EJN European Journal of Neuroscience

FENS Federation of European Neuroscience Societies

*European Journal of Neuroscience, Vol. 42, pp. 2097–2104, 2015* doi:10.1111/ejn.12971

BEHAVIOURAL NEUROSCIENCE

# Plasticity in striatopallidal projection neurons mediates the acquisition of habitual actions

Qiang Shan,[1] MacDonald J. Christie[2] and Bernard W. Balleine[1]
[1]Behavioural Neuroscience Laboratory, The University of Sydney, 100 Mallet Street, Camperdown, NSW 2050, Australia
[2]Discipline of Pharmacology, The University of Sydney, Sydney, NSW, Australia

*Keywords*: D2-green fluorescent protein mouse, dopamine D2 receptors, endocannabinoid signaling, habits, instrumental conditioning, overtraining

## Abstract

In instrumental conditioning, newly acquired actions are generally goal-directed and are mediated by the relationship between the action and its consequences or outcome. With continued training, however, the performance of such actions can become automatic, reflexive or habitual and under the control of antecedent stimuli rather than their consequences. Recent evidence suggests that habit learning is mediated by plasticity in the dorsolateral striatum (DLS). To date, however, no direct evidence of learning-related plasticity associated with overtraining has been reported in this region, nor is it known whether, or which, specific cell types are involved in this learning process. The striatum is primarily composed of two classes of spiny projection neurons, the striatonigral and striatopallidal spiny projection neurons, which express dopamine D1 and D2 receptors, and control direct and indirect pathways, respectively. Here we found evidence of a post-synaptic depression in DLS striatopallidal projecting neurons in the indirect pathway during habit learning in mice. Moreover, this training-induced depression occluded post-synaptic depression induced by co-activation of D2 receptors and transient receptor potential vanilloid 1 (TRPV1) channels, implying that this pathway is involved in habit learning. This hypothesis was further tested by disrupting this signal pathway by knocking out TRPV1 channels, resulting in compromised habit learning. Our findings suggest that post-synaptic plasticity at D2 neurons in the DLS mediates habit learning and, by implicating an interaction between the D2 receptor and TRPV1 channel activity, provide a potential drug target for influencing habitual action control.

## Introduction

Maintaining an optimal balance between flexible, goal-directed actions and reflexive habits is critical to our ability adaptively to explore and to exploit the environment in the service of our basic needs and desires (Balleine & O'Doherty, 2010). Whereas goal-directed actions provide the necessary flexibility to solve novel problems, the repetition of successful strategies can engage a second learning process through which actions become habitual and subject to a more efficient performance process requiring fewer cognitive resources (Dickinson, 1985). Although adaptive, a variety of conditions, such as substance abuse (Nelson & Killcross, 2006; Corbit *et al.*, 2012, 2014), eating disorders, including obesity (Johnson & Kenny, 2010; Furlong *et al.*, 2014), anxiety (Alvares *et al.*, 2014), and major psychiatric disorders (Morris *et al.*, 2015), can induce a dysexecutive syndrome characterised by rapid habit learning and poor goal-directed control (Godefroy *et al.*, 2010). Understanding the neural mechanisms of habit learning may therefore help to develop therapeutic strategies against these conditions.

Considerable evidence suggests that the critical plasticity mediating goal-directed and habitual actions involves distinct regions of the dorsal striatum, specifically the dorsomedial striatum (or caudate nucleus) and dorsolateral striatum (DLS or putamen), respectively (Yin *et al.*, 2004, 2005, 2008; Balleine *et al.*, 2007; Tricomi *et al.*, 2009; Balleine & O'Doherty, 2010; Quinn *et al.*, 2013). Nevertheless, although specific changes in synaptic plasticity in the dorsomedial striatum underlying goal-directed action have recently been reported (Shan *et al.*, 2014), the cellular and molecular mechanisms underlying the acquisition of habitual actions have yet to be established.

As with other regions of the striatum, 95% of neurons in the DLS are medium spiny projection neurons (SPNs), which can be divided into two functionally distinct groups: the D1 dopamine receptor-expressing SPNs (D1R-SPNs) and the D2 dopamine receptor-expressing SPNs (D2R-SPNs) (Gerfen & Surmeier, 2011; Cerovic *et al.*, 2013; Calabresi *et al.*, 2014). The D1R-SPNs form a direct pathway that projects directly to the internal globus pallidus, whereas the D2R-SPNs form an indirect pathway that projects to the internal globus pallidus via the external globus pallidus and subthalamic nucleus (Gerfen & Surmeier, 2011). Both D1R-SPNs and D2R-SPNs can express long-term potentiation and long-term depression (LTD); however, the underlying molecular bases for these forms of plasticity are quite distinct between the two neuronal populations (Centonze *et al.*, 2001; Shen *et al.*, 2008; Lovinger, 2010).

*Correspondence*: Bernard Balleine, as above.
E-mail: bernard.balleine@sydney.edu.au

*Received 10 February 2015, revised 14 May 2015, accepted 29 May 2015*

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

Here, we assessed plasticity associated with habit learning in these two distinct populations of neurons by measuring the spontaneous excitatory post-synaptic current (sEPSC) in the DLS using *ex-vivo* patch-clamp electrophysiological recording. Mice that express the enhanced green fluorescent protein under the control of the promoter for the D2 dopamine receptor (D2-GFP mice) were first trained to lever press for food and then overtrained to establish habits. In *ex-vivo* recording, we found that the amplitude of sEPSCs was selectively reduced in D2R-SPNs in the DLS after habit learning, and that this depression was mediated by co-activation of D2 dopamine receptors (D2Rs) and transient receptor potential vanilloid 1 (TRPV1) channels. Furthermore, disrupting this signal pathway by knocking out TRPV1 channels resulted in compromised habit learning. We conclude that habitual learning is mediated by post-synaptic depression in the DLS induced by co-activation of D2Rs and TRPV1 channels.

## Materials and methods

### Animals

For electrophysiological recording, male D2-GFP mice (C57Bl/6J–Swiss Webster hybrid, 7–9 weeks old) (Gong *et al.*, 2003) were used. The D2-GFP transgene is hemizygous. TRPV1-knockout (KO) mice were used for investigating the role of the TRPV1 in habitual action. Mice were housed in a 12 h light/12 h dark cycle in a temperature-controlled (21 °C) and humidity-controlled (50%) environment. All experiments were conducted according to the ethical guidelines approved by the University of Sydney Animal Care and Ethics Committee.

### Instrumental conditioning

Operant chambers (Med Associates) were used for instrumental conditioning. The food supply was controlled to maintain the weight of each mouse at 80–90% of its *ad-libitum* level for the duration of instrumental conditioning. Two sessions of magazine-entry training were given, during each of which 30 grain food pellets (each 20 mg, Bioserve Biotechnologies, USA) were delivered on a random time 60 s schedule. There followed nine instrumental training sessions in which mice had to press a lever to receive a grain pellet. Each session lasted until either 50 grain pellets were delivered or 1 h had passed, whichever came first. In the first two sessions, the pellets were delivered on a continuous reinforcement (CRF) schedule, whereas in sessions 3–4, 5–6 and 7–9, they were delivered on a random interval (RI) 15 s (RI15), 30 s (RI30) and 60 s (RI60) schedule, respectively. Each trained mouse had a yoked control mouse treated in the same way except that, during instrumental conditioning, the delivery of the pellet was determined by the mouse to which it was yoked rather than the lever. The next day following the final RI60 training, the mice were either decapitated (after anesthesia) for electrophysiological recording, or the trained mice proceeded to a satiety-specific outcome devaluation test or an omission test.

In the devaluation test, half of the mice were given 1 h free access to the reward grain pellets (devalued condition), which they normally earned during the training sessions, whereas the remainder were given free access for the same duration to the non-reward purified pellets (valued condition), which have a similar nutritional value to, but a distinct physical property from, the reward grain pellets. The purified pellets were provided as a control for the effects of general satiety. The devaluation effect was tested in a 5 min probe test conducted immediately after the satiety treatment. No pellets were delivered during the test. Mice were then retrained on an RI60 schedule the next day in order to return their lever press rates to the basal level before a second devaluation test was conducted on the third day. In the second devaluation test, the satiety (valued or devalued) conditions were reversed. Data obtained in two devaluation tests were pooled for analysis.

In the omission test, mice were exposed to the same training chamber for 30 min each day for 4 days. Reward grain pellets were delivered on an RI20 schedule only if no lever press occurred. This is a reversal of the original action–outcome contingency. Lever press rates over each 5 min interval were recorded for analysis.

### Electrophysiological recording

At 1 day after the last instrumental conditioning session, mice were killed for *ex-vivo* striatal slice recording. Mice were anesthetised using a ketamine and xylazine cocktail (210 and 14 mg/kg body weight, i.p., respectively) and decapitated. Parahorizontal striatal slices including surrounding cortical regions (300 μm thickness) were cut on a vibratome in an ice-cold dissection solution containing (in mM): 2.5 KCl, 0.5 $CaCl_2$, 7 $MgCl_2$, 1.2 $NaH_2PO_4$, 26 $NaHCO_3$, 10 glucose and 200 sucrose (saturated with 95% $O_2$/5% $CO_2$, osmolarity 295–305 mOsm). Slices were recovered for at least 1 h in an artificial cerebrospinal fluid (ACSF) solution containing (in mM): 126 NaCl, 2.5 KCl, 2 $CaCl_2$, 1.2 $MgCl_2$, 1.2 $NaH_2PO_4$, 26 $NaHCO_3$, and 10 glucose (saturated with 95% $O_2$/5% $CO_2$, osmolarity 310–320 mOsm).

During recording, slices were perfused continuously with ACSF (1–2 mL/min) at 31–32 °C. Picrotoxin (100 μM) was added to the ACSF solution to suppress any GABAergic inhibitory response. Whole-cell voltage-clamp recordings were carried out on the D1R-SPNs and D2R-SPNs. SPNs were identified by their medium size and lack of spontaneous firing. D1R-SPNs and D2R-SPNs were distinguished from each other by the absence or presence of GFP fluorescence, respectively. Recording pipettes (2.5–3.5 MΩ) were filled with an internal solution containing (in mM): 120 $CsMeSO_3$, 15 CsCl, 8 NaCl, 10 HEPES, 0.4 EGTA, 3 QX-314, 2 $Mg_2ATP$ and 0.33 $Na_3GTP$ (pH 7.3 and osmolarity 280–290 mOsm). Neurons were voltage-clamped at −70 mV. The sEPSCs were acquired for at least 3 min and analysed using AxoGraph X (AxoGraph).

In the pharmacological experiments, slices were perfused with ACSF containing the D2R agonist quinpirole (10 μM) and the TRPV1 agonist capsaicin (3 μM), alone or combined, for 15 min. After treatment, slices were washed in ACSF for at least 15 min before recording.

## Results

### *The amplitude of the spontaneous excitatory post-synaptic currents of the D2 dopamine receptor-expressing spiny projection neurons is reduced by habit learning*

To investigate habitual action-related plasticity in the DLS, we trained mice to acquire an action (lever press)–outcome (food pellets) association when food deprived (Fig. 1A). To achieve this, we employed extended training on interval schedules of reinforcement, a protocol known to result in the formation of habitual actions (cf. Figs 1 and 4B). In order to distinguish D1R-SPNs and D2R-SPNs, we used a transgenic mouse line that expresses GFP in the D2R-SPNs (D2-GFP mice) (Gong *et al.*, 2003). As the large majority of the GFP-negative cells are D1R-SPNs, we recorded from both the

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

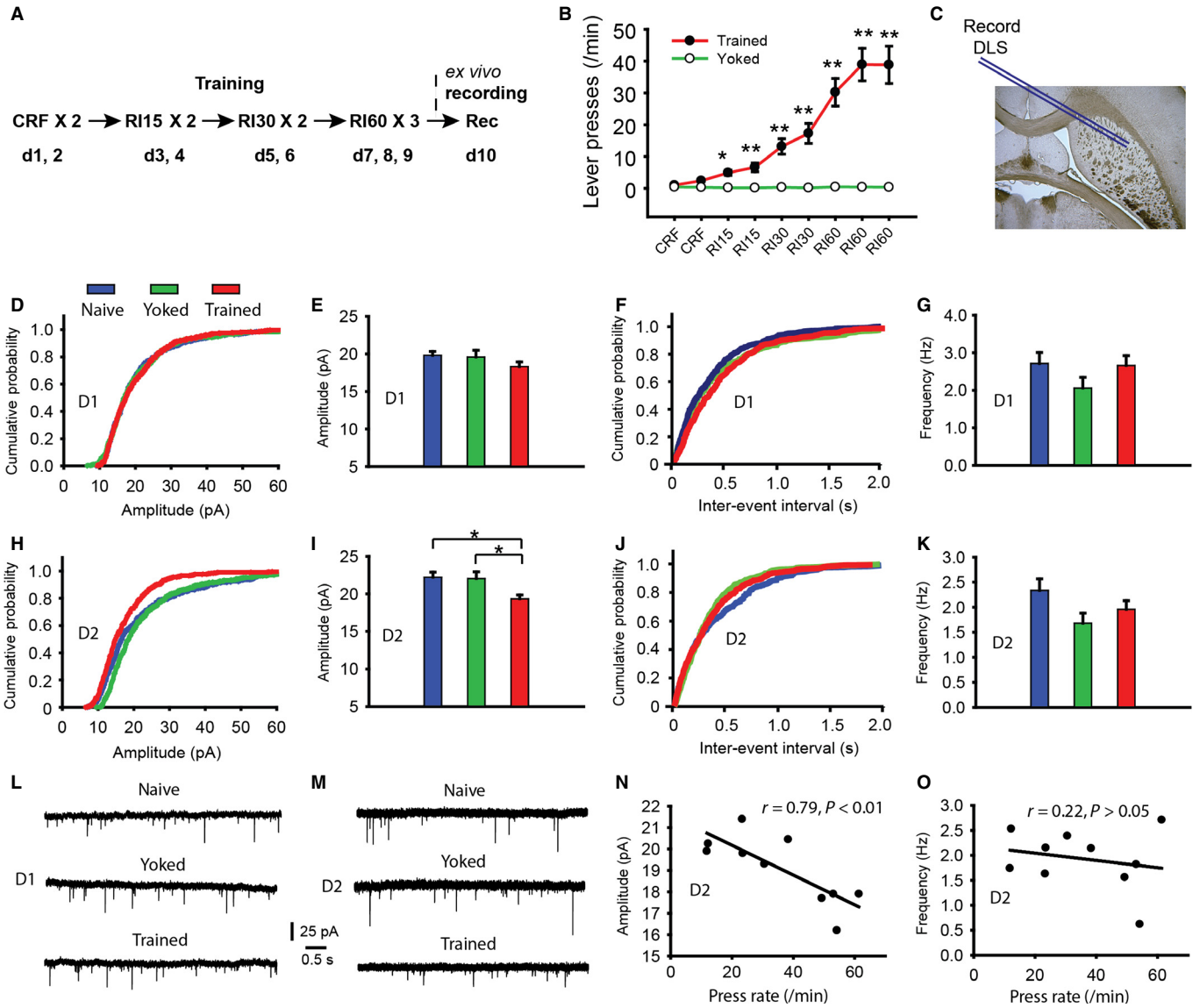


FIG. 1. *Ex-vivo* electrophysiological assessment of synaptic response in D1R-SPNs and D2R-SPNs in the DLS after habitual training. (A and B) D2-GFP mice were given prolonged habitual training sessions in which reward food pellet delivery was either paired with lever pressing (trained) or unpaired with lever pressing (yoked). The rate of lever presses in the trained mice is significantly higher than in the yoked mice (two-way repeated-measures ANOVA, $F_{1,288} = 128$, $P < 0.01$, post-hoc test, $^*P < 0.05$, $^{**}P < 0.01$, $n = 19$ and $n = 19$ for trained and yoked mice, respectively). (C) Parahorizontal striatal slices were prepared, and sEPSCs of either GFP-positive or GFP-negative neurons in the DLS were recorded. (D–G and L) In the GFP-negative neurons (D1R-SPNs), no significant difference is found between the naive, yoked and trained mice for either sEPSC amplitude or frequency (ANOVA, $F_{2,44} = 1.4$, $P > 0.05$, $F_{2,44} = 1.7$, $P > 0.05$, $n = 15$–$16$). (H–K and M) In contrast, in the GFP-positive neurons (D2R-SPNs), the sEPSC amplitude in the trained mice is significantly lower than that in either the naive or yoked mice (ANOVA, $F_{2,42} = 4.6$, $P < 0.05$, post-hoc test, $^*P < 0.05$, $n = 14$–$16$), and the sEPSC frequency demonstrates no difference between the naive, yoked and trained mice (ANOVA, $F_{2,42} = 2.7$, $P > 0.05$). (N-O) Moreover, in the trained mice, the amplitude but not the frequency of the sEPSCs of the D2R-SPNs is negatively correlated with the lever press rates of the last RI60 training session. CRF, continuous reinforcement.

GFP-negative and GFP-positive neurons to sample both the D1R-SPNs and D2R-SPNs, respectively.

Two groups of mice were established, one trained with action and outcome delivery paired, and a second serving as a control group given yoked exposure to the reward outcome and for which lever pressing and reward delivery were unpaired (Fig. 1B). As expected, lever pressing in the trained group increased relative to the yoked group over 9 days (Fig. 1B; magazine entry data are presented in Supporting Information Fig. S1). To measure the synaptic changes after habit training, striatal slices were prepared at 1 day after the final training session and the sEPSCs of the D1R-SPNs (GFP-negative) and D2R-SPNs (GFP-positive) in the DLS region were recorded (Fig. 1C). In the D1R-SPNs, both the amplitude and frequency of the sEPSCs were similar between the yoked and trained mice (Fig. 1D–G and I), indicating that no changes in synaptic plasticity occurred in the direct pathway as a consequence of habit training. In contrast, in the D2R-SPNs, the amplitude of the sEPSCs was lower in the trained mice than in the yoked mice, whereas the frequency was similar between the two groups (Fig. 1H–K and M). Moreover, we also found that the amplitude but not the frequency of sEPSCs in the D2R-SPNs was negatively correlated to the lever press rate in the final training session (Fig. 1N and O), confirming a relationship between habitual lever press performance and the reduced amplitude of sEPSCs in the indirect pathway. As changes

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

in the amplitude of the sEPSCs generally reflect a post-synaptic mechanism, we conclude that a post-synaptic LTD-like plasticity was formed in the indirect pathway of the DLS after habitual training.

In order to eliminate the possibility that non-specific behavioral factors contributed to the synaptic plasticity observed, slices from a third group of naive mice without any behavioral manipulation were also recorded. No significant differences were found between the naive and yoked mice for any of the measurements described above (Fig. 1D–M), verifying that post-synaptic LTD-like plasticity found in the indirect pathway of the DLS reflects learning.

It should be noted that there is a possibility that this synaptic change resulted from goal-directed rather than habitual action, because any animal given overtraining to form habitual actions must have undergone a goal-directed phase in the early stage of training, and the synaptic changes found in the D2R-SPNs of the DLS might therefore have been a residual trace of this earlier stage of learning. However, we previously found no plasticity in either the D1R-SPNs or D2R-SPNs of the DLS after goal-directed action training (Shan *et al.*, 2014). Therefore, we conclude that the depression found in the D2R-SPNs of the DLS was specific to habit learning.

Post-synaptic changes in α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) currents are sometimes accompanied by a replacement of GluR2 subunits (Conrad *et al.*, 2008), which are likely to be reflected in the changes of decay times of AMPA currents (Thiagarajan *et al.*, 2005). However, measuring the decay times of naive, yoked and trained mice found no significant difference between groups (2.815 ± 0.123, 2.930 ± 0.171 and 2.895 ± 0.111 ms, respectively; ANOVA, $F_{2,42} < 1$, $P > 0.05$; $n = 14$–16). This implied that the subunit composition of AMPA receptors stayed constant during training. Similarly, input resistance did not differ between groups. The mean input resistance for each group was: 88.14 ± 7.44, 101.56 ± 7.44 and 111.57 ± 10.11 MΩ in D2 neurons of the DLS for naive, yoked and trained mice, respectively. ANOVA conducted on these data found no significant between-group effect ($F_{2,42} = 1.89$, $P > 0.05$, $n = 14$–16).

### *The depression induced by the co-activation of the D2 dopamine receptor and the transient receptor potential vanilloid 1 is occluded in the habitual mice*

We next investigated the molecular basis for the post-synaptic LTD-like plasticity in the indirect pathway of the DLS after habit training. LTD-like plasticity in the indirect pathway has been widely reported previously; however, it is mediated by a pre-synaptic mechanism in most cases (Kreitzer & Malenka, 2008). Nevertheless, Grueter *et al.* (2010) reported that the application of the TRPV1 agonist capsaicin induced a post-synaptic LTD in the indirect pathway, but not in the direct pathway, of the ventral striatum. Considering the similarities of the structure and function between the dorsal and ventral striata, we hypothesised that TRPV1-mediated LTD may also be induced in the indirect pathway of the dorsal striatum and mediates habits (Hopf *et al.*, 2010). We tested this hypothesis by pre-treating brain slices from naive mice with 3 μM capsaicin, but, surprisingly, did not find any depression in the D2R-SPNs of the DLS (Fig. 2B–E). However, pre-treating the slices with the D2R agonist quinpirole (10 μM) together with 3 μM capsaicin reliably suppressed the amplitude but not the frequency of the sEPSCs of the D2R-SPNs (Fig. 2A–E), implying a post-synaptic LTD-like plasticity as found in the ventral striatum. This depression was only induced by co-activation of D2Rs and TRPV1s; pre-treatment with the D2R agonist alone did not cause any effect (Fig. 2B–E).

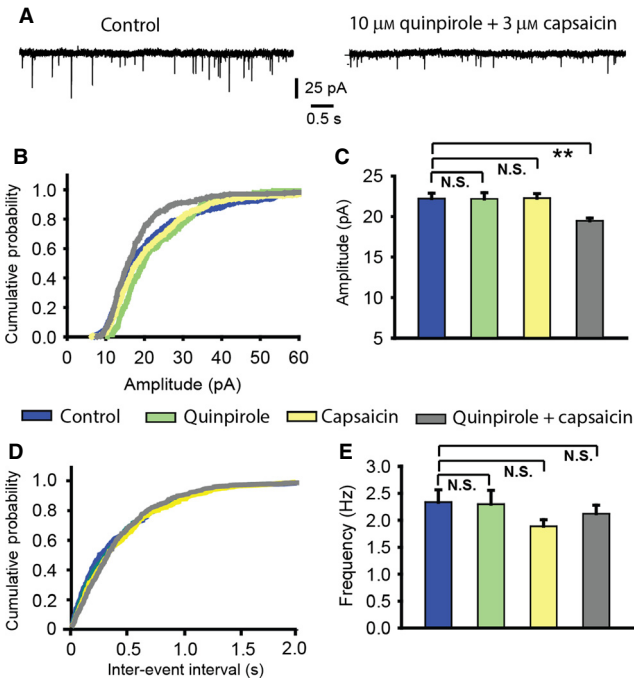


FIG. 2. Co-activation of the D2R and TRPV1 suppresses post-synaptic response in the D2R-SPNs. (A–E) Pre-treatment of striatal slices with the D2R agonist quinpirole (10 μM) and the TRPV1 agonist capsaicin (3 μM) suppresses the amplitude, but not the frequency of sEPSCs of the D2R-SPNs. However, pre-treatment with quinpirole or capsaicin alone has no significant effect (amplitude: ANOVA, $F_{3,54} = 6.0$, $P < 0.05$, post-hoc test, $^{**}P < 0.01$, N.S., not significant, $n = 10$–17; frequency: ANOVA, $F_{3,54} = 1.3$, $P > 0.05$, post-hoc test, N.S., not significant, $n = 10$–17).

As both the co-activation of D2Rs and TRPV1s and habit training induce similar changes in the synaptic response of D2R-SPNs in the DLS, i.e. a post-synaptic LTD-like plasticity, we hypothesised that habit training also induces post-synaptic LTD in the D2R-SPNs of the DLS via co-activation of D2Rs and TRPV1s. If this is true, then habit training might occlude any subsequent attempt to induce LTD in D2R-SPNs of the DLS by the co-activation of the D2R and TRPV1. Our experiments found clear evidence for this prediction. Co-application of the D2R agonist quinpirole and the TRPV1 agonist capsaicin after habit training did not induce any change in the amplitude of the sEPSCs of the D2R-SPNs of the DLS, in stark contrast to the effects in naive mice described above (Fig. 3F and G). This effect appeared to be specific to habit training, and was not induced by non-specific behavioral exposure; the same drug application reliably reduced the sEPSC amplitude in yoked mice (Fig. 3B and C), which received behavioral treatment similar to those given habit training except for the lack of contingency between action and outcome. As expected, the co-application of the D2R agonist quinpirole and the TRPV1 agonist capsaicin did not change the frequency of the sEPSCs of the D2R-SPNs in the DLS in either trained or yoked mice (Fig. 3D, E, H and I).

### *Knockout of transient receptor potential vanilloid 1 compromises habitual action*

We have demonstrated that habitual training induces an LTD-like plasticity via the co-activation of the D2R and TRPV1. We next tested the necessity of this process in habit learning by disrupting this signal pathway using TRPV1-KO mice. Both TRPV1-KO and TRPV1-wild-type (WT) mice were trained as described previously

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

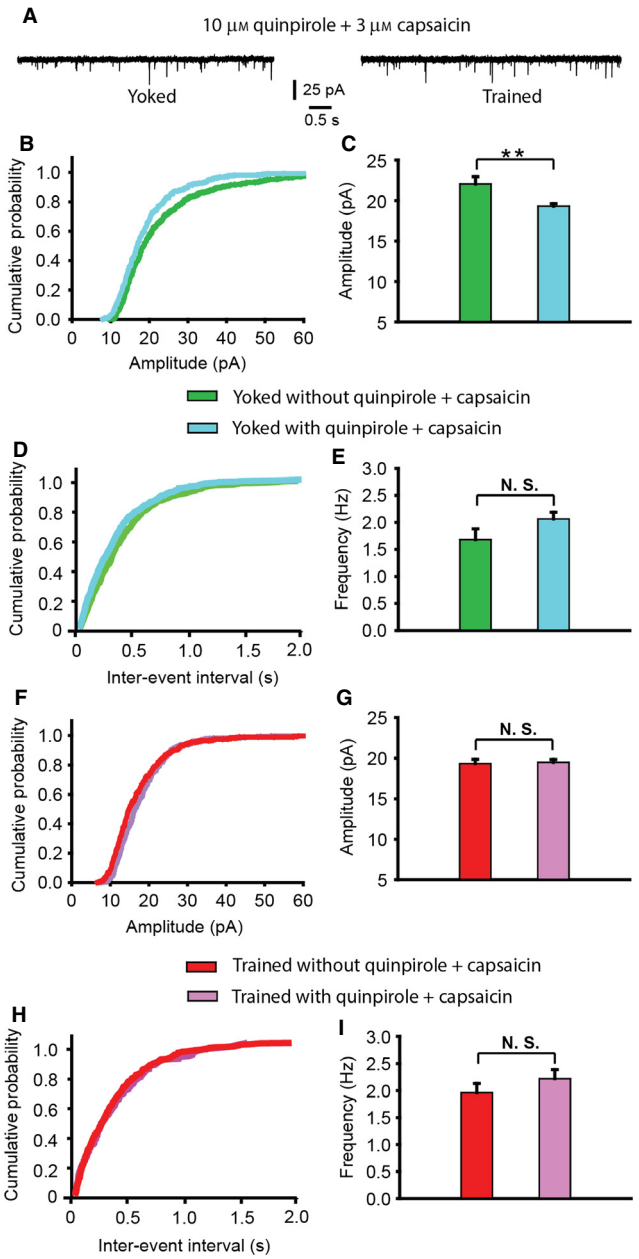


FIG. 3. The post-synaptic depression induced by the co-activation of the D2R and TRPV1 in the D2R-SPNs is occluded in the habitual mice. Mice received either habitual training or yoked treatment, and their striatal slices were pre-treated with the D2R agonist quinpirole (10 μM) and the TRPV1 agonist capsaicin (3 μM) before recording. (B–E) In the yoked mice, as in the naive mice, pre-treatment suppresses the amplitude, but not the frequency, of sEPSCs of the D2R-SPNs (amplitude: ANOVA, $F_{1,38} = 11.0$, $P < 0.01$, post-hoc test, $^{**}P < 0.01$, $n = 16–24$; frequency: ANOVA, $F_{1,38} = 3.1$, $P > 0.05$, post-hoc test, N.S. not significant, $n = 16–24$). (A and F–I) In contrast, in the trained mice, no suppression is found in either the amplitude or the frequency of sEPSCs of the D2R-SPNs (amplitude: ANOVA, $F_{1,50} < 1$, $P > 0.05$, post-hoc test, N.S., not significant, $n = 14–38$; frequency: ANOVA, $F_{1,50} < 1$, $P > 0.05$, post-hoc test, N.S., not significant, $n = 14–38$).

(Fig. 1A; magazine data for this study are presented in Supporting Information Fig. S2) and their degree of habitual control was then assessed. Compared with goal-directed actions, habitual actions tend to be less sensitive to changes in either the value of the outcome associated with an action or the causal relationship between action and outcome. To assess these two aspects of habitual control in this experiment, we conducted a satiety-specific outcome devaluation test and an omission test, respectively.

In the devaluation test, despite overtraining sufficient to produce habitual control in the TRPV1-WT mice (Fig. 4B), the TRPV1-KO mice demonstrated reliable sensitivity to outcome devaluation (Fig. 4B), implying that (i) their actions were goal-directed and (ii) they had failed to acquire a habit.

In contrast to the simple pattern exhibited in the devaluation test, the pattern of results in the omission test was complicated by the fact that the TRPV1-KO mice tended to press the lever more frequently than the TRPV1-WT mice, which was reflected in their RI60 training sessions and their first omission test session (Fig. 4A and C). Because pellet delivery depended on withholding the lever press action in the omission test, the tendency to press the lever at a higher rate resulted in the TRPV1-KO mice earning fewer food pellets in the first omission training session than the TRPV1-WT mice (Fig. 4D). Nevertheless, the lever press rate of the TRPV1-KO mice declined at a faster rate than that of the TRPV1-WT mice and reached a lower point of performance than that of the TRPV1-WT mice between 20 and 25 min of training in the second omission test session. This trend was sustained in the third and fourth test sessions. As a consequence, the TRPV1-KO mice earned more food pellets than the TRPV1-WT mice in the second, third and, especially, the fourth omission test session, in which the difference was statistically significant and the relationship between the two groups of mice was significantly reversed from that of the first omission training session (Fig. 4D). This implies that, in spite of a higher basal lever press rate, the TRPV1-KO mice were more sensitive to the reversal of the original action–outcome contingency, i.e. the association between lever press and food pellet delivery, than the TRPV1-WT mice. Indeed, normalising the lever press rates of mice over the omission test sessions against the respective lever press rates of the last RI60 training session revealed that, during the omission tests, the TRPV1-KO mice pressed the lever less frequently than the TRPV1-WT mice (Fig. 4E), confirming that the TRPV1-KO mice reversed the original action–outcome contingency faster than the TRPV1-WT mice. We conclude that the original lever press acquisition in the TRPV1-KO mice resulted in much weaker habit learning than in the TRPV1-WT mice and, as a consequence, they were more sensitive to any treatment that reversed the action–outcome contingency than the TRPV1-WT mice.

## Discussion

We found that post-synaptic depression is formed in the indirect pathway of the DLS associated with habit learning. We also found that this depression was mediated by the co-activation of D2Rs and the TRPV1 channel, and that disrupting this signal pathway compromised the acquisition of habit learning. SPNs receive excitatory glutamatergic inputs from the cortex and thalamus, and modulatory dopaminergic inputs from the midbrain (Gerfen & Surmeier, 2011; Cerovic *et al.*, 2013; Calabresi *et al.*, 2014). The depression of sEPSC amplitude without a change in frequency suggests that the mechanism is post-synaptic. This could potentially be confirmed by examining the paired-pulse ratios of electrically-evoked synaptic currents but our preliminary experiments established a very large variability of ratios in D2 SPNs (−40% to +45%) precluding this approach. It is uncertain whether changes in the composition of AMPA receptor subunits or another mechanism contributes to the depression but our finding of unchanged decay time constants of sEPSCs suggests that such changes are minor (Thiagarajan *et al.*,

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

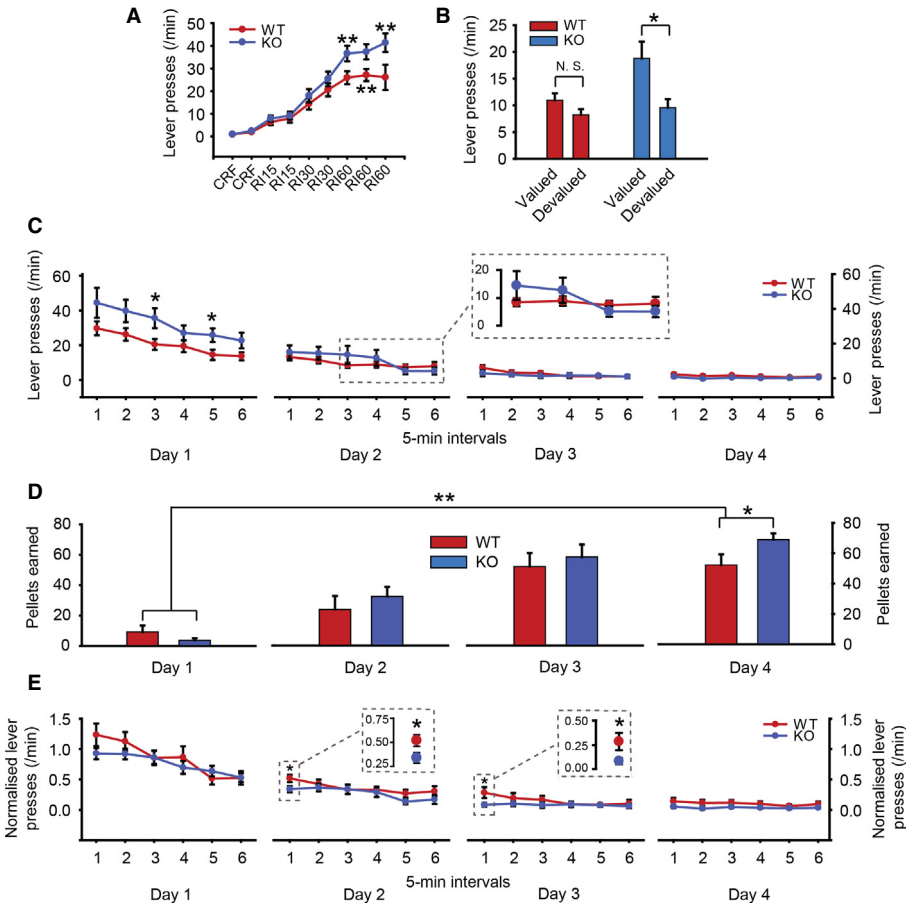


FIG. 4. KO of the TRPV1 compromises habitual action. TRPV1-KO mice and the TRPV1-WT control mice were trained to press a freely available lever to earn food pellets across nine sessions of training using increasing RI schedules, following which one of two tests were administered: a satiety-specific outcome-devaluation test or an omission test. (A) In the initial training, the TRPV1-KO mice pressed the lever significantly more frequently than the TRPV1-WT mice, especially in the three RI60 training sessions (two-way repeated-measures ANOVA, $F_{1,320} = 4.2$, $P < 0.05$, post-hoc test, $^*P < 0.05$; $^{**}P < 0.01$, $n = 21$ and $n = 21$ for trained and yoked mice, respectively). (B) In the subsequent outcome-devaluation test, which compares performance after satiation on the reward pellets vs. satiation on the non-reward pellets, the TRPV1-WT mice did not differentiate reward (devalued group) and non-reward pellets (valued group) (ANOVA, $F_{1,20} = 3.4$, $P > 0.05$, post-hoc test, N.S., not significant, $n = 11$), implying a habitual property of their action. In contrast, the TRPV1-KO mice demonstrated a lower response towards reward pellets than non-reward pellets (ANOVA, $F_{1,18} = 6.7$, $P < 0.05$, post-hoc test, $^*P < 0.05$, $n = 10$), indicating that they are still sensitive to outcome devaluation and their action is goal-directed. (C) In the omission test, which is independent of the outcome-devaluation test, mice were exposed to a reversed action–outcome contingency, i.e. the association between non-lever press and food pellet delivery. Initially, the TRPV1-KO mice pressed levers much more frequently than the TRPV1-WT mice (Day 1, ANOVA for individual 5 min intervals, $^*P < 0.05$, $n = 9$ and $n = 9$ for TRPV1-WT and TRPV1-KO mice, respectively). However, this difference became increasingly smaller in the following tests until the fifth 5 min interval of Day 2, in which the TRPV1-KO mice pressed levers less frequently than the TRPV1-WT mice (inset), and this trend remained in the following tests (Days 2–4). (D) Although this difference in lever press rates is not statistically significant, the amount of reward pellets that TRPV1-KO mice earned on Days 2–4 is higher than that of the TRPV1-WT mice (Day 4, ANOVA, $^*P < 0.05$). This is a significant reversal of what occurred on Day 1, when the TRPV1-KO mice earned fewer reward pellets than the TRPV1-WT mice (Genotype × Day interaction, two-way repeated-measures ANOVA, $F_{1,20}$=8.3, $^{**}P < 0.01$). (E) In a more direct demonstration, normalising the lever press rates of mice over the omission test sessions against the respective lever press rates of the last RI60 training session revealed that the TRPV1-KO mice tend to press levers less frequently than the TRPV1-WT mice (insets, ANOVA for individual 5 min intervals, $^*P < 0.05$). CRF, continuous reinforcement.

2005). To confirm this, measurement of surface membrane expression of AMPA receptor subtypes would be required in future experiments.

The D2Rs are readily activated by dopamine released from midbrain inputs. Our results suggest that the interaction of dopamine and D2R is important for habit learning. Indeed, manipulations that suppress the plasticity of the dopaminergic neuron, such as lesion (Faure *et al.*, 2005) or cell type-specific KO of *N*-methyl-D-aspartate receptors (Wang *et al.*, 2011), have been shown to inhibit habit learning. However, the D2R has also been implicated in playing a positive role in habit formation. For example, in humans, increased D2R density is associated with deficits in action flexibility and enhanced action automaticity (Stelzel *et al.*, 2010). In mice, Yin *et al.* (2009) reported that a D2 antagonist blocked the performance of motor skill learning at a late training stage but not at an early training stage. In this study, skill learning early and late in training was compared and found to involve the dorsomedial striatum and DLS, respectively, and appeared generally to parallel goal-directed and habitual action.

In contrast to D2R activation, how TRPV1 is activated *in vivo* is less certain. It was previously proposed that one of the endocannabinoids, anandamide, might serve as an endogenous agonist of the TRPV1 (Zygmunt *et al.*, 1999; Di Marzo *et al.*, 2002; Ross, 2003). In D2R-SPNs, anandamide is readily synthesised by the activation of group I metabotropic glutamate receptors and the elevation of intracellular $Ca^{2+}$ generated by L-type voltage-gated calcium channels and intracellular $Ca^{2+}$ stores (Lerner & Kreitzer, 2012). Both processes require glutamate, because L-type voltage-gated calcium channels and intracellular $Ca^{2+}$ stores are gated by post-synaptic

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

depolarization, which is usually induced by AMPA receptor activation by glutamate. Therefore, the activation of TRPV1 might originate from pre-synaptic glutamate release. In line with this suggestion, a previous study reported that the infusion of glutamate into the dorsal striatum accelerated stimulus–response habit learning (Packard, 1999).

Taken together, we hypothesise that D2R and TRPV1 co-activation during habit learning originates from the release of glutamate and dopamine in the DLS and that this co-activation leads to $Ca^{2+}$ entry through the TRPV1 channels and cAMP reduction through the D2R-coupled signal pathways, resulting in the retraction of AMPA receptors from the post-synaptic membrane. Although we cannot be sure about the source of glutamate release, we hypothesise that it stems from cortical afferents, particularly those from sensorimotor cortices, which have often been proposed to contribute to habit learning in the past (Graybiel, 2008). It is possible, however, that afferents from other regions of the cortex also contribute to the acquisition of habits, particularly the rodent medial agranular cortex, which is thought to serve as the rodent supplementary motor area (Reep *et al.*, 1987; Van Eden *et al.*, 1992). We postulate that this region could mediate the hierarchical control of goal-directed and habitual actions and, if so, it should be expected to be involved in integrating plasticity in the DLS into that control process (Dezfouli & Balleine, 2012). To confirm the involvement of these distinct cortical afferents in the plasticity observed here will, however, require the stimulation of those afferents directly and examination of changes in evoked activity at D2, relative to D1, neurons in DLS in future studies.

Other signaling molecules acting on the indirect pathway by mechanisms distinct from TRPV1 have been implicated in habit learning. Yu *et al.* (2009) reported that adenosine A2A receptor KO in the striatum compromised habit learning, similar to that found in the TRPV1-KO mice in the current study. Intriguingly, however, unlike TRPV1 channels, whose activation induces a post-synaptic depression, activation of A2A receptors produced a post-synaptic potentiation (Shen *et al.*, 2008). In addition, Hilario *et al.* (2007) reported that systemic administration of a cannabinoid CB1 receptor antagonist or CB1 receptor KO also impaired habit learning. However, and again in contrast to TRPV1, CB1 receptor activation has been shown to generate a pre-synaptic depression in the indirect pathway (Cerovic *et al.*, 2013). It seems difficult to reconcile these signal molecules with TRPV1 in their contributions to habit formation. Nevertheless, it is worth noting that, in these latter two studies, no electrophysiological recordings were performed on the indirect pathway of the mice and therefore it remains to be confirmed whether post-synaptic potentiation mediated by A2A receptor activation or pre-synaptic depression mediated by CB1 receptor activation occurs in the indirect pathway of the DLS in habit trained mice. In contrast, our data suggest that post-synaptic depression in the indirect pathway of the DLS is associated with habit learning and this behaviorally-induced synaptic depression occluded further pharmacological depression triggered by co-activation of D2R and TRPV1. It is also worth noting that our data do not eliminate the possibility that the A2A and CB1 receptors contribute to habit formation; we recorded the synaptic response at 1 day after the final training session, which therefore reflects a relatively long-term change in plasticity. In contrast, A2A and CB1 receptors could affect habit formation by regulating basal neurotransmission and short-term plasticity in the DLS.

More broadly, rather than suggesting that plasticity selects and drives the acquisition and performance of specific habitual actions in the striatum through an increase in excitation in the direct pathway, the finding that habits are associated with reduced activity in the striatopallidal indirect pathway implicates a reduction in the inhibition of a subset of specific movements, relative to others, in the acquisition of habits. It has recently been suggested that, whereas D1 neurons control the selection of specific actions, D2 neurons work simultaneously to sharpen that selection by inhibiting competing or extraneous responses (Mink, 2003; Cui *et al.*, 2013). Such a mechanism could result in the acquisition and performance of habitual actions if an initially varied movement form is generally inhibited by D2 activation in the DLS and if subsequent plasticity reduces the activity of only a subset of D2 neurons to allow the consistent disinhibition of a set of specific movements (Tang *et al.*, 2007). Assessing what may be called an 'off-center on-surround' theory of D2 function in habit acquisition will require the ability to record from a large and dispersed set of D2 neurons in the DLS or at least obtain measures of neural activity in those neurons, which, although difficult, will have the merit of providing a definitive test of this hypothesis.

## Supporting Information

Additional supporting information can be found in the online version of this article:

Fig. S1. (A) D2-GFP mice were given prolonged habitual training sessions in which reward food pellets delivery was either paired with lever pressing (trained) or unpaired with lever pressing (yoked). The rate of magazine entries in the trained mice is not significantly different from the yoked mice (Two-way repeated measures ANOVA, $F(1,144) < 1$, $P > 0.05$; $n = 10$ each). Note only mice on which recordings in D2 neurons were performed, were analysed. (B, C) Moreover, unlike the lever press rate (shown in Fig. 1N), the magazine entry rate of the last RI60 training session, is not significantly correlated with the amplitude (or the frequency) of sEPSCs of the D2R-SPNs in the trained mice.

Fig. S2. (A) TRPV1-knockout (TRPV1-KO) mice and their wild-type (TRPV1-WT) control mice were trained to press a freely available lever to earn food pellets across nine sessions of training using increasing random interval schedules. The magazine entry rate of the trained mice is not significantly different from that of the yoked mice (Two-way repeated measures ANOVA, $F(1,160) < 1$, $P > 0.05$; $n = 11$ each).

## Acknowledgements

We thank Dr S. Brierley for kindly sharing TRPV1-KO mice. This research was supported by funding from the Australian Research Council (grant no. DP110105636) and both a Laureate Fellowship from the Australian Research Council to B.W.B. and NHMRC Senior Principal Research Fellowships to both B.W.B. and M.J.C. The authors declare no conflict of interest.

## Abbreviations

ACSF, artificial cerebrospinal fluid; AMPA, α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid; D1R-SPN, D1 dopamine receptor-expressing spiny projection neuron; D2R, D2 dopamine receptor; D2R-SPN, D2 dopamine receptor-expressing spiny projection neuron; DLS, dorsolateral striatum; GFP, green fluorescent protein; KO, knockout; LTD, long-term depression; RI, random interval; sEPSC, spontaneous excitatory post-synaptic current; SPN, spiny projection neuron; TRPV1, transient receptor potential vanilloid 1; WT, wild-type.

## References

Alvares, G.A., Balleine, B.W. & Guastella, A.J. (2014) Impairments in goal-directed actions predict treatment response to cognitive-behavioral therapy in social anxiety disorder. *PLoS One*, **9**, e94778.

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

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.
Balleine, B.W., Delgado, M.R. & Hikosaka, O. (2007) The role of the dorsal striatum in reward and decision-making. *J. Neurosci.*, **27**, 8161–8165.
Calabresi, P., Picconi, B., Tozzi, A., Ghiglieri, V. & Di Filippo, M. (2014) Direct and indirect pathways of basal ganglia: a critical reappraisal. *Nat. Neurosci.*, **17**, 1022–1030.
Centonze, D., Picconi, B., Gubellini, P., Bernardi, G. & Calabresi, P. (2001) Dopaminergic control of synaptic plasticity in the dorsal striatum. *Eur. J. Neurosci.*, **13**, 1071–1077.
Cerovic, M., d'Isa, R., Tonini, R. & Brambilla, R. (2013) Molecular and cellular mechanisms of dopamine-mediated behavioral plasticity in the striatum. *Neurobiol. Learn. Mem.*, **105**, 63–80.
Conrad, K.L., Tseng, K.Y., Uejima, J.L., Reimers, J.M., Heng, L.J., Shaham, Y., Marinelli, M. & Wolf, M.E. (2008) Formation of accumbens GluR2-lacking AMPA receptors mediates incubation of cocaine craving. *Nature*, **454**, 118–121.
Corbit, L.H., Nie, H. & Janak, P.H. (2012) Habitual alcohol seeking: time course and the contribution of subregions of the dorsal striatum. *Biol. Psychiat.*, **72**, 389–395.
Corbit, L.H., Chieng, B.C. & Balleine, B.W. (2014) Effects of repeated cocaine exposure on habit learning and reversal by N-acetylcysteine. *Neuropsychopharmacology*, **39**, 1893–1901.
Cui, G., Jun, S.B., Jin, X., Pham, M.D., Vogel, S.S., Lovinger, D.M. & Costa, R.M. (2013) Concurrent activation of striatal direct and indirect pathways during action initiation. *Nature*, **494**, 238–242.
Dezfouli, A. & Balleine, B.W. (2012) Habits, action sequences and reinforcement learning. *Eur. J. Neurosci.*, **35**, 1036–1051.
Di Marzo, V., De Petrocellis, L., Fezza, F., Ligresti, A. & Bisogno, T. (2002) Anandamide receptors. *Prostag. Leukotr. Ess.*, **66**, 377–391.
Dickinson, A. (1985) Actions and habits: the development of behavioural autonomy. *Philos. T. R. Soc. Lond.*, **B308**, 67–78.
Faure, A., Haberland, U., Conde, F. & El Massioui, N. (2005) Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. *J. Neurosci.*, **25**, 2771–2780.
Furlong, T.M., Jayaweera, H.K., Balleine, B.W. & Corbit, L.H. (2014) Binge-like consumption of a palatable food accelerates habitual control of behavior and is dependent on activation of the dorsolateral striatum. *J. Neurosci.*, **34**, 5012–5022.
Gerfen, C.R. & Surmeier, D.J. (2011) Modulation of striatal projection systems by dopamine. *Annu. Rev. Neurosci.*, **34**, 441–466.
Godefroy, O., Azouvi, P., Robert, P., Roussel, M., LeGall, D. & Meulemans, T.; Groupe de Reflexion sur l'Evaluation des Fonctions Executives Study, G. (2010) Dysexecutive syndrome: diagnostic criteria and validation study. *Ann. Neurol.*, **68**, 855–864.
Gong, S., Zheng, C., Doughty, M.L., Losos, K., Didkovsky, N., Schambra, U.B., Nowak, N.J., Joyner, A., Leblanc, G., Hatten, M.E. & Heintz, N. (2003) A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature*, **425**, 917–925.
Graybiel, A.M. (2008) Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.*, **31**, 359–387.
Grueter, B.A., Brasnjo, G. & Malenka, R.C. (2010) Postsynaptic TRPV1 triggers cell type-specific long-term depression in the nucleus accumbens. *Nat. Neurosci.*, **13**, 1519–1525.
Hilario, M.R.F., Clouse, E., Yin, H.H. & Costa, R.M. (2007) Endocannabinoid signaling is critical for habit formation. *Front. Integr. Neurosci.*, **1**, 6.
Hopf, F.W., Seif, T., Mohamedi, M.L., Chen, B.T. & Bonci, A. (2010) The small-conductance calcium-activated potassium channel is a key modulator of firing and long-term depression in the dorsal striatum. *Eur. J. Neurosci.*, **31**, 1946–1959.
Johnson, P.M. & Kenny, P.J. (2010) Dopamine D2 receptors in addiction-like reward dysfunction and compulsive eating in obese rats. *Nat. Neurosci.*, **13**, 635–641.
Kreitzer, A.C. & Malenka, R.C. (2008) Striatal plasticity and basal ganglia circuit function. *Neuron*, **60**, 543–554.
Lerner, T.N. & Kreitzer, A.C. (2012) RGS4 is required for dopaminergic control of striatal LTD and susceptibility to Parkinsonian motor deficits. *Neuron*, **73**, 347–359.
Lovinger, D.M. (2010) Neurotransmitter roles in synaptic modulation, plasticity and learning in the dorsal striatum. *Neuropharmacology*, **58**, 951–961.
Mink, J.W. (2003) The basal ganglia and involuntary movements: impaired inhibition of competing motor patterns. *Arch. Neurol.-Chicago*, **60**, 1365–1368.
Morris, R.W., Quail, S., Griffiths, K.R., Green, M.J. & Balleine, B.W. (2015) Corticostriatal control of goal-directed action is impaired in schizophrenia. *Biol. Psychiat.*, **77**, 187–195.
Nelson, A. & Killcross, S. (2006) Amphetamine exposure enhances habit formation. *J. Neurosci.*, **26**, 3805–3812.
Packard, M.G. (1999) Glutamate infused posttraining into the hippocampus or caudate-putamen differentially strengthens place and response learning. *Proc. Natl. Acad. Sci. USA*, **96**, 12881–12886.
Quinn, J.J., Pittenger, C., Lee, A.S., Pierson, J.L. & Taylor, J.R. (2013) Striatum-dependent habits are insensitive to both increases and decreases in reinforcer value in mice. *Eur. J. Neurosci.*, **37**, 1012–1021.
Reep, R.L., Corwin, J.V., Hashimoto, A. & Watson, R.T. (1987) Efferent connections of the rostral portion of medial agranular cortex in rats. *Brain Res. Bull.*, **19**, 203–221.
Ross, R.A. (2003) Anandamide and vanilloid TRPV1 receptors. *Brit. J. Pharmacol.*, **140**, 790–801.
Shan, Q., Ge, M., Christie, M.J. & Balleine, B.W. (2014) The acquisition of goal-directed actions generates opposing plasticity in direct and indirect pathways in dorsomedial striatum. *J. Neurosci.*, **34**, 9196–9201.
Shen, W., Flajolet, M., Greengard, P. & Surmeier, D.J. (2008) Dichotomous dopaminergic control of striatal synaptic plasticity. *Science*, **321**, 848–851.
Stelzel, C., Basten, U., Montag, C., Reuter, M. & Fiebach, C.J. (2010) Frontostriatal involvement in task switching depends on genetic differences in d2 receptor density. *J. Neurosci.*, **30**, 14205–14212.
Tang, C., Pawlak, A.P., Prokopenko, V. & West, M.O. (2007) Changes in activity of the striatum during formation of a motor habit. *Eur. J. Neurosci.*, **25**, 1212–1227.
Thiagarajan, T.C., Lindskog, M. & Tsien, R.W. (2005) Adaptation to synaptic inactivity in hippocampal neurons. *Neuron*, **47**, 725–737.
Tricomi, E., Balleine, B.W. & O'Doherty, J.P. (2009) A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.*, **29**, 2225–2232.
Van Eden, C.G., Lamme, V.A. & Uylings, H.B. (1992) Heterotopic cortical afferents to the medial prefrontal cortex in the rat. a combined retrograde and anterograde tracer study. *Eur. J. Neurosci.*, **4**, 77–97.
Wang, L.P., Li, F., Wang, D., Xie, K., Shen, X. & Tsien, J.Z. (2011) NMDA receptors in dopaminergic neurons are crucial for habit learning. *Neuron*, **72**, 1055–1066.
Yin, H.H., Knowlton, B.J. & Balleine, B.W. (2004) Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.*, **19**, 181–189.
Yin, H.H., Ostlund, S.B., Knowlton, B.J. & Balleine, B.W. (2005) The role of the dorsomedial striatum in instrumental conditioning. *Eur. J. Neurosci.*, **22**, 513–523.
Yin, H.H., Ostlund, S.B. & Balleine, B.W. (2008) Reward-guided learning beyond dopamine in the nucleus accumbens: the integrative functions of cortico-basal ganglia networks. *Eur. J. Neurosci.*, **28**, 1437–1448.
Yin, H.H., Mulcare, S.P., Hilario, M.R.F., Clouse, E., Holloway, T., Davis, M.I., Hansson, A.C., Lovinger, D.M. & Costa, R.M. (2009) Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. *Nat. Neurosci.*, **12**, 333–341.
Yu, C., Gupta, J., Chen, J.F. & Yin, H.H. (2009) Genetic deletion of A2A adenosine receptors in the striatum selectively impairs habit formation. *J. Neurosci.*, **29**, 15100–15103.
Zygmunt, P.M., Petersson, J., Andersson, D.A., Chuang, H.H., Sørgård, M., Di Marzo, V., Julius, D. & Högestätt, E.D. (1999) Vanilloid receptors on sensory nerves mediate the vasodilator action of anandamide. *Nature*, **400**, 452–457.

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd