

"This is the peer reviewed version of the following article: Shan, Q., Christie, M. J. and Balleine, B. W. (2015), Plasticity in striatopallidal projection neurons mediates the acquisition of habitual actions. *European Journal of Neuroscience*, 42: 2097–2104. doi: 10.1111/ejn.12971, which has been published in final form at <http://onlinelibrary.wiley.com/doi/10.1111/ejn.12971/abstract>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving."

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

## Authors

Qiang Shan<sup>1</sup>, MacDonald J Christie<sup>2</sup>, and Bernard W Balleine<sup>1</sup>

## Affiliations

<sup>1</sup> Behavioural Neuroscience Laboratory, Brain and Mind Research Institute, The University of Sydney, NSW 2006, Australia

<sup>2</sup> Discipline of Pharmacology, The University of Sydney, NSW 2006, Australia

## Running title.

Habit-related plasticity in striatum

## Correspondence

Bernard Balleine  
Brain & Mind Research Institute  
Sydney University  
100 Mallet St  
Camperdown, NSW 2050, Australia

Tel: + 61 2 9114 4011  
Fax: +61 2 9114 4034  
Email: [bernard.balleine@sydney.edu.au](mailto:bernard.balleine@sydney.edu.au)

## Acknowledgements (footnote to first page)

We thank Dr S. Brierley for kindly sharing TRPV1 knockout mice. This research was supported by funding from the Australian Research Council, grant# DP110105636, and both a Laureate Fellowship from the Australian Research Council to B.W. Balleine and NHMRC Senior Principal Research Fellowships to both B.W. Balleine and M.J. Christie. Correspondence concerning this article should be sent to Bernard Balleine, Brain & Mind Research Institute, Sydney University, 100 Mallet St, Camperdown, NSW 2050, Australia; [bernard.balleine@sydney.edu.au](mailto:bernard.balleine@sydney.edu.au). The authors declare no conflict of interest.

**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 (SPNs), the striatonigral and striatopallidal SPNs, which express dopamine D1 and D2 receptors and control direct and indirect pathways, respectively. Here we found evidence of a postsynaptic depression in DLS striatopallidal projecting neurons in the indirect pathway during habit learning in mice. Moreover, this training-induced depression occluded postsynaptic depression induced by coactivation of D2 receptors and transient receptor potential vanilloid 1 (TRPV1) channels, implying 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.

**Keywords.** instrumental conditioning, D2-GFP mouse, overtraining, habits, dopamine D2 receptors, endocannabinoid signaling

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; Corbit *et al.*, 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 characterized 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 (DMS or caudate nucleus) and dorsolateral striatum (DLS or putamen), respectively (Yin *et al.*, 2004; Yin *et al.*, 2005; Balleine *et al.*, 2007; Yin *et al.*, 2008; Tricomi *et al.*, 2009; Balleine & O'Doherty, 2010; Quinn *et al.*, 2013). Nevertheless, although specific changes in synaptic plasticity in the DMS 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- and D2R-SPNs can express long-term potentiation (LTP) 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).

Here, we assessed plasticity associated with habit learning in these two distinct populations of neurons by measuring the spontaneous excitatory postsynaptic current (sEPSC) in the DLS using *ex vivo* patch-clamp electrophysiological recording. 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 D2Rs and 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 postsynaptic depression in the DLS induced by co-activation of D2Rs and TRPV1 channels.

## **Methods**

### **Animals**

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

### **Instrumental conditioning**

Med Associates operant chambers were used for instrumental conditioning. 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 hour of time 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 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 sacrificed 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-hr 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 mins each day for 4 days. Reward grain pellets were delivered on a 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**

One day after the last instrumental conditioning session, mice were sacrificed for *ex vivo* striatal slice recording. Mice were anaesthetized using a ketamine and xylazine cocktail (210 mg/ and 14 mg/kg- body weight, respectively, i.p.) and decapitated. Parahorizontal striatal slices including surrounding cortical regions (300  $\mu$ m thickness) were cut on a vibratome in an

ice-cold dissection solution containing (in mM): 2.5 KCl, 0.5 CaCl<sub>2</sub>, 7 MgCl<sub>2</sub>, 1.2 NaH<sub>2</sub>PO<sub>4</sub>, 26 NaHCO<sub>3</sub>, 10 glucose and 200 sucrose (saturated with 95% O<sub>2</sub>/5% CO<sub>2</sub>, osmolarity 295-305 mOsm). Slices were recovered for at least one hour in an artificial cerebrospinal fluid (ACSF) solution containing (in mM): 126 NaCl, 2.5 KCl, 2 CaCl<sub>2</sub>, 1.2 MgCl<sub>2</sub>, 1.2 NaH<sub>2</sub>PO<sub>4</sub>, 26 NaHCO<sub>3</sub>, and 10 glucose (saturated with 95% O<sub>2</sub>/5% CO<sub>2</sub>, 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<sub>3</sub>, 15 CsCl, 8 NaCl, 10 HEPES, 0.4 EGTA, 3 QX-314, 2 Mg<sub>2</sub>ATP and 0.33 Na<sub>3</sub>GTP (pH 7.3 and osmolarity 280-290 mOsm). Neurons were voltage-clamped at -70mV. Spontaneous EPSCs were acquired for at least 3 mins and analyzed 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 mins. After treatment, slices were washed in ACSF for at least 15 mins before recording.

## Results

### **The amplitude of the sEPSCs of the D2R-SPNs 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. Fig. 1; Fig. 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 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 nine days (Fig. 1B – magazine entry data are presented in Figure S1). To measure the synaptic changes after habit training, striatal slices were prepared one 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 (Fig1H-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 in the amplitude of the sEPSCs generally reflect a postsynaptic mechanism, we conclude that a postsynaptic 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 naïve mice without any behavioral manipulation were also recorded. No significant differences were found between the naïve and yoked mice for all the measurements described above (Fig1D-M), verifying that postsynaptic 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.

Postsynaptic changes in 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 naïve, yoked and trained mice found no significant difference between groups ( $2.815 \pm 0.123$  ms,  $2.930 \pm 0.171$  ms and  $2.895 \pm 0.111$ ms, respectively. ANOVA,  $F(2,42) < 1$ ,  $P > 0.05$ ;  $n = 14-16$ ). This implies that subunit composition of AMPA receptors stays constant during training. Similarly, input resistance did not differ between groups. The mean input resistance for each group was:  $88.14 \pm 7.44$  M $\Omega$ ,  $101.56 \pm 7.44$  M $\Omega$  and  $111.57 \pm 10.11$  M $\Omega$  in D2 neurons of DLS for naïve, yoked and trained mice, respectively. ANOVA conducted on these data found no significant between groups effect ( $F(2,42) = 1.89$ ,  $P > 0.05$ ,  $n = 14-16$ ).

### **The depression induced by the coactivation of the D2R and the TRPV1 is occluded in the habitual mice**

We next investigated the molecular basis for the postsynaptic 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 presynaptic mechanism in most cases (Kreitzer & Malenka, 2008). Nevertheless, Grueter *et al.* reported that the application of the TRPV1 agonist capsaicin induced a postsynaptic LTD in the indirect pathway, but not in the direct pathway, of the ventral striatum (Grueter *et al.*). Considering the similarities of the structure and function between the dorsal and ventral striata, we hypothesized 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 pretreating brain slices from naïve mice with 3  $\mu$ M capsaicin, but, surprisingly, did not find any depression in the D2R-SPNs of the DLS (Fig. 2B-E). However, pretreating the slices with the D2R agonist quinpirole (10  $\mu$ M) together with 3  $\mu$ M capsaicin reliably suppressed the amplitude but not the frequency of the sEPSCs of the D2R-SPNs (Fig 2A-E), implying a postsynaptic LTD-like plasticity as found in the ventral striatum. This depression was only

induced by coactivation of D2Rs and TRPV1s; pre-treatment with the D2R agonist alone did not cause any effect (Fig. 2B-E).

As both the coactivation of D2Rs and TRPV1s and habit training, induce similar changes in the synaptic response of D2R-SPNs in the DLS, i.e. a postsynaptic LTD-like plasticity, we hypothesized that habit training also induces postsynaptic LTD in the D2R-SPNs of the DLS via coactivation 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 coactivation of the D2R and the 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 naïve mice described above, (Fig. 3F, 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 sEPSC amplitude in yoked mice (Fig. 3B,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, I).

### **Knockout of TRPV1 compromises habitual action**

We have demonstrated that habitual training induces a LTD-like plasticity via the coactivation of the D2R and TRPV1. We next tested the necessity of this process in habit learning by disrupting this signal pathway using TRPV1 knockout (TRPV1-KO) mice. Both TRPV1-KO and wild type (TRPV1-WT) mice were trained as described previously (Fig. 1A – magazine data for this study are presented in Figure 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) that 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, C). Because pellet delivery depends 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 the TRPV1-WT mice and reached a lower point of performance than that of the TRPV1-WT mice between 20 and 25 mins 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, normalizing 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 that in the TRPV1-WT mice.

## Discussion

We found that postsynaptic depression is formed in the indirect pathway of the DLS associated with habit learning. We also found that this depression was mediated by the coactivation 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 the 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 postsynaptic. This could potentially be confirmed by examining 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 composition of AMPA receptor subunits or another mechanism contributes to the depression but our finding of unchanged decay time constants of sEPSCs suggest such changes are minor (Thiagarajan *et al.*, 2005). To confirm this, measurement of surface membrane expression of AMPA receptor subtypes would be required in future experiments.

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 knockout of NMDA receptors (Wang *et al.*, 2011), has been shown to inhibit habit learning. On the other hand, 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.* reported that a D2 antagonist blocked the performance of motor skill learning at a late training stage but not at an early training stage (Yin *et al.*, 2009). In this study, skill learning early and late in training was compared and found to involve the DMS 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 synthesized by the activation of the mGluR1/5 and the elevation of

intracellular  $\text{Ca}^{2+}$  generated by L-type VGCCs and intracellular  $\text{Ca}^{2+}$ -stores (Lerner & Kreitzer, 2012). Both processes require glutamate, because L-type VGCCs and intracellular  $\text{Ca}^{2+}$ -stores are gated by postsynaptic depolarization, which is usually induced by AMPA receptor activation by glutamate. Therefore, the activation of TRPV1 might originate from presynaptic 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 hypothesize 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  $\text{Ca}^{2+}$  entry through the TRPV1 channels and cAMP reduction through the D2R-coupled signal pathways, resulting in the retraction of AMPA receptors from the postsynaptic membrane. Although we cannot be sure about the source of glutamate release, we hypothesize that it stems from cortical afferents, particularly those from sensorimotor cortices, which have been often proposed to contribute to habit learning in the past (Graybiel, 2008). It is possible, however, that afferents from other regions of 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 examining 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.* reported that adenosine A2A receptor knockout in the striatum compromised habit learning (Yu *et al.*, 2009), similar to that found in the TRPV1 knockout mice in the current study. Intriguingly, however, unlike TRPV1 channels, whose activation induces a postsynaptic depression, activation of A2A receptors produced a postsynaptic potentiation (Shen *et al.*, 2008). In addition, Hilario *et al.* reported that systemic administration of a cannabinoid CB1 receptor antagonist or CB1 receptor knockout also impaired habit learning (Hilario *et al.*, 2007). However, and again in contrast to TRPV1, CB1 receptor activation has been shown to generate a presynaptic 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 postsynaptic potentiation mediated by A2A receptor activation or presynaptic depression mediated by CB1 receptor activation occurs in the indirect pathway of the DLS in habit trained mice. In contrast, our data suggest postsynaptic 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 one day after the final training session, which reflects, therefore, 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 striatum through an increase in excitation in the direct pathway, the finding that habits are associated with reduced activity in the striato-pallidal 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 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 dorsolateral striatum 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.

## 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.
- 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.
- 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.
- 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. *Neurobiology of learning and memory*, **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., Chieng, B.C. & Balleine, B.W. (2014) Effects of repeated cocaine exposure on habit learning and reversal by N-acetylcysteine. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, **39**, 1893-1901.
- Corbit, L.H., Nie, H. & Janak, P.H. (2012) Habitual Alcohol Seeking: Time Course and the Contribution of Subregions of the Dorsal Striatum. *Biological Psychiatry*, **72**, 389-395.
- 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. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, **66**, 377-391.
- Dickinson, A. (1985) Actions and habits: the development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London*, **B308**, 67-78.
- Faure, A., Haberland, U., Conde, F. & El Massioui, N. (2005) Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **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. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **34**, 5012-5022.
- Gerfen, C.R. & Surmeier, D.J. (2011) Modulation of Striatal Projection Systems by Dopamine. *Annual Review of Neuroscience*, **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. *Annals of neurology*, **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. Postsynaptic TRPV1 triggers cell type-specific long-term depression in the nucleus accumbens. *Nat Neurosci*, **13**, 1519.
- Hilario, M.R.F., Clouse, E., Yin, H.H. & Costa, R.M. (2007) Endocannabinoid signaling is critical for habit formation. *Frontiers in Integrative Neuroscience*, **1**.



- 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.
- Lerner, Talia N. & Kreitzer, Anatol 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.
- Mink, J.W. (2003) The Basal Ganglia and involuntary movements: impaired inhibition of competing motor patterns. *Archives of neurology*, **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 Psychiatry*, **77**, 187-195.
- Nelson, A. & Killcross, S. (2006) Amphetamine exposure enhances habit formation. *Journal of Neuroscience*, **26**, 3805-3812.
- Packard, M.G. (1999) Glutamate infused posttraining into the hippocampus or caudate-putamen differentially strengthens place and response learning. *Proceedings of the National Academy of Sciences*, **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. *British Journal of Pharmacology*, **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. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **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., 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.
- 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., 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.

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., Sørård, 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.

## Figure legends

### Figure 1: *Ex vivo* electrophysiological assessment of synaptic response in D1R-SPNs and D2R-SPNs in the DLS after habitual training.

(A, B) 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 lever presses in the trained mice is significantly higher than 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  $19$  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, L) In the GFP-negative neurons (D1R-SPNs), no significant difference is found between the naïve, 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, 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 naïve 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 naïve, 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.

### Figure 2: Coactivation of the D2R and the TRPV1 suppresses postsynaptic response in the D2R-SPNs.

(A-E) Pretreatment of striatal slices with the D2R agonist quinpirole ( $10 \mu\text{M}$ ) and the TRPV1 agonist capsaicin ( $3 \mu\text{M}$ ) suppresses the amplitude, but not the frequency of sEPSCs of the D2R-SPNs. However, pretreatment 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;  $n = 10-17$ ).

**Figure 3: The postsynaptic depression induced by the coactivation of the D2R and the TRPV1 in the D2R-SPNs is occluded in the habitual mice.**

Mice received either habitual training or yoked treatment, and their striatal slices were pretreated with the D2R agonist quinpirole (10  $\mu$ M) and the TRPV1 agonist capsaicin (3  $\mu$ M) before recording. (B-E) In the yoked mice, like in the naive mice, pretreatment 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, 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;  $n = 14-38$ ).

**Figure 4: Knockout of the TRPV1 compromises habitual action**

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, 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,  $**p < 0.01$ ;  $n = 21$  and  $21$  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 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 they are still sensitive to outcome devaluation and their action is goal-directed. (C) In the omission test, which is independent from 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  $9$  for TRPV1-WT and TRPV1-KO mice, respectively). However, this difference became increasingly smaller in the following tests until the 5th 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 (Day 2-4). (D) In spite that this difference in lever press rates is not statistically significant, the amount of reward pellets TRPV1-KO mice earned in Day 2-4 is higher than that of the TRPV1-WT mice (Day 4, ANOVA  $*P < 0.05$ ). This is a significant reversal of what occurs in Day 1, in which the TRPV1-KO mice earned fewer reward pellets than the TRPV1-WT mice (Genotype X Day interaction, Two Way Repeated Measures ANOVA,  $F(1,20) = 8.3$ ,  $**P < 0.01$ ). (E) In a more direct demonstration, normalizing 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,  $*P < 0.05$ ).

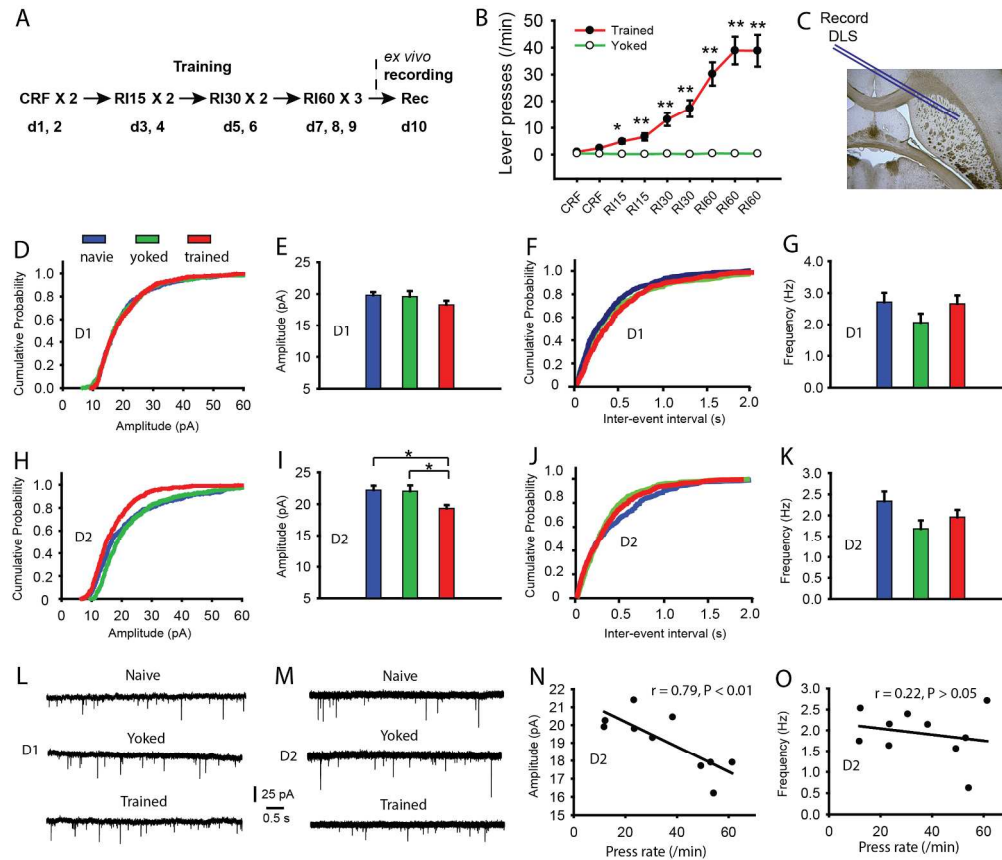


Figure 1: Ex vivo electrophysiological assessment of synaptic response in D1R-SPNs and D2R-SPNs in the DLS after habitual training.

(A, B) 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 lever presses in the trained mice is significantly higher than 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  $19$  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, L) In the GFP-negative neurons (D1R-SPNs), no significant difference is found between the naïve, yoked and trained mice for either sEPSC amplitude or frequency (ANOVA,  $F(2,46) = 1.4$ ,  $P > 0.05$ ;  $F(2,46) = 1.7$ ,  $P > 0.05$ ;  $n = 12-16$ ). (H-K, 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 naïve or yoked mice (ANOVA,  $F(2,44) = 4.6$ ,  $P < 0.05$ ; post hoc test,  $*P < 0.05$ ;  $n = 12-15$ ), and the sEPSC frequency demonstrates no difference between the naïve, yoked and trained mice (ANOVA,  $F(2,44) = 2.7$ ,  $P > 0.05$ ). (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.

211x180mm (300 x 300 DPI)

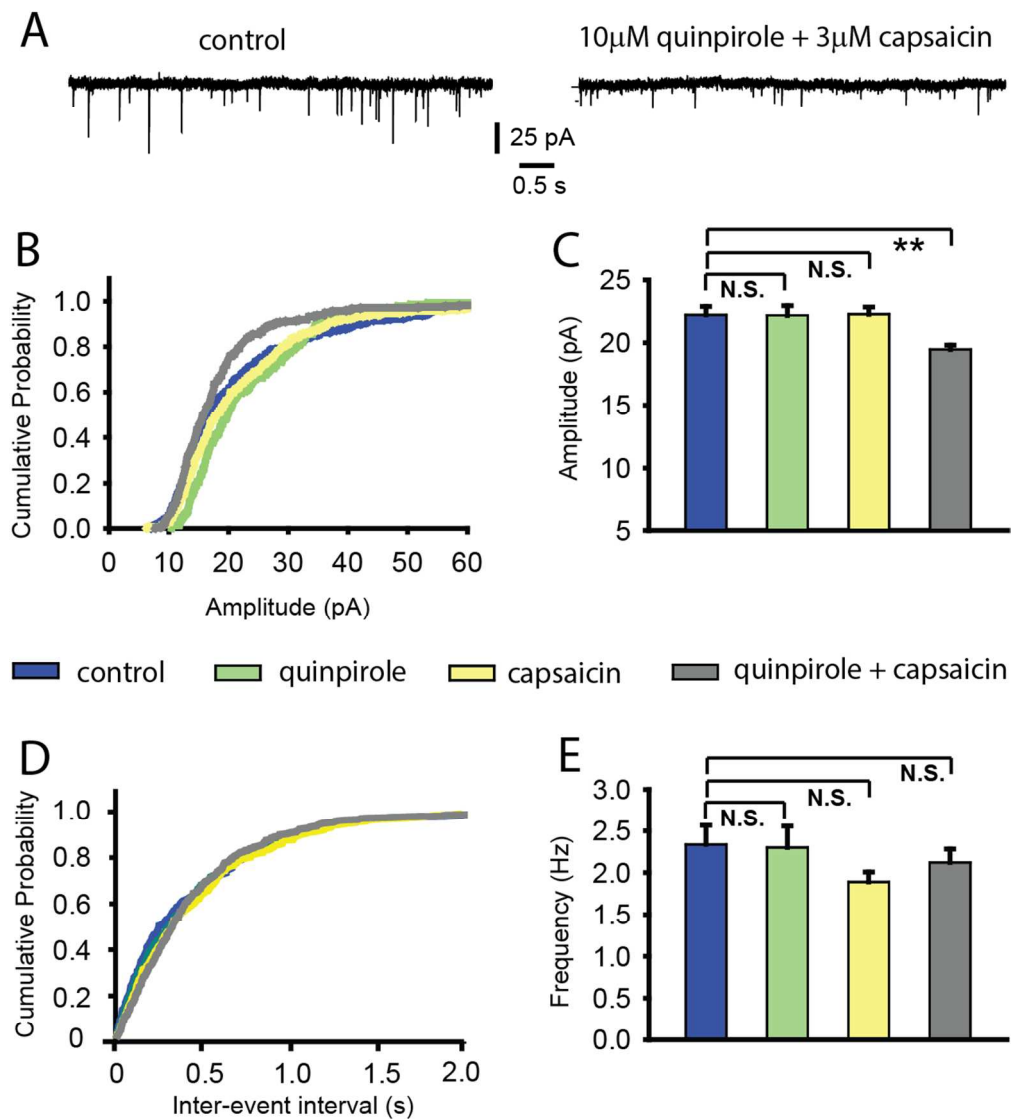


Figure 2: Coactivation of the D2R and the TRPV1 suppresses postsynaptic response in the D2R-SPNs. (A-E) Pretreatment of striatal slices with the D2R agonist quinpirole (10 µM) and the TRPV1 agonist capsaicin (3 µM) suppresses the amplitude, but not the frequency of sEPSCs of the D2R-SPNs. However, pretreatment 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;  $n = 10-17$ ).

106x118mm (300 x 300 DPI)



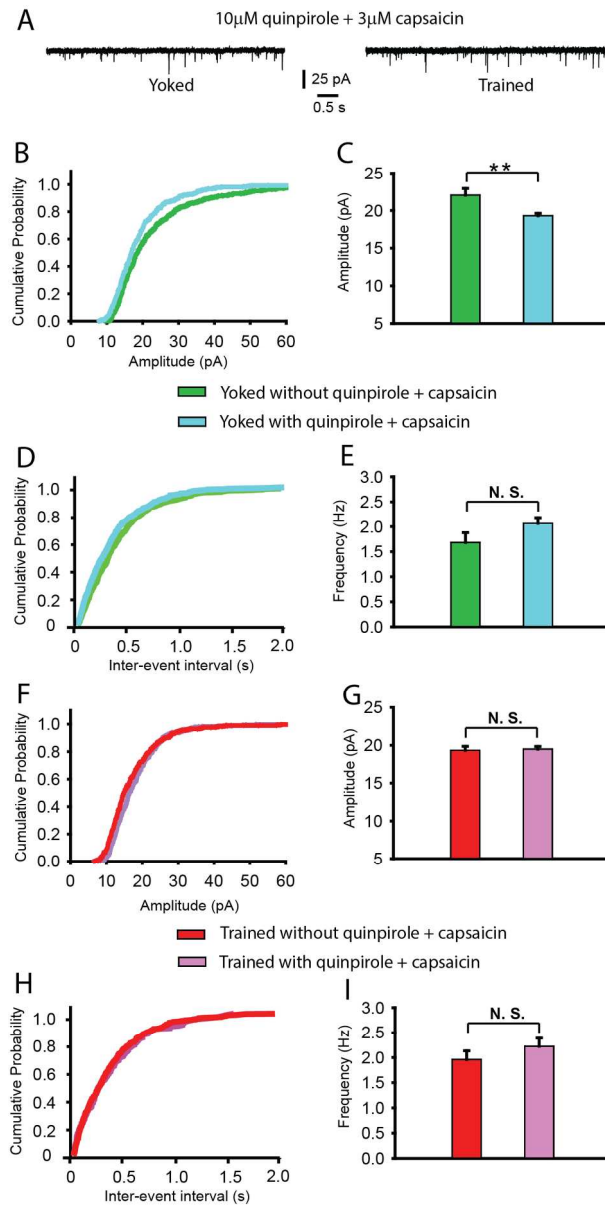


Figure 3: The postsynaptic depression induced by the coactivation of the D2R and the TRPV1 in the D2R-SPNs is occluded in the habitual mice.

Mice received either habitual training or yoked treatment, and their striatal slices were pretreated with the D2R agonist quinpirole (10  $\mu$ M) and the TRPV1 agonist capsaicin (3  $\mu$ M) before recording. (B-E) In the yoked mice, like in the naive mice, pretreatment 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, 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;  $n = 14-38$ ).

106x212mm (300 x 300 DPI)

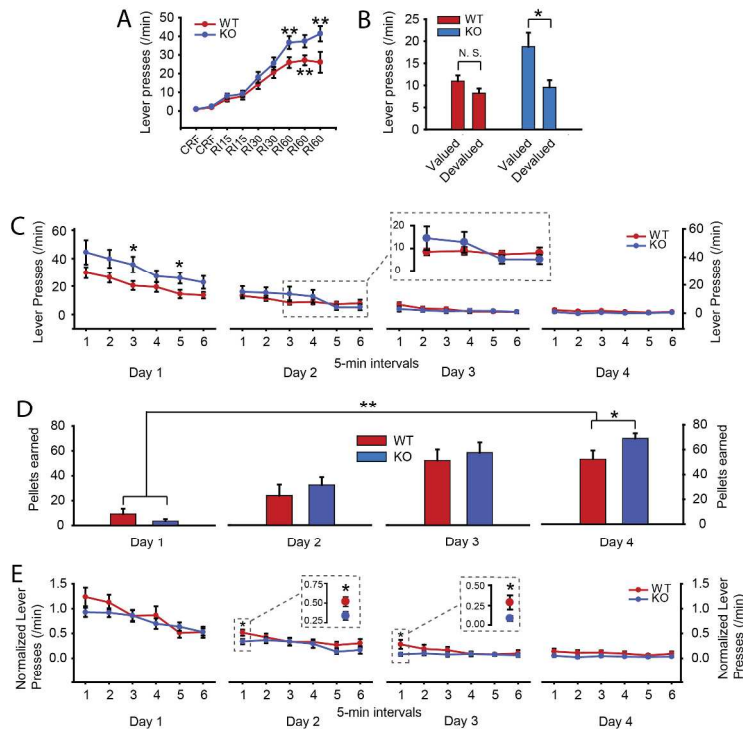


Figure 4: Knockout of the TRPV1 compromises habitual action

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, 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,  $**p < 0.01$ ;  $n = 21$  and 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 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 they are still sensitive to outcome

devaluation and their action is goal-directed. (C) In the omission test, which is independent from 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  $9$  for TRPV1-WT and TRPV1-KO mice, respectively). However, this difference became increasingly smaller in the following tests until the 5th 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 (Day 2-4). (D) In spite that this difference in lever press rates is not statistically significant, the amount of reward pellets TRPV1-KO mice earned in Day 2-4 is higher than that of the TRPV1-WT mice (Day 4, ANOVA  $*P < 0.05$ ). This is a significant reversal of what occurs in Day 1, in which the TRPV1-KO mice earned fewer reward pellets than the TRPV1-WT mice (Genotype X Day interaction, Two Way Repeated Measures ANOVA,  $F(1,20) = 8.3$ ,  $**P < 0.01$ ). (E) In a more direct demonstration, normalizing 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,  $*P < 0.05$ ).

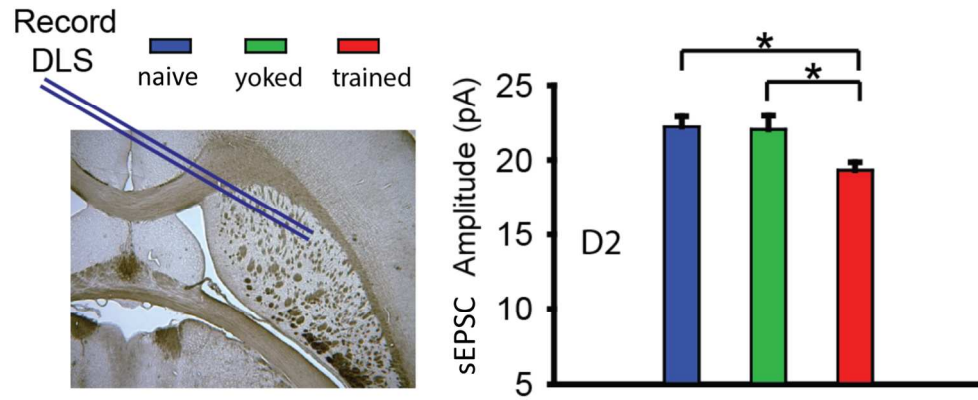
199x323mm (300 x 300 DPI)

For Peer Review

Graphical abstract text (493 characters including spaces)

When over trained, actions become habits. We found evidence that such overtraining induced postsynaptic depression in dorsolateral striatopallidal projecting neurons in the indirect pathway associated with habit learning in mice. Training-induced depression occluded postsynaptic depression produced by coactivation of D2 receptors and TRPV1 channels implying this interaction mediates habit-related plasticity at D2 neurons and providing a potential drug target to influence habitual control.

For Peer Review



150x62mm (300 x 300 DPI)

Peer Review