fig 2C**). We recorded dopamine release in the pDMS in one hemisphere (left or right, counterbalanced) with presses categorized according to whether the lever was located ipsilateral (IPSI) or contralateral (CONTRA) to the recording hemisphere (**Fig 1B)**. The mean average z-scored $\Delta F/F$ signals aligned to IPSI and CONTRA presses averaged across instrumental training are presented in **Fig 1C-D**.

There was a clear difference in the profile of dopamine release during IPSI relative to CONTRA presses. To quantify the relative change in release over acquisition, we subdivided the training into 4 blocks and calculated the area under the curve (AUC; trapezoidal method) within each Action window (**Fig 1E).** Dopamine release was progressively suppressed during IPSI presses but maintained during CONTRA presses across training. There was no difference in release during IPSI and CONTRA presses in Block 1 ($p=0.6$), but a significant difference in Blocks 2-4 ($k=6$, p-critical=0.008, p's < 0.0002). Importantly, there was no change in the magnitude of release during CONTRA presses across blocks ($p>0.4$), but there was a significant reduction across blocks during IPSI presses (Block 1 vs. Block 4: $p=2.17E-09$).

Next we classified IPSI and CONTRA presses according to the events immediately preceding them: i.e., whether they (1) followed another press in sequence (Ipsi→IPSI or Contra→CONTRA); (2) followed a magazine entry (Mag→IPSI or Mag→CONTRA); or (3) when neither a press or entry occurred in the preceding window (Ø→IPSI or Ø →CONTRA) based on a time window that was 1 SD of the inter-press-interval on that lever for that session. The total number of each press type in each training block is presented in **Fig 2D** (left), which shows presses were predominantly part of Press→Press or Mag→Press sequences. The mean AUCs for each action category are presented in **Fig 1F.** Hemispheric lateralization of dopamine release increased across training for each press category: there were no differences between IPSI and CONTRA presses of any type in Block 1 ($k=48$, p-critical =0.001, p's >0.2); in Block 2, this difference was significant for Mag→Presses (i.e. Mag→IPSI vs. Mag→CONTRA $p=0.0003$); in Block 3 this difference was significant for both Mag→Presses and Ø→Presses (p's < 2.80E-09); and by Block 4, CONTRA presses were significantly higher than IPSI presses for all subcategories (p's < 0.0002).

To assess whether this emergent lateralization was related to increased response vigor as training progressed, we assessed the correlation between the SD of the inter-press interval against the session-average AUC during the Action window for each press type occuring in the final training block (Block 4; Fig 2E). Surprisingly, we found that response vigor was either unrelated or negatively related to the magnitude of dopamine lateralization (i.e. dopamine release during CONTRA presses and dopamine inhibition during IPSI presses); for Mag→Presses, dopamine release was more lateralized (i.e. greater during CONTRA and more inhibited during IPSI) for sessions with lower press rate (Mag→IPSI, $r=-0.22$, $p<0.0001$; Mag→CONTRA, $r=0.16$, $p=0.0016$), this was also true of Ø→CONTRA presses ($r=0.4$, $p<0.0001$), and there were no other significant correlations (p's>0.1).

We also assessed whether dopamine release in the Action window differed according to the

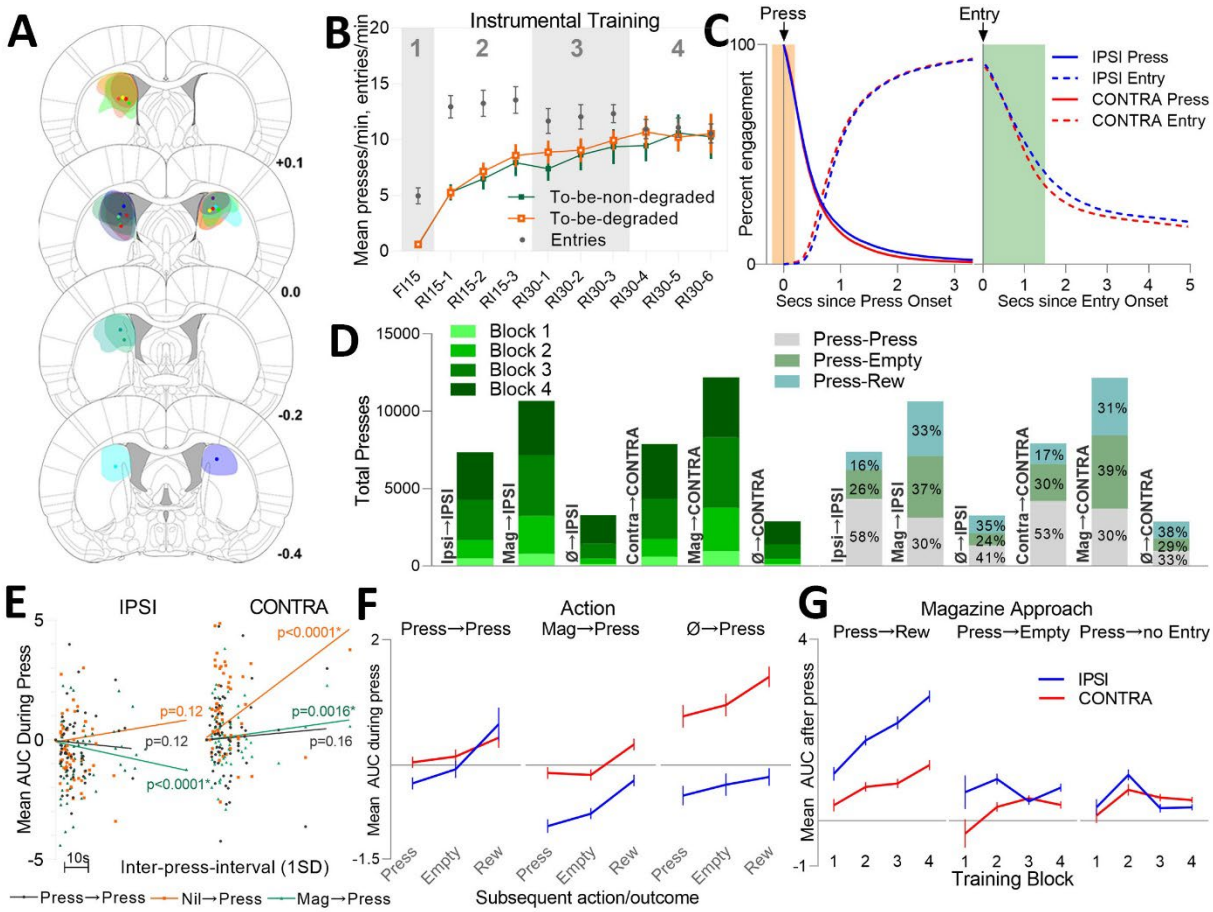

Fig. 2 (A) Placement of fiber photometry cannula tips and spread of dLight1.1 in the pDMS for all rats included in the analysis. Numbers indicate anterior/posterior co-ordinates (mm) from bregma. (B) Mean (±SEM) presses per minute on the to-be-degraded and to-be-nondegraded levers, and mean (±SEM) magazine entries per minute across each day of instrumental training. Numbers in grey indicate training blocks. (C) Percent of animals engaged in a press (solid lines) or entry after a press (dashed lines) that was IPSI (blue) or CONTRA (red) across time aligned to press (left) and entry (right) onset, across all presses in instrumental training. Action window shaded in orange, Outcome window shaded in green. (D) Left: Total number of each press subtype within each training block across instrumental training. Right: Number of each press subtype followed by either another press (grey), an empty entry (green) or a rewarded entry (blue) summed across instrumental training. (E) Mean (session-average) AUC during IPSI (left) and CONTRA (right) press subtypes (y-axis) against 1SD of the inter-press-interval for that session (x-axis). Each dot represents one session for one rat. Lines indicate the best-fit slopes (simple linear regression), p-values indicated. (F) Mean (±SEM) AUC of the ΔF/F signal during the Action window for each press subtype, divided according to following event (as in D right). (G) Mean (±SEM) AUC of the ΔF/F signal during the Magazine Approach window (0.2s-0.4s after press) for each press subtype, averaged within training blocks.

event following its performance: Fig 2D (right panel) shows the percentage of each press type followed by another press (Press-Press), a non-rewarded magazine entry (Press-Empty) or a rewarded magazine entry (Press-Rew). Press sequences (Ipsi→IPSI or Contra→CONTRA) were least likely to be rewarded, and most likely to be followed by another press, whereas spaced presses (i.e. Ø→IPSI or Ø→CONTRA) were most likely to be rewarded. The mean AUCs during the Action window for each of these press subtypes are presented in Fig 2F. There were no differences in release for actions followed by either another press or an empty magazine check (k=18, p-critical=0.003, p's > 0.15). There was, however, significantly greater release during rewarded actions that had followed magazine entries (Mag→IPSI and Mag→CONTRA, p's < 0.0002; p's > 0.01 for other press subtypes), suggesting that reward predictions were enhanced at such moments.

Dopamine release during the outcome retrieval reflects both motor output and reward

To establish whether this lateralized dopamine signal was specific to the Action window, we assessed dopamine release during the window 0.2-0.4s after the press, during which the reward was delivered (if there was reward) and animals often turned in the opposite direction from the lever to approach the magazine (illustrated in **Fig 1B**). These responses were categorized according to both the preceding press and the immediately following outcome: i.e., whether IPSI and CONTRA presses followed by magazine entries were either rewarded or not (Press→Rew, Press→Empty), or were followed by other presses with no magazine entry (Press→no entry). The mean AUCs across this Magazine Approach window are presented in **Fig 2G**. There was a pronounced dopamine release spike in this window on rewarded trials, corresponding to the reward delivery (starting 0.2s after press), the magnitude of which increased across training blocks (k=52, p-critical=0.001; IPSI→Rew and CONTRA→Rew, Block 1 vs Block 4 p's < 1.04E-06; other press types Block 1 vs Block 4 p's > 0.06). However, importantly, whereas there was heightened dopamine during CONTRA presses (relative to IPSI presses) during the Action window, this pattern reversed during the Magazine Approach window: dopamine release was significantly higher following IPSI presses, as rats made the contralateral turn towards the magazine (IPSI→Rew vs. CONTRA→Rew, p's < 0.0008 across all training blocks). Therefore, dopamine release increased across training as rats learned to detect the reward delivery and to approach the magazine, and this was augmented when the approach involved a contralateral turn from the lever to magazine, in accord with the general lateralization of release during motor performance. Importantly, therefore, the lateralization of dopamine during the lever press was temporally restricted to the Action window and clearly distinguishable from subsequent dopamine signals associated with magazine approach.

Finally, we assessed dopamine release during the Outcome window. The mean z-scored $\Delta F/F$ signals averaged across instrumental training aligned to magazine entry are presented in **Fig 1G-H**. The dopamine release profile differed for entries that followed IPSI vs. CONTRA presses: after IPSI presses (and contralateral turns; IPSI→Rew and IPSI→Empty) dopamine release was elevated at the time of magazine entry and remained so throughout the first half of the window, then declined. By contrast, dopamine after CONTRA presses (and ipsilateral turns; i.e. CONTRA→Rew and CONTRA→Empty) was relatively low at the point of magazine entry, and had a sharp dip before a sharp peak, returning to below baseline across the second half of the window. The relative dopamine release averaged across the Outcome window (0-1.5s after entry onset) is reflected in the mean AUCs (**Fig 1I**), showing significantly greater dopamine release during rewarded than empty magazine entries (k=32

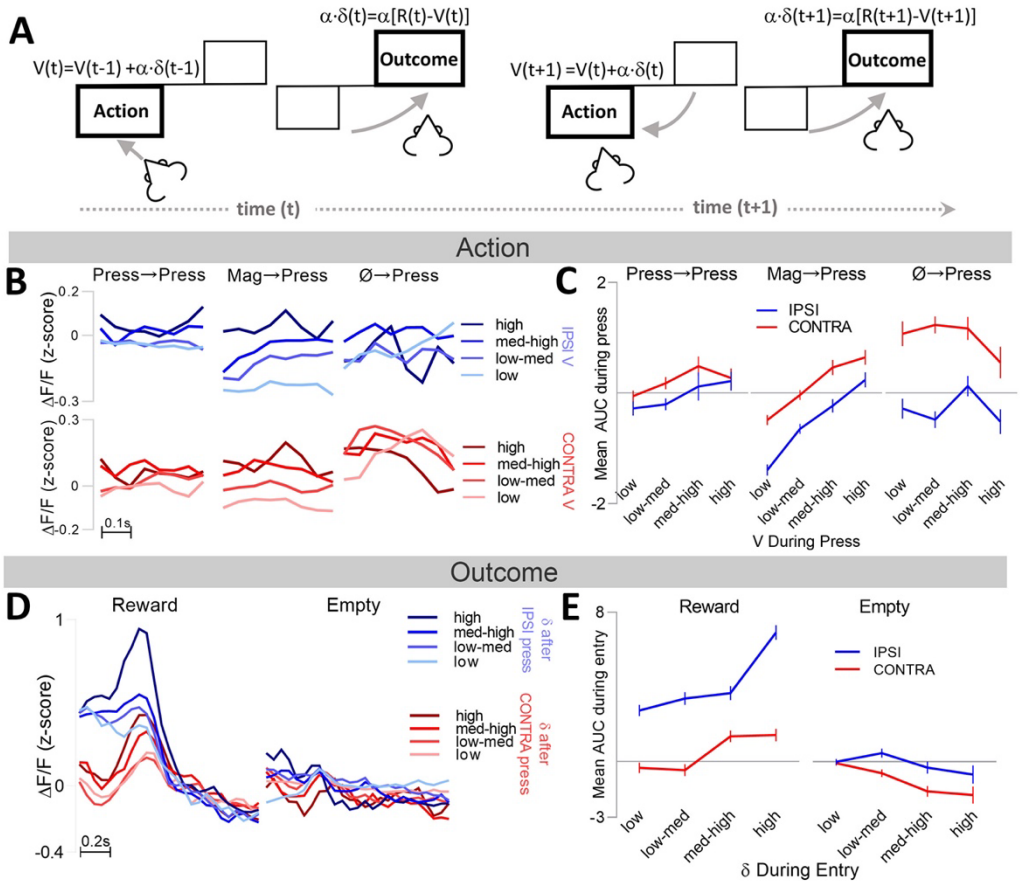

Fig. 3. Dopamine reflects action values and outcome reward prediction errors. (**A**) Action values are updated by summing the previous action value with the reward prediction error calculated during the preceding magazine entry (Outcome). (**B**) Mean ΔF/F signals during the Action window (-0.2s before press to 0.2s after press) for IPSI (blues) and CONTRA (reds) press subtypes, subdivided according to action value. (**C**) Mean AUC (±SEM) for signals shown in (B). (**D**) Mean ΔF/F signals during the Outcome window (0s to 1.5s after entry) for rewarded (left) and empty (right) magazine entries after IPSI (blues) and CONTRA (reds) press subtypes, subdivided according to reward prediction error calculated during that same entry. (**E**) Mean AUC (±SEM) for signals shown in (D).

p-critical =0.0015, IPSI→Rew vs. IPSI→Empty Blocks 1-4, p's<5.8E-05; CONTRA-Rew vs CONTRA-Empty Blocks 2-4 p's<0.0002). The direction of the immediately preceding turn also significantly modulated the magnitude of dopamine release during magazine entry: the contralateral turn after an ipsilateral press resulted in greater dopamine release than the ipsilateral turn after a contralateral press (IPSI→Rew vs. CONTRA→Rew Blocks 1-4 p's <8.6E-07, IPSI→Empty vs. CONTRA→Empty Blocks 3-4 p's<0.0012). Finally, unlike the Action window, there was no evidence during the Outcome window for a significant change in dopamine release across training blocks (Block 1 vs Block 4 p's>0.09).

Dopamine release during instrumental acquisition reports the updated action value

Our results suggest that, during instrumental training, dopamine release is consistently greater in the hemisphere contralateral to motor output, however the evolution of this signal across training differs in each stage of the action-outcome sequence: Dopamine release diverges between the ipsilateral and contralateral hemispheres during the action, whereas, during outcome retrieval, it is lateralized from the outset of training and remains relatively unchanged as training progresses. As such, these results suggest that changes in dopamine release across training reflect learning engaged by the action rather than the outcome. To assess this hypothesis more directly, we next sought to examine whether dopamine release during goal-directed actions is modulated by a reward prediction error generated within the Action and Outcome windows described above.

We calculated reward prediction-errors for each action-outcome combination across training for each animal, whereby action values (V) were updated using a simplified temporal difference learning rule (Sutton and Barto, 1998). According to this model, at the time of magazine entry, prediction error is calculated for subsequent action selection, at which point in time action values are updated according to: $V(t+1)=V(t)+\alpha.\delta(t)$ (**Fig 3A**), where $\alpha.\delta$ is the prediction error (δ) modulated by a learning rate parameter (α) according to: $\delta(t)= R(t) - V(t)$, where R is the reward value, which is set at 0 for no reward or 1 for reward to generate a negative δ on non-rewarded entries and a positive δ on rewarded entries. Importantly, this simplified model assumes that the reward predictions are only updated by experience with the consequences of the action (i.e., outcomes; rewarded or empty magazine entries); therefore, during press sequences in which the rat doesn't enter the magazine, V will remain unchanged. **Fig 3B** shows the mean z-scored $\Delta F/F$ signals during the Action window across a range of V values (low-high: 0-1) corresponding to the calculated reward prediction at the time of the action.

The mean AUCs during the Action window for each action subtype are presented in **Fig 3C.** There was a clear relationship between dopamine release and V but only when calculated for actions that followed magazine entries (k=6, p-critical=0.008; V(low) vs. V(high): Mag→IPSI p=2.21E-22, Mag→CONTRA p=2.79E-12), and not for those that followed other presses, or occurred in isolation (V(low) vs. V(high): Ipsi→IPSI and Ø→IPSI p's > 0.02; Contra→CONTRA and Ø→CONTRA p's > 0.1). Note that, as our model assumed that prediction error is only generated during magazine entries, this finding suggests that these changes in dopamine release reflect the updating of action values. However, to ensure that this assumption within the model didn't drive the pattern of results, we modified the model to allow V to be updated on every press. In this case, all non-rewarded presses resulted in negative δ (i.e., presses followed by empty magazine checks and presses followed by other presses), and only presses followed by rewarded entries accrued a positive δ. The mean

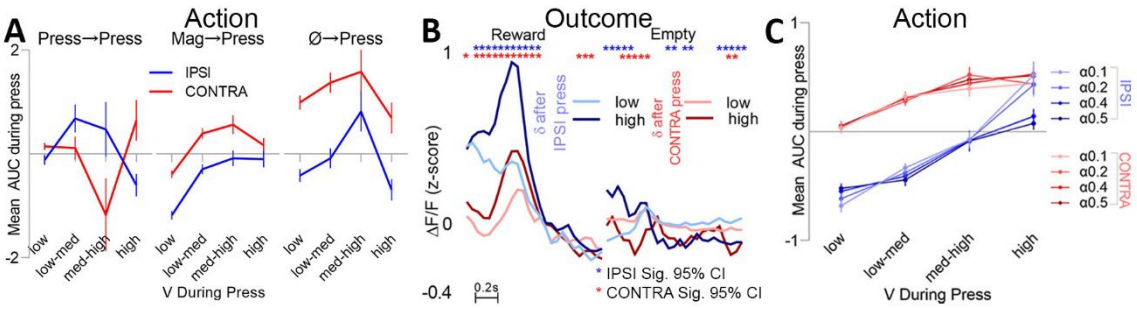

Fig. 4. (A) Mean (±SEM) AUC of the ΔF/F signal during the Action window for IPSI (blue) and CONTRA (red) press subtypes, subdivided according to action value when actions values are updated on every press. (B) ΔF/F signals during the Outcome window for rewarded (left) and empty (right) magazine entries, after IPSI (blues) and CONTRA (reds) presses, subdivided according to high vs low reward prediction errors, which are positive for rewarded entries and negative for empty entries. Asterisks indicate where high and low traces are significantly non-zero using bootstrapping confidence intervals procedure. (C) Mean (±SEM) AUC during the Action window for all IPSI (blues) and CONTRA (reds) presses during instrumental training, subdivided according to action values, for varying alpha values (0.1-0.5).

AUCs during presses across binned values of V using this model are presented in **Fig 4A**. Although there was generally more variability, the findings remained the same: dopamine release during presses that had followed magazine entries was positively related to V (k=6, p-critical=0.008; V(low) vs. V(high): Mag→IPSI p=3.36E-08, Mag→CONTRA p=0.0008), and there was no significant relationship between dopamine release and V during other press subtypes (i.e. V(low) vs. V(high): Ipsi→IPSI, Ø→IPSI , Contra→CONTRA and Ø→CONTRA p's > 0.05).

Dopamine release during outcomes reports a reward prediction error

The previous results confirm that dopamine release in the pDMS most consistently reflects the value of a goal-directed action but only for the action following a magazine entry, suggesting that it was only after contact with reward (or non-reward) – and the consequent exposure to the prediction error – that action values are updated. To further examine this finding, we analyzed the magnitude of both reward prediction error (δ) and dopamine release during the Outcome window. These data are presented in **Fig 3D-E**. We binned magazine entries according to the magnitude of the error (δ) generated by the model (high-low), assuming δ is only calculated during entries. The mean z-scored ΔF/F signals during the Outcome window across differing δ values, which are positive for rewarded entries and negative for unrewarded entries, are presented in **Fig 3D.** As previously noted, the release profile during the Outcome window differs for entries following IPSI and CONTRA presses and was maintained across varying δ values. During the release peak for rewarded entries (**Fig 3D, left**), there was a graded increase in dopamine release according to δ and this was reflected in the mean AUCs, **Fig 3E.** A positive relationship was observed between dopamine release and positive prediction errors (δ) during the Outcome window when rewards were delivered (k=4 p-critical=0.01, δ(low) vs. δ(high): IPSI→Rew p=2.35E-17; CONTRA→Rew p=5.22E-05). Equally importantly, there was a significant relationship between negative prediction errors and unrewarded magazine entries (δ(low) vs. δ(high): CONTRA→Empty p=0.0005), however this relationship wasn't significant for IPSI entries (IPSI→Empty, p=0.16). To directly compare waveform profiles for retrieved outcomes with high δ vs. low δ, we used a bootstrapping confidence intervals procedure (Methods, **Fig 4B**), which confirmed that dopamine release was significantly greater for (positive) high δ vs. low δ during the release peak for rewarded entries and significantly lower for (negative) high δ vs. low δ during the comparable timepoint for nonrewarded CONTRA entries. By contrast, for IPSI entries, dopamine release was inversely related to δ (negative) at the very start of the Outcome window (i.e. δ(high) > δ(low)) and this pattern reversed at the very end of the window (i.e. δ(low) > δ(high); positive). Together these results confirm that dopamine release during reward reports a positive prediction error and during nonreward a negative prediction error.

Contingency degradation reverses Action-specific but not Outcome-specific hemispheric lateralization.

If dopamine release reflects the action-outcome association then it should increase during acquisition and decline when the action-outcome contingency is degraded. To investigate the relationship between dopamine activity and the strength of the action-outcome association more directly, therefore, we assessed release during degradation of one of the two action-outcome contingencies. Both levers continued to be presented in the same fashion and earned the same outcomes on the same interval schedule as during training,

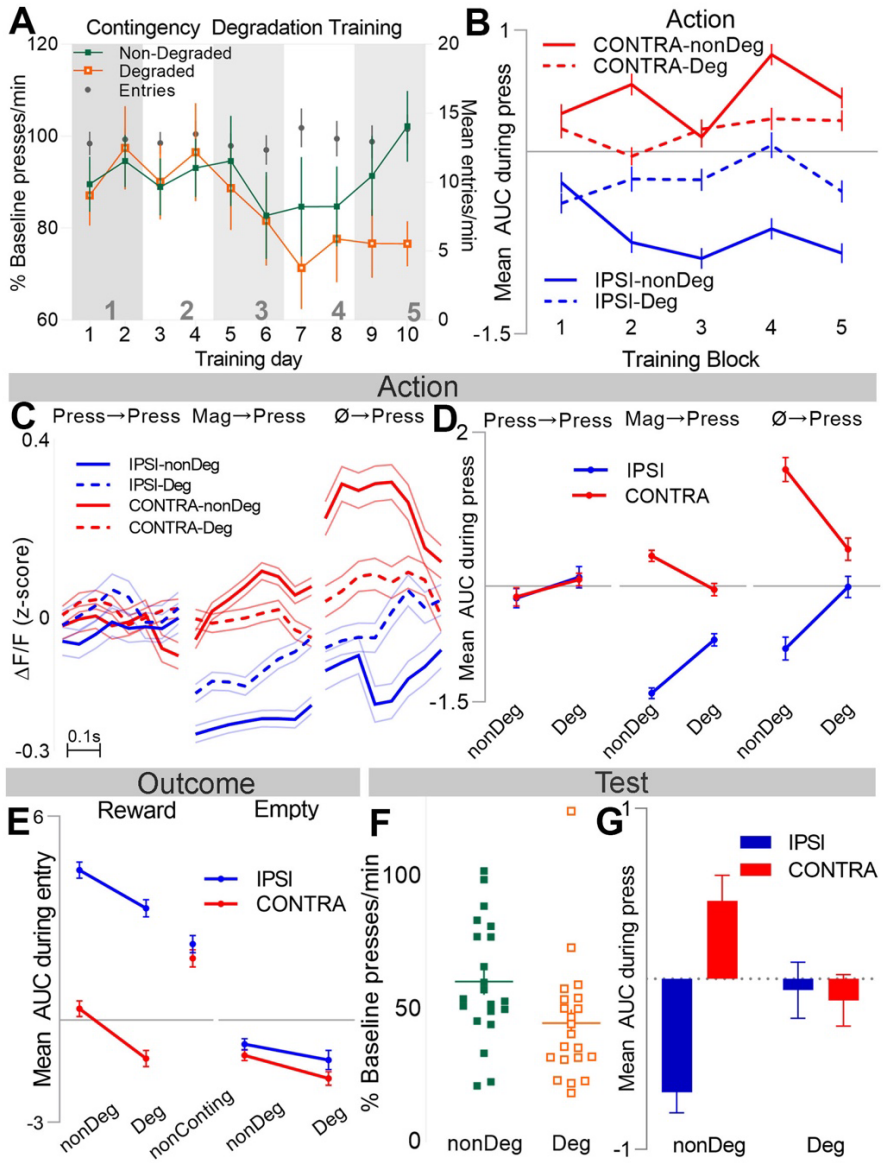

Fig. 5. Dopamine lateralization during the Action but not the Outcome reflects instrumental contingency. (**A**) Mean (±SEM) percent of baseline presses per minute on the non-degraded (nonDeg) and degraded (Deg) levers, and mean (±SEM) entries per minute (Entries) across 10 days of contingency degradation training. Numbers in grey indicate training blocks. (**B**) Mean (±SEM) AUC of the ΔF/F signal during the Action window for IPSI (blue) and CONTRA (red) presses that were degraded or non-degraded, averaged across training blocks of contingency degradation. (**C**) ΔF/F signals during the Action window for IPSI (red) and CONTRA (blue) press subtypes, averaged across Blocks 2-5 of contingency degradation training. Thin lines indicate ±SEM. (**D**) Mean (±SEM) AUC for signals shown in (C). (**E**) Mean (±SEM) AUC of the ΔF/F signal during the Outcome window, averaged across contingency degradation training. Left: Rewarded entries that were contingent on IPSI (blue) or CONTRA (red) presses that were degraded (Deg) or nondegraded (nonDeg), or noncontingent rewards (nonConting). Right: Empty entries following degraded or nondegraded IPSI or CONTRA presses. (**F**) Percent of baseline presses per minute on the degraded and non-degraded levers for each animal during the contingency degradation test. Lines indicate mean ±SEM. (**G**) Mean (±SEM) AUC of the ΔF/F signal during the Action window for IPSI (blue) and CONTRA (red) presses that were degraded or nondegraded, averaged across the contingency degradation test.

however, in addition, throughout each session rats received non-contingent, i.e., response independent, deliveries of one of the earned outcomes delivered on a random time (RT) schedule. It has previously been demonstrated that, for goal-directed actions, response independent outcomes 'degrade' previously established associations between an action and that specific outcome, resulting in the selective reduction in performance of the action relative to actions delivering other outcomes (Colwill and Rescorla, 1986; Dickinson and Charnock, 1985; Rescorla, 1992). Responding relative to baseline (Block 4 of instrumental training, see Methods) on the degraded and non-degraded levers across the 10 days of contingency degradation is presented in **Fig 5A**. We compared responding on the degraded and non-degraded levers averaged across training blocks (i.e. blocks 1-5, 2-5, 3-5, 4-5 and 5 only). There were no differences between response rates on the two levers averaged across all days of training (Deg vs nonDeg Blocks 1-5 $p>0.9$), however responding on the degraded lever was gradually suppressed from Block 3 onward, confirmed by a marginal difference between response rates on the degraded and non-degraded levers on training Blocks 4-5 ($k=5$, p-critical=0.01, $p=0.02$), and a significant difference between degraded and non-degraded levers in Block 5 ($p=0.005$).

We measured dopamine release in the pDMS across the 10 days of contingency degradation training. The mean AUCs averaged over the Action window for IPSI and CONTRA presses that were degraded or non-degraded are presented across training blocks in **Fig 5B** and show that the effect of degradation on dopamine was apparent on both IPSI and CONTRA presses: Degradation increased dopamine release during IPSI presses, and reduced dopamine release during CONTRA presses, and this effect emerged across degradation training. Specifically, there was no difference between degraded and nondegraded levers during the first training block **(Fig 5B**; IPSI and CONTRA Deg vs. nonDeg, Block 1 p's > 0.1), but a significant difference was observed across all subsequent blocks for IPSI presses ($k=10$, p-critical = 0.005, IPSI Deg vs. nonDeg, Blocks 2-5, p's <0.0001). A significant effect of contingency degradation was also apparent for CONTRA presses in Blocks 2 and 4 (CONTRA Deg vs. nonDeg, p's <3.18E-05), but not for Blocks 3 and 5 (p's >0,1). To assess the possible source of this variability, we subdivided presses according to their preceding events; the mean z-scored $\Delta F/F$ signals within the Action window for each press type averaged across the last 4 training blocks is shown in **Fig 5C**, and the mean AUCs in **Fig 5D**. The effect of contingency degradation was to reverse the hemispheric lateralization observed for Mag→Presses and Ø→Presses; for contralateral presses this manifested as a reduction in dopamine release during degraded relative to nondegraded presses (**Fig 5D;** $k=6$, p-critical=0.008; Deg vs. NonDeg: Mag→CONTRA p=7.47E-05, Ø→CONTRA p=1.27E-06), whereas for ipsilateral presses, this meant an increase in dopamine with degradation ($k=6$, p-critical=0.008, Deg vs. nonDeg: Mag→IPSI p=1.20E-10, Ø→IPSI p=0.0001). Interestingly, there was no evidence of hemispheric lateralization in press sequences (Ipsi→IPSI vs. Contra→CONTRA, p's>0.1), and likewise, no difference in dopamine release during degraded and nondegraded sequence presses (Ipsi→IPSI and Contra→CONTRA: Deg vs. NonDeg, p's > 0.1). This pattern of results was maintained when averaging across all of contingency degradation training (i.e., including Block 1, **Fig 6A**), except that Deg vs. nonDeg Ø→IPSI presses weren't significantly different ($p=0.02$). Taken together, these results indicate that contingency degradation reversed the hemispheric lateralization of dopamine release clearly seen during the initial training period.

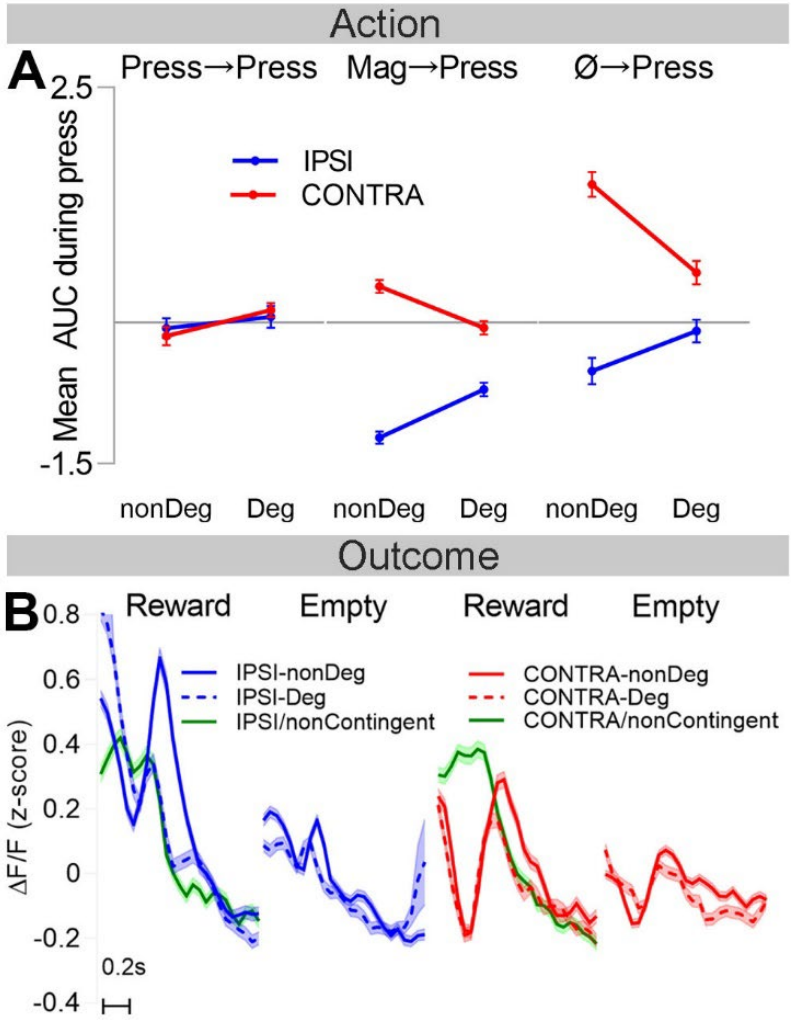

Fig. 6. (A) Mean (±SEM) AUC of the ΔF/F signal during the Action window for IPSI (blue) and CONTRA (red) degraded (Deg) or non-degraded (nonDeg) press subtypes, averaged across all blocks of contingency degradation. (B) Mean ΔF/F signals during the Outcome window following IPSI (blue, left) or CONTRA (red, right) presses that were rewarded (left) or empty (right), entries for noncontingent rewards shown in green.

Next, we assessed dopamine release during the Outcome window, presented in **Fig 5E and Fig 6B**. The distinct profiles of dopamine release during magazine entries following IPSI and CONTRA presses (**Fig 6B)** were maintained for degraded presses, and critically, these shapes looked very different for entries following noncontingent outcome deliveries (IPSI/noncontingent and CONTRA/noncontingent), which (by design) didn't follow a press but were delivered outside of IPSI or CONTRA presses. The mean AUCs during the Outcome window following each press type and following noncontingent rewards are presented in **Fig 5E**. There was no difference in dopamine release during response-independent rewards delivered during IPSI or CONTRA lever blocks (IPSI/noncontingent vs. CONTRA/noncontingent, k=9, p-critical=0.006, p=0.13), however there was a general reduction in release during degraded outcome retrieval: For empty magazine checks this effect wasn't significant (Deg vs. nonDeg: IPSI→Empty p=0.1; CONTRA→Empty, p=0.0064), however, this reduction was significant for degraded rewards (IPSI→Rew and CONTRA→Rew, Deg vs. nonDeg, p's < 0.002), indicating that contingency degradation significantly depressed reward-related dopamine release.

To confirm that contingency degradation training induced a long-term change in the action-outcome association and not just within-session suppression, we conducted a choice extinction test when rats were hungry on the day after the final contingency degradation training session (**Fig 5F**). Despite no rewards being delivered during this test, the suppression of performance on the degraded lever was maintained, with rats pressing the degraded lever significantly less than the non-degraded lever (k=1, p-critical=0.05, p=0.007), confirming the degradation effect. We measured dopamine release during this extinction test, presented in **Fig 5G.** The pattern of results observed during the Action window across contingency degradation training was maintained on test: There was significantly greater dopamine release (AUC) during CONTRA presses relative to IPSI presses on the nondegraded lever (k=4, p-critical = 0.01, IPSI-nondeg vs CONTRA-nonDeg, p=3.95E-09), however this difference was abolished on the degraded lever (IPSI-Deg vs CONTRA-Deg p>0.7). As in training, the loss of lateralization during these actions was due to a bidirectional change: There was significantly less dopamine released during degraded CONTRA presses (CONTRA-nonDeg vs CONTRA-Deg p=0.007), and significantly more dopamine (or, more accurately, significantly less inhibition) during degraded IPSI presses (IPSI-nonDeg vs IPSI-Deg p=0.003). Therefore, the action-specific lateralization of dopamine release was significantly reversed when the instrumental relationship was degraded, and this change was preserved under extinction conditions when no outcomes were delivered.

Discussion

Difficulty reconciling the functions of striatal dopamine release in reward prediction, movement and motivation has provided a significant barrier to understanding its role in goal-directed action. A prevailing view has been that reward prediction errors are signaled in rapid 'phasic' dopamine, whereas slow variations in 'tonic' dopamine control movement and motivation (Schultz, 2007; Niv et al., 2007). There has, however, been little or no evidence for such 'tonic' signals using techniques that allow fast timescale measurements (Mohebi et al., 2019). Furthermore, while there is evidence for anatomically distinct dopaminergic axonal signaling in the dorsal striatum related to locomotion and reward (Howe & Dombeck, 2016; Panigrahi et al., 2015; da Silva et al., 2018), it remains a challenge to discern how these signals are dissociated at the level of dopamine release; i.e., when increased dopamine release signals reward and when it signals action. Here, by

distinguishing hemispherically lateralized signals from net fluctuations in dopamine release, we have established that, during goal-directed actions, changes in dopamine release simultaneously convey distinct streams of information relating to action-reward predictions and outcome prediction errors, and the action-outcome association for performance. We found that trial-by-trial fluctuations in bilateral dopamine release during the action were associated with changes in action value whereas those changes during the outcome reported a reward prediction error. However, we found evidence of another signal that tracked the overall strength of the specific action-outcome relationship in the degree of lateralized release; i.e., in the difference between release in the hemisphere contralateral to the action and in the hemisphere ipsilateral to the action. In this way, dopamine signaling conveys both the strength of the action-outcome contingency and the current value of the outcome, the integration of which has long been supposed to underly the learning and performance of goal-directed actions (Dickinson & Balleine, 1994; Balleine, 2019).

Dopamine lateralization during action performance has been reported previously; Lee et al (Lee et al., 2019) and Parker et al. (Parker et al., 2016) described a contralateral bias in which dopamine neurons projecting to the DMS respond more strongly during contralateral (relative to ipsilateral) actions. Here, we were able to link this lateralized signal directly to the strength of the action-outcome association, and to show that this signal was distinct from the lateralized motor signals observed during the subsequent reward approach and retrieval. Dopamine release was similar during ipsilateral and contralateral instrumental actions during initial training and diverged over time, whereas release during reward retrieval was lateralized from the outset of training. Furthermore, whereas dopamine lateralization during reward retrieval persisted when the action-outcome relationship was degraded, lateralization diminished during performance of the degraded action. This action-specific lateralization reflects, therefore, a learning-induced change that specifically tracks the strength of the action-outcome association. Furthermore, the finding that this lateralized signal emerged rapidly and progressively increased in magnitude across training reconciles traditional functions ascribed to slow changes in dopamine release associated with motivation or response vigor (da Silva et al., 2018; Niv et al., 2007; Panigrahi et al., 2015; Parkinson et al., 2002; Schultz, 2007) with more recent demonstrations that such changes, when measured on a faster timescale, are comprised of fast fluctuations in dopamine signaling (Mohebi et al., 2019). In this context, it is interesting to note that Hamid et al (Hamid et al., 2021) have recently reported response-dependent modulation of the spatial distribution of dopamine release, which emerged as a 'wave' originating from medial to lateral dorsal striatum when a reward followed a response but from lateral to medial when it was presented independently of responding. These authors hypothesized that this signal reflects a general eligibility trace. The current results extend this claim by directly linking bilateral release to the modulation of action-outcome learning, resulting, over the course of training, in the development of a lateralized signal capable of directly modulating striatal output for goal-directed performance (Peak et al., 2020).

The trial-by-trial changes in bilateral dopamine release tracked the updated action values. Previous reports have suggested that GCaMP activity in DMS-projecting dopamine neurons reflects action values (Parker, et al., 2016; Lee et al., 2019), with dopamine signals in the ventral striatum encoding value and reward prediction-error (Mohebi et al., 2019). Importantly, however, we found that action value coding only occurred when an action's value was updated immediately following contact with reward or non-reward during

outcome retrieval. There was substantial evidence to support this conclusion: We found no evidence for action value coding during press-press sequences but, more importantly, there was clear evidence of both positive and negative outcome-induced prediction-error coding by dopamine release during reward retrieval, confirming that a dopamine error signal was indeed generated at this time to allow the value of the action that immediately followed that retrieval to be updated. These results confirm the longstanding assumption that, as in Pavlovian conditioning, dopamine release during exposure to the instrumental outcome is bilaterally modulated by the reward prediction error and demonstrate that it is this error that updates the action values controlling goal-directed action.

Materials and Methods

Animals

The experiment was conducted with 30 healthy, experimentally naïve wild-type outbred Long-Evans rats (13 female) aged between 8-16 weeks old prior to surgery (244-646g prior to surgery), obtained from the Randwick Rat Breeding Facility at the University of New South Wales. Rats were housed in transparent plastic boxes of 2-4 in a climate-controlled colony room and maintained on a 12 h light/dark cycle (lights on at 7:00 am). All experimental stages occurred during the light portion. Water and standard lab chow were continuously available prior to the start of the experiment. Counterbalancing was conducted within boxes to ensure that rats housed within the same box were allocated different levers and outcomes for degradation. All experimental and surgical procedures were approved by the Animal Ethics Committee at the University of New South Wales and are in accordance with the guidelines set out by the American Psychological Association for the treatment of animals in research.

Exclusions and counterbalancing

All training and lever allocations were fully counterbalanced between males and females, left and right hemisphere recordings and within housing boxes with regards to order of lever and outcome presentation and identity of lever and outcome being degraded. Following histological assessment for placement of dLight1.1 and fiber optic in the pDMS, there were 9 rats excluded for misplaced cannulas or virus injections, leaving a total of 21 rats in the experiment (10 female), with 6 right pDMS recordings (3 ipsilateral lever degraded) and 15 left pDMS recordings (7 ipsilateral lever degraded).

Surgery

Rats were anaesthetized with 3% inhalant isoflurane gas with oxygen, delivered at a rate of 0.5L/min throughout surgery. Anaesthetized rats were placed in a stereotaxic frame (Kopf), and an incision was made down the midline of the skull, and the scalp was retracted to expose the skull. A 1.0 uL glass Hamilton syringe was lowered into the brain for infusions of dLight1.1 (pAAV5-CAG-dLight1.1, AddGene #111067-AAV5), which was delivered at a total volume of 0.75uL (0.1uL/min), with the syringe left in place for an additional 3 min to allow for diffusion. A fiber optic cannula (400um, 6mm, Doric) was then implanted above the injection site, and secured in place with dental cement, attached to the skull with 3 jeweller's screws. Following surgery, rats were injected with a prophylactic (0.4 mL) dose of 300 mg/kg procaine penicillin immediately. Rats were given a minimum of 3 weeks of recovery time following surgery to allow sufficient viral expression.

Surgical co-ordinates were pre-determined from pilot studies and varied slightly between males and females, with dLight infused into males at the co-ordinates (mm from Bregma) A/P:-0.5, M/L:±2.6, D/V:-4.8 and for females A/P:-0.4, M/L:±2.55, D/V:-4.8, with fiber optic placement targeted 0.1mm dorsal to these injection sites.

Apparatus

Training was conducted in two MED Associates operant chambers enclosed in sound- and light-attenuating cabinets. Each chamber was fitted with a pellet dispenser capable of delivering a 45 mg grain food pellet (Bioserve Biotechnologies), as well a pump fitted with a syringe outside the chamber, capable of delivering 0.2 mL of 20% sucrose solution (white sugar, Coles, Australia) diluted in water, each delivered to separate compartments of a recessed magazine inside the chamber.

The chambers also contained two retractable levers that could be inserted individually on the left and right sides of the magazine. Head entries into the magazine were detected via an infrared photobeam. Unless otherwise stated, the operant chambers were fully illuminated during all experimental stages by a 3W, 24V house light located on the upper edge of the wall opposite to the magazine. All training sessions were pre-programmed on a computer through the MED Associates software (Med-PC).

Fiber Photometry

Fiber photometry recordings were conducted using a dedicated fiber photometry processor and software (RZ5P Processor, Synapse, Tucker-Davis Technologies, TDT), which were used to control LEDs (465nm excitation and 405nm isosbestic) via LED drivers with a Fluorescence Mini Cube, measured with Newport photoreceivers, all from Doric Lenses. Patch cords used for recordings (400um, Doric) were photobleached for a minimum of 45mins at the start of each recording day. Light was measured at the tip of the patch prior to recording sessions, and maintained at 10-15uW (for 465) and 2-5uW (for 205). LEDs were modulated and demodulated via Synapse software at 331 Hz (465) and 211 Hz (405) and low-pass filtered (6Hz). MedPC signals were sent to RZ5P/Synapse, to indicate reward deliveries, magazine entries and lever presses, which were timestamped into the fiber photometry recordings.

Behavioral Protocol and Food Restriction

Food restriction. Rats underwent 4 days of food restriction prior to the onset of magazine training and this continued throughout the duration of the experiment. During this time, they received 5 g each of chow daily in addition to 1 g each of grain pellets and 2hr/daily exposures to sucrose solution. Rats were then given 10-20 g of chow daily from the fourth day until the end of the experiment. Their weight was monitored to ensure it remained above 85% of their pre-surgery body weight at all times.

Magazine training. Rats were given two days of magazine training prior to the onset of instrumental conditioning, which consisted of 15 deliveries each of pellets and sucrose into the magazine, delivered intermingled (but non-overlapping) at random intervals of 60 s.

Instrumental training. Rats were trained on an alternating lever paradigm, in which each lever came out for a total of 15 minutes or 10 outcomes, whichever came first, in alternating fashion, such that each lever was presented twice (total of 20 outcomes available per lever).

On Day 1 of instrumental training (Block 1), rats were trained to press both levers on a fixed interval 15-sec schedule (FI15), where reinforcements were separated by a fixed interval of minimum 15 seconds, but every response spaced by more than 15 seconds was reinforced. This continued until rats reached criterion of earning 40 outcomes in a single session.

Over the following 3 days (Days 2-4 of instrumental training, Block 2), rats were trained on an RI15 sec schedule (presses were reinforced on average every 15 secs), on the same alternating levers paradigm, with outcome criterion increased to 30 of each outcome in each session (15 outcomes per lever presentation). Over the following 6 days (Days 5-10, Blocks 3-4), rats were trained on the same paradigm on an RI30 sec schedule (presses reinforced on average every 30 seconds).

Contingency degradation. Over the following 10 days (5 x 2-day blocks), rats were trained on a contingency degradation schedule, in which rewards were earned on the same RI30 schedule with alternating levers, with one outcome (pellets or sucrose) additionally delivered non-contingently on an RI45 sec schedule, explicitly unpaired with instrumental responses or magazine entries (no free outcomes could be delivered within 1 sec of a lever press or magazine entry), delivered during both lever presentations, but not during the 1min interval between levers. This schedule was chosen to approximately match the number of free and earned outcomes of the same type (rats received 31 free outcomes on average at the start of training and 36 on average by the end of training).

Test. On the final day after contingency degradation, rats given a choice test under extinction (i.e. no outcomes were delivered) with both levers for 20 minutes. Rats were placed in the chambers, and after a 1 min interval, the house lights were illuminated and both levers came out for 20 mins, after which the session ended.

Histology and Immunofluorescence Staining

Within a week after the final contingency degradation test, the rats were perfused with 4% paraformaldehyde in phosphate buffer (PB), and the brains sliced coronally on a vibratome at 40 um. A minimum of 6 sections each from the pDMS (from +1.2mm to -0.6mm A/P from Bregma) were stained for dLight1.1 and DARP-32. Sections were rinsed 3 times for 10 min each in 0.1 M phosphate buffered saline (PBS), then submerged in a blocking solution of PBS with 0.5% Triton and 10% normal horse serum (NHS) for 2 hrs at room temperature. Sections were then submerged in 600 uL of rabbit anti-GFP (1:1000, Invitrogen) and mouse anti-DARP-32 (1:1000, BD Biosciences) diluted in PBS with 0.2% Triton and 2% NHS for 48 hrs at 4 degrees C. Sections were then rinsed 3 times for 10 min in PBS, before being submerged in donkey-anti-rabbit Alexa 488 (1:1000, Invitrogen) and donkey-anti-mouse Alexa 546 (1:1000, Invitrogen), diluted in PBS with 0.2% Triton and 2% NHS for 2 hrs at room temperature. Sections were washed twice with PBS and twice with PB for 10 min each, and mounted with Vectashield mounting medium with DAPI (Vector Laboratories).

Placement of dLight1.1 injections and fiber cannulas were imaged 24-48 hours later under a confocal microscope (BX16WI, Olympus) with a 10x objective using the boundaries defined by Paxinos & Watson (2016). Rats were excluded from the analysis if the placement of the fiberoptic cannula or virus extended anterior to +0.2mm from bregma in the DMS, or cannula placement was outside of the boundaries of the pDMS.

Fiber Photometry Analysis

Fiber photometry data were analyzed in custom MATLAB scripts. To achieve a motion-artifact corrected dLight signal, the demodulated dLight signal (465nm) and isosbestic control signal (405nm) were extracted from the recording software (Synapse), and the isosbestic signal was fitted to the dLight signal using a least-squares linear fit model for each recording session. Change in fluorescence ($\Delta F/F$) was calculated as (dLight signal – fitted isosbestic)/fitted isosbestic. Peri-event $\Delta F/F$ signals were extracted for left lever presses, right lever presses and magazine entries, across a 10s window (5s before and 5s after event onset) and downsampled to 15.8 samples per second (i.e. non-overlapping windows of 64 samples were averaged into a single value). The $\Delta F/F$ signal was normalized within these windows to the first 4s of each 10s window (baseline period), to give z-scores indicating the mean change from baseline ($\Delta F/F$ minus the mean baseline $\Delta F/F$ divided by the standard deviation of the baseline $\Delta F/F$).

For event-related analysis, the area under the curve (AUC; calculated via the trapezoidal method for integral calculation; trapz function in MATLAB) was calculated for each event across event windows: Action: 0.2s before to 0.2s after press; and Outcome: 0 to 1.5s after magazine entry. These AUCs were averaged to give mean AUCs for each event type.

To examine the relationship between reward prediction error (δ) and the temporal dynamics of the dopamine release after magazine entry, we employed a bootstrapping confidence intervals procedure (bCI; Jean-Richard-dit-Bressel et al., 2020) on dLight signals aligned to the magazine entry for high δ vs low δ CONTRA-Rew, CONTRA-Empty, IPSI-Rew and IPSI-Empty entries. This approach allows for the direct comparison of waveform profiles for two groups of signals (high δ vs low δ within each category). For a given event category, the population of Z-scored $\Delta F/F$ signals were randomly resampled with replacement for the same number of trials in that category and the mean waveform calculated. This was repeated 1000 times (i.e. 1000 bootstraps) to create a distribution of bootstrapped means. The differences between mean resampled waveforms (i.e. high δ - low δ) were calculated at each timepoint, and 95% CIs were then calculated for each timepoint in the series (taking the 2.5 and 97.5 centiles from the ranked bootstrap distribution, expanded by $\sqrt{(n/(n-1))}$ to correct for the narrowness bias). CIs that did not contain zero were identified, with at least two such consecutive timepoints constituting a significant difference (imposed to minimize the detection of spurious transients and corresponding to a consecutive threshold period of ~0.13s).

Reinforcement Learning

Reinforcement learning was modelled using a simplified temporal difference learning rule, in which the action value, V, was updated according to an error term (δ) generated during magazine entry according to $\delta(t)=\alpha[R(t)-V(t)]$ whereby α was a learning rate parameter and R was set at 1 for rewarded entries and 0 for nonrewarded entries. The action value of the next action, V(t+1) was updated according to the sum of the preceding action value and its error $V(t+1) = V(t) + \alpha.\delta(t)$. Action values were calculated separately for each instrumental action (IPSI and CONTRA), all were set at 0 for the very first session, and for all subsequent sessions V-values started at the final value of V for that action of the preceding session. The learning rate (α), was set at 0.3, however the pattern of data was preserved when learning rates were varied over a range of values (from 0.1-0.5) (**Fig 4C**).

There was a modification to this model that was made to separately assess the possibility of action values being updated on every press, which allowed for negative prediction errors in the absence of magazine entries, whereby δ was calculated to generate a negative term following all presses in which there wasn't a rewarded magazine entry, and V was updated on every press.

Signal data was compiled across the Action window according to the value of V at the time of press: 0-0.3 (low); 0.3-0.6(low-med); 0.6-0.8(med-high) and 0.8-1(high). The same was done for $\alpha.\delta$ during Outcome window: 0-.075(low); 0.075-0.15(low-med);0.15-0.23(med-high) and 0.23-0.3(high), with the maximum error term being equivalent to α.

Statistical Analysis

Instrumental response rates during contingency degradation were converted into baseline scores, in order to account for pre-existing differences in response rates on each lever. This was taken as the response rate on each lever represented as a percentage of the mean response rate on that same lever across the final block of instrumental training (i.e. Block 4, Days 8-10).

Behavioral data were analyzed according to planned contrasts comparing response rates on the degraded (or to-be-degraded) and non-degraded lever across instrumental training, contingency degradation training, and on test, controlling the per contrast error rate at α/k for k contrasts according to the Bonferroni correction. These contrasts were designed according to the hypothesis that instrumental responding would gradually reduce on the degraded lever across contingency degradation, and this difference would persist in the extinction choice test.

Fiber photometry AUCs were analyzed according to planned pairwise comparisons (t-tests) controlling the Type I error rate with α/k for k comparisons according to the Bonferroni correction. Across training, pairwise comparisons were planned to test the difference between IPSI and CONTRA within each press type within each training block, each press type to itself across select training blocks (i.e. IPSI-Ipsi Block 1 vs Block 4), and presses when they were rewarded vs when they were unrewarded within each training block (i.e. IPSI-Rew vs IPSI-Empty). For assessment of prediction-error modulation of dopamine, we planned comparisons to test the magnitude of dopamine within each press type at the highest and lowest V or δ values (i.e. IPSI-Rew V(0-.3) vs V(.8-1). For contingency degradation, comparisons were planned to test the difference between degraded and non-degraded presses (i.e. IPSI-Row Deg vs nonDeg). Significance was set at $p<0.05/k$.

References and Notes

Balleine, B.W., 2019. The Meaning of Behavior: Discriminating Reflex and Volition in the Brain. Neuron 104, 47–62. https://doi.org/10.1016/j.neuron.2019.09.024

Balleine, B.W., O'Doherty, J.P., 2010. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology 35, 48–69. https://doi.org/10.1038/npp.2009.131

Balleine, B.W., Peak, J., Matamales, M., Bertran-Gonzalez, J., Hart, G., 2021. The dorsomedial striatum: an optimal cellular environment for encoding and updating goal-directed learning. Current Opinion in Behavioral Sciences 41, 38–44. https://doi.org/10.1016/j.cobeha.2021.03.004

Bradfield, L.A., Bertran-Gonzalez, J., Chieng, B., Balleine, B.W., 2013. The thalamostriatal pathway and cholinergic control of goal-directed action: interlacing new with existing learning in the striatum. Neuron 79, 153–166. https://doi.org/10.1016/j.neuron.2013.04.039

Colwill, R.M., Rescorla, R.A., 1986. Associative Structures In Instrumental Learning, in: Psychology of Learning and Motivation. Elsevier, pp. 55–104.

da Silva, J.A., Tecuapetla, F., Paixão, V., Costa, R.M., 2018. Dopamine neuron activity before action initiation gates and invigorates future movements. Nature 554, 244–248. https://doi.org/10.1038/nature25457

Dickinson, A., Balleine, B., 1994. Motivational control of goal-directed action. Animal Learning & Behavior 22, 1–18. https://doi.org/10.3758/BF03199951

Dickinson, A., Charnock, D.J., 1985. Contingency Effects with Maintained Instrumental Reinforcement. The Quarterly Journal of Experimental Psychology Section B 37, 397–416. https://doi.org/10.1080/14640748508401177

Dolan, R.J., Dayan, P., 2013. Goals and Habits in the Brain. Neuron 80, 312–325. https://doi.org/10.1016/j.neuron.2013.09.007

Fisher, S.D., Ferguson, L.A., Bertran-Gonzalez, J., Balleine, B.W., 2020. Amygdala-Cortical Control of Striatal Plasticity Drives the Acquisition of Goal-Directed Action. Current Biology 30, 4541-4546.e5. https://doi.org/10.1016/j.cub.2020.08.090

Gerfen, C.R., Surmeier, D.J., 2011. Modulation of striatal projection systems by dopamine. Annu. Rev. Neurosci. 34, 441–466. https://doi.org/10.1146/annurev-neuro-061010-113641

Hamid, A.A., Frank, M.J., Moore, C.I., 2021. Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. Cell 184, 2733-2749.e16. https://doi.org/10.1016/j.cell.2021.03.046

Hart, G., Bradfield, L.A., Balleine, B.W., 2018a. Prefrontal Corticostriatal Disconnection Blocks the Acquisition of Goal-Directed Action. J. Neurosci. 38, 1311–1322. https://doi.org/10.1523/JNEUROSCI.2850-17.2017

Hart, G., Bradfield, L.A., Fok, S.Y., Chieng, B., Balleine, B.W., 2018b. The Bilateral Prefronto-striatal Pathway Is Necessary for Learning New Goal-Directed Actions. Curr. Biol. 28, 2218-2229.e7. https://doi.org/10.1016/j.cub.2018.05.028

Holly, E.N., Davatolhagh, M.F., Choi, K., Alabi, O.O., Vargas Cifuentes, L., Fuccillo, M.V., 2019. Striatal Low-Threshold Spiking Interneurons Regulate Goal-Directed Learning. Neuron. https://doi.org/10.1016/j.neuron.2019.04.016

Klaus, A., Martins, G.J., Paixao, V.B., Zhou, P., Paninski, L., Costa, R.M., 2017. The Spatiotemporal Organization of the Striatum Encodes Action Space. Neuron 95, 1171-1180.e7. https://doi.org/10.1016/j.neuron.2017.08.015

Lee, J., Wang, W., Sabatini, B.L., 2020. Anatomically segregated basal ganglia pathways allow parallel behavioral modulation. Nat Neurosci 23, 1388–1398. https://doi.org/10.1038/s41593-020-00712-5

Lee, R.S., Mattar, M.G., Parker, N.F., Witten, I.B., Daw, N.D., 2019. Reward prediction error does not explain movement selectivity in DMS-projecting dopamine neurons. Elife 8. https://doi.org/10.7554/eLife.42992

Matamales, M., McGovern, A.E., Mi, J.D., Mazzone, S.B., Balleine, B.W., Bertran-Gonzalez, J., 2020. Local D2- to D1-neuron transmodulation updates goal-directed learning in the striatum. Science 367, 549–555. https://doi.org/10.1126/science.aaz5751

Mohebi, A., Pettibone, J.R., Hamid, A.A., Wong, J.-M.T., Vinson, L.T., Patriarchi, T., Tian, L., Kennedy, R.T., Berke, J.D., 2019. Dissociable dopamine dynamics for learning and motivation. Nature 570, 65–70. https://doi.org/10.1038/s41586-019-1235-y

Niv, Y., Daw, N.D., Joel, D., Dayan, P., 2007. Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology 191, 507–520. https://doi.org/10.1007/s00213-006-0502-4

Nonomura, S., Nishizawa, K., Sakai, Y., Kawaguchi, Y., Kato, S., Uchigashima, M., Watanabe, M., Yamanaka, K., Enomoto, K., Chiken, S., Sano, H., Soma, S., Yoshida, J., Samejima, K., Ogawa, M., Kobayashi, K., Nambu, A., Isomura, Y., Kimura, M., 2018. Monitoring and Updating of Action Selection for Goal-Directed Behavior through the Striatal Direct and Indirect Pathways. Neuron 99, 1302-1314.e5. https://doi.org/10.1016/j.neuron.2018.08.002

Panigrahi, B., Martin, K.A., Li, Y., Graves, A.R., Vollmer, A., Olson, L., Mensh, B.D., Karpova, A.Y., Dudman, J.T., 2015. Dopamine Is Required for the Neural Representation and Control of Movement Vigor. Cell 162, 1418–1430. https://doi.org/10.1016/j.cell.2015.08.014

Park, J., Coddington, L.T., Dudman, J.T., 2020. Basal Ganglia Circuits for Action Specification. Annu. Rev. Neurosci. 43, 485–507. https://doi.org/10.1146/annurev-neuro-070918-050452

Parker, N.F., Cameron, C.M., Taliaferro, J.P., Lee, J., Choi, J.Y., Davidson, T.J., Daw, N.D., Witten, I.B., 2016. Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. Nat. Neurosci. 19, 845–854. https://doi.org/10.1038/nn.4287

Parkinson, J.A., Dalley, J.W., Cardinal, R.N., Bamford, A., Fehnert, B., Lachenal, G., Rudarakanchana, N., Halkerston, K.M., Robbins, T.W., Everitt, B.J., 2002. Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive Pavlovian approach behaviour: implications for mesoaccumbens dopamine function. Behavioural Brain Research 137, 149–163. https://doi.org/10.1016/S0166-4328(02)00291-7

Patriarchi, T., Cho, J.R., Merten, K., Howe, M.W., Marley, A., Xiong, W.-H., Folk, R.W., Broussard, G.J., Liang, R., Jang, M.J., Zhong, H., Dombeck, D., von Zastrow, M., Nimmerjahn, A., Gradinaru, V., Williams, J.T., Tian, L., 2018. Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. Science 360, eaat4422. https://doi.org/10.1126/science.aat4422

Peak, J., Chieng, B., Hart, G., Balleine, B.W., 2020. Striatal direct and indirect pathway neurons differentially control the encoding and updating of goal-directed learning. eLife 9, e58544. https://doi.org/10.7554/eLife.58544

Peters, A.J., Fabre, J.M.J., Steinmetz, N.A., Harris, K.D., Carandini, M., 2021. Striatal activity topographically reflects cortical activity. Nature 591, 420–425. https://doi.org/10.1038/s41586-020-03166-8

Rescorla, R.A., 1992. Response-independent outcome presentation can leave instrumental R-O associations intact. Animal Learning & Behavior 20, 104–111. https://doi.org/10.3758/BF03200407

Reynolds, J.N., Hyland, B.I., Wickens, J.R., 2001. A cellular mechanism of reward-related learning. Nature 413, 67–70. https://doi.org/10.1038/35092560

Santos, F.J., Oliveira, R.F., Jin, X., Costa, R.M., 2015. Corticostriatal dynamics encode the refinement of specific behavioral variability during skill learning. eLife 4, e09423. https://doi.org/10.7554/eLife.09423

Schultz, W., 2007. Multiple Dopamine Functions at Different Time Courses. Annu. Rev. Neurosci. 30, 259–288. https://doi.org/10.1146/annurev.neuro.28.061604.135722

Shiflett, M.W., Balleine, B.W., 2011. Molecular substrates of action control in cortico-striatal circuits. Prog. Neurobiol. 95, 1–13. https://doi.org/10.1016/j.pneurobio.2011.05.007

Sutton, R.S., Barto, A.G., 1998. Reinforcement learning: an introduction, Adaptive computation and machine learning. MIT Press, Cambridge, Mass.

Jean-Richard-dit-Bressel, P. Clifford, CWG, McNally, G., 2020. Analyzing event-related transients: confidence intervals, permutation tests, and consecutive thresholds. Front. Mol. Neurosci. Feb; 13:14.

Acknowledgments: The authors thank Amir Dezfouli for helpful comments on the manuscript, Myles Billard from Tucker-Davis Technologies for technical assistance with the fiber photometry recordings, and Karly Turner for assistance in the lab.

Funding: Funding for this research was provided by a Discovery Project Grant from the Australian Research Council: DP160105070 (BWB) and by a Senior Investigator Grant from the National Health and Medical Research Council of Australia # GNT1175420 (BWB)

Author contributions:

Conceptualization: GH, TB, BWB

Methodology: GH, TB, BWB

Investigation: GH, TB

Funding acquisition: BWB

Writing: GH, TB, BWB

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All source data are published in Figshare https://figshare.com/articles/dataset/Striatal_dopamine_encodes_the_relationship_between_actions_and_rewards/19083647. dLight1.1 was obtained under material transfer agreement with AddGene. All other materials for fiber photometry recordings were obtained from Tucker-Davis Technologies and Doric Lenses.