**TITLE:**

# Direct and indirect striatal projecting neurons exert strategy-dependent effects on decision-making

**AUTHOR LIST:**

**Elena Chaves Rodriguez[1]*†, Jérémie Naudé[2,3]*, Daniel Rial[1,§] and Alban de Kerchove d'Exaerde[1,4,5]**

**AFFILIATION:**

**1. Neurophy Lab, ULB Neuroscience Institute, Université Libre de Bruxelles (ULB), 808 route de Lennik, 1070 Brussels, Belgium.**
**2. Institut de Génomique Fonctionnelle, UMR 5203 CNRS – U 1191 INSERM – Univ. Montpellier, Montpellier, France.**
**3. Sorbonne Université, Inserm, CNRS, Neuroscience Paris Seine; Institut de biologie Paris Seine (NPS - IBPS), 75005 Paris, France.**
**4. WELBIO, WEL Research Institute, Wavre, Belgium.**
**5. Corresponding author, adekerch@ulb. ac.be**
*** Both authors contributed equally to this work**
† Current address: Business & Decision Life Sciences, Brussels, Belgium.
§ Current address: Sygnature Discovery, In Vivo Department, Nottingham, UK.

**ABSTRACT**

The striatum plays a key role in decision-making, with its effects varying with anatomical location and direct and indirect pathway striatal projecting neuron (d- and i-SPN) populations. Using a mouse gambling task with a reinforcement-learning model, we described of individual decision- making profiles as a combination of three archetypal strategies: Optimizers, Risk-averse, Explorers. Such strategies reflected stable differences in the parameters generating decisions (sensitivity to the reward magnitude, to risk or to punishment) derived from a reinforcement-learning model of animal choice. Chemogenetic manipulation showed that dorsomedial striatum (DMS) neurons substantially impact decision-making, while the nucleus accumbens (NAc) and dorsolateral striatum neurons (DLS) have lesser or no effects, respectively. Specifically, DMS dSPNs decrease risk aversion by increasing the perceived value of risky choices, while DMS iSPNs emphasize large gains, affecting decisions depending on decision-making profiles. Hence, we propose that striatal populations from different subregions influence distinct decision-making parameters, leading to profile-dependent choices.

## INTRODUCTION

To adapt to dynamic environments, decision processes aim to select the course of action that will lead to the best outcome. Value-based decision-making can be described as the iterative process of selecting actions on the basis of their expected values and of evaluating outcomes to update action values(*1*, *2*). Decision-making is therefore fundamentally based on predictions of the probability of obtaining the different available outcomes (e.g., rewards, omissions or punishments), as well as estimations of their respective values. There has been a renewed interest in understanding how the nervous system weights the expected values of outcomes to choose between options.

Decision-making processes emerge from the mesocorticolimbic loop, which has a key, yet not fully understood, role in the striatum (*3–5*). The rodent striatum is divided into several regions, each of which is thought to have a particular function in decision-making. The dorsal striatum is responsible for motor and cognitive control associated with goal-directed actions (DMS) or the generation of habits (DLS), and the NAc (*6–15*) manages reward, motivation and Pavlovian associations. While habits, which consist of stimulus–response associations, are generally insensitive to the value of the outcome (*16*, *17*), Pavlovian (stimulus–outcome) and goal–directed (action–outcome) associations can affect value–based decisions (*13*, *16*). How these associations with outcomes (rewarding or punitive) rely on striatal computations has been the subject of much study (*18*). Striatal projection neurons (SPNs) account for 95% of all striatal neurons. There are two distinct SPN subpopulations: dopamine D1 receptor ($D_1R$)-expressing or direct SPNs (dSPNs) and dopamine D2 receptor/adenosine A2A receptor ($A_{2A}R$)-expressing indirect SPNs (iSPNs). It remains debated how these two neuronal groups interact during decision-making (*19*). Loss-of-function or gain-of-function experiments (*20–25*) support a go/no-go model (*26*, *27*) in which dSPNs play a prokinetic role and iSPNs play an antikinetic role. However, correlative approaches (*28–33*) revealed the coactivation of dSPNs and iSPNs during motor performance. This suggests a complementary encoding of behaviors, with congruent activations observed in dSPNs for multiple behaviors and iSPNs for suppressing competing behaviors (*34*).

In the context of decision-making, dSPNs and iSPNs may encode potential rewards and costs (*19*, *35*), respectively, or encode and update values in goal-directed learning (*36*). However, manipulations of striatal dopamine also affect risk seeking (*9*), which does not fit into a clear reward/cost dichotomy, and the relative influences of dSPNs and iSPNs on risk seeking remain to be assessed. Moreover, the attitude toward risk, as well as the sensitivity to rewards, costs, or punishments, is highly idiosyncratic (i.e., it reflects stable individual traits). Different decision profiles have been described in humans and animals. Interindividual variability in the choice behavior of healthy individuals has been described in humans, rats (*37–39*) and mice (*40–42*). How such differences are related to variations in mesocorticostriatal properties remains an open question. Furthermore, several substances, ranging from caffeine to cocaine, can have different effects on individuals according to their choice behavior. For example, impulsive choice has been recurrently described as a predictor of susceptibility to developing addictive disorders in humans, such as alcohol or cocaine use disorders (*43–46*). Risk-preferring rats are more prone to develop cocaine addiction and are also more sensitive to

cocaine craving during withdrawal (*47*). Additionally, in rats, caffeine, an agonist of $A_{2A}R$, can have a motivational effect on performing effortful tasks in low performers, whereas in naturally motivated rats, caffeine can disrupt effortful performance (*48*). These pathological cases highlight that the effect of modifying the weight of a given decision variable (e.g., reward size) depends on the sensitivity to the other decision variables (e.g., risk attitude, sensitivity to effort). In this framework, the different roles of dSPNs and iSPNs from subparts of the striatum might thus depend upon the decision-making traits of the animals.

Here, we used a rodent Iowa Gambling Task (IGT) adapted from Young et al. (*49*) to show how different decision-making profiles arise in mice. Furthermore, we targeted dSPNs and iSPNs in the DMS, DLS and NAc with a chemogenetic tool to assess how decision-making strategies can be differentially altered by changes in striatal excitability, depending on "basal" decision-making traits. Computational analyses allowed us to identify three different decision-making archetypes in mice (Explorers, Risk-averse and Optimizers). These cognitive profiles were characterized by distinct sensitivities to risks, the subjective utility of large rewards, and explore-exploit trade-offs. We then showed that increasing DMS excitability exerted the most profound effects on choices and motivation, which is consistent with the IGT being a goal-directed task. Specifically, we found that facilitating DMS dSPN activity decreased safe choices in all the mice by decreasing risk aversion, whereas facilitating DMS iSPN activity decreased safe choices only in the Optimizer mice through a decrease in reward saturation. Compared with their DMS counterparts, increasing NAc dSPN excitability induced similar, albeit blunted, effects. NAc iSPN and DLS i- and d-SPN manipulations had nonspecific effects on motivation but not on choice behavior. Overall, we highlighted how the striatal subpopulation exerts cognitive profile-dependent effects on choice behavior.

## RESULTS

### Mouse gambling strategies arise from varying sensitivities to task parameters.

To assess decision-making in a complex environment, we used a rodent-adapted version of the IGT (*49*), which we adapted for mice. The task takes place in an operant chamber in which animals can select among four nose-poke holes to obtain food pellets in a magazine on the opposite wall (Figure 1A, see Methods). As mice need to nose-poke on the holes between 5 s (shorter responses were not rewarded and counted as premature) and 10 s (longer responses were not rewarded and counted as omissions), this test measures both motivation/impulsivity and choice behavior. Each hole (P1-P4) is associated with a distinct reward probability $p$ to obtain a reward magnitude (number of pellets) and a related probability of 1-$p$ to obtain a punishment in the form of a time-out (TO). TOs correspond to a potential loss of reward because of the overall limit of time of the task, e.g., potential loss = (time-out duration × average pellet rate) × time-out probability, with the pellet rate taken as the number of pellets for the option divided by the task duration. As the potential (expected) gain is the number of pellets × reward probability, one can compute the (linear) "expected net return" for each option as follows: $E(X_i) = p_i R_i - (1 - p_i) T_i$ where $p_i$ is the probability of reward, $R_i$ the reward magnitude, and $T_i$ the punishment due to time-out, in terms of opportunity cost of time (time-out duration times the average number of pellet/s). This gives E(P1) = 0.89 pellets, E(P2) = 1.52 pellets, E(P3) =

0.60 pellets, E(P4) = -0.32 pellets. In this context, options P1 and P2 are labeled as "safe choices" because they deliver a higher net return through a smaller amount of reward than P3 and P4 but with a higher probability and shorter associated TOs. In the original context of the task(*40*, *50*, *51*), P3 and P4 are thus considered maladaptive: despite delivering more pellets on rewarded trials, the high probability of a long time-out renders them disadvantageous in the long run (see Methods). In our experiments, the mice indeed preferred, on average, the safe options (P1 and P2 compared with P3 and P4), and such choice behavior was stable across sessions (Figure 1B; repeated-measures (RM-)ANOVA: time: $F_{(3,668)}=92.2$, $p<1e^{-12}$; options $F_{(3,668)}=0.1$, n.s.). However, such average preferences often mask important variability. Hence, we leveraged the analysis of baseline (before chemogenetic/ Designer Receptor Exclusively Activated by Designer Drugs (DREADD) manipulations, see below) choice behavior from all the mice (n=168) to characterize such variability. There was marked variability both in terms of choice behavior, with P1 (the safest option) chosen as little as 0% or as much as 100% of the time by a given animal (Figure 1C), and in terms of motivation/impulsivity (number of trials, % of premature responses, and % of omissions; Figure 1D). We thus sought to delve into the origin of such variability.

A classical way to describe interindividual variability is by introducing a dichotomy between "good" (or "safe") and "poor" (or "risky") decision-makers. We used such a distinction (mice were considered "safe" if they had a preference for the advantageous options P1 and P2) so that our characterization of the roles of striatal subpopulations could be compared with other studies using a rodent gambling task. However, such categorizations hinge on the experimenter's predefined criteria for what is considered beneficial or detrimental to the animals. Moreover, dichotomies resulting from arbitrary thresholds to partition data may overlook the nuanced statistical intricacies of interindividual variability. Consequently, in addition to the safe/risky dichotomy, our approach aimed to identify behavioral profiles that not only captured (1) the statistical patterns within the behavioral data but also reflected (2) the generative properties of the underlying decision-making architecture.

As a starting point, we observed strong correlations between the different individual measures (choice behavior and motivation/impulsivity), e.g., the tendency to choose the safe choices (P1 and P2) was negatively correlated with the proportion of omissions ($R^2 = 0.27$, $p=6.10^{-13}$, Figure 2A, Supp. Figure 1A-C). This statistical structure hints at constraints in the mechanisms generating the data (i.e., the decision-making traits generating the animal behavior). We thus reduced the high-dimensional dataset (7 measures: preferences for the 4 options, i.e., P1-P2-P3-P4, as well as the number of trials, omissions and premature responses) via principal component analysis (PCA; Supp. Figure 1D). As there were no clearly separated clusters, we expressed individuals as a continuum between "archetypal" strategies via archetypal analysis, as previously developed (*52*, *53*) and applied to decision-making *(41*). Compared with classical clustering, which groups individual data around typical observations (cluster centers) that constitute a decision-making profile, archetypal analysis depicts individual behavior as intermediates between extreme strategies, i.e., archetypes (Figure 2B). This framework acknowledges that, contrary to experimenter-defined performance in each task, animals define their own success criteria, generally as a tradeoff between different objectives. Indeed, no

decision-making profile (being averse to risk, optimizing the expected reward, showing a low level of attentional control) can be optimal for all the possible environments (in terms of reward scarcity, risk of punishment, etc.) the animals might face. The best compromises between objectives lead to phenotypes that lie in low-dimensional polytopes in the trait space (*52*, *53*). Archetypal strategies corresponding to the objectives lie at the apices (extrema) of these polytopes, and individual strategies are described as linear combinations of the archetypes. If the behavior in the mouse gambling task indeed emerges as an individual solution to a trade-off between archetypal strategies, then such a description should 1) be stable across sessions and 2) reflect underlying decision-making traits.

We first calculated the number *n* (n being less than the dimension of the data) of archetypes by fitting polytopes with *n* apices to the data and chose *n=3,* beyond which there was little improvement in the explained variance (Figure 2C, Supp. Figure 1E, F). The archetypes positions in the PCA space were robust to the use of only a subset of the data (jackknife resampling, Supp. Figure 1G). The archetype analysis explained more variance than did the classical safe/not safe dichotomy (Supp. Figure 1H, I). We then verified that individual strategies were stable by computing the distance of individual animals (expressed as coordinates in the PCA space) to the apices (the archetypes). This distance-to-archetype was highly correlated between consecutive days (Day 1 versus Day 2: $R^2$= 0.71, p=$3.10^{-46}$; Day 1 versus Day 3: $R^2$= 0.64, p=$3.10^{-38}$; Day 1 versus Day 4: $R^2$= 0.64, p=$8.10^{-39}$, n = 169 mice, Figure 2D, Supp. Figure 1J, K), indicating stable profiles. Finally, partitioning the animals into three groups, on the basis of their proximity to one of the archetypes, corresponded to significant differences in the behavioral data (choices: 2-way ANOVA: $F_{(6,66)}$=115.86, p=$1.10^{-99}$ ; trials: 1-way ANOVA, $F_{(2,165)}$=69.82, p=$1.10^{-22}$; premature responses: 1-way ANOVA, $F_{(2,165)}$=41.68, p=$1.10^{-15}$; omissions: 1-way ANOVA, $F_{(2,165)}$=5.92, p=$3.10^{-3}$). The first group (hereafter, "Risk-averse", R) displayed a greater preference for P1, i.e., the least risky option (Figure 2E), more premature responses, and fewer omissions (Figure 2F), than did the other two groups. The second group ("Explorers", E mice) displayed less marked preferences (the flatter pattern of choices in Figure 2E), initiated fewer trials, and made more omissions (Figure 2F). The third and last group ("Optimizers", O mice) preferred P2, i.e., the more advantageous option, over P1 (Figure 2E), initiated more trials, and made less premature (impulsive) responses (Figure 2F).

The fact that behavioral features were enriched at each archetype is only confirmatory (like significant differences between subgroups following a median split) because an archetypal analysis is performed on the PCA space derived from the same behavioral data. We thus sought to confirm that the statistical distribution in the "manifested" choices of the animals arises from differences in underlying ("hidden") decision-making strategies. To test for putative differences in their subjective valuation/decision processes, we fit the animals' choices with a reinforcement-learning model (Figure 3A). In this model, the probability of choosing the hole Px (e.g., P1) depends on its expected value compared with the sum of values from all the other options. The model assumes that the value of an option depends on its expected (subjective) reward, punishment, and risk (see the Methods for details). The expected reward (or punishment) equals the reward (or punishment) probability times the magnitude of the subjective reward (or punishment), such as for the "expected net return". Nevertheless, a major

difference between the IGT in rodents and those in humans is that the reward takes the form of food pellets (instead of points or money), and punishments are time-outs. Hence, classical models considering that animals compute an "expected net return" rely on the hypothesis of linear preferences (or time perception), e.g., that winning 4 pellets (or waiting 40 s) from one choice is similar to having 4 times 1 pellet from 4 choices (or waiting 4 times 10 s, respectively). In contrast, animals often display nonlinear subjective values or time perceptions. Hence, we introduced an additional parameter ρ that depicts how the subjective value saturates with the number of pellets: $\rho<1$ corresponds to a value of 4 pellets lower than 4 times the value of 1 pellet, henceforth favoring small gains (i.e., a small ρ favors small gains, whereas a high ρ favors large gains). Similarly, the parameter $T$ depicts how subjective punishment depends on the time-out duration. The option value also incorporates a risk-sensitive term (φ) that can be positive (risk prone) or negative (risk averse). Finally, we included an inverse temperature parameter (β) to account for the exploitation–exploration tradeoff: animals choosing their preferred option nearly all the time were considered exploitive (high β), whereas animals with a "flat" distribution of choices (no marked preference) were labeled explorative (low β). We checked that this model provided a better fit than simpler models used in the literature (Figure 3B); in particular, nonlinear reward and time perception resulted in a better fit (Figure 3B, Supp. Figure 2A) and generation (Figure 3C) of choice patterns. We also verified that model parameters could be recovered with the relatively small number of trials performed by the animals, demonstrating a reasonable level of accuracy (Supp. Figure 2B). By fitting this model to individual data, we could express the choices from each animal as if it had been generated by a decision-maker with a given set of parameters (β, ρ, $T$, φ; Figure 3C). This decision model allowed us to better characterize the archetypal strategies (Figure 3D, β parameter: 1-way ANOVA, $F_{(2,165)}=13.67$, $p=3.10^{-6}$; ρ parameter: 1-way ANOVA, $F_{(2,165)}=15.73$, $p=6.10^{-7}$; $T$ parameter: 1-way ANOVA, $F_{(2,165)}=8.83$, $p=2.10^{-4}$; φ parameter: 1-way ANOVA, $F_{(2,165)}=24.49$, $p=4.10^{-10}$). The R group displayed the lowest φ parameter (indicating strong aversion to risk) but also a high sensitivity to time-out $T$ and a low sensitivity to reward ρ. The E group was characterized by a low inverse temperature β, indicative of high exploration (or low exploitation), in line with its absence of a clear preference among the four options. The O group had more balanced decision parameters, with a high β (exploitive), an insensitivity to risk on average, and near linear saturation functions (ρ and $T$ close to 1) for reward and punishment. Expressing individual fits as intermediates between the archetypes extrema yielded the same interpretation (Supp. Fig 3A). suggesting that the computational characterization does not arise from partitioning the data into clusters. The conjunction of the archetypal analysis with the computational model further suggests that the stable interindividual variability observed in the mouse gambling task reflects stable (Supp. Figure 3B) decision-making profiles.

**Dorsomedial dSPNs decrease safe choices by promoting risk-seeking.**

We next sought to assess the contributions of both SPN populations to decision-making in the main striatal subdomains (DMS, DLS and NAc). We thus used DREADD activation, rather than optogenetic activation (see Discussion section), to facilitate neuronal activity from a

261 specific subpopulation after the values of the different options were learned by the animal to
262 focus on the decision process. For this purpose, we tested 6 different groups of mice, one per
263 SPN population (direct or indirect pathway SPNs) and striatal area (DMS, DLS or NAc). We
264 bilaterally injected a Cre-dependent hM3Dq DREADD virus (AAV5 hSyn-DIO-hM3D(Gq)-
265 mCherry) or a control fluorophore lacking the hM3Dq construct (AAV5 hSyn-DIO-mCherry)s
266 into the striatal zone of interest (DMS, DLS or NAc) in $D_1$R-Cre (*54*) or $A_{2A}$R-Cre mice (*22*)
267 to selectively express hM3Dq or mCherry on either dSPNs or iSPNs, respectively (Supp. Figure
268 4).
269
270 As the mouse gambling task relies on instrumental behavior, we first focused on DMS d-SPNs
271 (Figure 4A). Given that mice received DREADD ligand (CNO, 1 mg/kg) injection on the last
272 day of behavioral testing (see below), we verified ex vivo the effectiveness of the hM3Dq
273 DREADDs in increasing the intrinsic excitability of targeted neurons by collecting slices from
274 mice (hM3Dq and mCherry controls) just after the behavioral testing. The marker of neuronal
275 activation Fos was colocalized with mCherry expression only in hM3Dq animals (Figure 4B,
276 Supp.Table 1, $t_{(16)}$=8.5, p<0,0001; unpaired t test), confirming that CNO facilitated neuronal
277 activation during the mouse gambling test. Next, we measured that neuronal activation was
278 followed by an increase in intrinsic excitability (i.e., active membrane properties), as expected
279 from hM3Dq recruitment (*55*). We did not find any changes in the passive properties of the
280 neurons via patch-clamp recordings: the resting membrane potential did not change following
281 CNO injection (Supp. Figure 5A, $t_{(10)}$=0.2, n.s.). We did not observe any modifications in
282 (putatively corticostriatal or thalamicostriatal) synapse strength as measured by the
283 AMPA/NMDA ratio ($t_{(10)}$=0.93, n.s., Supp. Figure 5B), suggesting that Fos activity was not
284 due to broad changes in synaptic excitation. Owing to the intrinsic excitability of the cells, the
285 rheobase (minimal current needed to elicit an action potential) was not changed following
286 hM3Dq + CNO treatment ($t_{(10)}$=0.96, n.s., Figure 4C left). However, the number of action
287 potentials elicited by intermediate current intensities increased following hM3Dq + CNO
288 treatment, indicating an increase in neuronal gain (in the linear range of the frequency intensity,
289 or f-I curve, 2-way ANOVA $F_{(14,140)}$ = 13, p<0.0001, Figure 4C right). This finding suggests
290 that neuronal Fos activation arose in neurons receiving a notable drive (putatively task-related)
291 amplified by increased neuronal gain rather than from nonspecific electrical activity (which
292 would follow from changes in resistance or rheobase). Together, the patch clamp recordings
293 confirmed that CNO injection before behavioral testing had lasting effects (detectable ex vivo
294 after the session) on neuronal excitability.
295
296 Having verified that hM3Dq expressed in DMS-dSPNs favored their activation during the
297 mouse gambling test, we next assessed how this increase in neuronal activity affected decision-
298 making. We compared the differences in preferences for the four options (P1-P4) in the mouse
299 gambling task under CNO and at baseline for hM3Dq animals and mCherry controls (Figure
300 4D). Safe mice (i.e., mice with a preference for the advantageous options P1 and P2) in the
301 hM3Dq + CNO conditions displayed a decrease in the preference for P1 and P2 (Figure 4E,
302 RM-ANOVA, $F_{(1,38)}$=5.83, p = 0.02; CNO effect for the mCherry group: n.s.; for the Gq group:
303 p=0.017). However, this decrease could arise for different reasons: worst decision-making,
304 increased exploration, decreased aversion to punishment, etc. There was no change in overall

305 performance (i.e., average number of pellets per trial; RM-ANOVA, $F_{(1,38)}$=1.48, Figure 4F);
306 hence, we sought a more precise characterization in terms of a modification in decision-making
307 traits. We also observed a decrease in the number of trials and an increase in omissions (Figure
308 4G; Trials: RM-ANOVA, $F_{(1,38)}$=100, p = $3.10^{-12}$; CNO effect for the mCherry group: n.s.; for
309 the Gq group: p=$2.10^{-11}$; Omissions: RM-ANOVA, $F_{(1,38)}$=18.08, p = $1.10^{-4}$; CNO effect for the
310 mCherry group: n.s.; for the Gq group: p=$2.10^{-6}$), i.e., a decrease in the pace of the instrumental
311 action. However, this decrease in decision frequency was not due to motor control, as
312 locomotion was found to increase after DMS-dSPN facilitation (Supp. Figure 5). The combined
313 modification of preferences and instrumental pace suggested a global effect of DMS-dSPNs
314 facilitating archetypal strategies. Indeed, we observed a global shift in the PCA space for
315 hM3Dq animals under CNO, with a displacement away from the risk-averse apex (Figure 4H,
316 RM-ANOVA, $F_{(1,38)}$=5.16, p = 0.03; CNO effect for the mCherry group: n.s.; for the Gq group:
317 p=0.006) that could be observed for all archetypes (Supp. Figure 5). However, the effect on
318 safe choice (P1+P2 choices) was observed only in Optimizer mice (Figure 4H, 2-way ANOVA:
319 $F_{(1,34)}$=5.25, p = 0.01; R+E mice: n.s.; O mice: p=0.005). We thus analyzed the full choice
320 pattern with the computational model. The model-based analysis indicated that the change in
321 the full choice pattern in the hM3Dq-CNO condition was better explained by an increase in the
322 risk sensitivity parameter (RM-ANOVA, $F_{(1,38)}$=4.84, p = 0.03; CNO effect for the mCherry
323 group: n.s.; for the Gq group: p=0.01, Figure 4I, Supp. Figure 5). Overall, the behavioral data
324 and computational analyses suggest that facilitating DMS-dSPN activity decreases the choice
325 of P1 and P2 by increasing proneness to risk.
326
327 **Dorsomedial iSPNs favor large gains, exerting strategy-dependent effects.**
328
329 We then compared the above facilitation of DMS-dSPNs with that of DMS-iSPNs (Figure 5A),
330 as these two subpopulations are hypothesized to act in opposition (*20–27*), in synergy (*28–*
331 *33*) or in a complementary manner, supporting a dual selection-suppression function (*34, 56*).
332 In this subpopulation, selective hM3Dq DREADD expression followed by CNO injection
333 during the mouse gambling test effectively increased intrinsic excitability (f-I gain, 2-way
334 ANOVA $F_{(14,140)}$ = 13), driving Fos activation (p<0,0001, U(6)=0, Mann–Whitney) in targeted
335 neurons (Figure 5B) without any effects on resting membrane potential ($t_{(8)}$=0.39, n.s., rheobase
336 : $t_{(8)}$=0.63, n.s., and AMPA/NMDA ratio: $t_{(8)}$=0.15, n.s., Fig. 5C and Supp. Fig. 6). In contrast
337 to the facilitation of DMS dSPN activity, CNO did not significantly influence preferences in
338 hM3Dq animals compared with mCherry controls when all mice or only safe mice were
339 considered (Figure 5D). However, there was a significant decrease in the performance
340 (pellets/trials) of the animals under DMS-iSPN facilitation (Figure 5E, RM-ANOVA,
341 $F_{(1,33)}$=172.69, p = $1.10^{-14}$; CNO effect for mCherry group: n.s.; for the Gq group: p=$1.10^{-7}$),
342 suggesting that decision-making had shifted away from reward maximization (due to poorer
343 decision-making or a shift in preferences). Moreover, facilitating the activity of DMS iSPNs
344 with DREADDs decreased the overall number of trials (RM-ANOVA, $F_{(1,33)}$=25.6, p = $8.10^{-7}$;
345 CNO effect for the mCherry group: n.s.; for the Gq group: p=$1.10^{-4}$) and increased the
346 percentage of omissions (RM-ANOVA, $F_{(1,33)}$=25.08, p = $2.10^{-5}$; CNO effect for the mCherry
347 group: n.s.; for the Gq group: p=0.001; Figure 5F). We verified in a subset of mice that DMS-
348 iSPNs facilitated decreased locomotion in the operant box (Supp. Figure 6), in contrast with

349 DMS-dSPN manipulation, which was consistent with changes in omissions or in the number of
350 trials independent of locomotor effects. We then looked for an effect on archetypal profiles.
351 Indeed, CNO induced a global shift in the PCA space in hM3Dq animals toward the Explorers
352 (E) apex (Figure 5G, RM-ANOVA, $F_{(1,33)}$=37.69, p = $6.10^{-7}$; CNO effect for the mCherry group:
353 n.s.; for the Gq group: p=$2.10^{-4}$). The effect of DREADD-mediated facilitation of DMS-iSPN
354 neuronal activity was thus different from that of DMS-dSPNs (i.e., a shift away from the R
355 apex). This led us to examine more closely the effects of CNO on the preferences of hM3Dq
356 animals, depending on their proximity to archetypes rather than on the safe/risky dichotomy.
357 Facilitating DMS-iSPN activity had opposite effects on the preference for safe (P1 and P2)
358 options in Optimizer mice (a decrease in P1+P2 choices) and in the other two archetypes
359 (increased P1+ P2 choices in R+I animals). Hence, the DREADD effect on choices depended
360 on the baseline strategy of the animals. We assessed the state-dependent effects of CNO with
361 the decision model after verifying that the decrease in trial number, observed under CNO
362 (Figure 5F), did not lead to a systematic bias in model parameter estimations, although it
363 introduced more noise in the parameter recovery (Supp. Figure 6). We then characterized which
364 modifications of the decision-making parameters could best account for the differential effect
365 of DMS-iSPN facilitation on choices. We reasoned that a decrease in performance due to poorer
366 decision-making would appear in the model as a decrease in the exploitation (β) parameter,
367 whereas animals could also become less optimal on average because of nonlinear sensitivity to
368 rewards (ρ) or time-outs (T). A modification in the ρ parameter (decrease in the reward
369 saturation with the pellet number) best explained the differential DREADD effect on choices
370 (Figure 5H, Supp. Figure 6, RM-ANOVA, $F_{(1,33)}$=14.23, p = $6.10^{-4}$; effect of CNO for the
371 mCherry group: n.s.; for the Gq group: p=0.003). Indeed, in reinforcement-learning models
372 such as ours, the alteration of one decision-making parameter can induce opposite behavioral
373 patterns depending on the values of the other parameters (Supp. Figure 6). Here, an increase in
374 reward saturation (decreasing the value of large gains) increased the preference for P1 and P2
375 when animals were risk averse but decreased the preference for P1 and P2 in risk-neutral mice
376 (Supp. Figure 6). Furthermore, decreases in reward saturation (i.e., preferring 4 pellets over 4
377 times 1 pellet) deviated the animals from maximizing the average reward, which is consistent
378 with the observed decrease in global performance following DMS-iSPN facilitation.
379 Overall, the computational analyses suggest that facilitating DMS-iSPN activity decreases how
380 the reward saturates with reward size (favoring large gains), exerting a decision profile-
381 dependent effect on choices.
382
383 **Nucleus accumbens SPNs are less involved in gambling task than DMS.**
384
385 Even if the mouse gambling task is instrumental, the nucleus accumbens (NAc) may also
386 influence choice behavior. Facilitating NAc-dSPN (Figure 6A) activity did not exert effects
387 similar to those of facilitating DMS-dSPN activity. hM3Dq expression coupled with CNO
388 treatment efficiently drove Fos activity and changes in intrinsic excitability in NAc-dSPNs
389 (Figure 6B). In the mouse gambling task, such facilitation of neuronal activity strongly
390 decreased the number of trials (Figure 6C, RM-ANOVA, $F_{(1,28)}$=16.68, p = $3.10^{-4}$; CNO effect
391 for the mCherry group: n.s.; for the Gq group: p=$4.10^{-4}$), in contrast to DMS-dSPN facilitation,
392 which did not exert any effect. Facilitating NAc-dSPN activity did not alter the proportion of

393 premature or omission trials (Figure 6C), again in contrast with the reduction in these measures
394 following DMS-dSPN manipulation. We did not observe any differences in choice behavior,
395 either following the safe/risky dichotomy or when considering archetypes or overall
396 performance (Supp. Figure 7). However, facilitating NAc-dSPNs increased locomotion (Supp.
397 Figure 7). In the PCA, the overall effect corresponded to a modest (but significant) shift away
398 from the Optimizer (O) apex (Figure 6D, RM-ANOVA, $F_{(1,28)}$=13.53, p = $1.10^{-3}$; CNO effect
399 for the mCherry group: n.s.; for the Gq group: p=0.03). We did not find any differences in the
400 parameters from the computational model of decision-making when we fitted the choices under
401 the hM3Dq+CNO treatment (Supp. Figure 7). Overall, facilitating NAc-dSPNs had an effect
402 on the decision task, but that was less specific than that of facilitating DMS-dSPNs.
403
404 By contrast, CNO injection in animals expressing hM3Dq in NAc-iSPNs (Figure 6E) increased
405 intrinsic excitability and drove efficient Fos activation in targeted cells (Figure 6F, p<0,0001,
406 unpaired t test). The behavioral effect of DREADD facilitation on NAc-iSPN activity was
407 relatively similar to what we observed when facilitating DMS-iSPNs. Preferences did not
408 change on average, but the overall number of trials decreased (RM-ANOVA, $F_{(1,22)}$=25.43, p =
409 $5.10^{-5}$; CNO effect for the mCherry group: n.s.; for the Gq group: p=0.001), whereas the
410 proportion of omissions increased (RM-ANOVA, $F_{(1,22)}$=25.6, p = $8.10^{-7}$; CNO effect for the
411 mCherry group: n.s.; for the Gq group: p=$1.10^{-4}$, Figure 6G). This pattern was reflected in the
412 PCA space, where CNO induced a shift of hM3Dq animals toward the E apex, albeit smaller
413 than what we observed following DMS-iSPN manipulation (RM-ANOVA, $F_{(1,22)}$=33.22, p =
414 $2.10^{-4}$; CNO effect for the mCherry group: n.s.; for the Gq group: p=0.005, Figure 6H). Like
415 DMS-iSPNs, facilitation by NAc-iSPNs decreased overall performance and locomotion (Supp.
416 Figure 7). Similarly, the behavioral effects of NAc-iSPN facilitation could be accounted for by
417 an increase in reward saturation in the computational model (RM-ANOVA, $F_{(1,22)}$=26.03, p =
418 0.03; CNO effect for the mCherry group: n.s.; for the Gq group: p=0.0076, Figure 6G, Supp.
419 Figure 7). Hence, facilitating NAc-iSPNs appeared to exert similar, but smaller in magnitude,
420 effects than facilitating DMS-iSPNs did.
421
422 **Both dorsolateral SPN populations have no specific effects on gambling task.**
423
424 Finally, we evaluated the influence of the DLS on choice behavior. CNO treatment drove Fos
425 activity and increased the intrinsic excitability of targeted neurons in hM3Dq animals in the
426 DLS-dSPN (p<0,0001, unpaired t-test, Figures 7A, 7B). We observed small effects on the
427 proportions of omission trials (Figure 7C). However, in the mouse gambling task, there was no
428 change in preferences (in safe mice or when archetypes were considered) under hM3Dq +CNO
429 treatment (Figure 7D). Similarly, CNO treatment drove Fos activity and increased the intrinsic
430 excitability of targeted neurons in hM3Dq animals in the DLS-iSPN (p<0,0001, unpaired t-test,
431 Figure 7E,F). We observed small effects on the proportions of omission trials (Figure 7G), with
432 a decrease in premature trials (Figure 7G) and in locomotion (Supp. Fig 8). As for DLS-
433 dSPN, there was no change in preferences under hM3Dq +CNO treatment (Figure 7H).
434 Consistently, we did not find any significant differences in model parameters when the choices
435 were fitted; under any (DLS-iSPN or DLS-dSPN) conditions (Supp. Figure 8), for both

manipulations (DLS-iSPNs and DLS-dSPNs), consistent with the lesser involvement of the DLS in decision-making.

## DISCUSSION

Consistent with other studies in human and rat versions of the IGT, we found marked interindividual variability in preferences. However, we departed from usual classifications in terms of experimenter-defined performance (*37*, *39*). Instead, we chose an unsupervised approach (archetypal analysis) to characterize individuals as intermediate between extrema rather than clearly separated clusters. Whether explicitly or implicitly, assessing interindividual variability on the single scale of "task performance" assumes that (1) animals have evolved to optimize a given quantity, e.g., long-term reward relative to motor costs, and that (2) the experimental setup measures the fitness of the animals on this scale (i.e., the setup is "factor pure"(*57*)). However, organisms have likely evolved to trade priorities across various objectives. In particular, most animals are equipped with a cognitive architecture for deliberative decisions between goals (*58*). Decision-making requires assigning values to options, which is a fundamentally subjective process (*59*, *60*). Computational decision-making models constitute powerful tools designed to infer subjective valuations from the temporal series of choices. However, these models have limitations when applied to experimental data, particularly due to the limited number of choices a rodent can make before reaching satiety, which introduces noise into model parameter estimation (*61, 62*)- especially relevant when interpreting CNO sessions with fewer trials.

Overall, we could describe patterns of interindividual variability in preferences in terms of "cognitive profiles" with animals distributed between extrema (archetypes) rather than strongly defined clusters. Considering several cognitive dimensions rather than a single "adaptive" scale, recasts the "maladaptive" decisions (*50*) concept as "atypical" instead. We further show how cognitive profiles from the archetypal analysis relate to the decision processes in the model (i.e., reward, risk and time-out sensitivities) that generate choices. By moving along these dimensions of decision parameters, animals can exhibit different preference patterns based on their social and physical context, age, and previous experiences (*42*, *63*, *64*). Importantly, as computational models of decision-making are nonlinear, affecting one decision parameter (e.g., reward sensitivity) through chemogenetics may exert effects that depend upon the values of the other parameters. This reinforces the need to infer generative processes from the data rather than focusing purely on overt measures (e.g., performance or a given choice).

Specific manipulations of striatal subpopulations affecting cognitive profiles underscore that decision processes "parameters" are emergent properties from neural interactions. Brain markers correlate with individual variability in choice behavior, such as prefrontal serotonin (*40*) or the balance between striatal and prefrontal excitability (*37*). Nevertheless, much work remains to be done to determine the causality between brain markers and cognitive profiles. Here we detected marked variability in decision-making among a quasiclonal population of inbred mice. Instead of being the origin of interindividual variability in decision-making, brain markers may act as mediators of individual experience and the social context of cognitive strategies (*42*).

To study the causal implications of striatal subpopulations in choice behavior, we preferred DREADD over optogenetics because a temporal window for manipulation was not needed. Additionally, decision-making likely occurs during the whole task. Furthermore, we wanted to avoid the risk of nonphysiological synchronization of striatal neurons under specific optogenetic stimulation protocols. Indeed, striatal neurons are involved in action selection (*29*, *65–68*), and neurons promoting distinct goals (or weighting different decision parameters) are not activated synchronously (*30*, *33*, *34*). DREADDS partially circumvents this issue as a gain-of-function approach in this study. However, one limitation of our DREADD approach is that the CNO ligand may be converted to clozapine (*69*), an antipsychotic with sedative effects potentially decreasing behavioral impulsivity (*70–72*), but our mCherry controls excluded DREADD-independent effects.

Our results precise and extend the literature on the differential involvement of subparts of the striatum (*73*). The role of the DMS in sustaining instrumental associations (action-reward) has been widely proven before (*6*, *7*, *13*, *16*, *74*). Here, we show that the DMS is critical not only for maintaining the instrumental response of mice but also for determining their preferences under risk. We further provide a computational rationale (and iSPN versus dSPN distinction) for the effect of the DMS on risky choices (*75*) in terms of sensitivity to reward variance and reward saturation. While the NAc has also been implicated in risky decision-making, with specific involvement of iSPNs, as in our study(*15*), we found a lower NAc effect than the that of the DMS. This may be due predominantly to the instrumental nature of the task, as the NAc supports the acquisition of stimulus–outcome associations (*76–79*). There was no effect of DLS manipulations in the IGT, which is consistent with the known dissociation of the DMS and DLS (*12*, *13*, *17*) in controlling goal-directed behavior and habits, respectively. Despite extensive training, the mice did not develop rigid preferences (i.e., most mice had balanced choice behavior; Figure 1), suggesting that they did not develop habits (although we did not test choices in extinction).

We also provide novel data on the major anatomical and functional distinctions in the basal ganglia between the direct (dSPNs) and indirect (iSPNs) pathways. Concurrent views have proposed that dSPNs and iSPNs may work either in opposite or complementary ways to promote and oppose (or refine) action selection (*33*, *64*, *80–82*). However, in the DMS, we found that DREADD-mediated activation of both iSPNs and dSPNs enhanced risky choice through distinct computational mechanisms (i.e., risk sensitivity and reward saturation, respectively). This points to more complex interactions between pathways. The activity of iSPNs has been related to the encoding of nonrewarding events or changes in reward value (prediction errors), leading to the updating of action value (*36*) and henceforth task switching (*82*, *83*). Our computational account is consistent with this view, as reward saturation (i.e., how the subjective value saturates with reward magnitude) may arise from learning effects. However, we could not test this hypothesis directly, as including a learning rate induced correlations with explore/exploit and risk-seeking parameters and led to poor parameter recovery, probably due to the relatively small number of trials in the mice. As we could disentangle reward saturation from sensitivity to uncertainty in our multiple-choice setup, our results extend and explain those obtained in tasks involving binary choices. For example, activating iSPNs may increase seemingly stochastic choices in serial choice tasks in which the

exploratory choice may also constitute a risky (high variance) choice (*84*) because reward saturation affects the perception of risk (Supp. Figure 5).

Our findings reframe previous results on dopamine manipulation. DAT-knockdown mice make riskier choices in the IGT(*49*), and "safe" decision-makers show lower dopamine levels in the dorsal striatum. This suggests dopamine's effects on risk-taking arise from distinct D1R- and D2R-mediated mechanisms. The next step will be to use correlative approaches such as in vivo calcium imaging to characterize the physiological responses of DMS-dSPNs during choice preference. Indeed, concerted and cooperative activity between both striatal pathways is needed for proper action initiation and execution (*28–34*, 56, *66*). It will then be possible to compare how decisions are encoded in the different types of SPNs and compare such encoding between the Risk-averse, Explorers, and Optimizers mice. In addition, a correlation between the variables of the computational model and the activity patterns of certain dSPNs or iSPNs can also be found (*85*). Moreover, we could also measure whether, as our results suggest, the activity of dSPNs when mice are making a choice differs among the DMS, DLS and NAc. Using in vivo techniques, our results suggest that increasing neuronal excitability rather than directly activating targeted cells is beneficial. These questions should be carefully investigated in future work.

Understanding and aiding individuals in the context of decision-related disorders, such as pathological gambling (*86*) and drug addiction (*87*), necessitates a shift in perspective. Impairments in decision-making should be seen as discrepancies between anticipated outcomes and actual choices. We propose a framework where decision-making "profiles" arise from evolutionary trade-offs, shaped by parameters from the deliberative machinery leading to goal-directed choice. Biological and social (*42*) factors can shift the spectrum of potential strategies, likely governed by "meta-learning" rules (*88*), offering promising avenues for translational research.

549 **METHODS**
550
551 **Animal care and use**
552 All procedures were performed according to the Institutional Animal Care Committee
553 guidelines and were approved by the Local Ethical Committee (Comité d'Ethique et de Bien-
554 Être Animal du pôle santé de l'Université Libre de Bruxelles (ULB), Ref. No. 646 N). The mice
555 were maintained on a 12-hour dark/light cycle (lights on at 8 pm) with *ad libitum* access to food
556 and water. The room temperature was set to 22 ± 2 °C with constant humidity (40–60%). The
557 behavioral tests were performed during the dark photoperiod. Both male and female transgenic
558 mice (≥ 8 weeks) were used in all the behavioral experiments.
559
560 **Generation of transgenic mice**
561 The genetic background of all the transgenic mice used in this study was C57BL/6J. The mice
562 were heterozygous and maintained by breeding with C57BL/6 mice. Two transgenic mouse
563 lines were used: $A_{2A}$-Cre (*22*) and $D_1$-Cre (EY262; GENSAT)(*54*). Simple transgenic $A_{2A}$-Cre
564 or $D_1$-Cre mice were used for the virus-mediated targeting of iSPNs or dSPNs, respectively.
565
566
567 **Viral injections**
568 Under Avertin anesthesia (2,2,2-tribromoethaol 1.25%, 2-methyl-2-butanol 0.78%; 20 µL/g,
569 i.p.; Sigma Aldrich), male $A_{2A}$-Cre and $D_1$-Cre mice (≥ 8 weeks old), which allowed us to target
570 A2A- and D1- expressing neurons, respectively, received two injections (at 100 nL/min) under
571 stereotaxic control in the DLS (AP +0,8 L ±2,42 DV -3,2), DMS (AP +1,2 L ±1,33 DV -3,2)
572 or NAc (AP +1,95 L ±1,2 DV -4,85) of a Cre-dependent virus encoding hM3Dq (AAV5-hSyn1-
573 DIO-hM3Dq-mCherry, Addgene) or mCherry alone as a control (AAV5-hSyn1-DIO-mCherry,
574 Addgene). The coordinates were relative to Bregma according to the Franklin and Paxinos atlas
575 third edition. The injection volumes were as follows: DLS 0.4 µl, DMS 0.35 µl, and NAc 0.2
576 µl. The injections were delivered through a cannula connected to a Hamilton syringe (10 µL)
577 placed in a syringe pump (KDS-310-PLUS, KDScientific). The cannulas were lowered into the
578 brain and left in place for 10 min after infusion. A minimum of 3 weeks elapsed between the
579 stereotaxic injections to ensure optimal protein expression levels. For all the mice, the accuracy
580 of the injections was checked under a microscope after behavioral testing. The transfected area
581 could be identified via mCherry staining (see Supp. Figure 4). Behavioral data from animals
582 whose targeted area was not accurate or unilaterally injected were excluded from the analyses
583 (see table below).
584

585

| Mouse strain | Striatal area | Injected virus | Total n° of mice injected | **Excluded**(bad stereotaxic targeting) |
|---|---|---|---|---|
| $D_1$-Cre | DMS | mCherry | 20 | 0 |
| | | hM3Dq | 21 | 1 |
| | DLS | mCherry | 10 | 0 |
| | | hM3Dq | 10 | 1 |
| | NAc | mCherry | 23 | 6 |
| | | hM3Dq | 20 | 7 |
| $A_{2A}$-Cre | DMS | mCherry | 22 | 0 |
| | | hM3Dq | 15 | 2 |
| | DLS | mCherry | 9 | 0 |
| | | hM3Dq | 11 | 0 |
| | NAc | mCherry | 15 | 0 |
| | | hM3Dq | 12 | 3 |

586
587
588 **Mouse gambling task**
589 Behavioral testing was performed in 6 identical operant chambers (IMETRONIC, Pessac,
590 France). Four circular holes were available in a curved wall (the central hole was obstructed,
591 and only 4 holes were actually available), with LED lights at the rear of the chamber and a food
592 dispenser of 20 mg chocolate pellets (purified dustless precision pellets for rodents, Bioserv,
593 New Jersey, USA) on the opposite wall. The infrared beams in the nose-poke holes and the
594 magazine allowed the detection of entries to the nose-poke holes. Additionally, a house light
595 located in the ceiling of the chamber was used under specific conditions. These chambers were
596 in individual soundproof closets to avoid any external disturbances during training. Video
597 cameras placed on the top of the chambers allowed tracking of every session. The chambers
598 were connected to a computer and controlled by POLY software (IMETRONIC).
599 We used a rodent version of the Iowa Gambling Task (IGT) adapted from Young and colleagues
600 (2011)(*49*). Mice have to choose among 4 different options associated with a certain gain and
601 a certain loss (Figure 1). To achieve that goal, mice must undergo several phases of training.
602 The mice were placed on a reduced diet two days before training, aiming to reach 85–90% of
603 their original weight to enhance performance, given that food motivation drives the IGT.
604 Initially, they received 2.5 grams of food per mouse, which was adjusted the next day on the
605 basis of individual weight loss. During IGT training, food amounts remained stable but could
606 be adjusted if the mice fell out of the 85–90% weight range. Once the final training phase
607 commenced, the food quantities were fixed to avoid performance interference. The whole
608 protocol lasted between 1 month and a half and 3 months, depending on the animal. The mice
609 were trained daily at 9 am in a reverse light/dark cycle.

610 *Magazine habituation***.** The mice were trained for 10 minutes per day for 3 consecutive days.
611 During these sessions, every light was turned off. Fifteen seconds after the start of the session,
612 the pellet dispenser was activated to distribute a single pellet. The LED lights of the magazine
613 were then lit. The lights were turned off when the mice obtained the pellet, and the 15 second
614 cycle started again. The magazine would stay lit until retrieval of the pellet to allow an
615 association between the lit magazine and food to be formed.
616 *Nose–poke habituation.* Four nose-pokes holes were available, and a lit nose-poke meant that
617 it was an active nose-poke. This training was divided into two phases. In the initial stage, which
618 lasted 4 days, the mice responded to one active nose-poke out of 4, which was chosen randomly
619 in two daily sessions. If successful, 3 pellets were dispensed; otherwise, the sessions could last
620 as long as 10 minutes. For the first 2 days, the pellets were placed in active nose-pokes holes to
621 encourage exploration. In the second phase, all 4 nose-pokes holes remained active throughout
622 the 30-minute sessions, delivering a single pellet upon visitation. Training was continued until
623 the mice achieved 40 responses in two consecutive days, with individual progress varying.
624 Upon reaching this criterion, the mice progressed to the next training phase.
625 *Forced-choice IGT***.** Different nose-pokes holes had specific rewards and punishments. The
626 mice underwent 3 days of forced 30-minute sessions, where a random nose-poke hole was
627 activated, leading to either a reward or punishment afterward.
628 *Final IGT*. The mice underwent 30-minute sessions with up to 100 trials. Sessions began with
629 a lit magazine that the mice had to approach. A 5-second interval preceded the illumination of
630 4 nose-poke hole lights; premature responses triggered a time-out. After the interval, mice had
631 10 seconds to complete a nose poke, and omissions led to new trials. Correct responses
632 extinguished lights and rewarded mice on the basis of specific nose-poke hole probabilities (P1,
633 P2, P3, and P4). Rewards involved pellet delivery, followed by an 8-second period before the
634 next trial began. Punishments initiated a time-out with flashing lights, after which the mice
635 could start a new trial. Perseverative responses were noted but not punished. The mice were
636 trained until their preference stabilized over four consecutive days, establishing a baseline.
637
638 **CNO treatment**
639 When the mice reached a stable pattern of preference over 4 days of IGT (see below), we started
640 treatment with 0.9% NaCl for 2 days. All injections were applied 30 minutes before the start of
641 the session. After receiving the CNO injection (1 mg/kg), the mice underwent their last session
642 of IGT. Immediately after the end of this session, the mice were euthanized. A part of the group
643 was used to perform the electrophysiological recordings. In this group, there was a balanced
644 representation of hM3Dq-expressing mice and control mice.
645
646 **Immunohistochemistry**
647 After the behavioral experiments were completed, the mice were deeply anesthetized with
648 avertin (2,2,2-tribromoethaol 1.25%, 2-methyl-2-butanol 0.78%; 20 µL/g, i.p.; Sigma Aldrich)
649 and transcardially perfused with PBS followed by 4% paraformaldehyde in PBS. The brains
650 were removed and postfixed overnight at 4 °C. Then, 30-µm coronal slices containing the
651 striatum were cut with a vibratome (VT1000 S; Leica) and stored in PBS. The sections were
652 incubated overnight with a dilution (1/2000 in 1% NHS 0.1% PBST) of a rabbit Fos primary
653 antibody (Santa Cruz). Next, the slices were washed twice for 5 min with PBST and incubated

654 for 1.5 h with a dilution (1/200 in 1% NHS 0.1% PBST) of donkey anti-rabbit secondary
655 antibody Alexa 647 (far red) (Jackson ImmunoResearch). The slices were then washed twice
656 for 5 minutes with PBST. Finally, DAPI nuclear staining was performed (10 minutes in a
657 1/5000 DAPI solution in 0.01 M PBS), and the samples were washed for 5 min at 0.01 M PBS,
658 mounted on glass slides and cover slipped with Fluoromount.
659 For the GFP immunostaining, the protocol was the same as that for the chicken anti-GFP
660 primary antibody (1/2000, Santa Cruz) and the goat anti-chicken secondary antibody Alexa 647
661 (1/400, Jackson ImmunoResearch).
662 Image acquisition was performed with a Zeiss Axioimager Z1 at 20X magnification, and the
663 images were processed with AxioVision (Zeiss) software. Cell counting was performed with
664 the open-source software FIJI. Additionally, we used Axio Zoom.V16 (Zeiss) and its tiling tool
665 to obtain whole-slice surface images.
666
667 **Acute brain slice preparation.** Recordings were made ex vivo on brain slices that were kept
668 alive. The animals were anesthetized with halothane and then euthanized via decapitation. Their
669 brains were quickly collected, adhered to a vibratome plate and submerged in a vibratome
670 container filled with cutting solution at 4 °C. Brain slices were cut (220 µm thick) and then
671 transferred to an incubator chamber filled with artificial cerebrospinal fluid (aCSF) (Table MM-
672 1) at 34 °C for at least 60 minutes before the recordings started. This resting period is required
673 for the recovery of proper neuronal metabolic activity. Both the cutting solution and aCSF were
674 constantly oxygenated, as they were supplied with a 95% $O_2$ and 5% $CO_2$ mixture to avoid
675 hypoxia. After recovery, the brain slices were transferred one by one to the recording chamber,
676 which was constantly perfused with oxygenized aCSF at room temperature. The setup was
677 equipped with a camera and a Zeiss upright microscope (Axioskop 2FS Plus, Zeiss), which first
678 allowed for the identification of different areas in the striatum in the slices (5x/0,15 EC
679 PlanNEOFLUAR objective lens, Zeiss) and secondly, for the identification of different cell
680 populations (63x water-immersion objective, Zeiss). Depending on the group of mice tested,
681 we recorded from either the DLS, DMS or NAc (according to the injection site of the group).
682
683 **Recordings.** The patch pipette was obtained with a vertical two-stage puller (PIP 5, HEKA),
684 and during the recordings, it was filled with intracellular solution (Table MM-1). The Ag/AgCl
685 electrode in the pipette was connected to an EPC-10 (Heka) amplifier, whose signal was
686 recorded with Patchmaster software (Heka). Intrinsic excitability was studied via a current
687 clamp. First, the resting membrane potential was measured (without applying any current).
688 Afterward, a negative amount of current was injected into the cell to maintain its membrane
689 potential at a value of −80 mV. Increasing intensity currents (increase of 10 pA) were then
690 injected to depolarize the cell and trigger action potentials. Rheobase is defined as the minimum
691 amount of injected current that allows the induction of an action potential. Evoked excitatory
692 postsynaptic currents (EPSCs) were recorded with the internal solution supplemented with
693 spermine (0.1 mM). Spermine blocks calcium-permeable AMPA-Rs at positive potentials and
694 thus allows us to distinguish calcium-permeable from noncalcium-permeable AMPA-R
695 currents(*89*). Evoked EPSCs were isolated from GABAergic currents via the application of
696 gabazine (or SR-95531, 10 µm). We do not make any reference to kainate-Rs because of the
697 difficulties in distinguishing them from AMPA-Rs, as they cannot be easily separated

pharmacologically(*90*). Thus, we will use shortcut AMPA-R currents to refer to EPSCs resulting from both AMPA-Rs and kainate-Rs. Evoked EPSCs were recorded in the striatum (DLS, DMS or NAc) by placing a bipolar stimulating electrode (SNEX-200, Science Products GmbH) in the corpus callosum (white matter between the cortex and the striatum) to ensure specific stimulation of the corticostriatal pathway. The duration of the stimulation current pulses was constant and set at 0.2 ms. The intensity of the stimulation was adjusted for each cell and set at the minimal value needed to evoke the largest eEPSC (from 600 to 1000 μA). The slices were stimulated every 10 seconds. Evoked EPSCs were recorded successively at −70 mV, 0 mV and +40 mV to allow computation of the ratio between AMPA-R-mediated currents and NMDA-R-mediated currents. The AMPA-R/NMDA-R ratio provides a sensitive measure to detect differences in glutamatergic synaptic strength between two experimental groups(*91*). AMPA-R-mediated EPSCs were obtained at a holding potential of −70 mV, where no NMDA-R conductance was observed, due to the magnesium block. The sum of AMPA-R- and NMDA-R-mediated EPSCs was evoked at a holding potential of +40 mV. In MSNs, AMPA-R-mediated currents were pharmacologically isolated via the application of L689,560 (10 μM), a noncompetitive NMDA-R antagonist, to the solution. The NMDA-R current was obtained via off-line subtraction of the two traces. eEPSCs were digitized at a frequency of 10 kHz and filtered online with a low-pass Bessel filter, using a cutoff frequency set at 3 kHz. Ten eEPSC traces were averaged for each condition.

**Statistical analyses**

**General statistical analyses**. The results are plotted as the means ± sems. The total number (n) of observations in each group and the statistics used are indicated in the figure legends. Classical comparisons between means were performed via parametric tests (Student's t test or ANOVA for comparing more than two groups when the parameters followed a normal distribution (Shapiro test, $P > 0.05$) and nonparametric tests (here, Wilcoxon or Mann–Whitney tests) when the distribution was skewed. Multiple comparisons were corrected via a sequentially rejective multiple test procedure (Holm). Probability distributions were compared via the Kolmogorov–Smirnov (KS) test, and proportions were evaluated via the chi-square test ($\chi 2$). All the statistical tests were two-sided. $P > 0.05$ was considered not to indicate statistical significance.

**Archetypal analysis**. Computations were performed via the ParTI routine in MATLAB (*53*). Briefly, given an $n \times m$ matrix representing a multivariate dataset with n observations (n = number of animals) and m attributes (here, m = 7, the 4 preferences, the number of trials, the percentage of premature responses and the percentage of omissions), the archetypal analysis finds the matrix Z of k multidimensional archetypes (k is the number of archetypes). k was forced to be lower than m so that there cannot be more groups than dimensions. Z is obtained by minimizing $\| X-\alpha ZT \| \| 2$, with α representing the coefficients of the archetypes ($\alpha i,1...k \geq 0$ and $\sum \alpha i,1...k = 1$), and $| |.\|2$ representing a matrix norm. The archetype is also a convex combination of the data points $Z = XT\delta$, with $\delta \geq 0$, and their sum must be 1. The α coefficient depicts the relative archetypal composition of a given observation. For k = 3 archetypes and an observation i, $\alpha i,1$; $\alpha i,2$; $\alpha i,3 \geq 0$ and $\alpha i,1 + \alpha i,2 + \alpha i,3 = 1$. A ternary plot can then be used to visualize the data ($\alpha i,1$; $\alpha i,2$; $\alpha i,2$) used to assign an individual behavior to its nearest archetype

(i.e., k max($\alpha_{i,1}$; $\alpha_{i,2}$; $\alpha_{i,3}$)). $\alpha_{i,j}$ are also used as variables to estimate population archetypal composition. The pure archetype corresponds to 1, the archetypal composition decreases linearly with increasing distance from the archetype, and 0 corresponds to points on the opposite side.

**Reinforcement-learning model**

**Decision model**. Decision-making models describe the probability Pi of choosing the next state i as a function (the "choice rule") of a decision variable. We modeled decisions between the four options with a "softmax" decision rule, defined by $P_i = e{-}ß(\text{sum}(V_a)/(1 + \exp(-ß(\text{sum}(V_i)))$, where β is an inverse temperature parameter reflecting the sensitivity of choice to the difference of decision variables (values) $V_i$. A large β corresponds to exploitation, i.e., choosing the option that seems the best thus far, whereas a small β corresponds to exploration (at β=0, all options are chosen equally).

**Decision variable**. The decision variable or value $V$ of an option is modeled as $V = R + P + U$, i.e., the sum of the expected (average) reward, the expected punishment, and the expected uncertainty (risk). The expected reward was given by $R = p$*(Number of Pellets)^(ρ), where $p$ is the probability that the choice is rewarded, and ρ is a model parameter of how the subjective value of the option depends on the reward magnitude. A ρ value close to 1 corresponds to a linear subjective value (the subjective value equals the reward size). A value of ρ smaller than 1 corresponds to reward saturation, e.g., the subjective value of 4 pellets is less than 4 times the value of one pellet. The expected punishment depends on the opportunity cost of time $Tc$, i.e., the average reward to be lost if there is a time out: $P = (1\text{-}p)$*(TimeOut duration)^( $T$)*$Tc$; where $T$ scales the perception of the time-out duration. A $T$ value close to 1 corresponds to a linear perception of time, and a γ smaller (larger) than 1 corresponds to underweighting (respectively, overweighting) long durations. Risk corresponded to variance in the outcome, $V(X) = E(X^2) - E(X)^2 = pR^2 - (1-p)P^2 - (pR - (1-p)P)^2$ ; scaled by the free parameter φ (a negative φ corresponds to risk aversion, and a positive φ corresponds to risk seeking).

**Fitting the model**. The free parameters of the model (β, ρ, $T$, φ) were fitted by maximizing the data likelihood. Given a sequence of choices c = c1…T, the data likelihood is the product of their probability given by the softmax decision rule. We used the fmincon function in MATLAB to perform the model fitting, with the constraints that β ∈ ]0,10], ϕ ∈ [-10,10] and $Rs$ ∈ ]0,2] and $Rp$ ∈ ]0,2].

**Generative properties of the model**. To simulate the choice patterns for the different archetypes, we simply displayed the choice probabilities (toward which a simulation would converge at a large n) obtained for the average set of parameters fitted on the average of the individuals assigned to an archetype.

**Recovery analysis.** $N_{animals}$ series of $n_{trials}$ choices were generated for a set of parameters (β, ρ, $T$, φ) and then fitted with the above procedure to assess the quality of parameter recovery. In Supp. Figure 2, we used systematic variations of one parameter with the others held constant close to average fit from the data ($\beta_0$ = 2, $\rho_0$=0.8, $T_0$=1, $\varphi_0$=−2), and $N_{animals}$ =20; $n_{trials}$=50. For Supp. Figure 5, we performed $N_{animals}$ =200 pairs of simulations, one under a control $n_{trials}$ =50

and one under CNO $n_{trials}$ =20, each pair with the same set of parameters drawn from a normal distribution around the average parameters fitted from the data ($\beta_0$ = 2, $\rho_0$=0.8, $T_0$=1, $\varphi_0$=−2). We then fitted the model parameters for each condition separately.

**Model comparison**. We used the Bayesian information criterion (twice the negative log likelihood plus the number of parameters times the log of the number of trials) to test whether simpler versions of the model could provide more parsimonious fits. We also displayed some of the model parameter fits in simpler models with a divergence toward limit values to show poorer fits.

794 **References**

795


796 1. R. J. Dolan, P. Dayan, Goals and Habits in the Brain. *Neuron* **80**, 312–325 (2013).
797 2. A. Rangel, C. Camerer, P. R. Montague, A framework for studying the neurobiology
798 of value-based decision making. *Nat. Rev. Neurosci.* **9**, 545–556 (2008).
799 3. A. M. Graybiel, The basal ganglia and chunking of action repertoires. *Neurobiol.*
800 *Learn. Mem.* **70**, 119–136 (1998).
801 4. O. Hikosaka, Neural systems for control of voluntary action--a hypothesis. *Adv.*
802 *Biophys.* **35**, 81–102 (1998).
803 5. P. Redgrave, T. J. Prescott, K. Gurney, The basal ganglia: a vertebrate solution to the
804 selection problem? *Neuroscience* **89**, 1009–1023 (1999).
805 6. G. Hart, B. K. Leung, B. W. Balleine, Dorsal and ventral streams: the distinct role of
806 striatal subregions in the acquisition and performance of goal-directed actions. *Neurobiol.*
807 *Learn. Mem.* **108**, 104–118 (2014).
808 7. D. A. Kupferschmidt, K. Juczewski, G. Cui, K. A. Johnson, D. M. Lovinger, Parallel,
809 but Dissociable, Processing in Discrete Corticostriatal Inputs Encodes Skill Learning. *Neuron*
810 **96**, 476-489.e5 (2017).
811 8. B. Mai, S. Sommer, W. Hauber, Dopamine D1/D2 Receptor Activity in the Nucleus
812 Accumbens Core But Not in the Nucleus Accumbens Shell and Orbitofrontal Cortex
813 Modulates Risk-Based Decision Making. *Int. J. Neuropsychopharmacol.* **18**, pyv043 (2015).
814 9. C. M. Stopper, S. Khayambashi, S. B. Floresco, Receptor-Specific Modulation of
815 Risk-Based Decision Making by Nucleus Accumbens Dopamine. *Neuropsychopharmacology*
816 **38**, 715–728 (2013).
817 10. Y. Vandaele, N. R. Mahajan, D. J. Ottenheimer, J. M. Richard, S. P. Mysore, P. H.
818 Janak, Distinct recruitment of dorsomedial and dorsolateral striatum erodes with extended
819 training. *eLife* **8**, e49536 (2019).
820 11. J.-H. Yang, R.-M. Liao, Dissociable contribution of nucleus accumbens and
821 dorsolateral striatum to the acquisition of risk choice behavior in the rat. *Neurobiol. Learn.*
822 *Mem.* **126**, 67–77 (2015).
823 12. H. H. Yin, B. J. Knowlton, B. W. Balleine, Lesions of dorsolateral striatum preserve
824 outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.*
825 **19**, 181–189 (2004).
826 13. H. H. Yin, S. B. Ostlund, B. J. Knowlton, B. W. Balleine, The role of the dorsomedial
827 striatum in instrumental conditioning. *Eur. J. Neurosci.* **22**, 513–523 (2005).
828 14. H. H. Yin, S. P. Mulcare, M. R. F. Hilário, E. Clouse, T. Holloway, M. I. Davis, A. C.
829 Hansson, D. M. Lovinger, R. M. Costa, Dynamic reorganization of striatal circuits during the
830 acquisition and consolidation of a skill. *Nat. Neurosci.* **12**, 333–341 (2009).
831 15. K. A. Zalocusky, C. Ramakrishnan, T. N. Lerner, T. J. Davidson, B. Knutson, K.
832 Deisseroth, Nucleus accumbens D2R cells signal prior outcomes and control risky decision-
833 making. *Nature* **531**, 642–646 (2016).
834 16. B. W. Balleine, M. Liljeholm, S. B. Ostlund, The integrative function of the basal
835 ganglia in instrumental conditioning. *Behav. Brain Res.* **199**, 43–52 (2009).
836 17. H. H. Yin, B. J. Knowlton, The role of the basal ganglia in habit formation. *Nat. Rev.*
837 *Neurosci.* **7**, 464–476 (2006).
838 18. J. Peak, G. Hart, B. W. Balleine, From learning to action: the integration of dorsal
839 striatal input and output pathways in instrumental conditioning. *Eur. J. Neurosci.* **49**, 658–671
840 (2019).
841 19. S. Bariselli, W. C. Fobbs, M. C. Creed, A. V. Kravitz, A competitive model for striatal
842 action selection. *Brain Res.* **1713**, 70–79 (2019).
843 20. H. S. Bateup, E. Santini, W. Shen, S. Birnbaum, E. Valjent, D. J. Surmeier, G. Fisone,

E. J. Nestler, P. Greengard, Distinct subclasses of medium spiny neurons differentially regulate striatal motor behaviors. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 14845–14850 (2010).

21. F. Carvalho Poyraz, E. Holzner, M. R. Bailey, J. Meszaros, L. Kenney, M. A. Kheirbek, P. D. Balsam, C. Kellendonk, Decreasing Striatopallidal Pathway Function Enhances Motivation by Energizing the Initiation of Goal-Directed Action. *J. Neurosci. Off. J. Soc. Neurosci.* **36**, 5988–6001 (2016).

22. P. F. Durieux, B. Bearzatto, S. Guiducci, T. Buch, A. Waisman, M. Zoli, S. N. Schiffmann, A. de Kerchove d'Exaerde, D2R striatopallidal neurons inhibit both locomotor and drug reward processes. *Nat. Neurosci.* **12**, 393–395 (2009).

23. P. F. Durieux, S. N. Schiffmann, A. de Kerchove d'Exaerde, Differential regulation of motor control and response to dopaminergic drugs by D1R and D2R neurons in distinct dorsal striatum subregions. *EMBO J.* **31**, 640–653 (2012).

24. A. V. Kravitz, B. S. Freeze, P. R. L. Parker, K. Kay, M. T. Thwin, K. Deisseroth, A. C. Kreitzer, Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature* **466**, 622–626 (2010).

25. E. A. Yttri, J. T. Dudman, Opponent and bidirectional control of movement velocity in the basal ganglia. *Nature* **533**, 402–406 (2016).

26. G. E. Alexander, M. D. Crutcher, Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends Neurosci.* **13**, 266–271 (1990).

27. C. R. Gerfen, T. M. Engber, L. C. Mahan, Z. Susel, T. N. Chase, F. J. Monsma, D. R. Sibley, D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. *Science* **250**, 1429–1432 (1990).

28. G. Barbera, B. Liang, L. Zhang, C. R. Gerfen, E. Culurciello, R. Chen, Y. Li, D.-T. Lin, Spatially Compact Neural Clusters in the Dorsal Striatum Encode Locomotion Relevant Information. *Neuron* **92**, 202–213 (2016).

29. G. Cui, S. B. Jun, X. Jin, M. D. Pham, S. S. Vogel, D. M. Lovinger, R. M. Costa, Concurrent activation of striatal direct and indirect pathways during action initiation. *Nature* **494**, 238–242 (2013).

30. A. Klaus, G. J. Martins, V. B. Paixao, P. Zhou, L. Paninski, R. M. Costa, The Spatiotemporal Organization of the Striatum Encodes Action Space. *Neuron* **95**, 1171-1180.e7 (2017).

31. J. E. Markowitz, W. F. Gillis, C. C. Beron, S. Q. Neufeld, K. Robertson, N. D. Bhagat, R. E. Peterson, E. Peterson, M. Hyun, S. W. Linderman, B. L. Sabatini, S. R. Datta, The Striatum Organizes 3D Behavior via Moment-to-Moment Action Selection. *Cell* **174**, 44-58.e17 (2018).

32. J. G. Parker, J. D. Marshall, B. Ahanonu, Y.-W. Wu, T. H. Kim, B. F. Grewe, Y. Zhang, J. Z. Li, J. B. Ding, M. D. Ehlers, M. J. Schnitzer, Diametric neural ensemble dynamics in parkinsonian and dyskinetic states. *Nature* **557**, 177–182 (2018).

33. M. Weglage, E. Wärnberg, I. Lazaridis, D. Calvigioni, O. Tzortzi, K. Meletis, Complete representation of action space and value in all dorsal striatal pathways. *Cell Rep.* **36**, 109437 (2021).

34. C. Varin, A. Cornil, D. Houtteman, P. Bonnavion, A. de Kerchove d'Exaerde, The respective activation and silencing of striatal direct and indirect pathway neurons support behavior encoding. *Nat. Commun.* **14**, 4982 (2023).

35. A. G. E. Collins, M. J. Frank, Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychol. Rev.* **121**, 337–366 (2014).

36. J. Peak, B. Chieng, G. Hart, B. W. Balleine, Striatal direct and indirect pathway neurons differentially control the encoding and updating of goal-directed learning. *eLife* **9**, e58544 (2020).

37. A. Fitoussi, P. Renault, C. Le Moine, E. Coutureau, M. Cador, F. Dellu-Hagedorn, Inter-individual differences in decision-making, flexible and goal-directed behaviors: novel insights within the prefronto-striatal networks. *Brain Struct. Funct.* **223**, 897–912 (2018).
38. T. Kurikawa, T. Haga, T. Handa, R. Harukuni, T. Fukai, Neuronal stability in medial frontal cortex sets individual variability in decision-making. *Nat. Neurosci.* **21**, 1764–1773 (2018).
39. M. Rivalan, E. Coutureau, A. Fitoussi, F. Dellu-Hagedorn, Inter-Individual Decision-Making Differences in the Effects of Cingulate, Orbitofrontal, and Prelimbic Cortex Lesions in a Rat Gambling Task. *Front. Behav. Neurosci.* **5** (2011).
40. E. Pittaras, J. Callebert, M. Chennaoui, A. Rabat, S. Granon, Individual behavioral and neurochemical markers of unadapted decision-making processes in healthy inbred mice. *Brain Struct. Funct.* **221**, 4615–4629 (2016).
41. M. Dongelmans, R. Durand-de Cuttoli, C. Nguyen, M. Come, E. K. Duranté, D. Lemoine, R. Brito, T. Ahmed Yahia, S. Mondoloni, S. Didienne, E. Bousseyrol, B. Hannesse, L. M. Reynolds, N. Torquet, D. Dalkara, F. Marti, A. Mourot, J. Naudé, P. Faure, Chronic nicotine increases midbrain dopamine neuron activity and biases individual strategies towards reduced exploration in mice. *Nat. Commun.* **12**, 6945 (2021).
42. N. Torquet, F. Marti, C. Campart, S. Tolu, C. Nguyen, V. Oberto, M. Benallaoua, J. Naudé, S. Didienne, N. Debray, S. Jezequel, L. Le Gouestre, B. Hannesse, J. Mariani, A. Mourot, P. Faure, Social interactions impact on the dopaminergic system and drive individuality. *Nat. Commun.* **9**, 3081 (2018).
43. D. Belin, A. C. Mar, J. W. Dalley, T. W. Robbins, B. J. Everitt, High impulsivity predicts the switch to compulsive cocaine-taking. *Science* **320**, 1352–1355 (2008).
44. A. Elton, C. T. Smith, M. H. Parrish, C. A. Boettiger, Neural Systems Underlying Individual Differences in Intertemporal Decision-making. *J. Cogn. Neurosci.* **29**, 467–479 (2017).
45. M. Spoelder, J. P. Flores Dourojeanni, K. C. G. de Git, A. M. Baars, H. M. B. Lesscher, L. J. M. J. Vanderschuren, Individual differences in voluntary alcohol intake in rats: relationship with impulsivity, decision making and Pavlovian conditioned approach. *Psychopharmacology (Berl.)* **234**, 2177–2196 (2017).
46. T. J. Hynes, K. M. Hrelja, B. A. Hathaway, C. D. Hounjet, C. S. Chernoff, S. A. Ebsary, G. D. Betts, B. Russell, L. Ma, S. Kaur, C. A. Winstanley, Dopamine neurons gate the intersection of cocaine use, decision making, and impulsivity. *Addict. Biol.* **26**, e13022 (2021).
47. J.-M. N. Ferland, C. A. Winstanley, Risk-preferring rats make worse decisions and show increased incubation of craving after cocaine self-administration. *Addict. Biol.* **22**, 991–1001 (2017).
48. N. SanMiguel, M. Pardo, C. Carratalá-Ros, L. López-Cruz, J. D. Salamone, M. Correa, Individual differences in the energizing effects of caffeine on effort-based decision-making tests in rats. *Pharmacol. Biochem. Behav.* **169**, 27–34 (2018).
49. J. W. Young, J. van Enkhuizen, C. A. Winstanley, M. A. Geyer, Increased risk-taking behavior in dopamine transporter knockdown mice: further support for a mouse model of mania. *J. Psychopharmacol. Oxf. Engl.* **25**, 934–943 (2011).
50. M. Rivalan, S. H. Ahmed, F. Dellu-Hagedorn, Risk-prone individuals prefer the wrong options on a rat version of the Iowa Gambling Task. *Biol. Psychiatry* **66**, 743–749 (2009).
51. A. Bechara, H. Damasio, D. Tranel, A. R. Damasio, Deciding advantageously before knowing the advantageous strategy. *Science* **275**, 1293–1295 (1997).
52. A. Cutler, L. Breiman, Archetypal Analysis. *Technometrics* **36**, 338–347 (1994).
53. Y. Hart, H. Sheftel, J. Hausser, P. Szekely, N. B. Ben-Moshe, Y. Korem, A. Tendler,

A. E. Mayo, U. Alon, Inferring biological tasks using Pareto analysis of high-dimensional data. *Nat. Methods* **12**, 233–235 (2015).
54. S. Gong, M. Doughty, C. R. Harbaugh, A. Cummins, M. E. Hatten, N. Heintz, C. R. Gerfen, Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J. Neurosci. Off. J. Soc. Neurosci.* **27**, 9817–9823 (2007).
55. B. L. Roth, DREADDs for Neuroscientists. *Neuron* **89**, 683–694 (2016).
56. C. Varin, A. de Kerchove d'Exaerde , Neuronal encoding of behaviors and instrumental learning in the dorsal striatum. *Trends Neurosci.* **48**, 77-91 (2025).
57. B. W. Balleine, The Meaning of Behavior: Discriminating Reflex and Volition in the Brain. *Neuron* **104**, 47–62 (2019).
58. S. Budaev, C. Jørgensen, M. Mangel, S. Eliassen, J. Giske, Decision-Making From the Animal Perspective: Bridging Ecology and Subjective Cognition. *Front. Ecol. Evol.* **7** (2019).
59. C. Hultman, N. Tjernström, S. Vadlin, M. Rehn, K. W. Nilsson, E. Roman, C. Åslund, Exploring decision-making strategies in the Iowa gambling task and rat gambling task. *Front. Behav. Neurosci.* **16,** 964348 (2022).
60. A. Gomez-Marin, A. A. Ghazanfar, The Life of Behavior. *Neuron* **104**, 25–36 (2019).
61. N.D. Daw, "Trial-by-trial data analysis using computational models" in *Decision Making, Affect, and Learning: Attention and Performance XXIII* (Oxford University Press, Oxford (2011)), pp. 3-38 .
62. L. Danwitz, D. Mathar, E. Smith, D. Tuzsus, J. Peters, Parameter and model recovery of reinforcement learning models for restless bandit problems, *Computational Brain & Behavior* **5**, 547-563 (2022).
63. F. D. Zeeb, C. A. Winstanley, Functional disconnection of the orbitofrontal cortex and basolateral amygdala impairs acquisition of a rat gambling task and disrupts animals' ability to alter decision-making behavior after reinforcer devaluation. *J. Neurosci. Off. J. Soc. Neurosci.* **33**, 6434–6443 (2013).
64. M. Cassotti, A. Aïte, A. Osmont, O. Houdé, G. Borst, What have we learned about the processes involved in the Iowa Gambling Task from developmental studies? *Front. Psychol.* **5** (2014).
65. J. H. Shin, M. Song, S.-B. Paik, M. W. Jung, Spatial organization of functional clusters representing reward and movement information in the striatal direct and indirect pathways. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 27004–27015 (2020).
66. F. Tecuapetla, X. Jin, S. Q. Lima, R. M. Costa, Complementary Contributions of Striatal Projection Pathways to Action Initiation and Execution. *Cell* **166**, 703–715 (2016).
67. Y. Isomura, T. Takekawa, R. Harukuni, T. Handa, H. Aizawa, M. Takada, T. Fukai, Reward-Modulated Motor Information in Identified Striatum Neurons. *J. Neurosci.* **33**, 10209–10220 (2013).
68. X. Jin, F. Tecuapetla, R. M. Costa, Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences. *Nat. Neurosci.* **17**, 423–430 (2014).
69. D. F. Manvich, K. A. Webster, S. L. Foster, M. S. Farrell, J. C. Ritchie, J. H. Porter, D. Weinshenker, The DREADD agonist clozapine N-oxide (CNO) is reverse-metabolized to clozapine and produces clozapine-like interoceptive stimulus effects in rats and mice. *Sci. Rep.* **8**, 3840 (2018).
70. H. A. Haidary, R. K. Padhy, "Clozapine" in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2024; http://www.ncbi.nlm.nih.gov/books/NBK535399/).
71. C. Maurice-Gélinas, J. Deslauriers, C. Monpays, P. Sarret, S. Grignon, The 5α-reductase inhibitor finasteride increases suicide-related aggressive behaviors and blocks clozapine-induced beneficial effects in an animal model of schizophrenia. *Physiol. Behav.* **191**, 65–72 (2018).
72. B. Spivak, E. Shabash, B. Sheitman, A. Weizman, R. Mester, The effects of clozapine

994 versus haloperidol on measures of impulsive aggression and suicidality in chronic
995 schizophrenia patients: an open, nonrandomized, 6-month study. *J. Clin. Psychiatry* **64**, 755–
996 760 (2003).
997 73. J. Cox, I. B. Witten, Striatal circuits for reward learning and decision-making. *Nat.*
998 *Rev. Neurosci.* **20**, 482–494 (2019).
999 74. M. Malvaez, Neural substrates of habit. *J. Neurosci. Res.* **98**, 986–997 (2020).
1000 75. J. D. Schumacher, M. van Holstein, V. Bagrodia, H. B. Le Bouder, S. B. Floresco,
1001 Dorsomedial striatal contributions to different forms of risk/reward decision making.
1002 *Neurobiol. Learn. Mem.* **178**, 107369 (2021).
1003 76. E. S. Calipari, R. C. Bagot, I. Purushothaman, T. J. Davidson, J. T. Yorgason, C. J.
1004 Peña, D. M. Walker, S. T. Pirpinias, K. G. Guise, C. Ramakrishnan, K. Deisseroth, E. J.
1005 Nestler, In vivo imaging identifies temporal signature of D1 and D2 medium spiny neurons in
1006 cocaine reward. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 2726–2731 (2016).
1007 77. J. M. Gmaz, J. E. Carmichael, M. A. van der Meer, Persistent coding of outcome-
1008 predictive cue features in the rat nucleus accumbens. *eLife* **7**, e37275 (2018).
1009 78. J. D. Salamone, M. Correa, The mysterious motivational functions of mesolimbic
1010 dopamine. *Neuron* **76**, 470–485 (2012).
1011 79. H. H. Yin, S. B. Ostlund, B. W. Balleine, Reward-guided learning beyond dopamine
1012 in the nucleus accumbens: the integrative functions of cortico-basal ganglia networks. *Eur. J.*
1013 *Neurosci.* **28**, 1437–1448 (2008).
1014 80. Y. Li, Y. He, M. Chen, Z. Pu, L. Chen, P. Li, B. Li, H. Li, Z.-L. Huang, Z. Li, J.-F.
1015 Chen, Optogenetic Activation of Adenosine A2A Receptor Signaling in the Dorsomedial
1016 Striatopallidal Neurons Suppresses Goal-Directed Behavior. *Neuropsychopharmacol. Off.*
1017 *Publ. Am. Coll. Neuropsychopharmacol.* **41**, 1003–1013 (2016).
1018 81. T. Macpherson, M. Matsumoto, H. Gomi, J. Morimoto, E. Uchibe, T. Hikida, Parallel
1019 and hierarchical neural mechanisms for adaptive and predictive behavioral control. *Neural*
1020 *Netw. Off. J. Int. Neural Netw. Soc.* **144**, 507–521 (2021).
1021 82. S. Nonomura, K. Nishizawa, Y. Sakai, Y. Kawaguchi, S. Kato, M. Uchigashima, M.
1022 Watanabe, K. Yamanaka, K. Enomoto, S. Chiken, H. Sano, S. Soma, J. Yoshida, K.
1023 Samejima, M. Ogawa, K. Kobayashi, A. Nambu, Y. Isomura, M. Kimura, Monitoring and
1024 Updating of Action Selection for Goal-Directed Behavior through the Striatal Direct and
1025 Indirect Pathways. *Neuron* **99**, 1302-1314.e5 (2018).
1026 83. J. H. Shin, D. Kim, M. W. Jung, Differential coding of reward and movement
1027 information in the dorsomedial striatal direct and indirect pathways. *Nat. Commun.* **9**, 404
1028 (2018).
1029 84. K. Delevich, B. Hoshal, L. Z. Zhou, Y. Zhang, S. Vedula, W. C. Lin, J. Chase, A. G.
1030 E. Collins, L. Wilbrecht, Activation, but not inhibition, of the indirect pathway disrupts
1031 choice rejection in a freely moving, multiple-choice foraging task. *Cell Rep.* **40**, 111129
1032 (2022).
1033 85. K. Samejima, Y. Ueda, K. Doya, M. Kimura, Representation of action-specific reward
1034 values in the striatum. *Science* **310**, 1337–1340 (2005).
1035 86. R. van den Bos, W. Davies, F. Dellu-Hagedorn, A. E. Goudriaan, S. Granon, J.
1036 Homberg, M. Rivalan, J. Swendsen, W. Adriani, Cross-species approaches to pathological
1037 gambling: a review targeting sex differences, adolescent vulnerability and ecological validity
1038 of research tools. *Neurosci. Biobehav. Rev.* **37**, 2454–2471 (2013).
1039 87. W. W. Stoops, D. N. Kearns, Decision-making in addiction: Current knowledge,
1040 clinical implications and future directions. *Pharmacol. Biochem. Behav.* **164**, 1–3 (2018).
1041 88. N. Schweighofer, K. Doya, Meta-learning in reinforcement learning. *Neural Netw.*
1042 *Off. J. Int. Neural Netw. Soc.* **16**, 5–9 (2003).
1043 89. P. Jonas, N. Burnashev, Molecular mechanisms controlling calcium entry through

AMPA-type glutamate receptor channels. *Neuron* **15**, 987–990 (1995).
90. D. E. Jane, D. Lodge, G. L. Collingridge, Kainate receptors: pharmacology, function and therapeutic potential. *Neuropharmacology* **56**, 90–113 (2009).
91. S. Kourrich, P. E. Rothwell, J. R. Klug, M. J. Thomas, Cocaine experience controls bidirectional synaptic plasticity in the nucleus accumbens. *J. Neurosci. Off. J. Soc. Neurosci.* **27**, 7921–7928 (2007).

**Acknowledgments:** We thank Christophe Varin for their critical reading of the manuscript and Philippe Faure for the discussions on the archetype analysis. We thank Delphine Houtteman, Souad Laghmari, Perrine Hagué and Laetitia Cuvelier for their technical assistance and the mouse colonies in Brussels, and LIMIF for their assistance with the imaging techniques. ECR was a research fellow of FRS-FNRS-FRIA (Belgium). JN is a researcher of the CNRS, AKE is a Research Director of the FRS-FNRS and a Welbio investigator

**Funding:** This work was supported by grants from FRS-FNRS (#23587797, #33659288, #33659296), Welbio (#30256053), Fondation Simone et Pierre Clerdent (Prize 2018), ARC from FWB, Fondation ULB, AXA Research Fund (Chaire AXA) to AKE, the Agence Nationale de la Recherche (ANR-22-CE16-0013 LEARN) and the Fondation pour la Recherche Médicale (FRM, grant agreement No. EQU202303016311) to JN, and Fond David et Alice Van Buuren and Fond Lekime-Ropsy to ECR.

**Author contributions:** Conceptualization: ECR and AKE; methodology and investigation for the in vivo and ex vivo experiments: ECR and DR; methodology and investigation for the archetypal analysis; and reinforcement-learning model: JN; formal analysis: ECR, JN, DR; validation and supervision: AKE; writing: ECR, JN and AKE.

**Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

**List of supplementary materials:** Figures S1-S8

## FIGURE LEGENDS

**Figure 1. Mice exhibited different but stable preferences in the gambling task**. **A.** Task schema: upon initiating a trial by magazine entry, the mice had to choose. **B.** Preferences for the four options against days of training, showing that the mice learned the task and displayed a stable preference for the safe options. **C.** The preferences of all mice (all conditions) displayed substantial interindividual variability. **D.** The other behavioral measures in the gambling task (number of trials, proportions of premature responses, and omissions) also displayed substantial interindividual variability, suggesting different choice strategies. n.s. not significant; *** $p<0.001$.

**Figure 2. Variability in mouse behavior is a continuum between archetypal strategies**. **A.** Preference for the safe choice was correlated with the percentage of omissions. **B**. Archetypal analysis performed on 8 measures (number of trials, pellets obtained, % omission, % premature responses, and preferences for the 4 options) with n = 168 mice. Each point of the ternary plot represents the projection of an individual onto the two principal components (PCs) derived from the 8 measures. The apices of the polytope encompassing the data define the archetypal strategies (here, a triangle with 3 archetypes: Optimizers (O), Risk-averse (R), Explorers (E)). Individuals can be described as convex combinations of the 3 archetypes. **C.** Variability (in the 8 measures × 168 mice) explained as a function of the number of archetypes, showing that 3 archetypes constitute a parsimonious description of the data. **D.** The distance to archetypes (R apex) was highly correlated between consecutive days, suggesting that behavior in the task arose from stable cognitive profiles. **E**. Preferences from mice assigned to their closest archetype indicate that P1 (certain, small reward, short TO) was preferred by Risk-averse (R) mice, and that Optimizers (O) preferred P2 (larger, less certain reward, more advantageous relative to TO than P1), whereas Explorers (E) did not exhibit a preference. **F**. These choice strategies were associated with differences in other task measures: Risk-averse mice displayed fewer omissions, Optimizers earned more pellets and made fewer premature responses, and Explorers made more omissions and earned fewer pellets. *** $p<0.001$.

**Figure 3. Archetypal behaviors arise from different decision-making strategies. A**. Sketch of the computational model of decisions ("choice model"): Mice were assumed to assign a value to each of the four options (P1-P4) depending ons how much they positively valued a pellet ($\rho$ parameter), how much they considered the time-outs to be punishing ($T$ parameter), their attitude toward risk (i.e., toward outcome probabilities closer to 0.5, $\varphi$ parameter) and the precision in their decision process ("inverse temperature" $\beta$ parameter). **B**. Bayesian Information Criterion (top, model likelihood penalized for model complexity) and proportion of animals best explained (middle) for different versions of the Reinforcement-Learning model (bottom), with the 4-parameters model showing smaller (i.e. better) BIC. **C**. Average preferences for the 4 options (P1-P4) for the different archetypes (O for Optimizers, R for Risk-averse, E for explorers) can be generated through different combinations of decision-making parameters ($\beta$, $\rho$, $T$, and $\varphi$). Opaque bars display the model fit and shaded bars display the experimental data. **D.** The choice model suggests that the strategies unraveled by the archetypal analysis arise from different sensitivities to task variables: mice were in the Risk-averse archetype if their sensitivity to punishment was higher and their attitude toward risk was more negative; Explorers mainly displayed lower $\beta$ values (low precision, or high noise, in the decision process); and Optimizers exhibited higher sensitivity to reward and their attitude toward risk was neutral. *** $p<0.001$.

1121 **Figure 4. DMS d-SPNs promote risk-seeking**. **A.** Microscope fluorescence for control (top)
1122 and hM3Dq (bottom) individuals in D1-Cre mice injected in the dorsomedial striatum (DMS):
1123 DAPI (blue) indicates the nuclei, mCherry (red) indicates the transfected neurons and Fos
1124 (green) indicates the CNO-activated neurons. The arrowheads indicate hM3Dq-expressing
1125 neurons. **B**. Fos expression was significantly greater in the hM3Dq group than in the control
1126 group. Scale bar= 20 µm. The graph represents the mean±SEM. **C**. Electrophysiological
1127 recordings after CNO treatment during the IGT. From left to right: the rheobase did not differ
1128 between control and hM3Dq-expressing mice after CNO treatment, but hM3Dq-expressing
1129 neurons had increased intrinsic excitability compared with controls. **D.** Differences in all
1130 preferences for the four options between CNO and baseline for mCherry controls and hM3Dq
1131 mice. **E**. CNO activation of DMS D1-expressing neurons transfected with hM3Dq decreased
1132 the preference for the safe options (P1+P2). **F**. CNO treatment did not affect performance
1133 (number of pellets divided by the number of trials). **G.** CNO treatment decreased the number
1134 of trials and increased the percentage of omissions. **H**. Right: ternary plot showing the global
1135 effect of CNO on strategies, with each point (black, mCherry controls; pink, hM3Dq animals)
1136 representing an animal, with a connected line showing displacement in the ACP space from
1137 baseline to CNO. The larger arrows show the group averages (black, mCherry controls; pink,
1138 hM3Dq animals), with hM3Dq animals moving away from the R apex under CNO. Left: data
1139 split by archetypes indicate that the decrease in the safe choice was due mainly to an effect on
1140 Optimizers. **I**. Data fit by the choice model suggest that the decreased choice of the safe
1141 option reflects an increase in risk-seeking upon CNO treatment. n.s. not significant; * $p<0.05$.
1142 ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$.
1143
1144 **Figure 5. DMS iSPNs favor large gains**. **A.** Microscope fluorescence for control (top) and
1145 hM3Dq (bottom) individuals in A2A-Cre mice injected in the dorsomedial striatum (DMS):
1146 DAPI (blue) indicates the nuclei, mCherry (red) indicates the transfected neurons and Fos
1147 (green) indicates the CNO-activated neurons. The white arrowheads indicate hM3Dq-
1148 expressing neuronal somas. **B**. Fos expression was significantly greater in the hM3Dq group
1149 than in the control group. Scale bar= 20 µm. The graph represents the mean±SEM. **C**. From
1150 left to right: the resting membrane potential, rheobase and AMPA/NMDA ratio did not differ
1151 between control and hM3Dq-expressing mice after CNO treatment, but hM3Dq-expressing
1152 neurons had increased intrinsic excitability compared with controls. **C.** Differences in all
1153 preferences for the four options between CNO and baseline for mCherry controls and hM3Dq
1154 mice. **E**. CNO treatment decreased performance in hM3Dq animals. **F.** CNO treatment also
1155 decreased the percentage of trials and increased the percentage of omissions. **G**. Left: ternary
1156 plot showing the global effect of CNO on strategies, with each point (black, mCherry
1157 controls; pink, hM3Dq animals) representing an animal receiving CNO treatment, with a
1158 connected line showing displacement in the ACP space from baseline to CNO. The larger
1159 arrows show the group averages (black, mCherry controls; pink, hM3Dq animals) with
1160 hM3Dq moving toward the E apex under CNO. Right: the data split by archetypes indicate
1161 the opposite effect of DMS A2A-expressing neurons on the safe choice of Optimizers
1162 compared to Risk-averse and Explorers. **H**. Data fit by the choice model suggest that the
1163 decreased choice of the safe option reflects a decrease in reward saturation upon CNO
1164 treatment. n.s. not significant; * $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$.
1165
1166 **Figure 6. The NAc d- and i-SPNs affect mouse gambling, but less than their DMS**
1167 **counterparts. A.** Microscope fluorescence for control (top) and hM3Dq (bottom) individuals
1168 in D1-Cre mice injected in the nucleus accumbens (NAc): DAPI (blue) indicates the nuclei,
1169 mCherry (red) indicates the transfected neurons and Fos (green) indicates the CNO-activated
1170 neurons. The white arrowheads indicate hM3Dq-expressing neuronal somas. **B**. Fos

1171 expression was significantly greater in the hM3Dq group than in the control group. Scale bar=
1172 20 µm. The graph represents the mean±SEM. **C**. From left to right: CNO treatment decreased
1173 the number of trials initiated, increased the percentage of omissions, and decreased the reward
1174 saturation parameter from the model fit. **D**. Ternary plot showing the global effect of CNO on
1175 strategies, with each point (black, mCherry controls; pink, hM3Dq animals) representing an
1176 animal, with a connected line showing displacement in the ACP space from baseline to CNO.
1177 The larger arrows show the group averages (black, mCherry controls; pink, hM3Dq animals)
1178 with hM3Dq moving toward the E apex under CNO. **E,F.** Same as A but for A2A-Cre mice
1179 injected in the NAc. **G**. CNO treatment decreased only the number of trials, but not the
1180 percentage of premature responses or the percentage of omissions. **H**. Ternary plot (same as D
1181 but for A2A-Cre mice injected in the NAc) with hM3Dq moving away from the O apex under
1182 CNO.
1183
1184 **Figure 7. Dorsolateral i-SPN and d-SPN have no specific effects on mouse gambling task.**
1185 **A.** Microscope fluorescence for control (top) and hM3Dq (bottom) individuals in D1-Cre
1186 mice injected in the dorsolateral striatum (DLS): DAPI (blue) indicates the nuclei, mCherry
1187 (red) indicates the transfected neurons and Fos (green) indicates the CNO-activated neurons.
1188 The white arrowheads indicate hM3Dq-expressing neuronal somas. **B**. Fos expression was
1189 significantly greater in the hM3Dq group than in the control group. Scale bar= 20 µm. The
1190 graph represents the mean±SEM. **C**. CNO treatment did not affect the proportion of
1191 premature response (left), but increased the percentage of omissions (right). **D**. There was no
1192 difference in any preference for the four options between the CNO groups and the baseline
1193 groups for mCherry controls and hM3Dq DLS-injected D1-Cre mice. **E, F**. Same as A,B for
1194 A2A-DLS mice. **G**. CNO treatment decreased the proportion of premature response (left), and
1195 increased the percentage of omissions (right). **H**. There was no difference in any preference
1196 for the four options between the CNO groups and the baseline groups for mCherry controls
1197 and hM3Dq DLS-injected A2A-Cre mice. n.s. not significant; * $p<0.05$; ** $p<0.01$; ***
1198 $p<0.001$; **** $p<0.0001$.
1199

1200 **FIGURES**

1201

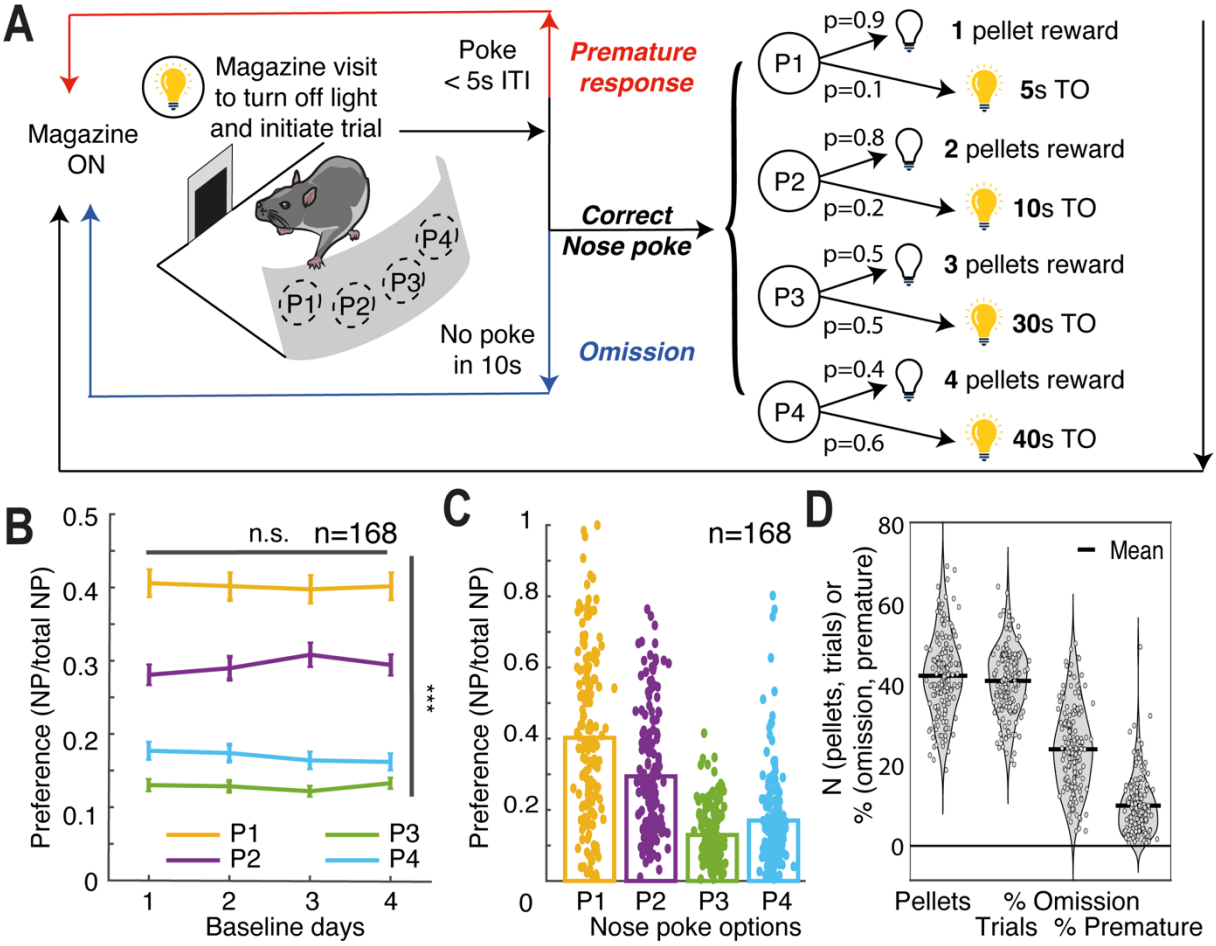



Figure 1

1202

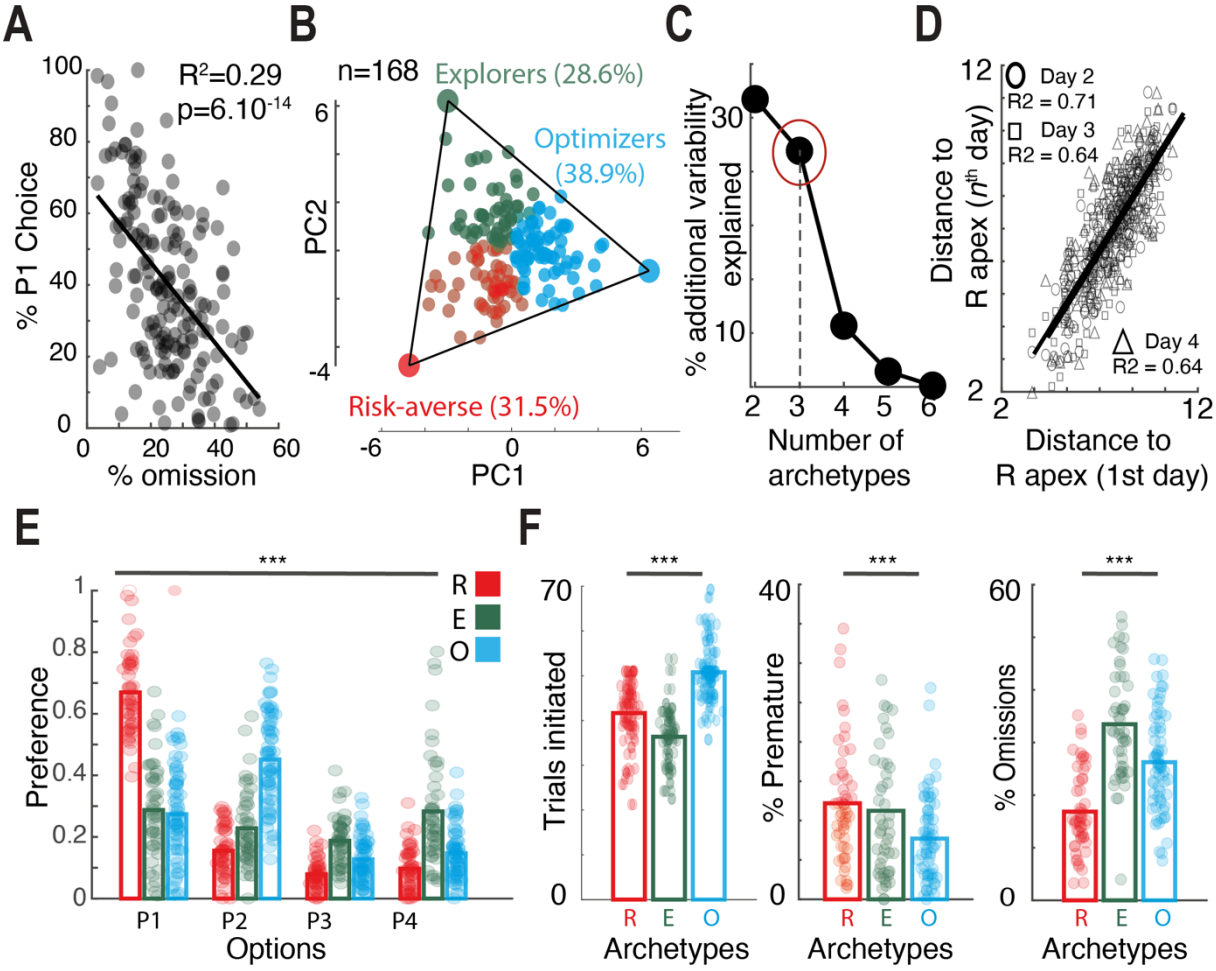



Figure 2

1203

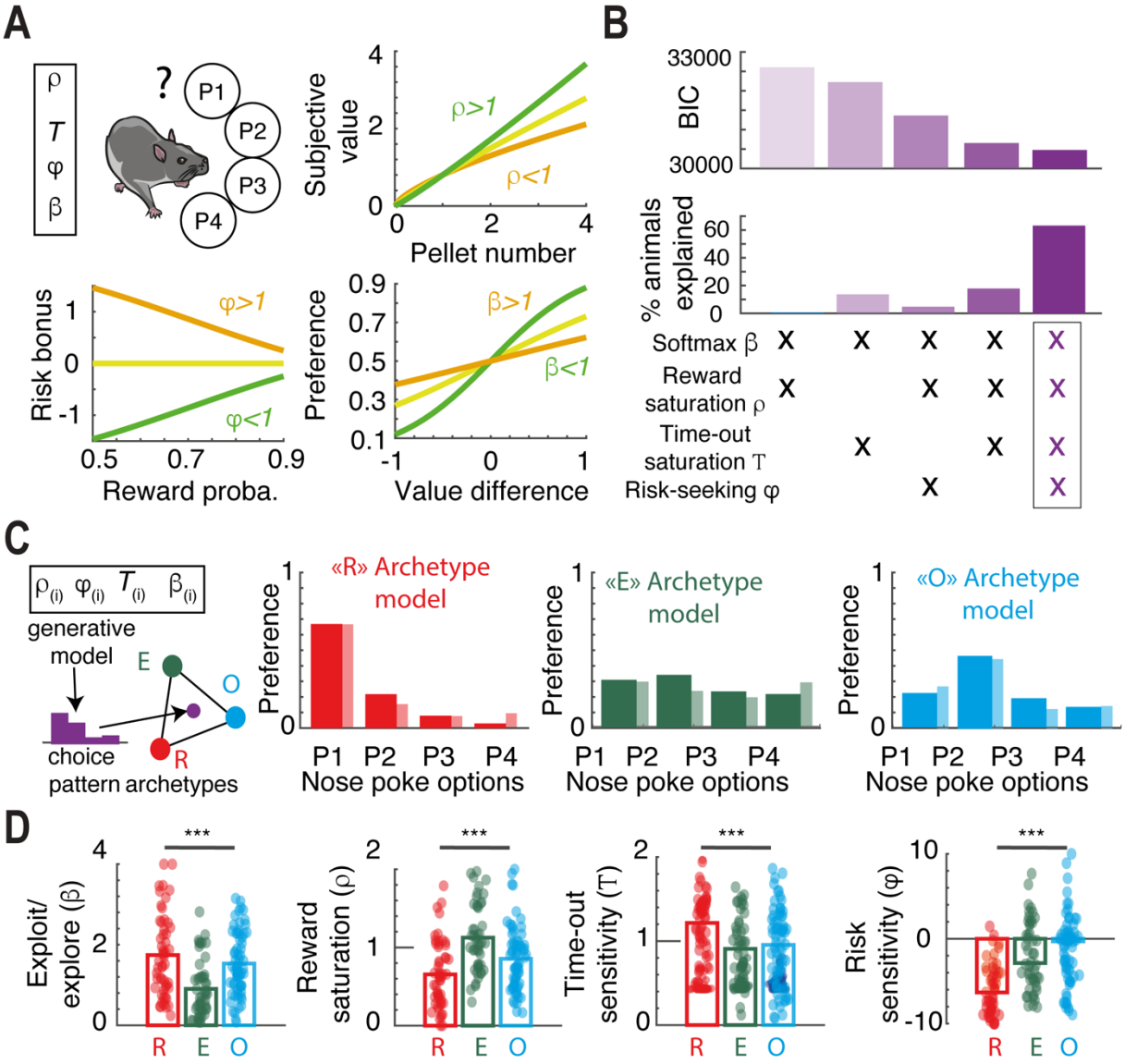

A
ρ
T
φ
β
?
P1
P2
P3
P4
Subjective value
ρ>1
ρ<1
Pellet number
Risk bonus
φ>1
φ<1
Reward proba.
Preference
β>1
β<1
Value difference
B
BIC
33000
30000
% animals explained
Softmax β
Reward saturation ρ
Time-out saturation T
Risk-seeking φ
C
generative model
choice pattern archetypes
«R» Archetype model
«E» Archetype model
«O» Archetype model
Preference
Nose poke options
D
Exploit/ explore (β)
Reward saturation (ρ)
Time-out sensitivity (T)
Risk sensitivity (φ)
***
R E O

Figure 3

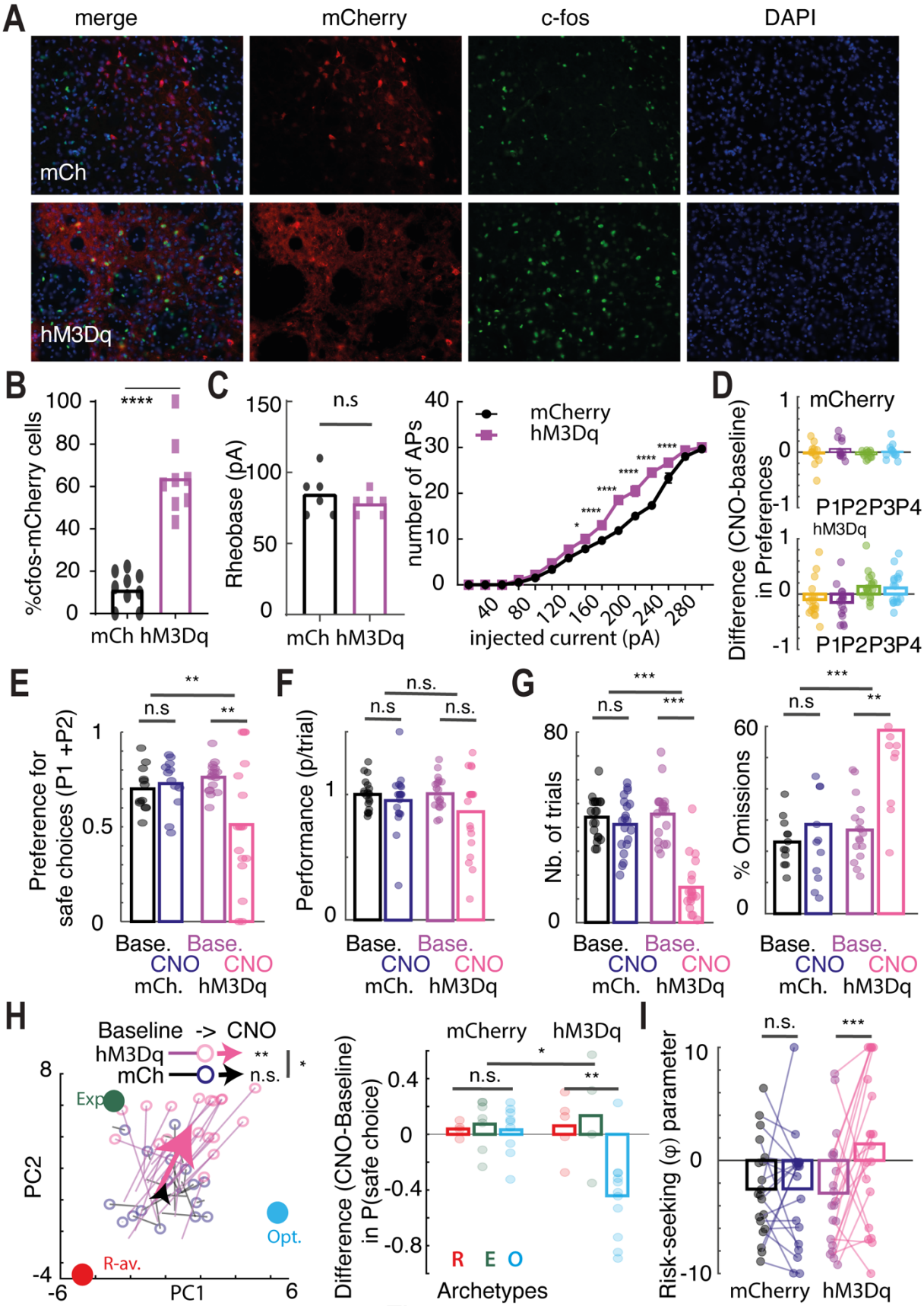



Figure 4

1206

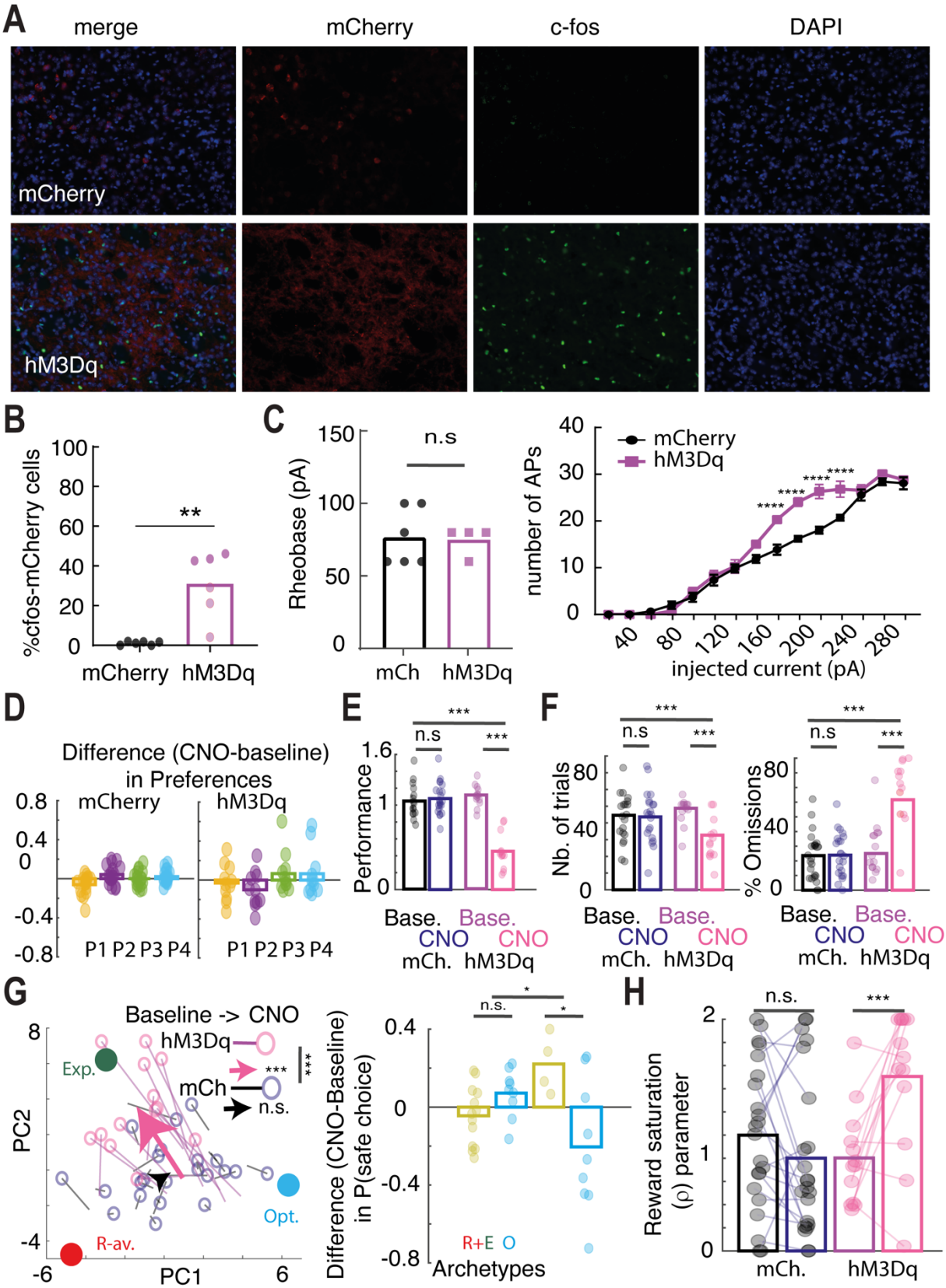

A
merge
mCherry
c-fos
DAPI
mCherry
hM3Dq
B
%cfos-mCherry cells
100
80
60
40
20
0
**
mCherry
hM3Dq
C
Rheobase (pA)
150
100
50
0
n.s
mCh
hM3Dq
number of APs
40
30
20
10
0
mCherry
hM3Dq
****
40
80
120
160
200
240
280
injected current (pA)
D
Difference (CNO-baseline)
in Preferences
mCherry
hM3Dq
0.8
0.4
0
-0.4
-0.8
P1 P2 P3 P4
P1 P2 P3 P4
E
Performance
1.6
1
0
***
n.s
***
Base.
CNO
Base.
CNO
mCh.
hM3Dq
F
Nb. of trials
80
40
0
% Omissions
80
40
0
***
n.s
***
Base.
CNO
Base.
CNO
mCh.
hM3Dq
G
Baseline -> CNO
hM3Dq
mCh
***
n.s.
Exp.
Opt.
R-av.
PC2
PC1
8
-4
-6
6
Difference (CNO-Baseline)
in P(safe choice)
0.4
0
-0.4
-0.8
n.s.
*
*
R+E
O
Archetypes
H
Reward saturation
(ρ) parameter
2
1
0
n.s.
***
mCh.
hM3Dq

Figure 5

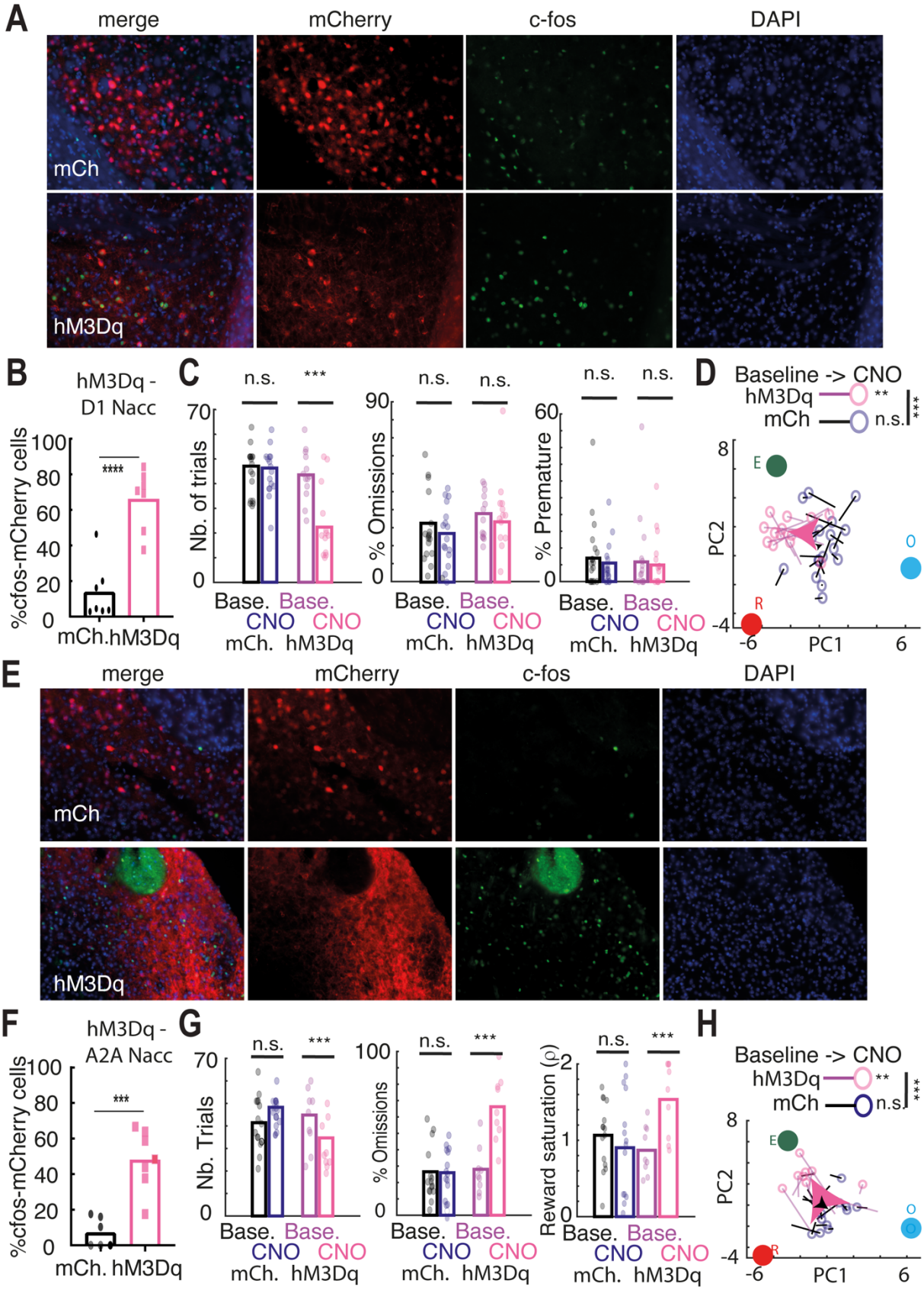

A
merge
mCherry
c-fos
DAPI
mCh
hM3Dq
B
hM3Dq -
D1 Nacc
%cfos-mCherry cells
mCh.hM3Dq
C
Nb. of trials
% Omissions
% Premature
n.s.
***
Base.
CNO
mCh.
hM3Dq
D
Baseline -> CNO
hM3Dq
mCh
PC1
PC2
E
merge
mCherry
c-fos
DAPI
mCh
hM3Dq
F
hM3Dq -
A2A Nacc
%cfos-mCherry cells
mCh. hM3Dq
G
Nb. Trials
% Omissions
Reward saturation (ρ)
H
Baseline -> CNO
hM3Dq
mCh
PC1
PC2

Figure 6

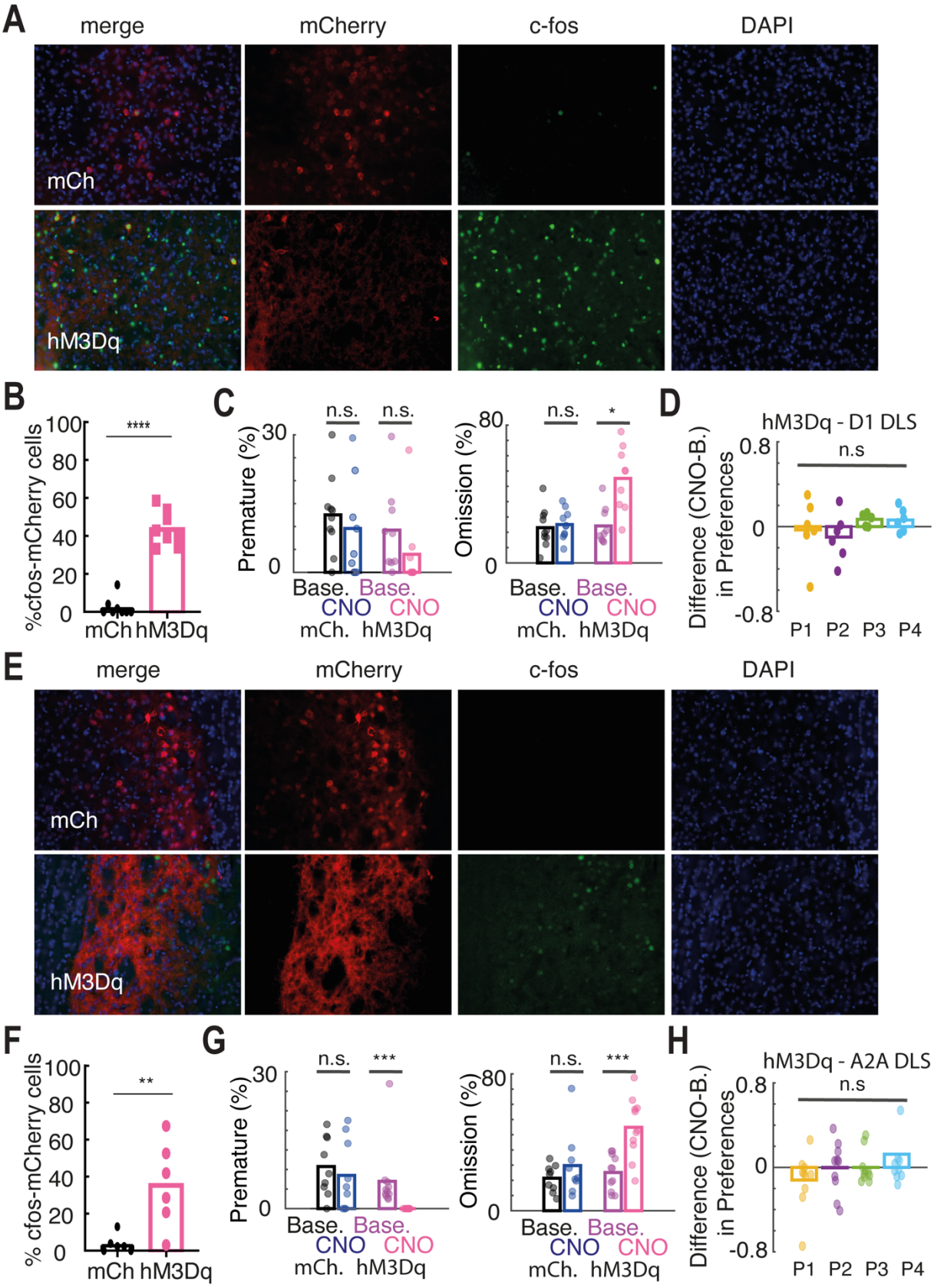



Figure 7

1215
1216

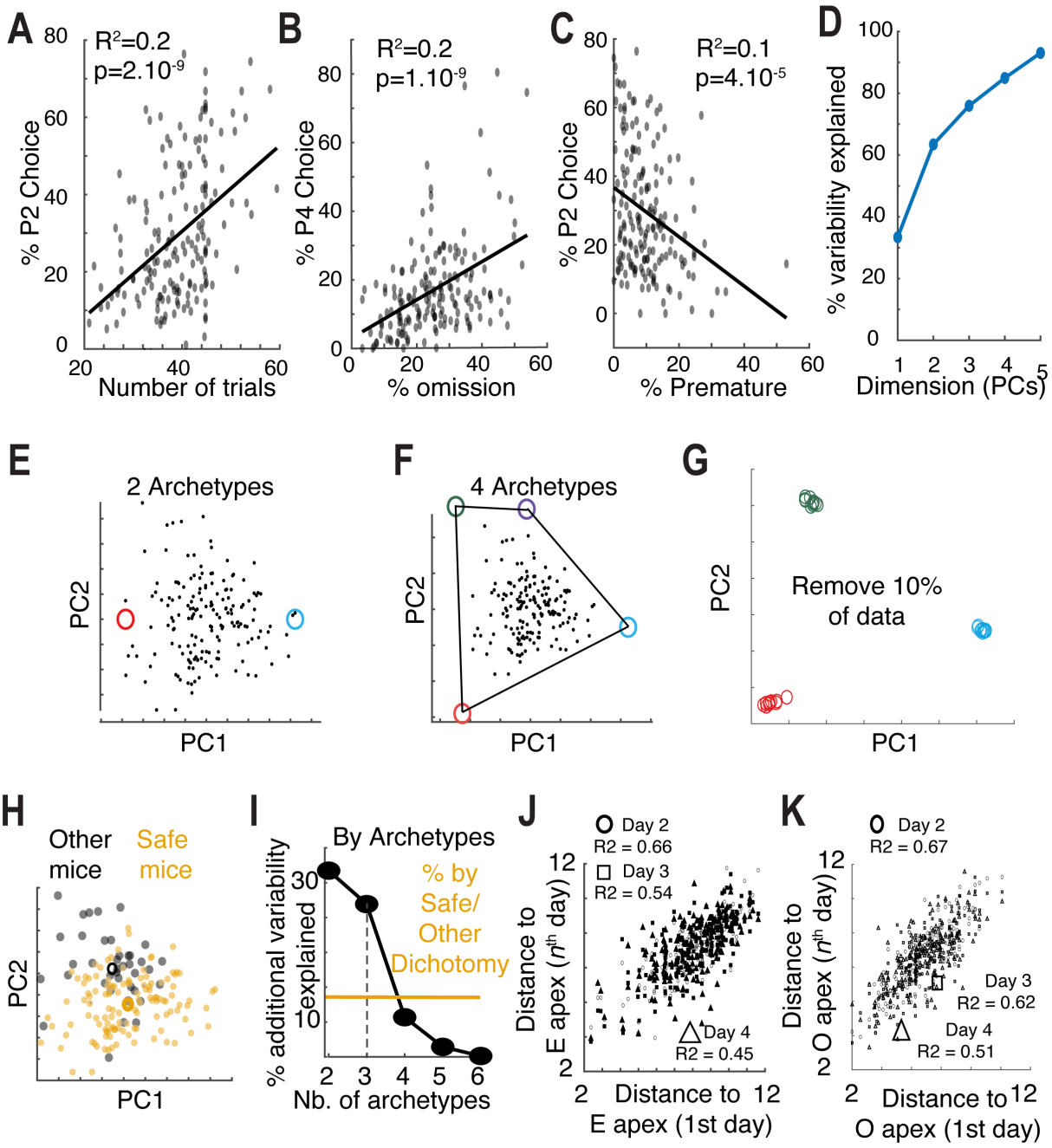
**Direct and indirect striatal projecting neurons exert strategy-dependent effects on decision-making**

Elena Chaves Rodriguez*, Jérémie Naudé*, Daniel Rial and Alban de Kerchove d'Exaerde

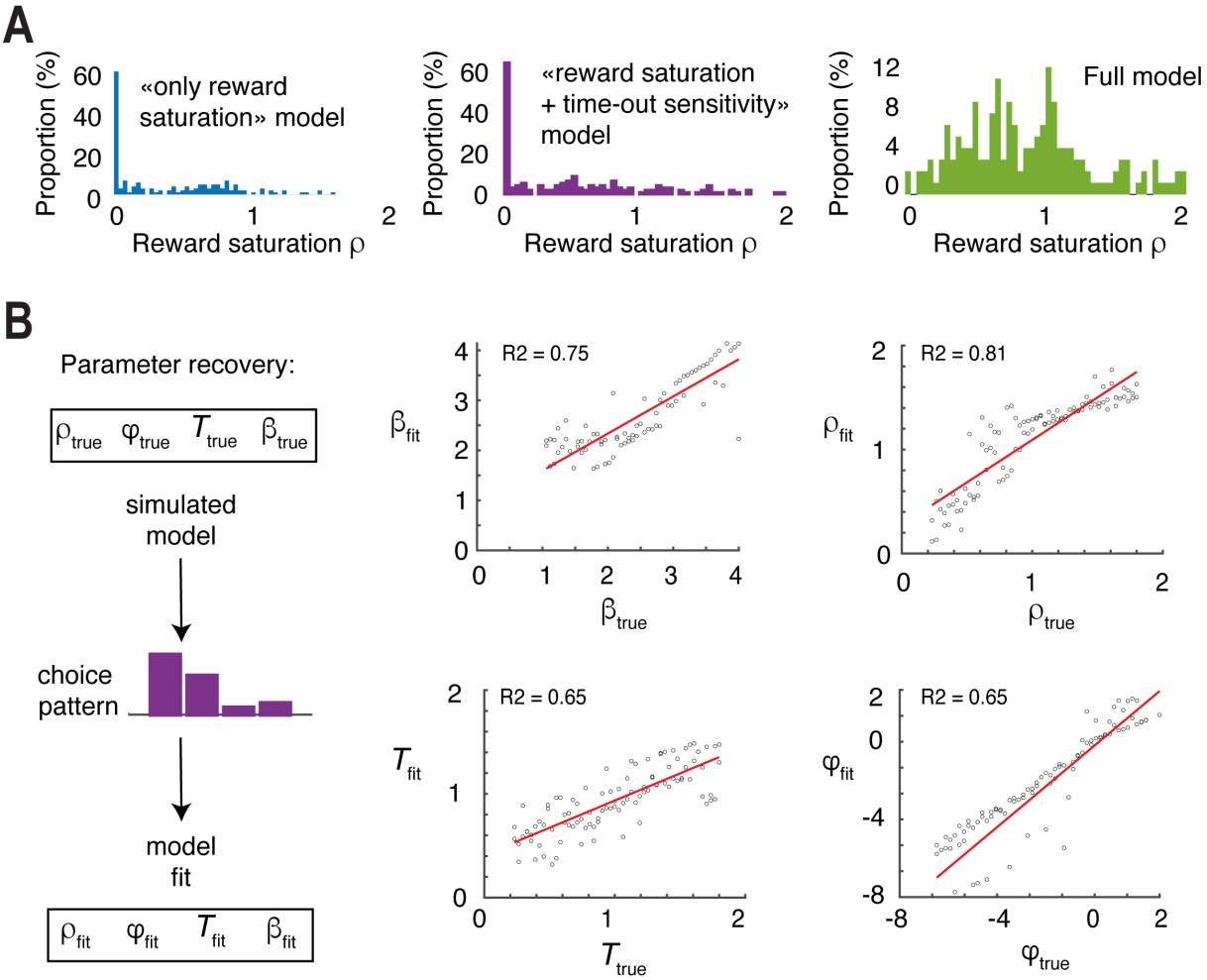
This PDF files includes:

Supplementary Figures 1 to 8

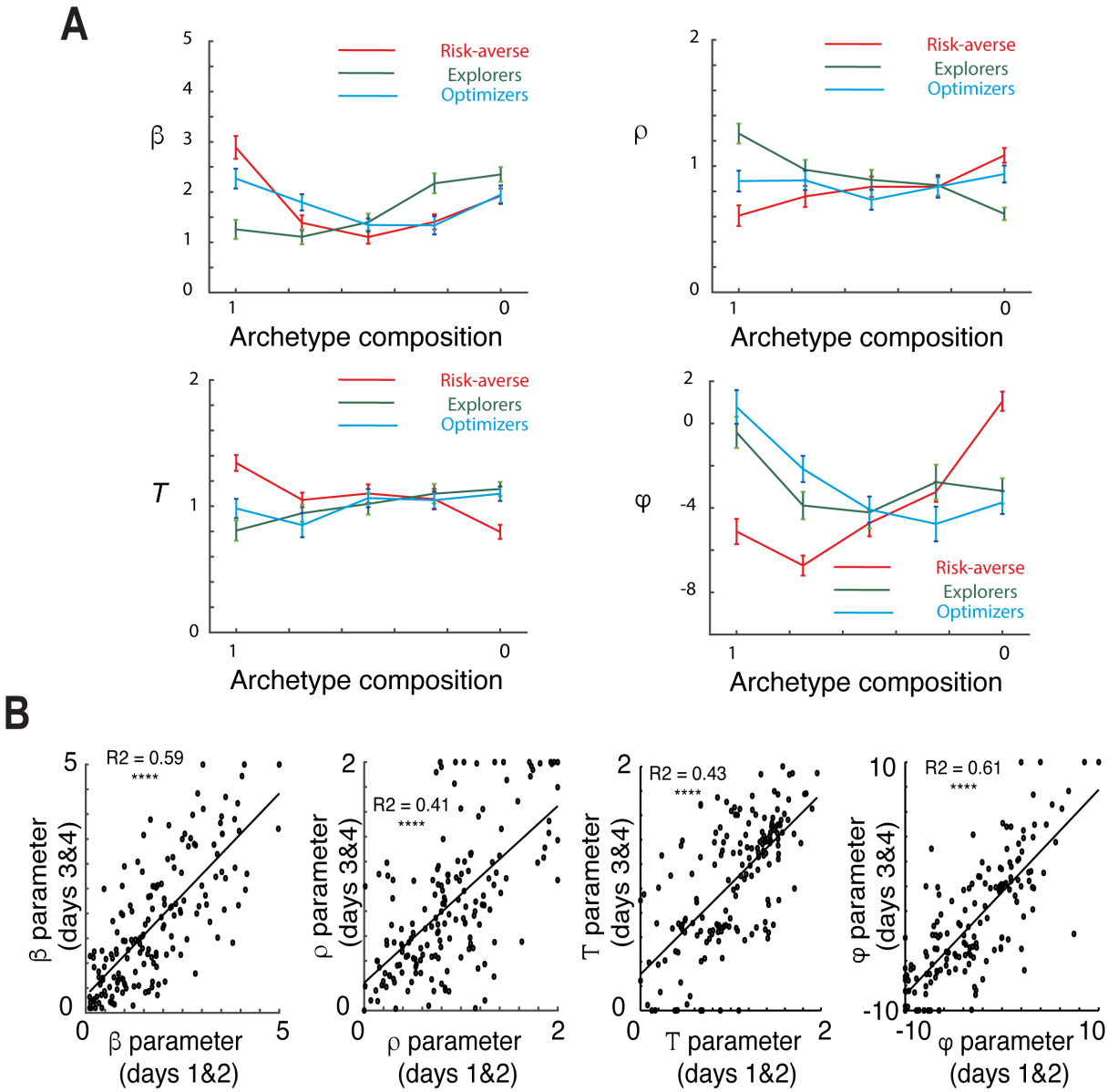
Supplementary Table 1



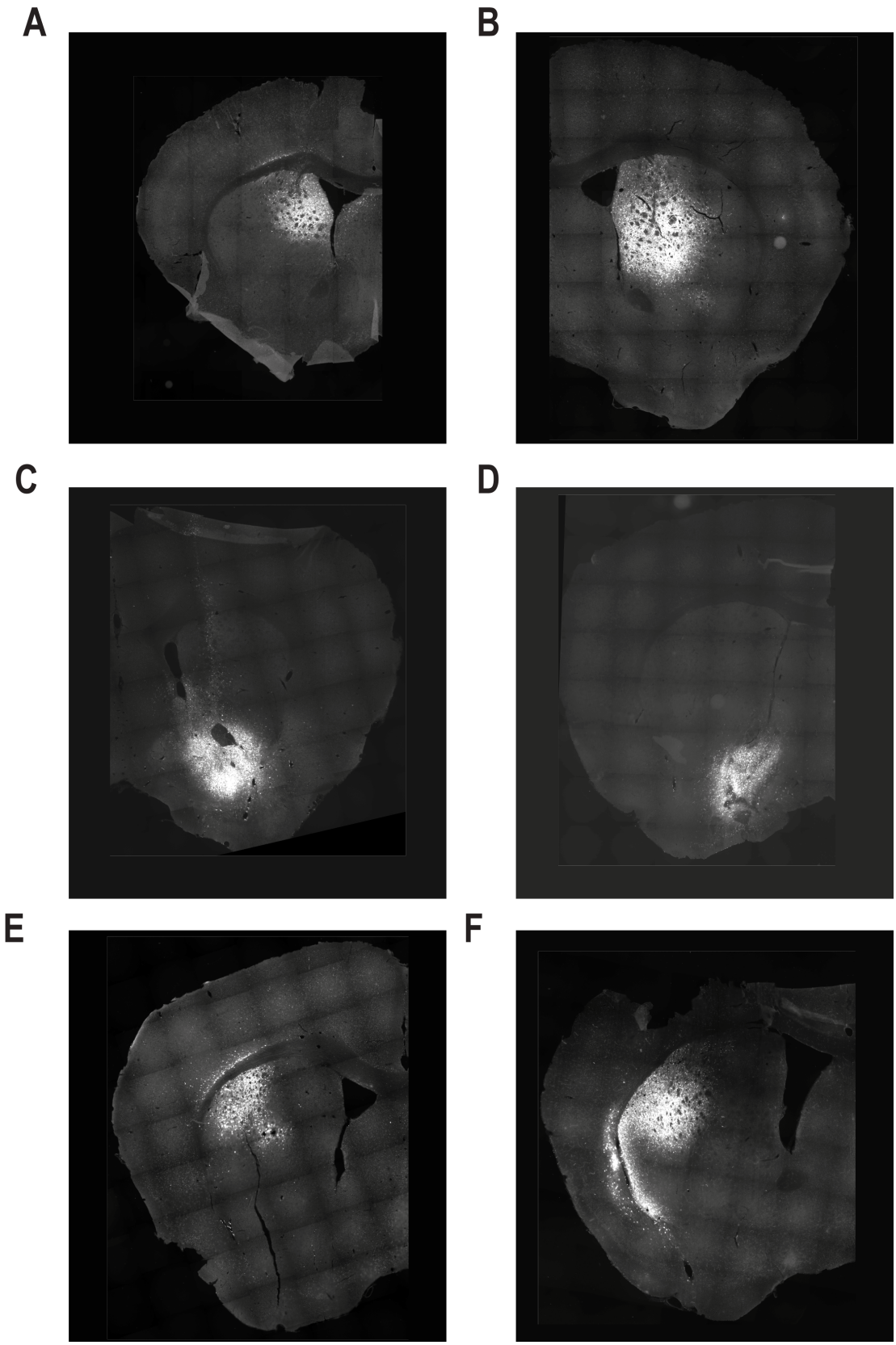

**Supplementary Figure 1. Other examples of correlations between behavioral measures**. **A**. Preference for the P2 choice was positively correlated with the number of trials ($R^2$ = 0.2, p=2.10$^{-9}$). **B**. Preference for the P4 choice was positively correlated with % of omissions ($R^2$ = 0.2, p=1.10$^{-9}$). **C**. Preference for the P2 choice was negatively correlated with % of premature responses ($R^2$ = 0.2, p=4.10$^{-5}$). (**D**) % of variance explained by the principal components. (**E,F**) Apices corresponding to 2 archetypes (**E**) and 4 archetypes (**F**). **G**. Apices found when 10% of the data was randomly removed. (**H**) Projection in the PC (1$^{st}$ and 2$^{nd}$) space of mice according to the safe/not safe classical dichotomy. (**I**) Comparison of the % of explained variance by archetypes (black, same as Fig. 2C) and by safe/not safe dichotomy (yellow). (**J,K**) Correlation over days between distances (same as Fig. 2D) to the E (**J**) and O (**K**) apices.



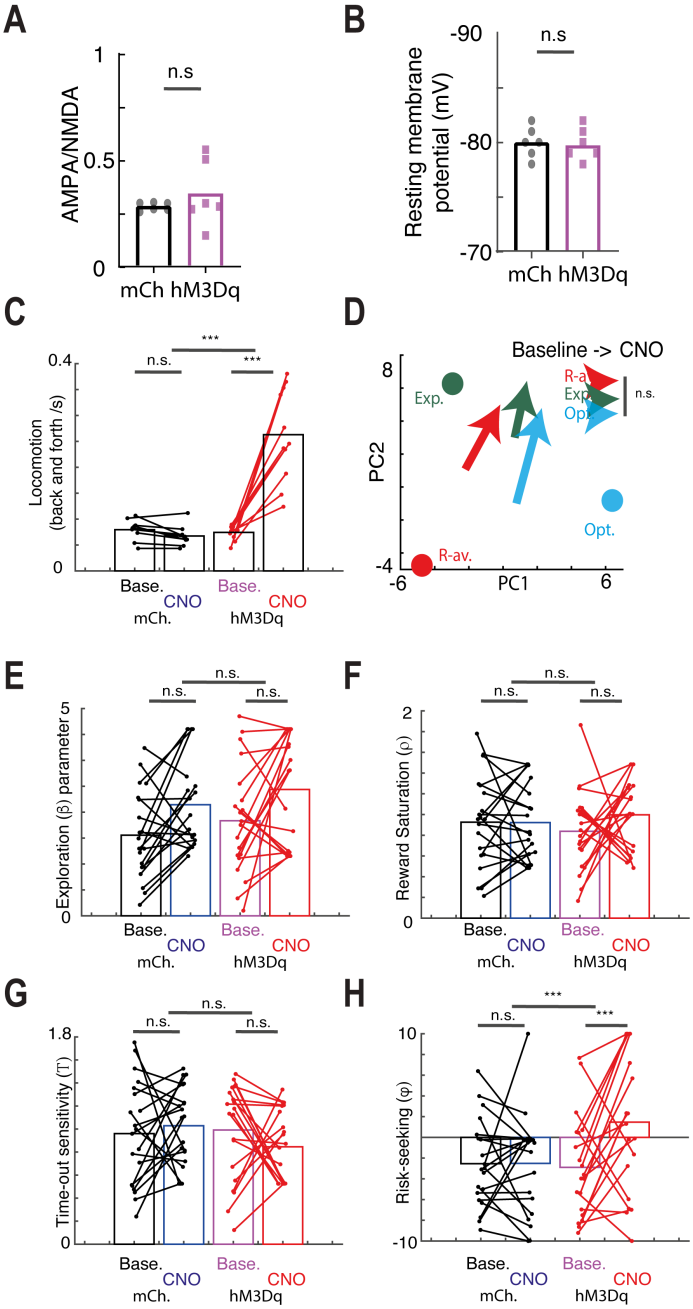

**Supplementary Figure 2**. **Model comparisons**. **A**. Distribution of fitted values for reward saturation parameter in simple model versus the full model; showing that simpler versions of the model fit the data with extreme, implausible values for the parameters (i.e. reward saturation is expected to lie to close to 1 or slightly below 1). **B**. Parameter recovery: choices were generated with multiple realization of a model with controlled parameter ("true" parameters) and these choice patterns were fitted with the model ("fit" parameters).



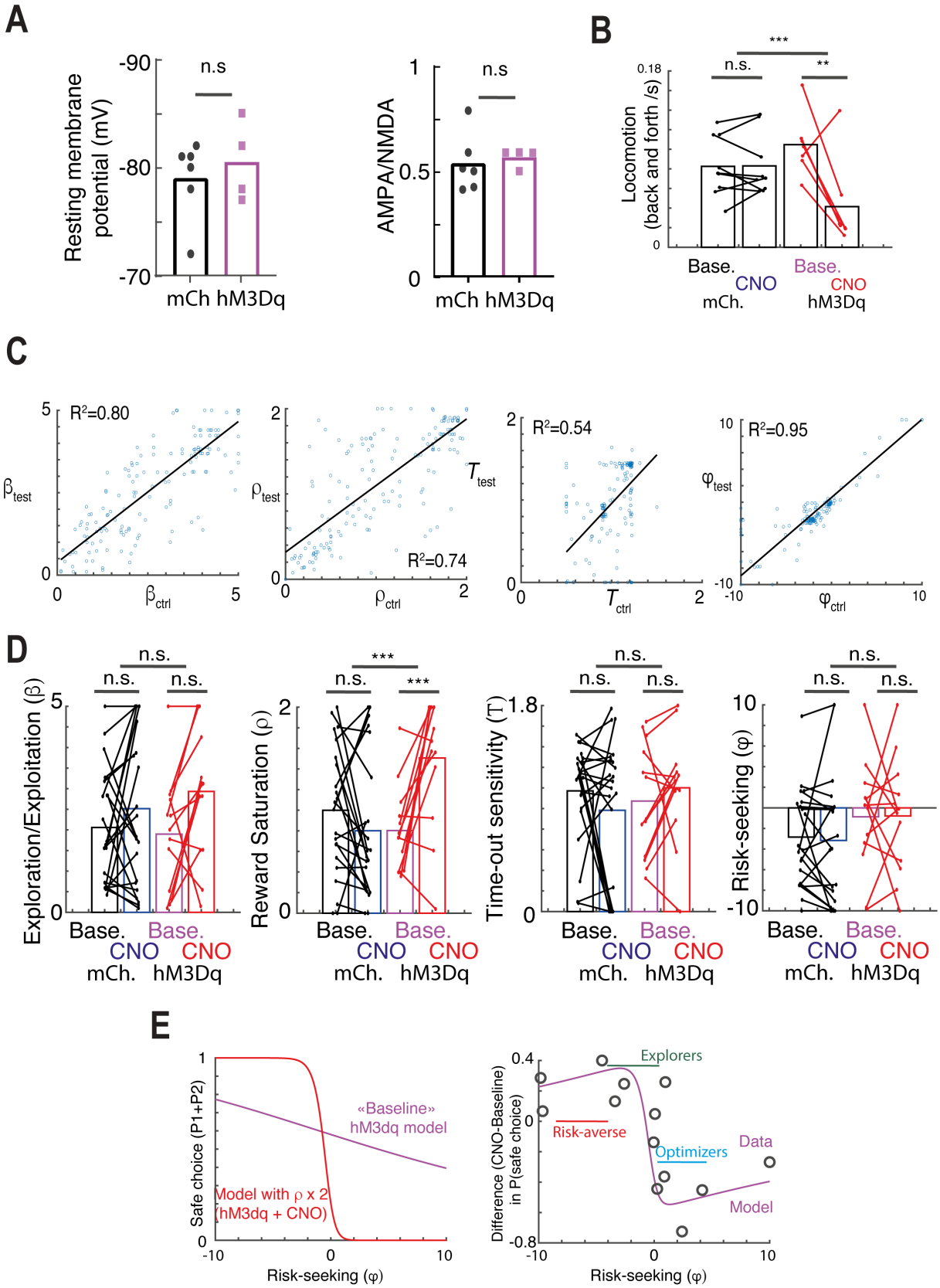

**Supplementary Figure 3**. **Additional model results**. **A**. Model parameters as a function of archetypal compositions. **B**. Parameter stability was determined by comparing the parameters fitted over the first 2 days of baseline compared to parameters fitted over the last two days.



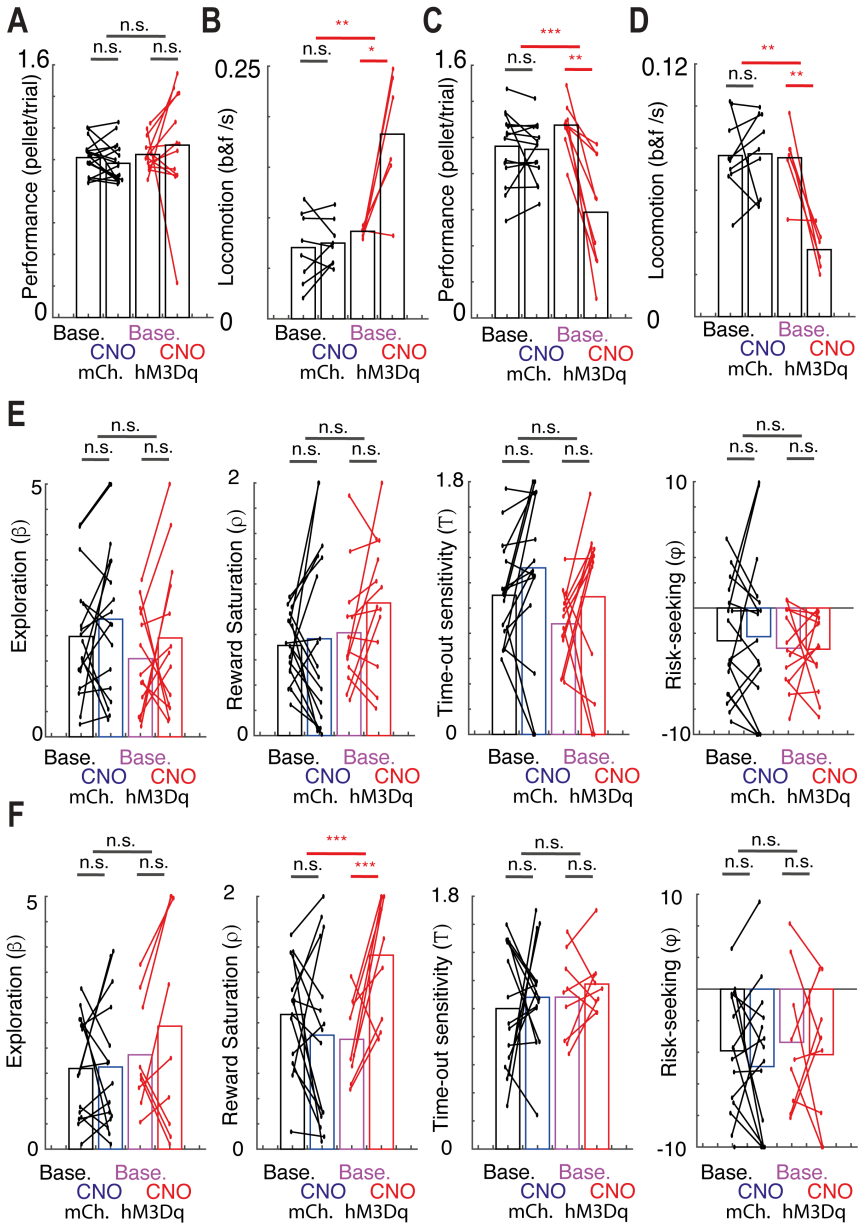

**Supplementary Figure 4. Initial targeting of three striatal regions** (A-B DMS, C-D NAc, and E-F DLS) in two mouse lines: Drd1-cre (A-C-E) and Adora2a-Cre (B-D-F), using the control viruses AAV5-hSyn1-DIO-mCherry.



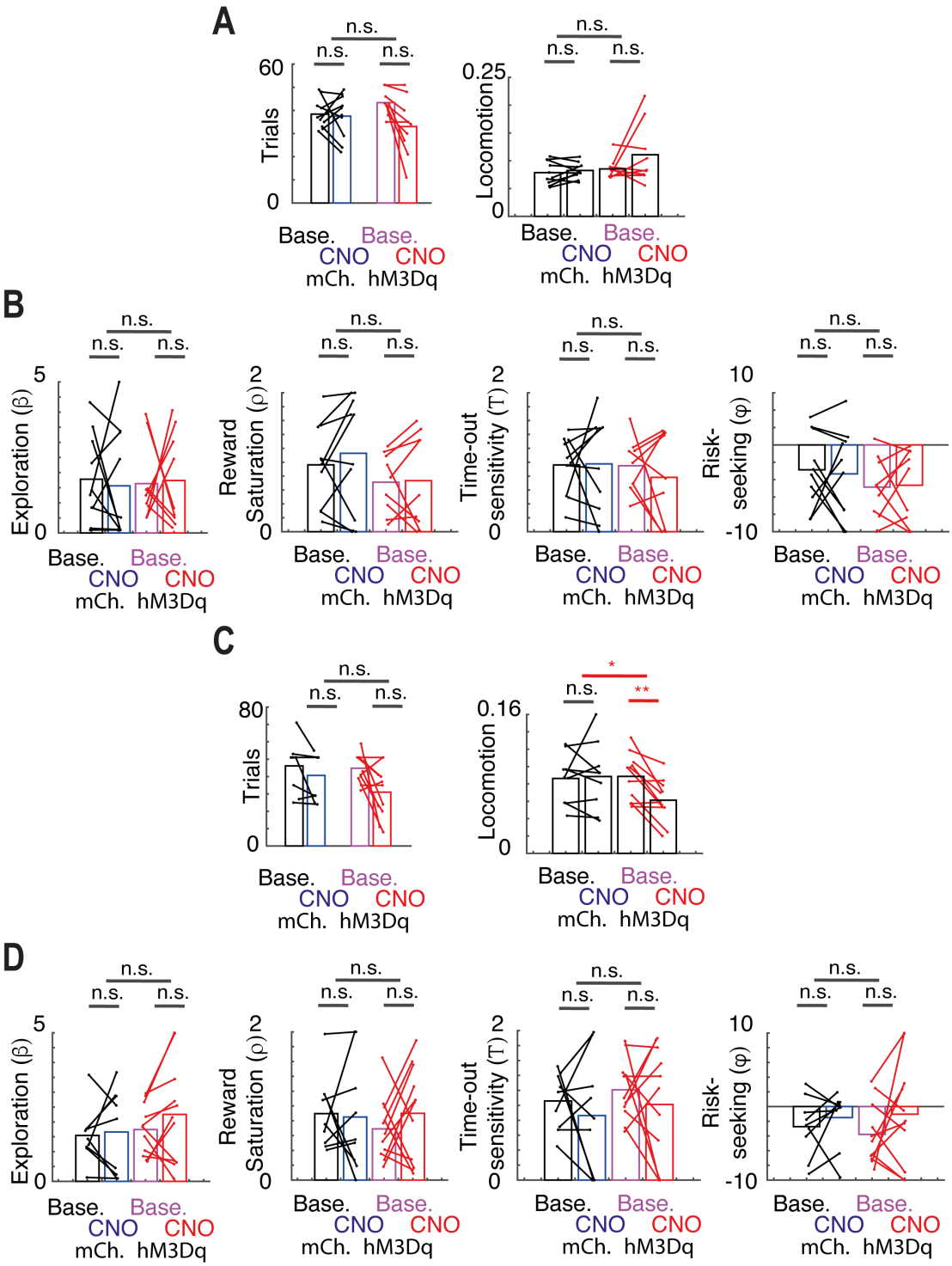

**Supplementary Figure 5. Additional measures and model for DMS-dSPN experiments.** **A**. The AMPA/NMDA ratio was not different between conditions. **B**. The resting membrane potential was not different between conditions. **C**. Locomotion (back and forth movements in the conditioning box) for (baseline, CNO) x (mCherry, hM3Dq) conditions show a specific increase in locomotion following CNO on hM3Dq animals. **D**. Ternary plot showing the global effect of CNO on strategies split by archetypes, with each arrow showing group averages for each archetype from h3MDq animals under CNO. **E-H**. Model parameters (explore/exploit; reward saturation; time-out sensitivity; risk-seeking) for DREADD (hM3Dq) animals and mCherry (mCh) controls, under baseline and CNO conditions. Only the risk-seeking was significantly different in the (hM3Dq, CNO) condition.


A
Resting membrane potential (mV)
-90
-80
-70
n.s
mCh hM3Dq
AMPA/NMDA
1
0.5
0
n.s
mCh hM3Dq
B
Locomotion (back and forth /s)
0.18
0
n.s.
***
**
Base.
CNO
mCh.
Base.
CNO
hM3Dq
C
R²=0.80
R²=0.74
R²=0.54
R²=0.95
β test
β ctrl
ρ test
ρ ctrl
T test
T ctrl
φ test
φ ctrl
D
Exploration/Exploitation (β)
Reward Saturation (ρ)
Time-out sensitivity (T)
Risk-seeking (φ)
n.s.
***
Base.
CNO
mCh.
Base.
CNO
hM3Dq
E
Safe choice (P1+P2)
«Baseline»
hM3dq model
Model with ρ x 2
(hM3dq + CNO)
Risk-seeking (φ)
Difference (CNO-Baseline) in P(safe choice)
Explorers
Risk-averse
Optimizers
Data
Model
Risk-seeking (φ)

**Supplementary Figure 6**. **Additional measures and model for DMS-iSPN experiments**. **A**. The the resting membrane potential (left) and the AMPA/NMDA ratio were not different between conditions. **B**. Locomotion (back and forth movements in the conditioning box) for (baseline, CNO) x (mCherry, hM3Dq) conditions show a specific decrease in locomotion following CNO on hM3Dq animals. **C**. Simulations showing the effects of a decrease in the number of trials on parameter recovery. Each dot depicts the parameter recovered from a simulation under the control number of trials (n=50) against the parameter recovered from a simulation under the h3Mdq number of trials. **D.** Model parameters (explore/exploit; reward saturation; time-out sensitivity; risk-seeking) for DREADD (hM3Dq) animals and mCherry (mCh) controls, under baseline and CNO conditions. Only the reward saturation was significantly different in the (hM3Dq, CNO) condition. **E**. Simulations of safe choices (P1+P2, as measured in Fig. 5) at baseline (purple) and following a decrease in reward saturation (increase in rho, red) depending on the risk seeking parameter (left); and data of safe choices under CNO for animals in the DMS-iSPN experiment against their risk-sensitivity parameter at baseline, showing a match with the model predictions (purple line, left).




**Supplementary Figure 7 Additional measures and model for NAc experiments**. **A.** CNO treatment decreased the performance in D1-NAc animals. **B.** Locomotion (back and forth movements in the conditioning box) for (baseline, CNO) x (mCherry, hM3Dq) conditions show a specific decrease in locomotion following CNO on A2A-NAc-hM3Dq animals. **C.** CNO treatment did not affect the performance in A2A-NAc animals **D.** Locomotion (back and forth movements in the conditioning box) for (baseline, CNO) x (mCherry, hM3Dq) conditions show a specific increase in locomotion following CNO on D1-NAc-hM3Dq animals. **E**. Model parameters (explore/exploit; reward saturation; time-out sensitivity; risk-seeking) for D1-NAc - DREADD (hM3Dq) animals and mCherry (mCh) controls, under baseline and CNO conditions. Only the risk-seeking was significantly different in the (hM3Dq, CNO) condition. **F**. Model parameters (explore/exploit; reward saturation; time-out sensitivity; risk-seeking) for A2A-NAc - DREADD (hM3Dq) animals and mCherry (mCh) controls, under baseline and CNO conditions, with no significant effect of CNO.




**Supplementary Figure 8**. **Model fits and behavioral measures for DLS animals**. **A** model parameters (from left to right : explore/exploit ; reward saturation; time-out sensitivity; risk-seeking) and **B** behavioral measures (from left to right : number of trials, % of omissions, % of premature responses, and locomotion) for A2A-DLS (hM3Dq) animals and mCherry (mCh) controls, under baseline and CNO conditions. CNO only affected % of omissions ($F_{(1,18)}$=5.32, p = 0.03; $T_{(10)}$ = -5.63, p = $2.10^{-4}$), % of premature responses ($F_{(1,18)}$=6.89, p = 0.02; $T_{(10)}$ = 2.76, p = 0.02) and locomotion ($F_{(1,18)}$=6.63, p = 0.02; $T_{(10)}$ = -4.21, p = 0.002) in A2A-DLS-hM3Dq animals. (**C,D**) same for D1-DLS animals. CNO only affected % of omissions ($F_{(1,17)}$=12.73, p = 0.02; $T_{(8)}$ = -3.94, p = 0.04).

| | A2A | | D1 | |
|---|---|---|---|---|
| | hm3Dq | mCherry controls | hm3Dq | mCherry controls |
| DMS | 31.1% (SD=16.4, n=6) | 1.1% (SD=0.9, n=6) | 64.4% (SD=17.8, n=9) | 13.3% (SD=7.2, n=9) |
| DLS | 35.7% (SD= 23.1, n=6) | 3.6% (SD=4.8, n=6) | 43.9% (SD=9.2, n=8) | 2.5% (SD=5, n=8) |
| NAc | 48.2% (SD= 17.2, n=7) | 7.3% (SD=8.4, n=6) | 66.1% (SD=16.9%, n=7) | 13.9% (SD=15.9, n=7) |

**Table 1. Quantification of % of cFos-positive cells among mCherry positive cells.**