Supporting Information

Schreiweis et al. 10.1073/pnas.1414542111

SI Materials and Methods

Animals. A total of 303 forkhead box P2, *Foxp2*$^{hum/hum}$, mice [5H10 line (1); P21–P53 for in situ electrophysiology; 1.8–15.2 mo for other experiments] and WT littermates were used. They were balanced for genotype and sex for behavioral tests (n = 160), for gene expression assays (n = 23), for dopamine measurements (n = 32), and for in situ electrophysiology (n = 88). Behavioral procedures were approved by the committee on animal care at the Massachusetts Institute of Technology. Experiments based in the United Kingdom were performed in accordance with the United Kingdom Animals (Scientific Procedures) Act of 1986. Electrophysiological procedures followed guidelines of the Max Planck Institute for Evolutionary Anthropology and federal regulations of Saxony, Germany.

Behavioral Experiments. ***Rotarod and tilted running wheel.*** A standalone accelerating rotarod was custom modified to allow the use of faster acceleration speeds than the standard model. The rod was set to accelerate from 8 rpm to 80 rpm over a 300-s time period. Mice were placed on the rod, and trials were deemed to have started when the rod began to turn. Trials ended when mice fell from the rod or after 300 s elapsed. Mice were trained for three consecutive days, with one daily session consisting of 10 trials separated by 300-s intertrial intervals spent in their home cages. Tilted running wheel experiments were conducted as previously described (2). Time spent running, average speed of running, as well as number and length of single running bouts were automatically monitored and recorded by a computer during 24-h periods.

T-maze task. The black acrylic T-shaped maze consisted of a long alleyway (55.2 cm × 3.2 cm) with two short alleys (39.4 cm × 3.2 cm), flanked by outward-sloping walls of a 60° angle. The floor was taped with a flat black tape to take off any possibly irritating shine and to provide a nonslippery walking ground for the animals. A sliding door prevented the mouse from leaving the start compartment until trial start, which was indicated by a warning click. Circular wells (1 cm in diameter, ∼1 mm deep, located 2 cm away from the end of right and left choice arms) could be filled from outside the maze, out of sight of the mouse, with chocolate milk delivered via a blunt-end metal syringe needle. The audio speaker, delivering the warning click, was located outside the apparatus, above the choice point of the maze. The tactile inserts were flat black to limit reflections and stress during training. Light conditions were dimmed during habituation and training periods to signal the start of the session and to create a less stressful environment while permitting vision of spatial cues; lights were turned up fully again to mark the end of the session. The cue-enriched environment contained numerous cues, among them the door, racks, and black posters on the walls to left side of the T-maze (seen from the start compartment, facing the long alleyway), as well as a bright monitor that emitted a 75-kHz ultrasound tone and was located diagonally behind the experimenter. To simulate a cue-deprived environment, we removed all cues we were able to remove, i.e., the black posters and the monitor.

A subset of mice was trained on a T-maze task as previously described (3). These mice were food-restricted (75–85% of free-feeding weight) and were first habituated to the apparatus and reward (chocolate milk). During the daily 40-trial training and testing session in the T-maze, mice received reward if they reached the correct goal location as instructed by tactile conditional cues (rough or smooth floor surface). Performance accuracy and running times of each trial were measured.

Cross-maze task. The cross-maze apparatus resembled the measurements and the angled design of the T-maze apparatus. However, it contained an additional long arm, which was closed off by an equally 60° angled Plexiglas wall so that the mouse was exposed to the choice of only two short arms at a time. The maze was constructed of white acrylic floor material, and equipped with transparent Plexiglas walls and a transparent door to separate off the start compartment to ensure good visibility of all provided cues. The floor of the cross-maze was elevated 40 cm above the ground and mounted on a revolvable construction. The apparatus was located in an test room adjacent to the T-maze room, which was equipped with a black squared cubicle (216 cm × 216 cm) of ripstop nylon panels, which were arranged along a squared framed metal construction (height, 220 cm from floor). A white ceiling of the same material was mounted to the metal frame to ensure an indirect light source providing equal light intensity inside the cubicle. There were again numerous cues in the room such as the door, light switches on the walls, and a table, as well as the nylon walls of the northwest, north, and southeast black nylon panels that were rolled down. Additionally, white cues in different shapes (triangle, cross) were hanging from black nylon strings in front of the black panels of the cubicle. Light conditions were kept at a moderate level to relax the animals and to ensure good visibility of all presented spatial cues. With a probe trial day in the egocentric learning task, during which the cross-maze apparatus was turned 90° clockwise to its original position, we confirmed that mice had acquired a procedural strategy and not learned to go to two different locations depending on the start arm (data not shown).

Mice were food-deprived and habituated as those in the T-maze task. Each mouse was placed on either of two opposing arms of the maze (north or south), and the other departure arm was closed. They were rewarded in the place-based version for a specific place (east or west), and in the response-based version for a particular turn (right or left turn). Mice received 10 daily trials, and performance accuracy was measured.

Statistical analysis. Behavioral data from the T-maze and cross-maze experiments were analyzed by using repeated-measures ANOVA (RMA) with training day, genotype, and sex as predictor variables. We used the Greenhouse–Geisser correction to adjust the degrees of freedom when the sphericity assumption was violated. Additionally, we applied a generalized linear mixed model (GLMM) approach on data from all behavioral experiments.

Detailed Statistical Analysis of Behavioral Data Using a GLMM. ***Accelerating rotarod.*** To test whether latencies to fall were influenced by genotype, day, trial, or age of the mice, we used a simple and a more differentiated GLMM approach (4, 5) into which we included these predictors as fixed effects and batch and subject as random effects. The models were fitted in R (6) using the function "lmer" of the R package lme4 (7). Before analyzing the data, we checked whether the assumptions of normally distributed and homogeneous residuals were fulfilled by visually inspecting a histogram of the plotted residuals, a Q-Q plot, and the residuals plotted against fitted values (all indicating that there were no obvious deviations from these assumptions). The significance of the full model including all predictors (genotype, trial, training day) as main effects, as well as all their interactions up to the third order and random effects (batch, subject), compared with the null model not comprising the factor genotype or

its interactions with other effects, was established by using a likelihood ratio test (R function "anova" with argument test set to "Chisq") (8). To achieve a more reliable *P* value, we fitted the models by using maximum likelihood (9).

Tilted running wheel. All parameters were analyzed by using the same GLMM and R packages as described earlier for the analysis of the rotarod assay. Here, day and genotype served as fixed effects; age and weight at beginning of testing were included as further fixed effects as well as in interactions with day and genotype. Furthermore, we corrected for individual variation in the change of the response as well as potential genotype differences in the batches (random slopes), which consistently remained in the built models (10). Again, before running the models, we checked the assumptions as described earlier and found that assumptions were not fulfilled for the bout length, bout number, and average speed of running. To amend this violation, length and number were transformed, applying the square root after subtracting the minimum value from all individual values, and the average running speed was squared. We compared the full model including all predictors with a model lacking genotype as factor of interest (likelihood ratio test with R function "anova," argument test set to "Chisq") and fitted models by using maximum likelihood.

T-maze task. We performed a GLMM (11) as described for the rotarod test. In the analyses of spatial cue-enriched and spatial cue-deprived setups, we included training day, genotype, and sex as fixed effects as in the RMA, and additionally trial as predictor of interest as well as age, stimulus type (rough or smooth), and maze-arm (arm 1 or 2) as control variables. Furthermore, we controlled for potential effects of batch and subject (random intercepts). We tested a model with day, trial, and genotype in all interactions up to the third order. Additionally, we included age, sex, stimulus type, and maze-arm and their interactions with all terms involving the test effect to control for their potential effects [model: training day * trial number * genotype * (sex + age + stimulus type + maze arm)]. Furthermore, random slope terms were included to control for genotype effects potentially differing between batches (random slope term encoded in R using the following construct: 1 + genotype | batch) or individual differences in performance across days or trials (1 + maze arm + training day * trial | subject). We first tested whether the full model containing all the mentioned factors was significant against a null model that lacked the factor genotype (cue-enriched environment, GLMM, df = 5, $P < 2.7 \times 10^{-14}$; cue-deprived environment, GLMM not significant) before reducing the model stepwise to reach the final model containing only the factors that contribute significantly to explain the observed data.

Cross-maze task. The same GLMM procedure as used for the T-maze was applied to the switch experiment in the cross-maze apparatus. We analyzed the responses to informative trials, i.e., the 50% of trials in which place-based rule and the response-based rule strategies led to opposite actions. For instance, a mouse that had learned to go to the location arm 1 will still be rewarded in 50% of the cases in the procedural task because the required turn will overlap with the specific location in half the trials.

For the GLMM analysis, the same functions and packages as those for the analyses of the T-maze task were applied. The contributions of fixed and random effects, as well as individual interactions between these effects, were tested by comparing a model including the term to be tested (main or interaction effect) with a model not containing the term (likelihood ratio test with R function "anova," argument test set to "Chisq," and models fitted by using maximum likelihood). Before testing, all continuous variables [training day "switch 1–3," informative trial number (n = 1–5), and age] were z-transformed, which, apart from the listed continuous variables, contained genotype as fixed effect and batch and subject as random effects.

Laser-Capture Microdissection and RNA Sequencing. Mice were 8 wk of age, and six genotype pairs per sex matched for litter were processed in random order. Brains were isolated, flash-frozen in −20 °C isopentane, and stored at −80 °C. Coronal cryosections (45 μm, −20 °C in a Microm HM 550 cryostat; Thermo Fisher) were put onto membrane slides (MembraneSlide NF 1,0 PEN; Zeiss) and incubated for 10 min at −20 °C in 70% (vol/vol) EtOH. Slides were stained for 1 min in 0.2% toluidine at room temperature and then washed three times in 100% EtOH. Slides were dried and stored together with Silica Gel (Roth) in slide boxes at −80 °C. Before microdissection, boxes containing the sample slides were brought to room temperature, and the samples were not processed further until humidity in the boxes was reduced by the silica gel which absorbed moisture. Striatal regions were cut with a laser (P.A.L.M. System; Zeiss), transferred manually into lysis buffer (Qiagen) and stored at −80 °C.

RNA was isolated by using an RNeasy Micro Kit (Qiagen), and RNA concentration and quality were measured on a PicoChip (2100 Bionalyzer; Agilent). We developed and applied an in-house protocol for library preparation of small amounts of RNA (15 ng total RNA per sample), which included mRNA isolation, RNA fragmentation, and cDNA synthesis similar to the TrueSeq protocol, and made indexed libraries from these cDNAs as described (12) The 48 libraries were pooled, and single-end 36-bp reads were sequenced on three flow cells on an Illumina GAIIx. Reads were processed as described previously (12), and one animal ($Foxp2^{hum/hum}$ female) was excluded as a result of contamination with cortical tissue. Gene expression analysis was performed by using the multifactor analysis of the DESeq package (13). Factors included in the generalized linear model were "Region," "Genotype," "Sex," and "Batch." false discovery rates were calculated by using the Benjamini–Hochberg correction (14). As this correction did not account for the dependency among genes, we also ran the same analysis on permutations of factors. To test the genotype effect, we permuted genotypes within matched pairs of animals, and, to test the region effect, we permuted region within each animal. $P_{\text{permutations}}$ values were calculated as the fraction of the 555 permutations that have as many or more significant genes ($P < 0.05$) as the data.

To summarize effects on the level of gene categories, we analyzed the dorsomedial and dorsolateral striatum separately and ranked genes according to a statistic that combines fold change and statistical significance (15). We then used the Wilcoxon rank-sum test implemented in the FUNC package (16) to identify Gene Ontology categories (version 3.6.2013) enriched for genes ranking low (high) in the list, corresponding to genes expressed high (low) in $Foxp2^{hum/hum}$ mice compared with $Foxp2^{wt/wt}$ mice. To estimate the overall significance of the enrichment, we used permutations of genotype as described earlier and compared the *P* value distribution of the permutations and the data as described in FUNC for the global *P* value. We used the average rank of genes in a category to calculate the enrichment.

Dopamine Content. Animals (n = 9–22 per genotype, age-matched, age between 9 and 37 wk) were decapitated under deep isoflurane anesthesia, and their brains were rapidly dissected and sliced into 1-mm slabs on ice. Biopsy specimens were taken from the dorsomedial and dorsolateral striatum by using disposable biopsy punches with 1.5 mm diameter at anterior–posterior bregma coordinates of +0.5 to −0.5, and were stored at −80 °C. Tissue samples from each region were homogenized by ultrasonication in 750 μL 0.1 N perchloric acid at 4 °C immediately after processing, and 100 μL of the homogenates were then added to equal volumes of 1 N sodium hydroxide for measuring protein content. The remaining homogenates were centrifuged at 17,000 × *g* and 4 °C for 10 min. Perchloric acid extracts were separated on a column (Prontosil 120–3-C18-SH; length 150 mm, inner diameter 3 mm; Bischoff Analysentechnik und Geräte) at

a flow rate of 0.55 mL/min. The mobile phase consisted of 80 mM sodium dihydrogen phosphate, 0.85 mM octane-1-sulfonic acid sodium salt, 0.5 mM EDTA disodium salt, 0.92 mM phosphoric acid, and 4% 2-propanol (all chemicals from Merck). Dopamine was detected by using an electrochemical detector (41000; Chromsystems Instruments and Chemicals) at an electrode potential of 0.8 V. Sample analysis was performed based on peak areas by using a computer-based chromatography data system (CSW 1.7; DataApex) in relation to the mean of two dopamine calibration solutions (0.1 M perchloric acid containing 1 mM DA injected into the HPLC system before and after sample analysis). Statistical analyses were done on log2-transformed dopamine amounts per milligram of protein normalized per region, sex, and batch.

In Situ Electrophysiology. ***Recording procedures.*** Brains of mice aged postnatal day (P)21–P53 were prepared into ice-cold sucrose-based cutting solution (85 mM sucrose, 60 mM NaCl, 3.5 mM KCl, 6 mM $MgCl_2$, 0.5 mM $CaCl_2$, 38 mM $NaHCO_3$, 1.25 mM NaH_2PO_4, 10 mM Hepes, and 25 mM glucose). Coronal slices of 250 μm thickness were cut (Vibroslice 7000smz; Campden Instruments), incubated for 30 min at 35 °C in artificial cerebrospinal fluid (120 mM NaCl, 3.5 mM KCl, 1 mM $MgCl_2$, 2 mM $CaCl_2$, 30 mM $NaHCO_3$, 1.25 mM NaH_2PO_4, and 15 mM glucose) supplemented with 5 mM Hepes, and allowed to recover for at least 40 min.

Medium spiny neurons (MSNs) were identified according to the methods used by Pawlak and Kerr (17). Their activity was recorded in the current-clamp configuration with the bridge mode enabled (EPC-10 amplifier; Patch- and Fitmaster software; HEKA). The internal solution contained K-gluconate (150 mM), NaCl (10 mM), MgATP (3 mM), GTP (0.5 mM), Hepes (10 mM), and EGTA (0.05 mM) adjusted to pH 7.3, and 310 mOsm with the liquid junction potential (15 mV) corrected online. Slices were perfused (2–3 mL/min, artificial cerebrospinal fluid, 21–24 °C) in presence of the GABA(A) receptor blocker SR-95531 (GABAzine). All solutions were continuously bubbled with carbogen (95% O_2 and 5% CO_2). All drugs were obtained from Sigma. Aliquots were diluted in water and used at the following final concentrations: SR-95531 10 μM, sulpiride 20 μM, APV 50 μM, and MK-801 1 mM.

Cortical excitatory afferents where stimulated with a saline solution-filled theta-glass electrode typically ∼100–150 μm from the MSN cell body. A bipolar voltage pulse (0.1 ms, 5–30 V) at 0.33–0.2 Hz induced subthreshold excitatory postsynaptic potentials (4–8 mV). Such low-amplitude stimulation was chosen to avoid inducing confounding voltage-gated calcium conductances present in MSNs (18, 19). Following 15 min baseline recording, MSNs were slightly depolarized to −70 mV by equivalent injection of positive current. Synaptic plasticity was then induced by a high-frequency protocol (four 100-Hz tetani, 3 s long, separated by 30 s), after which the cells were recorded under control conditions for at least another 30–40 min. For an enhanced depolarization to −15 mV in a subset of experiments, voltage-clamp was applied and stimulus duration doubled from 100 μs to 200 μs during tetani. Recordings were excluded from analysis when initial series resistance was more than 30 MΩ or input resistance changed by more than 25% during recordings. Input resistance was calculated from a voltage step of 1 mV (voltage clamp) or 50 pA (current clamp), respectively.

To determine NMDA to AMPA ratios synaptic currents were stimulated as described earlier and measured by using Cs-based electrode solution [Cs-gluconate (140 mM), NaCl (10 mM), Mg-ATP (4 mM), GTP (0.5 mM), Hepes (10 mM), and EGTA (0.05 mM)] supplemented with tetraethylammonium chloride (TEA) (5 mM) and QX-314 (5 mM, added fresh) to block confounding potassium and sodium currents. Cells were voltage-clamped at −60 and +60 mV and the ratio of the conductances was calculated after remaining voltage errors had been corrected.

Statistical analysis. Excitatory postsynaptic potential amplitudes were normalized to a mean baseline level at t −10 to 0 min. LTD magnitude of individual cells was calculated by averaging amplitudes 30–40 min after induction. Control data included seven cells per genotype of the dorsolateral striatum from a previous study (20), which were not significantly different from the cells of the according genotype recorded for this study (all $P > 0.3$ in RMA). Data were analyzed using one- or two-way ANOVA (Origin 9.0; Microcal) and RMA (SPSS version 16.00) with Greenhouse–Geisser correction implemented in SPSS to adjust the degrees of freedom when the sphericity assumption was violated.

Statistical Remarks About the Use of a RMA and a GLMM. All maze-learning data were analyzed by using a traditional RMA approach. However, such analysis does not allow for including more than one random effect (e.g., subject and batch), nor does it allow for missing cells (e.g., certain combinations of batch, genotype, and sex are missing in our data), and it may suffer from heavily unbalanced designs (e.g., different numbers of individuals per combination of genotype and sex). Most crucially, although, in an RMA with day included as a within-subjects factor (i.e., a categorical predictor), the question of learning (i.e., the degree of improvement over the succession of trials) is not addressed adequately because the analysis does not take the order of days into account and any rearrangement or relabeling of the days will leave results unaltered. Because of the various advantages of GLMMs compared with the traditionally applied RMA (4), we additionally used a GLMM approach. GLMMs allow for random effects with more complex structures (e.g., subject nested in genotype) and can deal better with unbalanced designs and empty cells. Furthermore, they allow for day being treated as a covariate (i.e., a numerical predictor) and, hence, can address hypotheses about learning directly. Furthermore, GLMMs allow for modeling responses with non-Gaussian error distribution (e.g., Poisson or binomial). This provides the opportunity to analyze a binary response (correct choice or not) (9) more appropriately than with the use of an ANOVA after having turned it into proportion correct responses (6).

1. Enard W, et al. (2009) A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell* 137(5):961–971.
2. Groszer M, et al. (2008) Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. *Curr Biol* 18(5):354–362.
3. Thorn CA, Atallah H, Howe M, Graybiel AM (2010) Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. *Neuron* 66(5):781–795.
4. Baayen RH (2009) *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R* (Cambridge Univ Press, Cambridge, UK).
5. Baayen RH, Davidson DJ, Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang* 59:390–412.
6. Jaeger TF (2008) Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *J Mem Lang* 59(4):434–446.
7. Bates D, Maechler M (2010) lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-35 (R Foundation for Statistical Computing, Vienna).
8. Dobson AJ (2002) *An Introduction to Generalized Linear Models* (Chapman and Hall/CRC, Boca Raton, FL).
9. Bolker BM, et al. (2009) Generalized linear mixed models: A practical guide for ecology and evolution. *Trends Ecol Evol* 24(3):127–135.
10. Schielzeth H, Forstmeier W (2009) Conclusions beyond support: Overconfident estimates in mixed models. *Behav Ecol* 20(2):416–420.
11. Baayen RH (2010) languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics." R package version 1.0. (R Foundation for Statistical Computing, Vienna).
12. Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* 2010(6):pdb prot5448.
13. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10):R106.

14. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc, B* 57(1):289–300.
15. Xiao Y, et al. (2014) A novel significance score for gene selection and ranking. *Bioinformatics* 30(6):801–807.
16. Prüfer K, et al. (2007) FUNC: A package for detecting significant associations between gene sets and ontological annotations. *BMC Bioinformatics* 8:41.
17. Pawlak V, Kerr JN (2008) Dopamine receptor activation is required for corticostriatal spike-timing-dependent plasticity. *J Neurosci* 28(10):2435–2446.
18. Carter AG, Sabatini BL (2004) State-dependent calcium signaling in dendritic spines of striatal medium spiny neurons. *Neuron* 44(3):483–493.
19. McRory JE, et al. (2001) Molecular and functional characterization of a family of rat brain T-type calcium channels. *J Biol Chem* 276(6):3999–4011.
20. Reimers-Kipping S, Hevers W, Pääbo S, Enard W (2011) Humanized Foxp2 specifically affects cortico-basal ganglia circuits. *Neuroscience* 175:75–84.

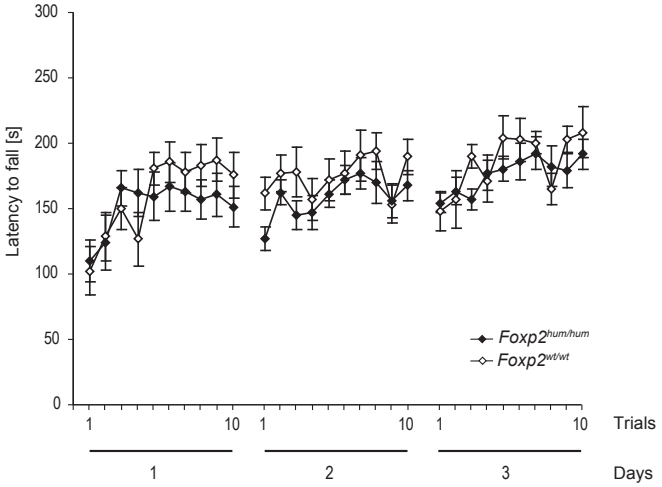

Fig. S1. Mice expressing humanized Foxp2 protein and their WT littermates acquire an accelerating rotarod task comparably. Across three consecutive days—the training periods that had previously been used to measure motor skill acquisition—with 10 daily trials (*x* axis), the *Foxp2*$^{hum/hum}$ (♦) and *Foxp2*$^{wt/wt}$ (◇) mice ($n = 9$ per genotype) improved their performance equally well (GLMM, likelihood ratio test for interaction of training day and genotype interaction: $\chi^2 = 6.15$, df = 4, $P = 0.19$), as indicated by increases in latencies to fall off the rod. Error bars indicate ±SEM.

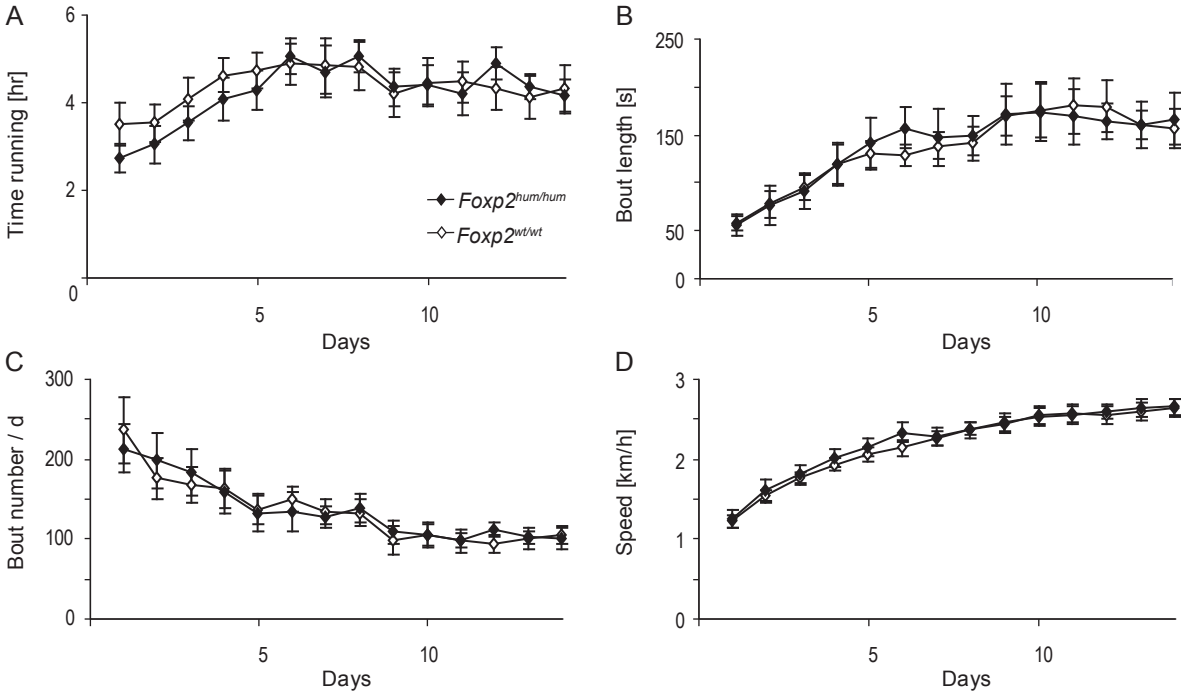

Fig. S2. Mutant and WT mice show comparable improvement on a voluntary motor skill learning task, the tilted running wheel. Across 14 d (*x* axis), *Foxp2*$^{hum/hum}$ (♦) and *Foxp2*$^{wt/wt}$ (◇) mice did not differ in the time spent running (*A*; GLMM, $\chi^2 = 7.26$, df = 8, $P = 0.51$), length of individual running bouts (*B*; GLMM, $\chi^2 = 5.95$, df = 8, $P = 0.65$), total number of running bouts per day (*C*; GLMM, $\chi^2 = 6.5$, df = 8, $P = 0.65$), or running speed (*D*; GLMM, $\chi^2 = 10.87$, df = 8, $P = 0.32$).

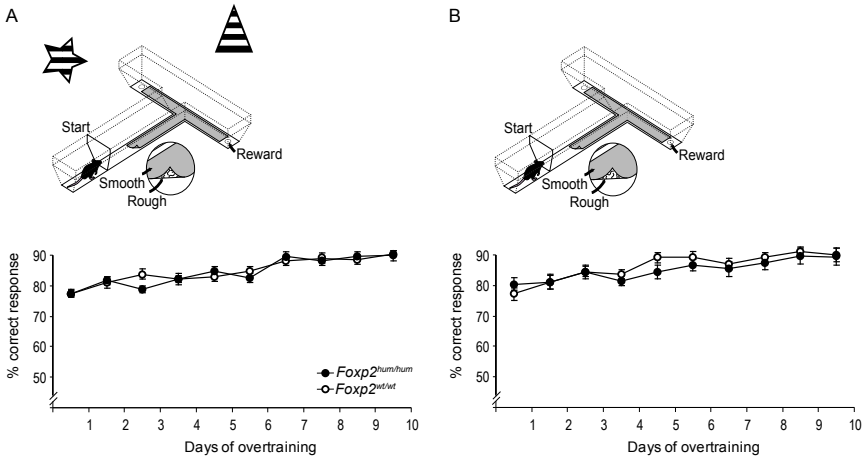

Fig. S3. Comparable performance of the T-maze task by mutant and control mice during overtraining. *Foxp2*$^{hum/hum}$ (♦) and *Foxp2*$^{wt/wt}$ (♢) mice did not differ in their performance during 10 daily training sessions with performance at or above the acquisition criterion (72.5% correct responses) in the cue-enriched setup (*A*; RMA overtraining days 1–10, $P_{day*GT} = 0.34$, $P_{GT} = 0.74$) and in the cue-deprived setup (*B*; $P_{day*GT} = 0.9$, $P_{GT} = 0.13$). Error bars indicate ±SEM.

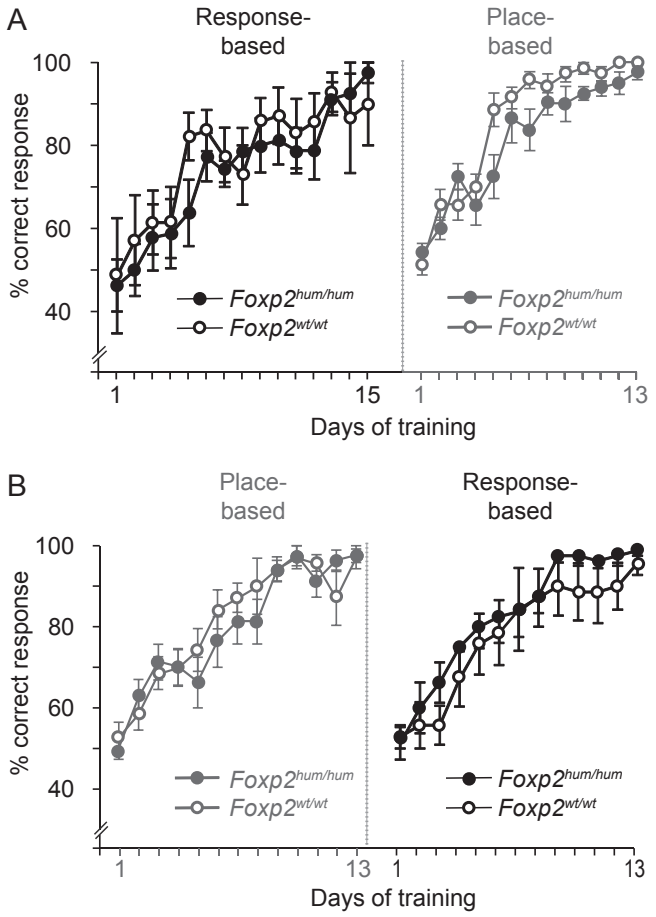

Fig. S4. Comparable declarative and procedural acquisition in the cross-maze task by mutant and control mice. *Foxp2*$^{hum/hum}$ (♦) and *Foxp2*$^{wt/wt}$ (♢) mice acquired the response-based/procedural (*A*; black; RMA, $n = 7–8$ per genotype; RMA, $F_{1,13} = 0.43$, $P_{GT} = 0.53$; $F_{4.6,13} = 0.56$, $P_{day*GT} = 0.72$) and the place-based/declarative (*B*; gray; RMA, $n = 19–20$ per genotype; RMA, $F_{1,37} = 0.45$, $P_{GT} = 0.51$; $F_{6.2,37} = 0.83$, $P_{day*GT} = 0.55$) cross-maze tasks per se comparably. Error bars indicate ±SEM.

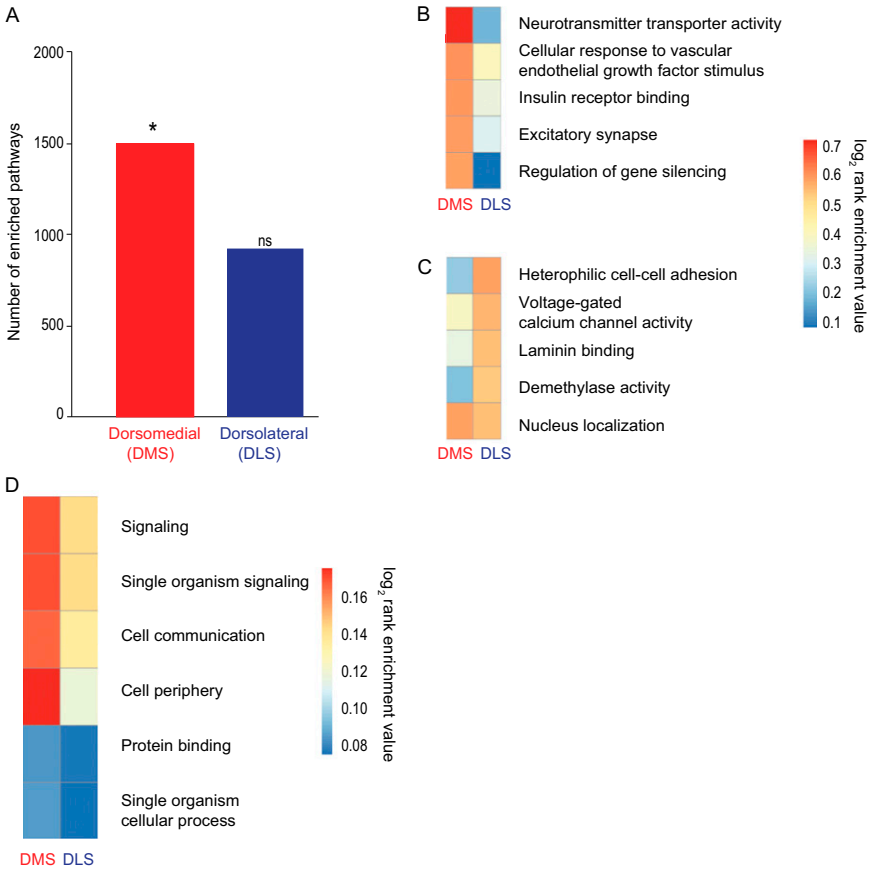

Fig. S5. Different gene expression patterns in the dorsomedial and dorsolateral striatum. (*A*) Functional categories showed significant enrichment (Benjamini–Hochberg-corrected false discovery rate < 0.05) only in the dorsomedial striatum (DMS) and not the dorsolateral striatum (DLS; $*P_{perm} < 0.05$). ns, not significant. To identify differentially enriched functional categories in dorsomedial striatum and dorsolateral striatum, we first ranked genes based on fold change as well as statistical significance and then applied a Wilcoxon rank-sum test for each region, differentiating between genes expressed higher or lower in *Foxp2*$^{hum/hum}$ mice. The average rank of genes in a category was used to calculate the rank enrichment value. (*B*) Results of the five most down-regulated categories in dorsomedial striatum according to rank enrichment value. Heat map represents logarithmically scaled rank enrichment values. (*C*) Corresponding heat map of rank enrichment values on a logarithmic scale of the five most down-regulated categories in dorsolateral striatum. (*D*) Logarithmically scaled heat map of the most significantly down-regulated categories according to significance resulting from permutation testing in dorsomedial striatum and dorsolateral striatum.

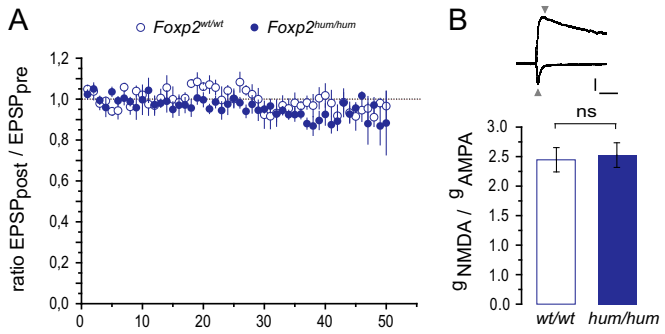

Fig. S6. (*A*) Control recordings obtained without any high-frequency stimulation were stable over a time course of 50 min and showed no significant reduction in amplitude. (*B*) NMDA/AMPA ratios recorded in presence of intracellular CsCl, TEA, and QX-314 were not altered in the *Foxp2*$^{hum/hum}$ mice. (*Inset*) Example traces at the given holding voltages. (Scale bars: 200 pA, 50 ms.). Symbols indicate time points at which AMPA (▼) and NMDA (▲) amplitudes were measured. Error bars indicate ±SEM. ns, not significant.

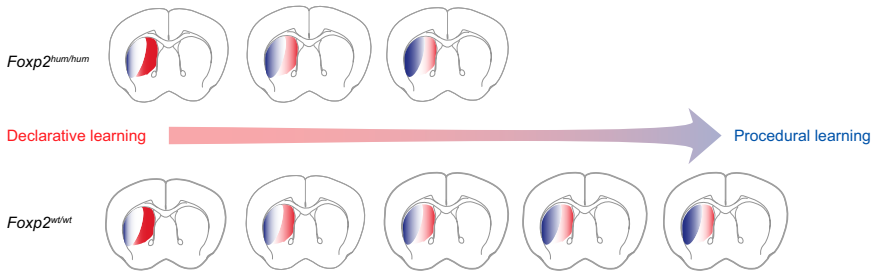

Fig. S7. Working model describing how humanized Foxp2 might alter striatum-dependent learning. Schematic diagrams showing transverse sections through the striatum of a mouse brain, symbolizing with different densities of color the activity of the dorsomedial (red) and dorsolateral (blue) striatum during the progression of learning with declarative and procedural strategies. The dorsomedial striatum is involved initially when declarative learning plays a predominant role, but as the behavior is progressively automatized, the dorsolateral striatum becomes increasingly more involved. Alterations of striatum-dependent functions in *Foxp2*$^{hum/hum}$ mice are proposed to rely on a more rapid transition from declarative to procedural learning than observed in WT littermates, which need longer times to reach the same state.

Table S1. Results of the GLMM analysis performed on the T-maze data in the environment enriched for spatial cues

Predictor term*	Estimate	SE	*z*-value	*P* value
Intercept*	0.74	0.07	†	†
Training day*	0.57	0.05	†	†
Trial*	0.05	0.02	†	†
Genotype*	−0.38	0.08	†	†
Sex*	−0.18	0.08	†	†
Stimulus*	0.16	0.04	†	†
Trial:stimulus*	0.08	0.03	2.63	0.009
Training day:sex*	−0.11	0.05	−2.17	0.03
Training day:trial*	0.05	0.02	3.28	0.001
Training day:genotype*	−0.2	0.05	−3.9	$< 10^{-4}$
Genotype:stimulus*	0.17	0.06	2.82	0.005

*Predictor variables and their interaction terms of the final statistical model. Interactions are marked by a colon between the names of the predictors.
†Not indicated because it has no meaningful interpretation.

Table S2. Results of the GLMM analysis performed on the T-maze data in the environment deprived of spatial cues

Predictor term*	Estimate	SE	*z*-Value	*P* Value
Intercept	0.76	0.10	†	†
Training day	0.54	0.06	†	†
Trial	0.13	0.03	†	†
Genotype	0.13	0.14	†	†
Stimulus	< −0.01	0.04	†	†
Training day:trial	0.07	0.02	3.31	< 0.001
Training day:genotype	0.17	0.07	2.24	0.025
Training day:stimulus	−0.28	0.04	−6.97	$< 10^{-11}$

*Predictor variables and their interaction terms of the final statistical model. Interactions are marked by a colon between the names of the predictors.
†Not indicated because it has no meaningful interpretation.

Table S3. Results of the GLMM analysis performed on the place-response-switch data in the cross maze

Predictor term*	Estimate	SE	*z*-Value	*P* Value
Intercept	−0.73	0.55	†	†
Training day	0.83	0.2	4.17	< 0.0001
Informative trial	0.34	0.19	1.83	0.067
Genotype	−1.79	0.84	−2.14	0.033

*Predictor variables and their interaction terms of the final statistical model.
†Not indicated because it has no meaningful interpretation. Shown are only the significant predictors; interactions are not shown as a result of nonsignificance.

Other Supporting Information Files

Dataset S1 (PDF)