GluN2B in corticostriatal circuits governs choice learning and choice shifting

Jonathan I Brigman1, Rachel A Daut1, Tara Wright1, Ozge Gunduz-Cinar1, Carolyn Graybeal1, Margaret I Davis2, Zhihong Jiang3, Lisa M Saksida4,5, Seiichiro Jinde3, Matthew Pease1, Timothy J Bussey4,5, David M Lovinger2, Kazu Nakazawa3 & Andrew Holmes1

A choice that reliably produces a preferred outcome can be automated to liberate cognitive resources for other tasks. Should an outcome become less desirable, behavior must adapt in parallel or it becomes perseverative. Corticostriatal systems are known to mediate choice learning and flexibility, but the molecular mechanisms of these processes are not well understood. We integrated mouse behavioral, immunocytochemical, in vivo electrophysiological, genetic and pharmacological approaches to study choice. We found that the dorsal striatum (DS) was increasingly activated with choice learning, whereas reversal of learned choice engaged prefrontal regions. In vivo, DS neurons showed activity associated with reward anticipation and receipt that emerged with learning and relearning. Corticostriatal or striatal deletion of Grin2b (encoding the NMDA-type glutamate receptor subunit GluN2B) or DS-restricted GluN2B antagonism impaired choice learning, whereas cortical Grin2b deletion or OFC GluN2B antagonism impaired shifting. Our convergent data demonstrate how corticostriatal GluN2B circuits govern the ability to learn and shift choice behavior.

The ability to make adaptive choices is fundamental to survival. When a given choice reliably produces a preferred outcome, it can be behaviorally efficient to automate execution of that choice and liberate cognitive processing for other tasks. However, if the value of that same outcome is lessened or a better choice becomes available, actions must adapt accordingly to prevent perseverative, intransigent patterns of behavior.

Choice learning and shifting are thought to depend on anatomically interconnected corticostriatal loops1. Animal lesion and single-unit recording experiments, together with human neuroimaging studies, have shown the ventromedial (vmPFC) and orbitofrontal (OFC) subregions of the prefrontal cortex (PFC) subserve decision-making, and the capacity to rapidly shift between actions2–4. The DS, by contrast, is posited to support the representation of reward–action relations to guide choice learning5, and to enable the automation and habitization of behavior6,7.

Plastic changes in corticostriatal circuits may allow the encoding and expression of stable choices. However, although there has been progress in elucidating neurochemical substrates of these processes8,9, the molecular mechanisms underlying such plasticity remain poorly understood. An excellent candidate in this regard is the NMDA receptor (NMDAR), given work showing that NMDARs, which are essential for certain forms of synaptic plasticity, subserve both PFC-mediated cognitive functions in rodent10 and DS-mediated motor11 and instrumental12,13 learning.

NMDARs are heteromers comprising an obligatory GluN1 subunit and modulatory GluN2A–GluN2D subunits. GluN2B-containing NMDARs are expressed throughout cortex and striatum14, and have slower channel kinetics and a lower channel open probability than GluN2A-NMDARs15. Pharmacological and mutation studies demonstrate that inactivating or overexpressing GluN2B, either systematically or specifically in forebrain regions, alters spatial reference and working memory, trace fear and extinction, attention, and conditional discrimination15–20. Together, these data suggest that GluN2B is crucial in mediating certain types of cognitive functions but do not isolate the specific contribution of corticostriatal GluN2B to choice learning and shifting.

Here we used a multitechnique approach to determine the function of GluN2B-expressing corticostriatal circuits in a simple pairwise choice behavior, as assayed in a visual discrimination and reversal model. We found dynamic patterns of PFC and DS engagement as reliable choice response developed by means of trial-and-error learning and then shifted to an alternative choice. In vivo single-unit recordings revealed dynamic changes in DS neuronal activity around reward anticipation and receipt that tracked learning and relearning. Choice relearning also drove alterations in DS synaptic plasticity. Using regionally restricted gene deletions and drug microinfusions, we found that GluN2B-expressing circuits in DS were crucial for choice learning, but not for flexibility. Conversely, GluN2B-expressing circuits in OFC mediated choice flexibility, but not learning. These data...
demonstrate highly dynamic patterns of corticostriatal activity mediating choice and reveal GluN2B as a key molecular mechanism underpinning this process.

**RESULTS**

**Structure of choice behavior**

We trained C57BL/6J mice on a touchscreen-based pairwise visual discrimination and reversal procedure21–23. Two distinct shapes were presented on a touchscreen in a spatially pseudorandomized manner. Responses to the ‘correct’ stimulus (conditioned stimulus +, CS+) resulted in food reward delivery (‘correct’). Responses to the ‘incorrect’ stimulus (conditioned stimulus −, CS−) produced a 15-s lights-out timeout period (‘error’). A repeat presentation of the previous trial (‘correction trial’) followed each error choice. Error choices on correction trials (‘correction error’) led to more correction trials until the mouse made a correct choice (which we did not record as a correct choice). There were 30 trials (excluding correction trials) per daily session. After a mouse achieved the discrimination criterion of >85% correct choices over two consecutive sessions, the CS+ versus CS− designation was reversed. Reversal training continued until the mouse attained criterion for each performance stage. The number of error choices made generally decreased across stages, with fewer errors on the Dlate than the Dearly stage (Fig. 1e). We used a stage-wise analysis25 to represent each of five major discrimination and reversal performance stages (Fig. 1a): (i) chance-level choice during the first session of discrimination (Dearly), (ii) high choice accuracy by the final session of discrimination (Dlate), (iii) very low choice accuracy on the first reversal session (Rearly), (iv) chance-level (50% correct) choice at the midpoint session of reversal (Rmid) and (v) high choice accuracy by the final reversal session (Rlate) (stage effect on percentage correct choice: F4,40 = 28.57, P < 0.01) (Fig. 1b).

Mice required an average of 8.0 ± 2 (s.e.m.) sessions to complete discrimination and approximately twice as many sessions (15.8 ± 2) to complete reversal, and hundreds of trials to form a robust discrimination and reverse the stimulus-reward contingencies (as consistent with refs. 22,23). This is illustrated by the number of choice errors made generally decreased across stages, with fewer errors on the Dlate than the Dearly stage, followed by an increase in errors on the Rearly and Rmid stages and a further decrease by the Rlate stage (F4,39 = 33.94, P < 0.01; Fig. 1e). Correction errors showed a similar pattern to that of errors (F4,40 = 69.29, P < 0.01) (Fig. 1f). One difference was that by far the highest number of correction errors made was on the Rearly stage. This is consistent with vigorous perseverative choice responding at the previous CS+ and indicates that correction errors are a sensitive measure of perseverative responding on this task. In contrast to the clear stage-wise changes in these measures, neither the time to make a choice nor the latency to retrieve rewards (simple measures of motivation and motor function) changed significantly across stages, although a nonsignificant decrease in both measures across sessions was apparent (time to choice F4,36 = 1.28, P = 0.29; latency to retrieve F4,36 = 2.38, P = 0.07; Supplementary Fig. 1a).

We next calculated the average length of continuous strings of either correct or error choices. Mice engaged in equal sampling of stimuli (strings of ~2 responses at each stimulus) during Dearly, but then, as they learned to make the correct choice by Dlate, there was a parallel shift to making long strings (~8) of consecutive correct choices (correct strings: F4,37 = 31.87, P < 0.01, errors strings: F4,37 = 41.39, P < 0.01) (Fig. 1g). We saw the same pattern for correct choice strings across stages Rearly to Rlate, as mice relearned the choice during reversal. Conversely, there were long strings (~9) of error choices during Rearly reflecting the high rates of perseveration during initial reversal. Of note, error strings remained elevated (~5) during Rmid (before decreasing by Rlate) even though overall choice accuracy had improved to chance levels by Rmid and was essentially at the same level as during the start of choice discrimination learning. Thus, whereas chance performance on the initial discrimination involved exploratory sampling of the choice options, chance performance at reversal was characterized by correct responses interspersed with blocks of perseverative responses, illustrating how ostensibly similar profiles of choice performance were associated with qualitatively distinct patterns of behavior. These analyses establish the main patterns of behavior across choice learning and relearning and provide a framework for studying the underlying neural and molecular mechanisms.

**Corticostriatal activation associated with choice learning**

Our next objective was to identify the principal brain regions activated during choice behavior. To this end, we trained mice to one of the five choice performance stages (Dearly–Rlate) and then immunocytochemically quantified the immediate-early gene (IEG) c-Fos in 13 different forebrain regions. The number of c-Fos–positive cells in various regions changed as a function of stage (Supplementary Table 1). The clearest patterns were seen in regions of the PFC and DS. There was increased c-Fos expression in the OFC (F4,36 = 4.17, P < 0.01) (Fig. 2a,b) and prelimbic area of vmPFC (F4,37 = 5.40, P < 0.01) (Fig. 2c,d) specifically during Rearly and, to a lesser extent, Rmid, as the original choice was reversed. These data demonstrate that choice shifting in our task activates the same subregions of PFC governing reversal learning and other measures of flexible decision-making in rats, nonhuman primates and humans2–4,26. Activation in these regions could reflect a number of processes that include, but may not be limited to, relearning of the change in stimulus–reward contingencies. For example, this could reflect a response to surprise or confusion at the contingency shift, which has been mainly linked to other prefrontal areas (for example, anterior cingulate)27. The type of reversal procedure we used in this study cannot readily parse these alternatives. Notwithstanding, engagement of OFC or PFC is not an epiphenomenon to relearning and clearly is functionally important, as demonstrated by previous evidence that lesioning these regions affects reversal learning in this task24.

Whereas OFC and the prelimbic area were most active during choice shifting, DS activation tracked choice learning and relearning. c-Fos expression in DS increased from Dearly to Dlate during choice learning (F4,38 = 11.02, P < 0.01) (Fig. 2e,f). c-Fos expression was then decreased on Rearly, notably, to levels that were lower than on Dearly, suggesting DS was not simply unengaged but may have been inhibited during initial reversal. As mice subsequently relearned choice, there was step-wise increase in DS c-Fos expression over the reversal stages. The close parallel between DS activation and choice performance imply that activation of this brain region may have supported choice learning and relearning. These data showed equivalent engagement of the lateral and medial aspects of DS, although previous rodent lesion studies specifically implicated the lateral DS in stimulus-reward and habit behavior and the medial DS in goal-directed behavior2.
However, the medial DS is also involved in reversal\textsuperscript{28}, and although inactivation of medial DS impairs goal-directed behavior\textsuperscript{29}, inactivation limited to the posterior part of medial DS can also impair habit-related, stimulus-outcome learning\textsuperscript{29}. Thus, the relative contributions of the medial and lateral DS to habit are complex and their respective roles in choice learning remain particularly unclear.

Further demonstrating the functional contribution of lateral DS to choice relearning in our task, DS lesions disrupt choice relearning\textsuperscript{24}. To confirm that initial choice learning was also DS-dependent, we made discrete bilateral lesions of lateral DS before discrimination (Fig. 2g) and found that lesioned mice made more discrimination errors ($t_{28} = 2.51, P < 0.01$) and correction errors ($t_{28} = 2.12, P < 0.01$) errors than sham controls (Fig. 2h,i). Given the importance of DS to motor behavior, we confirmed that choice learning deficits in lesioned mice were not an artifact of locomotor dysfunction by showing no difference in locomotor activity between sham and lesion mice in an open field (Supplementary Fig. 2a).

Given the contribution of DS to choice learning and relearning, we asked whether repeated DS engagement over the course of training in our task might have ‘primed’ the region in such a way as to positively transfer (for example, to improve) performance on other striatum-dependent forms of learning. Mice were trained to either the R\textsuperscript{late} or D\textsuperscript{late} stages or were given operant training with no choice learning (matched as to the total number of sessions but given the same number of sessions as the R\textsuperscript{late} group) and, the day after, assessed for DS-mediated motor learning on the rotarod task\textsuperscript{30}. Groups showed similar motor learning over ten training trials ($F_{1,15} = 27.02, P < 0.01$) (Supplementary Fig. 2b), indicating no demonstrable performance transfer between the two DS-dependent forms of learning.

A crucial role of DS in choice learning in the present setting extends previous findings obtained in a range of species, including human subjects, showing that DS mediates stimulus-response learning and automatized, habitual behaviors\textsuperscript{6,7}. Extended instrumental training can promote the development of habitual behavior\textsuperscript{31}, and we have previously shown\textsuperscript{24} that choice behavior during relearning (as early as R\textsuperscript{mid}) is insensitive to reinforcer devaluation, an operational measure of habit\textsuperscript{31}. Although it remains to be shown whether choice behavior in our task becomes habitual with training, present and previous IEG
and lesion data demonstrate that OFC and DS are crucial in choice learning and relearning.

**In vivo DS single-unit activity associated with choice learning**

Although our results thus far indicate that DS is crucial in choice learning and relearning, IEG and lesion approaches do not indicate which, if any, specific behavioral components of task performance are associated with DS function. To more directly test for DS activity in close temporal coincidence with behavior, we conducted *in vivo* neuronal recordings in DS in freely moving mice. We implanted multielectrode arrays in DS (Fig. 3a) and recorded from 402 putative neurons (84 ± 3 per stage) in eight mice during sessions corresponding to each of the five performance stages (Dearly through Rlate). Choice accuracy (F_{1,45} = 86.09, P < 0.01) (Fig. 3b) and the errors made (F_{1,45} = 23.94, P < 0.01) (Fig. 3c) differed between stages in the same manner as above.

We did not classify neurons on the basis of firing rate or waveform, although some studies have indicated a preferential contribution of DS fast-spiking interneurons in choice execution. The activity of all recorded neurons (5.89 ± 8.4 Hz) was sorted into 50-ms time bins and temporally aligned to four separate event-related 3-s epochs: after trial initiation, immediately before choice, after choice and after reward retrieval (Fig. 3d). To avoid overlap of activity across epochs, we terminated measurement of neuronal activity during one epoch at the time of next epoch. Activity was segregated for correct-choice and error-choice trials.

To measure the average activity of the recorded population, we Z-score normalized activity for each neuron to the average firing rate of that neuron across all four event epochs, with positive and negative values respectively indicating relatively higher and lower activity, relative to the average, at a given time point. For correct choice trials, there were stage × time interactions for trial initiation (F_{2,36,23364} = 1.61, P < 0.01), correct choice (F_{2,36,23364} = 2.40, P < 0.01) and reward (F_{2,36,23364} = 2.56, P < 0.01), but not prechoice, epochs (Fig. 3e,f). The effect for trial initiation was largely due to modest activity increases during Dearly. More marked stage-wise activity was evident during the latter seconds of the epoch after a correct choice was made and immediately before reward receipt. Specifically, a population-level inhibition of activity emerged with learning and relearning, peaking by Rlate. Conversely, marked excitation developed with learning and, particularly, with relearning after reward collection. Of note, although choice and reward-related activity developed in tandem across performance stages, there was only a weak correlation between the two (Pearson’s r = 0.154), suggesting largely segregating populations of DS neurons encoded each behavioral event (Fig. 3g).

To examine the network organization of DS neurons across stages, we examined the event-related firing of individual units. This clearly illustrated a learning- and relearning-related increase in population of units that inhibited their activity before reward (Supplementary Fig. 3a). To further quantify these shifts, we calculated the percentage of recorded cells exhibiting event-related Z-scores >1.0 (positively modulated) or <-1.0 (negatively modulated) at a given time bin. These data mirrored the individual single-unit data, showing during later reversal stages a higher percentage (~15%) of negatively modulated choice-related units for correct choice trials and a higher percentage of positively modulated reward-related units (Supplementary Fig. 3b). These stage-wise shifts were specific to correct choices and were not found for error trials. Although there was stage × time interactions for trial initiation (F_{2,36,23364} = 1.86, P < 0.01), prechoice (F_{2,36,23364} = 1.74, P < 0.01), choice (F_{2,36,23364} = 1.57, P < 0.01) and reward-omission (F_{2,36,23364} = 1.16, P < 0.01) error trial epochs, there...
was no clearly discernible stage-wise shift in activity associated with choice or reward omission other than modest post-error inhibition on Dearly and Rearly-to-Rmid and excitation after reward on Dearly (Fig. 3h and Supplementary Fig. 4a,b).

These data demonstrate dynamic changes in the activity of DS neurons around choice and reward receipt occurring in concert with improving choice performance. The pattern of changes was consistent with the emergence of inhibition of a significant population of DS units in anticipation of reward receipt. Because neuronal activity after reward receipt was restricted to learning and especially relearning, it cannot simply be an artifact of chewing or reflect a signal of the hedonic value or the reward. However, given the importance of the DS to motor functions, we asked whether the post–choice DS activity changes across stages reflected stage-related changes in movement timing, rather than reward anticipation. Post–correct choice activity data for each stage were sorted, using a median split, according to ‘fast’ or ‘slow’ choice-to-reward latencies. Activity did not differ as a function of the latency from choice to reward (Supplementary Fig. 5); while both the fast and slow response times were virtually equivalent at the Rearly (fast, 1.5 s; slow, 3.8) and Rlate (fast, 1.4 s; slow, 3.6) stages, there was only strong prereward activity inhibition for Rlate and not Rearly. Thus, the speed of choice-to-reward movement does not explain the stage-wise activity shifts in DS neuronal activity. Nonetheless, we cannot exclude a contribution of other behavioral states that vary coincident with learning.

The patterns of DS neuronal activity associated with choice learning echoes earlier examples of DS neurons exhibiting task-relevant shifts in activity during motor learning, formation of a motor habit or acquisition of stimulus-response learning. For example, DS neural activity associated with performance in a T-maze task shifts rapidly as responses are extinguished and reinstated. Our data add to this literature by showing clear and highly dynamic in vivo DS neuronal responses in the setting of an operant choice task and provide further support for the importance of this brain region in mediating choice behavior. It will be of interest to examine how these dynamic changes in DS activity relate to concurrent changes in OFC or vmPFC neuronal activity, given previous in vivo evidence that activity in the regions is closely coupled during learning.

DS plasticity associated with choice learning

Changes in DS unit activity across choice stages suggest that plasticity mechanisms may be engaged to shape and reshape behavior. We therefore tested whether choice stages were associated with alterations in DS plasticity using ex vivo slice electrophysiology. Mimicking the design of the experiments above, we trained mice to one of the five performance stages and, 2 h later, took brain slices that included the DS for electrophysiology recordings, along with those from a set of behaviorally naive control mice. Choice accuracy (F4,45 = 325.85, P < 0.01) (Fig. 4a) and the total number of errors made (F4,45 = 18.70, P < 0.01) (Fig. 4b) differed between stages in the now expected manner, closely replicating the patterns in the earlier experiments.

Evoked field potentials were recorded at DS synapses after local afferent stimulation by a locally placed bipolar twisted-tungsten electrode (Fig. 4c). We first measured the efficacy of synaptically driven neuronal output of DS neurons by measuring the population spike magnitude at increasing stimulation-amplitudes (0.1–1.5 mA) (F14,364 = 231.35, P < 0.01). There was no difference (F3,36 = 1.55, P = 0.23) in this input–output measure across stages, and although there was a trend for a leftward shift at the late reversal stage (Rlate) that could indicate an increase in the efficacy of activation, there was no discernible stage-wise pattern in these trends (Fig. 4d). Next we examined long-term depression (LTD) at these synapses using a
**Figure 4** Striatal synaptic plasticity changes with choice learning. *Ex vivo* recordings were conducted after attaining the five choice performance stages. (a,b) Percentage correct responding (a) and total cumulative errors from the start of learning and relearning (b) on the session immediately before *ex vivo* recordings. (c) Position of recordings of field potentials evoked by local afferent stimulation in coronal slices containing DS (black, stimulating electrode; gray, recording electrode). (d) Population spike (PS) amplitude did not vary with stage, but it was left-shifted in mice trained to R late. (e) Time course of PS amplitude at baseline and after three trains of HFS. (f) Average baseline and post-HFS values. All three HFS trains evoked LTD in test-naive mice, as compared to baseline. Mice trained to D early or R early showed LTD after the second or third, but not first, HFS trains, and mice trained to D late showed LTD after the third train only. LTD was occluded in mice trained to R early or R late. (g) Representative traces after the third stimulation train. Number of mice per stage: naive, 5; D early, 9; D late, 9; R early, 9; R mid, 9; R late, 7. For a,b, *P* < 0.05 versus D late, †*P* < 0.05 versus D early, ‡*P* < 0.05 versus R late, †‡*P* < 0.05 versus R early, *P* < 0.05 versus baseline.

High-frequency stimulation (HFS) protocol comprising sets of two 1-s trains of 100 pulses beginning 10 min after establishing baseline population spike amplitude and continuing once every 20 min. We used a procedure entailing three trains in order to test for graded alterations in LTD, as successive trains are expected to produce stronger LTD6. Data are presented in the full time course (Fig. 4e) and the averaged values for each train and baseline (Fig. 4f). Example traces are also shown (Fig. 4g).

In behaviorally naive mice, robust LTD was produced after the third (t6 = 6.56, *P* < 0.01), second (t6 = 6.70, *P* < 0.01) and first (t6 = 4.99, *P* < 0.01) sets of trains, as indicated by a decrease in population spike magnitude after HFS relative to baseline before the first set of trains (Fig. 4c,f). By contrast, LTD was partially impaired in mice trained to either D early or D late, being evident only after the third (t2 = 2.40, *P* < 0.05) or the second (t2 = 2.89, *P* < 0.01) sets of trains, respectively (Fig. 4e,f). Partial loss of LTD at the beginning of choice testing suggests that plasticity changes have already developed at this early point in training, possibly owing to some engagement of DS during pretraining.

More notably, LTD was essentially absent after training to R early and R late (that is, there was no LTD after any train) (Fig. 4c,f). This loss of plasticity corresponds to the stages where choice behavior is relatively rigid, either owing to perseveration during initial reversal or high choice accuracy after extensive training at late reversal. Loss of plasticity was not simply a function of the amount of choice training mice had undergone. This was evidenced by the ‘recovery’ of robust LTD during the R mid stage of reversal, when mice were shifting and relearning the choice: significant LTD after the third (t6 = 5.33, *P* < 0.01) and second (t6 = 3.80, *P* < 0.01) sets of trains and a trend toward LTD after the first (t5 = 2.62, *P* = 0.068) set of trains (Fig. 4e,f). Thus, there appeared to be a close association between plasticity at DS neurons and stages of maximal demands on choice flexibility, such that plasticity was highest when mice were relearning the choice and lowest when choice was either perseverative or well learned.

These *ex vivo* electrophysiological data are consistent with DS plasticity as a dynamic correlate of choice learning, but they do not establish a causal relationship between changes at DS neurons and choice performance. As in the case of the *in vivo* single-unit measures, this approach is unable to directly attribute plastic changes to learning and not some coincidental behavioral states. Nonetheless, because changes in LTD were evident in the absence of concomitant changes in synaptically driven DS neuronal firing (assessed by input–output analysis), we can conclude that they are unlikely to be an effect of a general enhancement of neuronal output. Instead, these data imply alterations in molecular mechanisms mediating plasticity at DS synapses. Previous findings implicate dopamine as one possible contributing mechanism. Phasic dopamine activity is crucial to reinforcement learning in various behavioral settings and is hypothesized to support learning in part by signaling reward uncertainty to regions that receive dense inputs, such as the DS. In addition, dopamine mediates DS LTD, and dopamine applied coincident with corticostriatal synaptic activation promotes synaptic plasticity. Shifts in dopamine input to DS during choice learning are a focus of our future studies.

**Impaired choice learning after corticostriatal Grin2b deletion**

We next examined the contribution of GluN2B to plasticity at DS neurons to choice learning and relearning. We began with a conditional mutant model, *Grin2b*^CoxNull*, in which *Grin2b* is postnatally deleted in forebrain principal neurons expressing calcium-calmodulin–dependent kinase II (CaMKII). To produce corticostriatal-wide loss of *Grin2b*, we took advantage of the observation that the Camk2a-promoter transgenic mouse (T29-1) produces increasingly widespread
deletion as mice age. Quantitative western blots confirmed loss of GluN2B protein in mutant tissue from DS (t(5) = 5.03, P < 0.01), medial prefrontal cortex (mPFC) (t(6) = 2.46, P < 0.05) and dorsal hippocampus (t(5) = 8.40, P < 0.01) (Supplementary Fig. 6a). In ~11-month-old mutants, Grin2b mRNA was decreased in the cortex, striatum and CA1 hippocampus but not in other forebrain regions that included thalamus and basolateral amygdala, relative to that in age-matched Grin2b<sup>fx</sup> littermate controls (Supplementary Fig. 6b). Comparison of ~11-month-old Grin2b<sup>fx</sup>StNull mutants and age-matched Grin2b<sup>fx</sup> controls found no differences in operant pretraining before discrimination training, indicating normal gross motor and motivational functions. However, choice learning was impaired in these mutants, as demonstrated by more errors (t<sub>16</sub> = 3.11, P < 0.01) (Fig. 5a) and correction errors (t<sub>16</sub> = 2.36, P < 0.05) (Fig. 5b) before attaining discrimination criterion, as compared to Grin2b<sup>fx</sup> control values. Choice-response and reward-retrieval latencies were no different between genotypes, further excluding a general performance deficit (Supplementary Fig. 1d). Grin2b<sup>StNull</sup> mutants also made more errors (t<sub>13</sub> = 3.30, P < 0.01) (Fig. 5c) and correction errors (t<sub>13</sub> = 4.08, P < 0.01) (Fig. 5d) than Grin2b<sup>fx</sup> controls before attaining reversal criterion but again had normal choice-response and reward-retrieval latencies (Supplementary Fig. 1e). Underscoring the severity of the learning deficit, five of the seven mutants failed to attain criterion even after extensive (60-session) reversal training.

Impaired choice learning in these mutants did not extend to other operant settings. We tested a naive cohort of ~11-month-old Grin2b<sup>StNull</sup> mutants on a task that required the mice to touch a small visual stimulus for reward, without having to make a choice between two options. Grin2b<sup>StNull</sup> mutants acquired (Supplementary Fig. 7a) and extinguished (Supplementary Fig. 7b) this behavior in the same number of trials as age-matched Grin2b<sup>fx</sup> controls. A similar dissociation between intact performance in this task and impaired choice learning has also been reported in the same choice task in mutants with brain-wide constitutive swap of the GluN2B and GluN2A C-terminal domains<sup>41</sup> or deletion of Grin2a (ref. 42), suggesting that NMDARs are dispensable for simple forms of operant learning. This set of experiments establishes that corticostriatal GluN2B is crucial in mediating choice learning and relearning.

Impaired choice learning after striatal Grin2b deletion

The data so far strongly implicate DS in choice learning and suggest that deficit in the Grin2b<sup>StNull</sup> mutants was due to GluN2B loss in DS. However, the corticostriatum-wide nature of deletion in the Grin2b<sup>StNull</sup> mutants precludes parsing of the relative contribution of GluN2B-expressing circuits in DS and cortex. We therefore generated a conditional mutant in which Grin2b is postnatally deleted in striatal cells expressing regulator of G-protein signaling 9 (RGS9) (Grin2b<sup>S<sub>N</sub>Null<sub>S</sub></sup>)<sup>11</sup>. Real-time-PCR confirmed significant loss of Grin2b in DS (t<sub>5</sub> = 4.51, P < 0.01), but not in mPFC or dorsal hippocampus, in 4–6-month-old Grin2b<sup>S<sub>N</sub>Null</sup> mice relative to age-matched Grin2b<sup>fx</sup> controls (Supplementary Fig. 8a). Western blots showed a modest but significant loss of GluN2B protein in tissue from DS (t<sub>7</sub> = 5.03, P < 0.01) but not from mPFC or dorsal hippocampus (Supplementary Fig. 8b).

There were no differences in operant pretraining before discrimination training in the Grin2b<sup>StNull</sup> mice, but choice learning was severely impaired, as demonstrated by more errors (t<sub>17</sub> = 5.57, P < 0.01) (Fig. 5e) and correction errors (t<sub>17</sub> = 3.65, P < 0.05) (Fig. 5f) before attaining discrimination criterion, relative to Grin2b<sup>fx</sup> littersmates. Five of the ten mutants failed to attain discrimination criterion even after extensive (60-session) training. The mutants also made more errors (t<sub>12</sub> = 4.21, P < 0.01) (Fig. 5g) and correction errors (t<sub>12</sub> = 2.62, P < 0.05) (Fig. 5h) than controls before attaining reversal criterion. There was no indication of a general performance deficit, as choice-response and reward-retrieval latencies were normal (Supplementary Fig. 1f,g). These data confirm that GluN2B-expressing striatal cells are crucial for choice learning and relearning.

Impaired choice learning after DS GluN2B antagonism

Although the Camk2a and Rgs9 promoters circumvent the confounding effects of lack of GluN2B during development (ref. 43), prolonged GluN2B loss may still have produced compensatory alterations in the expression or function of other subunits. Moreover, because Grin2b deletion was present at all testing stages, this approach cannot delineate the function of GluN2B in choice learning or shifting, and the expression of choice behavior once learned. Therefore, to complement the mutant data, we infused the selective GluN2B antagonist Ro 25-6981 into the DS at different stages of relearning.

We trained C57BL/6j mice through D<sub>late</sub> and assigned them to either vehicle or Ro 25-6981 groups, matching for number of trials to D<sub>late</sub>. In three separate experiments, 2.5 µg Ro 25-6981 (0.5 µl per hemisphere) or an equivalent volume of vehicle was infused bilaterally into DS 15 min before sessions corresponding to R<sub>early</sub>, R<sub>mid</sub> or R<sub>late</sub>. We then conducted three sessions without infusions to ensure that behavior was altered by GluN2B antagonism and was not an artifact of cannulation or infusion.

DS GluN2B blockade during R<sub>early</sub> (Fig. 6a) did not alter total errors (Fig. 6b) or correction errors (Fig. 6c) over three infusion sessions, relative to those in vehicle controls (Supplementary Fig. 9a).
This indicates that DS GluN2B is dispensable for initial choice shifting, presumably because DS-mediated choice relearning is not fully engaged at this stage and performance can be supported by other brain regions (for example, cortex). By contrast, in a separate experiment, blockade of GluN2B during R\textsuperscript{mid} (Fig. 6d) increased errors (t\textsubscript{14} = 3.04, P < 0.01) over three infusion sessions, relative to those in vehicle controls (Supplementary Fig. 9b). This confirms that GluN2B specifically in DS is crucial for choice relearning and extends previous evidence that systemic GluN2B antagonism impairs various forms of learning\textsuperscript{16–18,44–46}.

Finally, we asked whether DS GluN2B mediates choice behavior after relearning is complete. In mice trained to R\textsuperscript{late} (Fig. 6g), DS GluN2B blockade did not increase errors (Fig. 6h) or correction errors (Fig. 6i) over three infusion sessions, relative to those in vehicle controls (Supplementary Fig. 9c). Choice-response and reward-retrieval latencies were unaffected by GluN2B antagonism in this or any infusion experiment (Supplementary Fig. 1k–m). The absence of effects of GluN2B blockade on the expression of a learned choice is generally consistent with previous studies showing that infusion of a nonspecific NMDAR antagonist in DS does not alter a learned cue-driven cocaine-seeking response in rats\textsuperscript{47}. Other mechanisms, including AMPA and dopamine receptors, may be necessary for choice expression in our task, as found for other DS-mediated behaviors\textsuperscript{47}.

Taken together, these data demonstrate that functional inactivation of GluN2B in DS is sufficient to impair choice relearning and, moreover, that DS GluN2B is not necessary for either initial choice shifting or the expression of choice once learned. It remains to be shown whether this indicates that GluN2B is necessary for mediating learning-related plasticity at DS synapses or for regulation of the flow of crucial information from other regions either to or from the DS.

**Impaired choice shifting after cortical Grin2b deletion**

Our finding that choice relearning is mediated by GluN2B-expressing circuits in DS still leaves open the question of whether parallel circuits in PFC regions implicated by our c-Fos data mediate choice shifting. Our first approach was to generate a cohort of mutant mice, Grin2b\textsuperscript{CoxNull} from the same line described above but in which Grin2b deletion is largely restricted to CaMKII-expressing principal neurons in cortex by virtue of their younger age, as previously shown using in situ hybridization and quantitative immunoblot\textsuperscript{13}. We have previously reported loss of GluN2B protein and Grin2b mRNA throughout cortex, as well as the dorsal hippocampal CA1 subregion, in these mice\textsuperscript{15}. Here we replicated the in situ hybridization to show loss of Grin2b mRNA in these regions (Supplementary Fig. 8c).

We found that Grin2b\textsuperscript{CoxNull} mice (2 months old at start of testing) were no different from age-matched Grin2b\textsuperscript{lox} littermate controls on operant pretraining. These mutants also made the same number of errors (Fig. 7a) and correction errors (Fig. 7b) as controls before attaining discrimination criterion, indicating intact choice learning. Although Grin2b\textsuperscript{CoxNull} mutants also made a similar number of errors (Fig. 7c) as Grin2b\textsuperscript{lox} controls before attaining reversal criterion, they made more correction errors (t\textsubscript{14} = 3.63, P < 0.01) (Fig. 7d). Choice-response and reward-retrieval latencies were no different between genotypes for discrimination or reversal (Supplementary Fig. 1h–j). This selective increase in correction errors during reversal suggests that the Grin2b\textsuperscript{CoxNull} mutants are impaired on choice shifting. To explore this possibility further, we subdivided reversal performance into sessions in which choice accuracy was below chance (<50% correct) and above chance (>50% correct), equivalent to R\textsuperscript{early-to-Rmid}
and R\textsuperscript{mid}-to-R\textsuperscript{late} phases, respectively. This revealed that the higher rate of correction errors in the mutants was specific to the R\textsuperscript{early}-to-R\textsuperscript{mid} phase (t\textsubscript{14} = 2.75, P < 0.05), with no genotype difference at the R\textsuperscript{mid}-to-R\textsuperscript{late} phase (Fig. 7e) and no change in errors at either phase (Fig. 7f).

These mutant data are consistent with a selective impairment in choice shifting as a result of cortical GluN2B loss. Given the age-dependent nature of the loss of Grin2b in these mutants, it was, however, possible that the deficit was an artifact of the mutants being slightly older (with potentially some striatal loss) at the time of choice shifting than choice relearning. Excluding this possibility, we phenotyped another set of mice for discrimination at an older age (~5 months) and confirmed that there were no genotype differences in choice learning (errors from D\textsuperscript{early}-to-D\textsuperscript{late}: Grin2b\textsubscript{flox}, 102 ± 21; Grin2b\textsubscript{CxNull}, 131 ± 23; correction errors: Grin2b\textsubscript{flox}, 188 ± 30; Grin2b\textsubscript{CxNull}, 267 ± 67). Thus, we have demonstrated that cortical GluN2B is crucial in choice shifting but not choice learning. The specificity of the effects of cortical GluN2B inactivation echoes previously observed learning phenotypes in these conditional Grin2b null mutants. For example, Grin2b\textsubscript{CxNull} mutants show impaired corticohippocampal spatial memory, but normal striatally mediated cue-guided learning, in the Morris water maze\textsuperscript{15} (see also ref. 19).

Impaired choice shifting after OFC GluN2B antagonism

Some of the same caveats discussed above for the Grin2b\textsubscript{CxNull} mutants apply to the data in the Grin2b\textsubscript{CxNull} mutants; namely, Grin2b deletion is not temporally limited, nor is it spatially restricted to specific cortical subregions. We therefore sought to reinforce the mutant data with a pharmacological approach by testing the effects of OFC GluN2B blockade on choice performance.

OFC GluN2B blockade during R\textsuperscript{early} (Fig. 8a) did not alter the number of errors (Fig. 8b) but did increase correction errors (t\textsubscript{14} = 3.61, P < 0.001) (Fig. 8c), relative to those in vehicle controls (Supplementary Fig. 9a). By contrast, GluN2B blockade during R\textsuperscript{mid} (Fig. 8d) had no effect on either errors (Fig. 8e) or correction errors (Fig. 8f) compared to those in vehicle controls (Supplementary Fig. 9b). Stimulus-response and reward-retrieval latencies were unaltered (Supplementary Fig. 1n.o).

This pattern of deficits mimics the effect of non-subunit-selective NMDAR blockade on reversal in rats\textsuperscript{48} and the phenotype of the Grin2b\textsubscript{CxNull} mutants, and demonstrates that blockade of GluN2B in OFC is sufficient to disrupt choice shifting. This does not exclude a contribution from other PFC regions; for example, our c-Fos analysis indicated that the prelimbic area also showed activation during choice-shifting stages, and exposure to stress facilitates choice learning in a manner prevented by prelimbic area infusion of BDNF\textsuperscript{24}. Restricted re-expression of NMDARs in the prelimbic area also rescues impaired associative learning in mice lacking NMDARs on inputs to midbrain dopaminergic neurons\textsuperscript{49}. In turn, NMDARs expressed on dopaminergic neurons are crucial for habit behavior\textsuperscript{50}. Collectively, these various findings indicate that NMDARs and, per the present study, specifically GluN2B, are indispensable at several nodes in the corticostriatal circuitry subserving habit and other cognitive processes. An important avenue for future studies will be how NMDARs regulate the function and plasticity of circuits to integrate these various nodes of the system and regulate emergent behaviors such as choice.

DISCUSSION

Using several approaches, this study provides convergent support for a dynamic role of corticostriatal circuitry in choice learning and shifting. Our data also identify a crucial molecular mechanism subserving these functions by providing evidence of a double dissociation between OFC GluN2B in choice shifting and DS GluN2B in choice learning.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We are very grateful to G. Schoenbaum and A. Kravitz for discussions of the in vivo recording results and to G. Luo for mutant genotyping. J.L.B., C.G., T.W., M.I.D., R.A.D., M.P., O.G.-C., D.M.L. and A.H. were supported by the NIAAA Intramural Research Program, L.M.S. and T.J.B. were supported by The Wellcome Trust, Z.J. and K.N. were supported by the National Institute of Mental Health Intramural Research Program.

AUTHOR CONTRIBUTIONS

J.L.B. conducted behavioral, c-Fos, in vivo electrophysiological and in situ hybridization experiments and contributed to writing the manuscript; R.A.D., C.G. and M.P. conducted behavioral experiments; T.W. conducted the slice...
Subjects. Male C57BL/6J mice were obtained from The Jackson Laboratory (Bar Harbor, ME). Grin2bloxPlox mice were generated as previously described23. Briefly, the Grin2b gene was disrupted by inserting a loxp site downstream of the 599-bp exon 3 or exon 5 (depending on transcript) and a neomycin resistance gene cassette flanked by two loxp sites upstream of this exon. The 129 strain was used as the embryonic stem cell donor and C57BL/6J was used for blastocysts and as the genetic background for backcrossing. Grin2blox+mice were crossed with (C57BL/6J-congenic) transgenic mice expressing either Cre recombinase driven by the Camk2α promoter (T29-1 line) or Cre recombinase driven by the Rgyr promoter. With each Cre mutant line, Cre+ hemizygous Grin2blox mice (that is, Grin2b-exicised) were crossed with Cre+ Grin2b (non-exicised control) mice to produce mutant and control littersmates for experimentation. Male and female mutants were used. Mice were housed in same-sex groupings (two to four per cage, except for cannulated and implanted mice, which were one per cage) in a temperature- and humidity-controlled vivarium under a 12 h light/dark cycle (lights on 0600 h) and tested during the light phase. The number of mice used in each experiment is given in the figure legends. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in previous publications24. Experimenters were blind to all experimental conditions until all data were collected. Unless otherwise specified, mice were randomly assigned to experimental groups. All experimental procedures were performed in accordance with the National Institutes of Health Guide for Care and Use of Laboratory Animals and were approved by the local NIAAA Animal Care and Use Committee.

Operant apparatus. All operant behavior was conducted in a chamber measuring 21.6 x 17.8 x 12.7 cm (model no. ENV-307W, Med Associates, St. Albans, VT) housed in a sound- and light-attenuating box (Med Associates, St. Albans, VT). The grid floor of the chamber was covered with solid Plexiglas to make it easy for the mouse to walk. A pellet dispenser delivering 14 mg dustless pellets (F05684, Bioserv, Frenchtown, NJ) into a magazine was located at one end of the chamber. At the opposite end of the chamber there was a touch-sensitive screen (Light Industrial Metal Cased TFT LCD Monitor, Craft Data Limited, Chesham, UK), a house light and a tone generator. The touchscreen was covered by a black Plexiglas panel that had 2 x 5 cm windows separated by 0.5 cm and located 6.5 cm from the floor of the chamber. Stimuli presented on the screen were controlled by custom software (MouseCat; L.M.S.) and visible through the windows (one stimulus per window). Nose pokes at the stimuli were detected by the touchscreen and recorded by the software.

Stage analysis of discrimination and reversal performance. Pairwise visual discrimination and reversal learning was assessed in C57BL/6J mice (8–10 weeks at beginning of testing) as previously described22,24,42,51. Mice were first slowly reduced and then maintained at 85% free-feeding body weight. Before testing, mice were acclimated to the 14-mg pellet food reward by provision of about ten pellets per mouse in the home cage for 1–3 d. Mice were then acclimated to the operant chamber and to eating out of the pellet magazine by being placed in the chamber for 30 min with pellets available in the magazine. Mice eating ten pellets within 30 min were moved on to autoshaping.

Autoshaping consisted of variously shaped stimuli being presented in the touchscreen windows (1 per window) for 10 s (inter-trial interval (ITI), 15 s). The disappearance of the stimuli coincided with delivery of a single pellet food reward, concomitant with presentation of stimuli (2-s, 65-dB auditory tone and illumination of pellet magazine) that served to support instrumental learning. Pellet retrievals from the magazine were detected as a head entry and, at this stage of pretraining, initiated the next trial. To encourage screen approaches and touches at this stage, nose pokes at the touchscreen delivered three pellets into the magazine.

Mice retrieving 30 pellets within 30 min were moved on to pretraining. During pretraining, mice first obtained rewards by responding to a (variously shaped) stimulus that appeared in one of the two windows (spatially pseudorandomized) and remained on the screen until a response was made (‘respond’ phase). Mice retrieving 30 pellets within 30 min were next required to initiate each new trial with a head entry into the pellet magazine. In addition, responses at a blank window during stimulus presentation now produced a 15-s timeout (signaled by extinction of the house light) to discourage indiscriminate screen responding (‘punish’ phase). Errors were followed by correction trials in which the same stimulus and left/right position was presented until a correct response was made. Mice making ≥75% (excluding correction trials) of their responses at a stimulus-containing window over a 30-trial session were moved on to discrimination.

For discrimination learning, two novel, approximately equiluminous stimuli were presented in a spatially pseudorandomized manner over 30-trial sessions (15± ITI). Responses at one stimulus (correct) resulted in reward; responses at the other stimulus (incorrect) resulted in a 15± timeout (signaled by extinction of the house light) and were followed by a correction trial. Stimuli remained on screen until a response was made. Designation of the correct and incorrect stimuli was counterbalanced across groups. Mice were trained to a criterion of ≥85% correct responding (excluding correction trials) over two consecutive sessions.

Reversal training began on the session after discrimination criterion was attained. Here the designation of stimuli as correct versus incorrect was reversed for each mouse. Mice were trained on 30-trial daily sessions (as for discrimination) to a criterion of ≥85% correct responding (excluding correction trials) over two consecutive sessions.

The following dependent measures were taken during discrimination and reversal: percentage correct responding ((correct responses/30 session-trials) x 100), errors (incorrect responses made), correction errors (correction trials made), time to response (time from trial initiation to touchscreen response) and time to reward (time from touchscreen response to reward retrieval). In addition, for the initial experiment characterizing the main task performance stages in C57BL/6J mice (see Fig. 1), the average length of strings of consecutive errors or correct responses was also measured. Here and elsewhere in the study, behavior measures were compared across the five task performance stages using analysis of variance (ANOVA) followed by a statistically conservative post hoc test (Newman-Keuls). Data met the assumptions of normality and homogeneity of variance for analysis with parametric tests. No attempt was made to exactly equate the number of animals in each experimental group.

Regional mapping of neuronal activation. This experiment mapped patterns of regional neuronal activation associated with choice learning and shifting, through immunocytochemical staining for the immediate-early gene c-Fos. Separate groups of C57BL/6J mice were trained to one of five possible stages of discrimination or reversal performance (Fig. 1a): Dnew, first session of discrimination (performance at chance); Dend, final session of discrimination (performance at criterion); Rnew, first session of reversal (performance highly perseverative); Rend, reversal session when performance was around chance (50% correct); R0%ov, final session of reversal (performance at criterion). Two hours after the start of the final session, mice were deeply anesthetized with an overdose of ketamine/xylazine (200 mg/kg) and transcardially perfused with 4% formaldehyde in PBS (pH 7.4). Brains were removed and postfixd at 4°C overnight in 4% formaldehyde in PBS and then rinsed in PBS for 2–4 h.

Fifty-micrometer-thick coronal sections were cut into PBS on a vibratome and processed for c-Fos immunoreactivity as based on methods previously described52. Briefly, sections were permeabilized in PBS with 0.2% Triton X-100 (PBS-T) for 1 h, blocked with 5% BSA in PBS-T for 4 h and incubated on a platform rocker overnight at 4°C with rabbit polyclonal anti-Fos (sc-52, Santa Cruz Biotechnology, Santa Cruz, CA) diluted to 1 mg/ml in PBS-T. Negative controls were prepared by omitting the primary antibody. Sections were washed three times for 1 h in PBS-T and incubated overnight at 4°C with Alexa 488 goat anti-rabbit (A11034 Invitrogen, Carlsbad, CA) diluted 1:1 000 in PBS-T. They were then washed three times for 1 h in PBS-T and mounted.

Sections were imaged with a 32×, 0.4 NA objective using a Zeiss Axiowert 200 epifluorescence microscope (482/35 excitation filter, 505 dichroic, 540/25 emission filters). Images were collected using the same exposure time (determined by control signal intensity) using a CCD camera (Axiocam) combined with the Axiovision software (Carl Zeiss, Oberkochen GER). Care was taken not to repeatedly expose the sections in order to reduce photobleaching, and sections were stored in the dark during all procedures beginning with the secondary antibody treatment. Images were then adjusted using ImageJ (version 1.38x) by background subtraction and threshold adjustment, constant for each region. Circular particles larger than 20 μm in diameter were automatically counted and recorded. For each region, c-Fos was an average of counts from a 360 x 460 μm region, measured in duplicate sections.
Thirteen brain regions were analyzed: agranular insular cortex (AP = +2.10, ML = ±2.25, DV = −3.25), lateral orbitofrontal cortex (AP = +2.10, ML = ±1.50, DV = −3.25), medial orbitofrontal cortex (AP = +2.10, ML = ±0.25, DV = −3.25), primary motor cortex (AP = +2.10, ML = ±2.00, DV = −1.75), prelimbic cortex (AP = +1.54, ML = ±1.33, DV = −2.50), infralimbic cortex (AP = +1.54, ML = ±1.33, DV = −3.00), the dorsal CA1 subregion of the hippocampus (AP = −1.46, ML = ±1.00, DV = −1.50), dorsomedial striatum (AP = +1.10, ML = ±0.80, DV = −3.00), dorsolateral striatum (AP = +1.10, ML = ±2.10, DV = −3.00), nucleus accumbens shell (AP = +1.54, ML = ±0.50, DV = −4.80), nucleus accumbens core (AP = +1.54, ML = ±0.75, DV = −4.50), basolateral nucleus of the amygdala (AP = −1.46, ML = ±3.00, DV = −4.65) and central nucleus of the amygdala (AP = −1.46, ML = ±2.40, DV = −4.30). The number of c-Fos-positive cells was compared across the five task performance stages using ANOVA followed by Newman-Keuls post hoc tests.

Excitotoxic dorsolateral striatum lesions. This experiment assessed the functional contribution of the dorsolateral striatum to choice learning by making bilateral lesions of this region, before discrimination training. After completing pretraining, C57BL/6j mice were assigned to lesion or sham groups by matching to trials to complete pretraining. Mice were anesthetized with isoflurane and placed in a stereotaxic alignment system (Kopf Instruments, Tujunga, CA). The fiber-of-passage–sparing excitotoxic NMDA or saline was infused into four sites (two sites per hemisphere: one anterior and one posterior) at the coordinates AP = +1.18, +0.22, ML = ±2.4, ±3.0, DV = −2.5. After 7–10 d of recovery, body weight reduction resumed and mice were given postsurgery reminder sessions to ensure retention of pretraining criterion. Discrimination training was conducted as above. Behavioral measures (as above) were compared between sham and lesion groups using two-tailed (as elsewhere) Student’s t-test.

Given the role of dorsolateral striatum in controlling motor functions, after the completion of discrimination testing, C57BL/6j mice were assigned to lesion or sham groups by matching to trials to complete pretraining. Mice were anesthetized with isoflurane and placed in a stereotaxic alignment system (Kopf Instruments, Tujunga, CA). The fiber-of-passage–sparring excitotoxic NMDA or saline was infused into four sites (two sites per hemisphere: one anterior and one posterior) at the coordinates AP = +1.18, +0.22, ML = ±2.4, ±3.0, DV = −2.5. After 7–10 d of recovery, body weight reduction resumed and mice were given postsurgery reminder sessions to ensure retention of pretraining criterion. Discrimination training was conducted as above. Behavioral measures (as above) were compared between sham and lesion groups using two-tailed (as elsewhere) Student’s t-test.

Here changes in synaptic plasticity in dorsal striatum neurons were analyzed, ex vivo, as a function of choice learning and shifting. C57BL/6j mice were trained to one of the five discrimination or reversal stages defined above. Two hours after the start of the final session, mice were anesthetized by halothane or isoflurane inhalation. The brain was rapidly removed and placed in ice-cold cutting solution: in mM, 194 sucrose, 30 NaCl, 4.5 KCl, 26 NaHCO3, 1.2 NaH2PO4, 10 glucose, pH 7.3, equilibrated with 95% O2 and 5% CO2 at 34 °C. Field potentials were recorded with a bipolar twisted electrode (200 µm) made by passing 100 µA through the electrodes for 20 s using a current stimulator (S48 Square Pulse Stimulator, Grass Technologies, West Warwick, RI). Brains were removed and 50-µm coronal sections were cut with a vibratome (Classic 1000 model, Vibratome, St. Louis, MO) and then stained with cresyl violet. Estimates of the maximum and minimum extent of lesions were estimated with reference to a mouse brain atlas and the aid of a microscope. Mice with lesions outside the DLS were excluded from the analysis.

Performance transfer from operant training to motor learning. Training on cognitive and motor tasks that heavily recruit certain brain regions can facilitate performance on separate tasks that are mediated by the same regions54–56. This experiment tested for performance transfer from choice testing to a DLS-mediated motor learning task45. Separate groups of C57BL/6j mice were trained (as above) to either (i) pretraining criterion (PTC), (ii) discrimination criterion (DC), or (iii) reversal criterion (RC). The aim was to test the differential effects of previous experience with operant training but no-choice learning, training + choice learning or training + choice learning + choice shifting, rather than the accumulated amount of operant testing or reinforcement. Therefore, the three groups were matched for the total number of sessions from the beginning of operant training until the motor learning test (26.1 ± 0.7 sessions) by giving the pretraining group an additional 20.0 ± 0.7 sessions after reaching criterion, and the discrimination groups an additional 6.8 ± 1.3 sessions.

One day after the completion of operant testing, motor learning was assessed using the accelerating rotarod, as previously described57. Mice were placed on a 7-cm-diameter dowel (Med Associates rotarod model ENV-577) rotating at 4 r.p.m. and accelerating at a constant rate of 8 r.p.m. min−1 up to 40 r.p.m. The latency to fall to the floor 10.5 cm below was recorded by photocell beam. Mice were given ten consecutive training trials (30-s inter-trial interval), with a cutoff latency of 300 s for a given trial. Motor learning was calculated as the difference in latency from trial 1 to 10. Groups were compared using ANOVA.

In vivo dorsal striatum neuronal recordings. The specific role of dorsal striatum in choice behavior was investigated via in vivo neuronal recordings made during choice learning and shifting. After completing pretraining, C57BL/6j mice were anesthetized with isoflurane and placed in a stereotaxic alignment system (Kopf Instruments, Tujunga, CA) for implantation of a microelectrode array48. The array (fabricated by Innovative Neurophysiology, Durham, NC) comprised 16 tungsten microelectrodes 35 µm in diameter arranged into 2 rows of 8 (150 µm spacing between microelectrodes within a row, 1,000 µm spacing between rows). One row was placed in lateral dorsal striatal and the other central-medial (Fig. 3a), with rows running lengthwise anterior to posterior (targeting coordinates for center of array: AP = +0.75, ML = ±1.60, DV = −2.75). After 7–10 d of recovery, body weight reduction resumed and mice were given postsurgery reminder sessions to ensure retention of pretraining criterion. Discrimination training, followed by reversal training, was conducted as above.

Neuronal activity was recorded using the Plexon Inc (Dallas, Texas) Multichannel Acquisition Processor during one session corresponding to each of the five performance stages described above. Extracellular waveforms exceeding a set voltage threshold were digitized at 40 kHz and stored on a PC. Waveforms were manually sorted using principal component analysis of spike clusters and visual inspection of waveform and inter-spike interval49. Neuronal activity was timestamped for 5-s neuron event epochs (trial initiation, prechoice, choice, reward receipt), separately for correct and error trials. Spike and timestamp information was integrated and analyzed using NeuroExplorer (NEX Technologies, Littleton, MA).

To measure the average activity of the recorded population, activity for each neuron was Z-score normalized to the average firing rate of that neuron across all events and presented in 50-ms time bins. Changes in Z-scored firing across performance stages were analyzed using two-way ANOVA, with repeated measures for time. Z-scored firing of individual cells is also shown. To examine the event-related firing of individual units, units with Z-scores either >1.0 or <−1.0 were designated as event-related. The percentage of recorded units classified as event-related were calculated at each 50-ms time bin. The Pearson’s R correlation between firing during correct responses and reward-retrieval was measured by summing the Z values during the entirety of each epoch.

At the completion of testing, array placement was verified by electrolytic lesions made by passing 100 µA through the electrodes for 20 s using a current stimulator (S48 Square Pulse Stimulator, Grass Technologies, West Warwick, RI). Brains were removed and 50-µm coronal sections were cut with a vibratome (Classic 1000 model, Vibratome, Bannockburn, IL) and stained with cresyl violet. Placement was estimated with reference to a mouse brain atlas and the aid of a microscope. Mice with placements outside the DLS were excluded from the analysis.

Ex vivo dorsal striatum slice recordings. Here changes in synaptic plasticity in dorsal striatum neurons were analyzed, ex vivo, as a function of choice learning and shifting. C57BL/6j mice were trained to one of the five discrimination or reversal stages defined above. Two hours after the start of the final session, mice were anesthetized by halothane or isoflurane inhalation. The brain was rapidly removed and placed in ice-cold cutting solution: in mM, 194 sucrose, 30 NaCl, 4.5 KCl, 26 NaHCO3, 1.2 NaH2PO4, 10 glucose, pH 7.3, equilibrated with 95% O2 and 5% CO2, Osin 320. Coronal sections (250 µm thick) were cut with an Integraslice 7500 vibratome (Campden Instruments, Loughborough, UK) and incubated in ice-cold modified aCSF and transferred immediately to normal aCSF: in mM, 124 NaCl, 4.5 KCl, 1 MgCl2, 26 NaHCO3, 1.2 NaH2PO4, 10 glucose, 2 CaCl2, pH 7.3, equilibrated with 95% O2 and 5% CO2 at 34 °C for 30 min and then at room temperature for at least 30 min before the experiment. Slices were maintained at 28–32 °C during the experiment.

Extracellular field recordings were performed with microelectropettes (2.5–5 MΩ) filled with 1 M NaCl solution, as previously described45. Field potentials were evoked by constant current stimulation delivered through a bipolar twisted Teflon-coated tungsten electrode placed in the striatum. Individual stimulus pulses of 0.01 ms duration were generated by a Grass 44 stimulator through a Grass optical isolator. Input–output relationship was examined by stimulating at intensities from 0.1 to 1.5 mA (two stimuli at each intensity), with an interstimulus interval of 30 s, and recording population spike (PS) amplitude. To measure
PS amplitude before and after high-frequency stimulation (HFS), responses to stimuli (1/30 s) that evoked a PS that was approximately half the amplitude of the maximal evoked response were recorded for at least 10 min before the first HFS trains (baseline period) and for 20 min after each set of trains. LTD was induced via HFS consisting of three 1-s trains of 100 pulses (each pulse 0.01 ms) delivered at 100 Hz (10-s inter-train interval) with the stimulus intensity set at 1.5 mA during the trains. The peak amplitude of the negative-going PS was measured relative to the positive-going field potential component just before PS onset, using cursors in Clampfit v8.0.

**Corticostratial Grin2b deletion.** The contribution of corticostrial GluN2B circuits was first assessed in mutant mice lacking GluN2B in neurons in these brain regions, as well as dorsal CA1 hippocampus. This was achieved by crossing Grin2b<sup>flx</sup> mice with Camk2a-driven cre transgenic mice and testing the progeny at an age when the deletion has spread from cortex and CA1 hippocampus to striatum (see main text). *Grin2b<sup>CstrNull</sup>* mice and age-matched controls were tested for choice discrimination (age range, 28–36 weeks) and then choice reversal (age range, 30–42 weeks) as above.

**Western blot.** To confirm and quantify loss of GluN2B, another set of 11-month-old mice were used to quantify GluN2B protein by western blot. Tissue from mPFC, dorsal hippocampus and dorsal striatum was dissected from frozen brains with a 2-mm-diameter micropunch. Tissue was homogenized by sonication on ice in lysis buffer (10 mM sodium phosphate, pH 7.5, 20 mM Tris-HCl, pH 8.0, 2 mM EDTA, 2 mM EGTA, 1 mM NaF, 1 mM Na<sub>2</sub>VO<sub>4</sub>, 2 mM DTT) with protease inhibitor cocktail (Sigma–Aldrich, St. Louis, MO). Supernatant was obtained by centrifugation at 1,700g for 10 min at 4 °C. Protein concentration was determined with Micro BCA protein assay kit (Pierce, Rockford, IL). Protein extracts were denatured in 2× Laemmli buffer (Sigma–Aldrich, St. Louis, MO) and 20 µg protein per well were loaded for SDS-PAGE.

After electrophoresis, the proteins were transferred to Immobilon-P membranes (Millipore, Billerica, MA), blocked with 5% milk in TBS with 0.05% Tween-20, and blotted with rabbit polyclonal anti-GluN2B (1:2,000, Millipore, catalog number AB1557P) followed by the HRP-labeled anti-rabbit secondary antibody and blotted with Immobilon-P (Millipore, Billerica, MA) or an equivalent volume of saline vehicle was infused bilaterally into dorsolateral striatum before the first three reversal sessions; that is, when choice shifting was most strongly taxed. Solutions were infused with the aid of dorsolateral striatal glun2B antagonism.

**Dorsolateral striatal GluN2B antagonism.** This series of experiments were conducted to delineate the role of GluN2B-expressing neurons in dorsolateral striatum in choice shifting, choice learning and the expression of learned choice behavior, during reversal testing.

**Effects on choice shifting.** After attaining discrimination criterion, C57BL/6J mice were assigned to drug or vehicle groups by matching to trials to complete discrimination. Mice were anesthetized with isoflurane and placed in a stereotaxic alignment system (Kopf Instruments, Tujunga, CA). Guide cannulas (Plastics One, Roanoke, VA) were inserted bilaterally (AP +0.85, ML ±0.15, DV -2.5) and stabilized with dental cement. After 7–10 d of recovery, body weight reduction and mice were given post-surgery reminder sessions to ensure retention of discrimination criterion. Reversal testing was conducted as above, except that Ro 25-6981 (2.5 µg per side in a volume of 0.5 µl) (Sigma–Aldrich, St. Louis, MO) or an equivalent volume of saline vehicle was infused bilaterally into dorsolateral striatum before the first three reversal sessions; that is, when choice shifting was most strongly taxed. Solutions were infused with the aid of a dual syringe pump (Harvard Apparatus, Holliston, MA) over 5 min through injectors that projected into the tissue 1 mm beyond the tip of the cannula. The injectors were left in place for 3 min to ensure full diffusion. Mice were tested 15 min later. From the fourth session onwards, reversal testing continued to criterion with no further infusions.

**Effects on choice learning.** A set of naive C57BL/6J mice were trained to discrimination criterion and implanted with guide canulas, as above. To test the effects of dorsolateral striatal GluN2B blockade when choice learning was evident, mice were trained to chance performance and Ro 25-6981 vehicle was infused...
before the next three reversal sessions. Thereafter, reversal testing continued to criterion without further infusions.

Effects on learned choice expression. Another set of naive C57BL/6J mice were trained to discrimination criterion and implanted with guide cannulas, as above. These mice were trained to reversal criterion and then infused with drug or vehicle over another three sessions, to test whether dorsolateral striatal GluN2B blockade affects the expression of the choice behavior, once learned. After the infusion sessions, reversal testing continued for another three no-infusions sessions to ensure retention of learned choice.

For all three experiments, trials per session were doubled from 30 to 60 to minimize the number of potentially tissue-damaging infusions. The sum of errors and correction errors, and the average stimulus-response and reward-retrieval times, during the infusion sessions were compared between drug and vehicle groups using Student’s t-test. In addition, the effect of drug treatment on choice accuracy on each of the three infusion sessions and subsequent three (no-infusion) sessions was analyzed using ANOVA, with repeated measures for session, followed by Newman–Keuls post hoc tests.

At the completion of testing, brains were removed and 50-μm coronal sections were cut with a vibratome (Classic 1000 model, Vibratome, Bannockburn, IL) and stained with cresyl violet. Cannula placements were estimated with reference to a mouse brain atlas and the aid of a microscope. Mice with placements outside the DLS were excluded from the analysis.

Orbitofrontal cortical GluN2B antagonism. These experiments were conducted to examine the contribution of GluN2B-expressing neurons in orbitofrontal cortex to choice shifting and choice learning, using the same pharmacological approach as described for dorsolateral striatum. Procedures were the same as above with the guide cannulas bilaterally targeted to orbitofrontal cortex (AP +2.80, ML ± 1.35, DV –1.80).

Two experiments were conducted, in separate cohorts of naive C57BL/6J mice. In the first experiment, Ro 25-6981 (1.0 μg per side in a volume of 0.2 μl) or an equivalent volume of saline vehicle was infused before the first three sessions of reversal, to assess effects on choice shifting. In the second experiment, to assess effects on choice learning, mice were trained to chance performance and infusions made before the next three sessions. Given the absence of treatment effects on choice learning (see main text), the effects on the expression of learned choice behavior was not assessed. Behavior was analyzed and cannula placements verified, as described above for dorsolateral striatal pharmacological experiments.

51. Izquierdo, A. et al. Genetic and dopaminergic modulation of reversal learning in a touchscreen-based operant procedure for mice. Behav. Brain Res. 171, 181–188 (2006).
52. Hefner, K. et al. Impaired fear extinction learning and cortico-amygdala circuit abnormalities in a common genetic mouse strain. J. Neurosci. 28, 8074–8085 (2008).
53. Karlsson, R.M., Tanaka, K., Heilig, M. & Holmes, A. Loss of glial glutamate and aspartate transporter (excitatory amino acid transporter 1) causes locomotor hyperactivity and exaggerated responses to psychotomimetics: rescue by haloperidol and metabotropic glutamate 2/3 agonist. Biol. Psychiatry 64, 810–814 (2008).
54. Dahlín, E., Neely, A.S., Larsson, A., Backman, L. & Nyberg, L. Transfer of learning after updating training mediated by the striatum. Science 320, 1510–1512 (2008).
55. Erickson, K.I. et al. Training-induced functional activation changes in dual-task processing; An fMRI study. Cereb. Cortex 17, 192–204 (2007).
56. Olesen, P.J., Westerberg, H. & Klingberg, T. Increased prefrontal and parietal activity after training of working memory. Nat. Neurosci. 7, 75–79 (2004).
57. Boyce-Rustay, N. & Holmes, A. Ethanol-related behaviors in mice lacking the NMDA receptor NR2A subunit. Psychopharmacology (Berl.) 187, 455–466 (2006).
58. Holmes, A. et al. Chronic alcohol remolds prefrontal neurons and disrupts NMDAR-mediated fear extinction encoding. Nat. Neurosci. 15, 1359–1361 (2012).