TITLE

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

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

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. WELBIO, WEL Research Institute, Wavre, Belgium.
4. Corresponding author, adekerch@ulb. ac.be
*** Both authors contributed equally to this work**
[§] Current address: Sygnature Discovery, In Vivo department, Nottingham, UK.

ABSTRACT

Striatum constitutes one of the key nodes of the decision-making circuit, with differing roles depending on the striatal localization and both direct and indirect pathway Striatal Projecting Neurons (d- and i-SPNs) populations. In a gambling task adapted to mice, we leveraged a computational characterization of the diversity in cognitive profiles to delve further into the striatal roles in decision-making. Individual decision-making profiles could be described as a combination between archetypal strategies (Optimizers, Risk-averse, Explorers). Such strategies reflected stable differences in the parameters generating decisions (sensitivity to reward magnitude, to risk or to punishment) derived from a reinforcement-learning fit of animal choices. Using a chemogenetic method, we found that dorsomedial striatum (DMS) neurons had the most significant impact on decision-making, while Nucleus accumbens (NAc) had a minor effect and dorsolateral striatum (DLS) had none. Specifically, d- SPNs reduce risk aversion by enhancing the perceived value of a risky choice in mice. On the other hand, i-SPNs amplified the importance of large gains, exerting opposite effects on mice depending on their decision-making profiles. Hence, we propose that striatal manipulations exert profile-dependent effects on choices because striatal subpopulations affect distinct parameters that interact to generate decisions.

INTRODUCTION

To adapt to dynamic environments, decision processes aim to select the course of action that will bring the best outcome. Value-based decision-making can be described as the iterative process of selecting and executing actions based on their values, and of evaluating outcomes to update action values[1,2]. Decision-making is therefore fundamentally based on estimates (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 therefore been a renewed interested in understanding how the nervous system estimates the probabilities and values of outcomes, and how it weights these estimates to choose between options.

Decision-making processes emerge from the meso-cortico-limbic loop with a key, yet not fully understood, role for 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 that sets goal-directed actions (DMS) or generates habits (DLS), and NAc [6–15] manages reward, motivation and pavlovian associations. While habits, which consist of stimulus-response (or state-response) associations, are generally insensitive to the value of the outcome[16,17], Pavlovian (stimulus-response) and goal-directed (action-outcome) associations can affect value-based decision[13,16]. How these associations with outcomes (rewarding or punitive) do or do not rely on shared computations between striatal structures has been the subject of much study[18]. The conserved distinction between SPNs (accounting for 95% of striatal neurons) from the Dopamine D1 receptor (D_1R)-expressing, direct (dSPNs) or Adenosine A2A receptor ($A_{2A}R$)-expressing, indirect (iSPNs) pathway suggests an anatomical pattern supporting canonical computations. It remains debated what are the respective role of dSPNs and iSPNs[19] : antagonist prokinetic dSPNs and antikinetic iSPNs , as demonstrated by loss-of-function or gain-of-function experiments[20–25], supporting a go/no-go model [26,27], or coactivation of the dSPN and iSPN during locomotion or actions, as supported by correlative approaches[28–33]. However, a recent report indicates that there is 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, antagonists model propose that dSPNS and iSPNs respectively encode potential reward and costs[19,35] or encoding and updating in goal directed learning[36]. It is less clear which decision variables would be encoded by synergistic dSPNs and iSPNs. Last, it has been shown that manipulations of striatal dopamine affects risk-seeking[9], which does not fit in a clear reward/cost dichotomy, and the relative influences of dSPNs and iSPNs on risk remains to determine. However, the attitude toward risk, as well as the sensitivity to reward, costs, or punishments, is highly idiosyncratic (i.e. it reflects stable individual traits). Different decision profiles have been described in humans and animals. Inter-individual variability in the choice behaviour of healthy individuals has been described in humans, rats [37–39] and mice [40–42], and often related to differences in mesocorticostriatal properties. Furthermore, the effect of several substances, ranging from caffeine to cocaine, can have a different impact on individuals according to their choice behavior: for example, impulsive choice has been recurrently described as a predictor of susceptibility to develop addictive disorders in humans, like 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]. Also in rats, caffeine, an agonist of $A_{2A}R$, can have a motivational effect to perform 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 sensitivities 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. (2011)[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 differently 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 sensitivity to risk, subjective utility of large rewards, and explore-exploit trade-off. We then show that increasing DMS excitability exerted the most profound effects on choices and motivation, consistent with the IGT being a goal-directed task. Specifically, we found that facilitating DMS dSPNs activity decreased safe choices in all mice by decreasing risk aversion, while facilitating DMS iSPNs activity decreased safe choices only in Optimizer mice, through a decrease in reward saturation. Increasing NAc dSPNs excitability induced similar, albeit blunted, effects compared to their DMS counterparts. NAc iSPN and DLS i- and d-SPN manipulations exerted nonspecific effects on motivation, but not on choice behavior. Overall, we highlight how striatal subpopulation exert cognitive profile-dependent effects on choice behavior.

RESULTS

Idiosyncratic strategies in the mouse gambling task emerge from different sensitivities to reward, punishment, and risk.

To assess decision-making in a complex environment, we used a rodent-adapted version of the Iowa Gambling Task[49], that we adapted for mice. The task takes place in an operant chamber in which animals can select among four nose-poke holes to get food pellets in a magazine on the opposite wall (Figure 1A, see Methods). As mice need to nose-poke on the holes between 5s (shorter responses were not rewarded and counted as premature) and 10s (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 get a reward magnitude (number of pellets), and a related probability 1-p to get 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 x average pellet rate) x 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 (number of pellets x reward probability), one can compute the (linear) "expected net return" for each option (E(P1) = 0.89 pellets, E(P2)= 1.52 pellets, E(P3)= 0.60 pellets, E(P4)= -0.32 pellets). In this context, it often considered that

options P1 and P2 constitute “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 thus considered maladaptive: despite delivering more pellets on rewarded trials, the high probability of a long time-out renders them disadvantageous on the long run (see Methods). In our experiments, mice indeed preferred on average the safe options (P1 and P2 compared to P3 and P4) and such choice behavior was stable across sessions (Figure 1B; all statistics are given in the figure legends). However, such average preferences often mask important variability. Hence, we leveraged the analysis of baseline (before chemogenetic / DREADD manipulations, see below) choice behavior from all mice (n=168) to characterize such variability. There was a marked variability both in terms of choice behavior, with e.g. P1 (the safest option) chosen as little as 0% or as high 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 inter-individual variability is by introducing a dichotomy between “good” (or “safe”) and “poor” (or “risky”) decision makers. We used such distinction (mice were considered “safe” if they had a significant preference for 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 inter-individual 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 safe choices (P1 and P2) was negatively correlated with the proportion of omissions (Figure 2A, Supp. Figure 1A-C). Such statistical structure hinted at constraints in mechanisms generating the data (i.e. the decision-making traits generating 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) with a principal component analysis (PCA, Supp. Figure 1D), and then described inter-individual variability using an archetypal analysis[41,52,53]. Compared to classical clustering that group individual data around typical observations (the cluster centers) that would constitute a decision-making profile, archetypal analysis depicts individual behavior as a continuum between extreme strategies, the archetypes (Figure 2B). This framework acknowledges that contrary to experimenter-defined performance in each task, animals define their own success criterions, 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

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 description should 1) be stable over sessions and 2) reflect underlying decision-making traits.
We first determined the number *n* 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 2E, F). We also checked that archetype analysis explained more variance than the classical safe/not safe dichotomy (Supp. Figure 2G, H). Then, we 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). Such distance-to-archetype was highly correlated between consecutive days (Figure 2D, Supp. Figure 2I, J), indicating stable profiles. Last, we checked that partitioning the animals into three groups, based on their proximity to one of the archetypes, corresponded to significant differences in behavioral data. Indeed, we found that the first group (hereafter, "Risk-averse", R) displayed higher preference for P1, i.e. the least risky option (Figure 2E), more premature responses, and less omissions (Figure 2F), than the two other groups. The second group ("Explorers", E mice) displayed less marked preferences (flatter pattern of choices in figure 2E), initiated less trials, and made more omissions. 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.

That behavioral features were enriched at each archetype is only confirmatory (like statistical differences between subgroups following a median split) as archetypal analysis is performed on the PCA space derived from the same behavioral data. We then sought to find differences between archetypes that would characterize the "cognitive objectives" that archetypes represent, in terms of the (hidden) decision-making strategies that underlie the overt choices of the animals. Indeed, the "expected net return" above considers that mice would have linear preferences and would be neutral to risk (i.e. they only optimize the average reward), e.g. that winning 4 pellets (or waiting 40 seconds) with a choice is similar as having 4 times 1 pellets with 4 choices (or waiting 4 times 10 seconds, respectively). By contrast, animals often display non-linear subjective values, and are sensitive not only to expected (mean) rewards and punishments, but also to reward/punishment risk (variance). To test for putative differences in their subjective valuation/decision processes, we fit the animals' choices with a reinforcement-learning model (Figure 2G). In this model, the probability of choosing the hole PX (e.g. P1) depends on its expected value compared to 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 Methods for details). Expected reward (or punishment) equals the reward (or punishment) probability times the subjective reward (or punishment) magnitude, like for the "expected net return". Additionally, parameter ρ depicts how subjective value saturates with the number of pellets: $\rho<1$ corresponds to a value of gaining 4 pellets lower than 4 times the value of 1 pellet, henceforth favoring small gains (i.e. small ρ favors small gains, while high ρ favors large gains). Similarly, parameter T depicts how subjective punishment

depends on time-out duration. Option value also incorporated a risk-sensitive term (φ) that could 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) while animals with "flat" distribution of choices (no marked preference) were labelled as explorative (low β). We checked that this model provided better fits than simpler models (Supp. Figure 2A,B). 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). Such model allowed to better characterize the archetypal strategies (Figure2H, Supp. Figure 2C). 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 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. The conjunction of the archetypal analysis with the computational model thus suggests that the stable interindividual variability observed in the mouse gambling task reflects stable (Supp. Figure 2D) decision-making profiles.

Dorsomedial dSPNs decrease safe choices by promoting risk-seeking.

We next sought to assess the contributions of both SPN populations in decision-making in the main striatal subdomains (DMS, DLS and NAc). We thus used DREADDs activation, rather than optogenetic (see discussion), to facilitate neuronal activity from a specific subpopulation, after the value of the different options were learned by the animal, to focus on the decision process. For this purpose, we tested 6 different groups of mice, one per SPNs population (direct or indirect pathway SPNs) and striatal area (DMS, DLS or NAc). We bilaterally injected a Cre-dependent hM3Dq DREADDs virus (AAV5 hSyn-DIO-hM3D(Gq)-mCherry) or a control fluorophore lacking the hM3Dq construct (AAV5 hSyn-DIO-mCherry) into the striatal zone of interest (DMS, DLS or NAc) in D_1R-Cre (Gong et al., 2007) or A_{2A}R-Cre mice (Durieux et al., 2009), to selectively express hM3Dq or mCherry on either in dSPNs or iSPNs respectively.

As the mouse gambling task relies on instrumental behavior, we first focused on DMS d-SPNs. As mice received the DREADD ligand (CNO, 1mg/kg) injection on the last day of behavioral testing (see below), we verified ex vivo the effectiveness of hM3Dq DREADDs in increasing the intrinsic excitability of targeted neurons by collecting slices from mice (hM3Dq and mCherry controls) just after the behavioral testing. The marker of neuronal activation cfos colocalized with the mCherry expression only in hM3Dq animals (Figure 3A), checking that CNO had facilitated neuronal activation during the mouse gambling test. Next, we controlled that neuronal activation followed from an increase in intrinsic excitability (i.e. active membrane properties), as expected from hM3Dq recruitment[54]. We did not find any changes in neurons' passive properties in patch-clamp recordings: the resting membrane potential was not changed following the CNO injection (Figure 3B). We did not observe any modifications of (putatively

corticostriatal or thalamicostriatal) synapses strength as measured by the AMPA/NMDA ratio, suggesting that cFos activity was not due to broad changes in synaptic excitation. For the intrinsic excitability of the cells, the rheobase (minimal current needed to elicit an action potential) was not changed following the hM3Dq + CNO treatment, nor the number of action potentials at which the neurons saturate. However, the number of action potentials elicited by intermediate current intensities was increased following the hM3Dq + CNO treatment, indicating an increase in neuronal gain (in the linear range of the frequency-intensity, or f-I, curve). This suggests that neuronal cFos activation arose in neurons receiving a significant drive (putatively task-related) amplified by an increased neuronal gain, rather than from nonspecific electrical activity (that would follow from changes in resistance or rheobase). Together, patch clamp recordings confirmed that CNO injection before the behavioral testing exerted lasting effects (detectable ex vivo after the session) on neuronal excitability.

Having verified that hM3Dq expressed on DMS-dSPNs favored their activation during the mouse gambling test, we next assessed how this enhancement of neuronal activity affected decision-making. We compared the difference in preferences for the four options (P1-P4) in the mouse gambling task under CNO and at baseline, for hM3Dq animals and mCherry controls (Fig. 3C). Safe mice (i.e. mice with a preference for advantageous options P1 and P2) in the hM3Dq + CNO conditions displayed a decrease in the preference for P1 and P2 (Fig. 3D). However, this decrease could arise from different reasons: worst decision-making, increased exploration, decreased aversion for punishment, etc. Hence, we sought for a more precise characterization in terms of a modification in decision-making traits. We also observed a decrease in the number of trials and an increase in omissions (Fig. 3E), i.e. a decrease in the pace of the instrumental action. This combined modification of preferences and instrumental pace suggested a global effect of DMS-dSPNs facilitation on the archetypal strategies. Indeed, we observed a global shift in the PCA space for hM3Dq animals under CNO, with a displacement away from the risk-averse apex (Fig. 3F). This was confirmed by the computational model-based analysis, as the change in preferences in the hM3Dq-CNO condition was better explained by an increase of the risk sensitivity parameter (Fig. 3F, Supp. Fig 3A-D). Overall, the behavioral data and computational analyses suggest that facilitating DMS-dSPNs activity decreased P1 and P2 choices by increasing proneness to risk.

Dorsomedial iSPNs favor large gains, exerting opposite effects on safe choices depending on the baseline strategy.

We then compared the above facilitation of DMS-dSPNs with that of DMS-iSPNs, as these two subpopulations are hypothesized to act either in opposition[20–27] or in synergy[28–33] or in a complementary manner supporting a dual selection-suppression function[34]. In this subpopulation too, selective hM3Dq DREADDs expression followed by CNO injection during the mouse gambling test was effective in increasing the intrinsic excitability (f-I gain), driving cfos activation in targeted neurons (Figure 4A and 4B). Contrary to the facilitation of DMS dSPNs activity, CNO did not exert any significant influence on preferences in hM3Dq animals compared to mCherry controls, when considering all mice or only safe mice (Fig. 4C). Nevertheless, facilitating the activity of DMS iSPNs with DREADDs decreased the overall

number of trials and the proportion of premature responses, and increased the proportion of omissions (Fig. 4D). This hinted at an effect on archetypal profiles. Indeed, CNO induced a global shift in the PCA space on hM3Dq animals, toward the Explorers (E) apex. The effect of DREADD-mediated facilitation of DMS-iSPNS neuronal activity was thus different from that of DMS-dSPNs (i.e. a shift away from the R apex). This led us to examine more closely the effects of CNO on the preferences from hM3Dq animals, depending on their proximity to archetypes rather than based on the safe/risky dichotomy. Facilitating DMS-iSPNs activity exerted opposite effects on the preference for safe (P1 and P2) options in Optimizers mice (a decrease in P1+P2 choices) and in the other two archetypes (increased P1+P2 choices in R+I animals). Hence the DREADD effect on choices depended on the baseline strategy of the animals. We thus characterized which modifications of the decision-making parameters could account best for this differential effect. Surprisingly, a single modification in the ρ parameter (decrease in the reward saturation with pellet number) could explain the differential DREADD effect on choices (Fig. 4F). Indeed, in reinforcement-learning models such as ours, the alteration of one decision-making parameter can induce opposite behavioral patterns depending on the value of the other parameters. Here, an increase in reward saturation (decreasing the value of large gains) increases the preference for P1 and P2 when animals are risk-averse but can decrease the preference for P1 and P2 in risk-neutral mice (Supp. Figure 4E,F). Hence, the computational analyses suggest that facilitating DMS-iSPNs activity decreased how reward saturates with reward size (favoring large gains), exerting a decision profile-dependent effect on choices.

Lesser involvement of Nucleus accumbens and dorsolateral SPNs in instrumental decisions.

To extend the understanding of the role(s) of d- and i-SPNs in decision making we assessed their roles in the 2 other main regions of the striatum, NAc and DLS. Even if the mouse gambling task is instrumental, the nucleus accumbens (NAc) and dorsolateral striatum (DLS) may also influence choice behavior. Given the importance of iSPNs in the DMS, we began by assessing their role in the NAc. The CNO injection in animals expressing hM3Dq in NAc-iSPNS drove efficient cfos activation (Fig. 5A) and increase in intrinsic excitability. The behavioral effect of DREADD facilitation of in NAc-iSPNs activity was relatively similar to what we observed when facilitating DMS-iSPNs. Preference was not changed on average, but the overall number of trials and the proportion of premature responses decreased, while the proportion of omissions increased (Fig. 5B). This pattern was reflected in the PCA space, where CNO induced a shift of hM3Dq animals toward the E apex, albeit smaller than what we observed following DMS-iSPNs manipulation (Fig. 5C). Similarly, the behavioral effects of NAc-iSPNs facilitation could be accounted by an increase in reward saturation in the computational model (Fig. 5B). Hence, facilitating NAc-iSPNs appeared to exert similar, but smaller in magnitude, effects as DMS-iSPNs facilitation.

On the contrary, facilitating NAc-dSPNs activity did not exert effects similar to that of DMS-dSPNs facilitation. hM3Dq expression coupled with CNO treatment was efficient in driving cfos activity and changes in intrinsic excitability in NAc-dSPNs (Fig. 5C). In the mouse

gambling task, such facilitation of neuronal activity strongly decreased the number of trials (Fig. 5D), contrary to DMS-dSPNs facilitation, which did not exert any effect. Facilitating NAc-dSPNs activity did not alter the proportion of premature or omission trials (Fig. 5D), again in contrast with the reduction in these measures following DMS-dSPNs manipulation. We did not observe any differences in the choice behavior, neither following the safe/risky dichotomy, nor when considering archetypes. In the PCA, this corresponded to a modest (but significant) shift away from the Optimizers (O) apex (Fig. 5D). We did not find any differences in the parameters from the computational model of decision-making when fitting the choices under the hM3Dq+CNO treatment. Overall, facilitating NAc-dSPNs exerted a specific effect not comparable to DMS-dSPNs facilitation (nor to that of NAc-iSPNs), by only inducing a reduction in the number of trials.

Last, we evaluated the influence of DLS on choice behavior. We checked that the CNO treatment drove cfos activity and increased intrinsic excitability of targeted neurons in hM3Dq animals from the DLS-iSPNs (Fig 5E) and DLS-dSPNs condition (Fig 5F). However, in the mouse gambling task, there was no change in preferences (in safe mice or when considering archetypes) under hM3Dq +CNO treatment (Fig 5E, F). We did not find any significant differences in model parameters when fitting the choices, in any (DLS-iSPNs or DLS-dSPNs) conditions (Supp Fig.) We observed small, nonspecific effects on the proportions of premature and omission trials (Supp Fig.) for both the manipulations of DLS-iSPNs and DLS-dSPNs activity. Consistent with the lesser involvement of DLS in decision-making, this highlights the specificity of the effects of DMS-dSPNs and iSPNs on decision-making parameters we unravel.

DISCUSSION

Consistent with other studies in human and rat versions of the Iowa Gambling Task, we found marked inter-individual variability in preferences. However, we departed from usual classifications in terms of experimenter-defined performance[37,39] and chose instead an unsupervised approach (Archetypal analysis) to characterize subgroups. Whether explicitly (e.g. optimal foraging or life history theory) or implicitly, assessing inter-individual variability on the single scale of "task performance" assumes that animals have evolved to optimize a given quantity, e.g. long-term reward relative to motor costs, and that the experimental setup measures the fitness of the animals on this scale (i.e. the setup is "factor pure"[55]). However, organisms have likely evolved to trade priorities across various objectives, and, accordingly, animals are equipped with a cognitive architecture for deliberative, goal-directed decisions[56]. Hence valuation, i.e. how mice assign a value to options, remains fundamentally subjective[57,58]. Computational models of decision-making constitute powerful tools designed to infer subjective valuation from the temporal series of choices. We could thus explain patterns of inter-individual variability in preferences in terms of "cognitive profiles", i.e. the decision-processes (i.e. reward, risk and time-out sensitivities) generating choices. These cognitive strategies consisted in tradeoffs between decision parameters, suggesting that evolution may have selected a space of objectives among which animals are distributed. Indeed, depending on the social and physical context, age, the history of previous experiences and states, animals can

show different patterns of preferences[42,59,60]. Importantly, as computational models of decision-making are non-linear, affecting one decision parameter (e.g. reward sensitivity) through chemogenetics may exert effects that depend upon the values of the other parameters (e.g. risk or time-out sensitivities). This is in line with the idea that biological spaces are concave[61] and that experience, disease, or experimental manipulations, may exert differential (and even opposite) effects on a given task measures. 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). The fact that specific manipulations of striatal subpopulation can affect such cognitive profiles also highlight that decision-processes (and the "cognitive architecture" in general) are emergent properties, emerging from the interaction between neural processes. Brain markers correlate with individual variability in choice behaviour, such as prefrontal serotonin[40] or the balance between striatal and prefrontal excitability[37], thereby affecting the landscape of possible neuronal trajectories in prefronto-striatal networks, leading to different choices (Kurikawa et al., 2018). Nevertheless, a lot of work remains to be done to determine the causality between brain markers and cognitive profiles. While an "endophenotype" view may consider brain markers as the origin of inter-individual variability in decision-making; brain markers may instead constitute mediators of individual and social context and experience on cognitive strategies[42]. The fact that we found marked variability in decision-making among a quasi-clonal population of inbred mice goes in this line. Considering decision-making in terms of subjective valuation and evolutionary tradeoff between strategies, rather than on a single "adaptive" scale, also recast the concept "maladaptive" decisions[50] as "atypical" instead.

To study the causal implication of striatal subpopulation in choice behavior, we preferred DREADD over optogenetics, because it was not needed to restrict a temporal window for manipulation (and decision-making likely occurs during the whole task). Furthermore, we wanted to avoid the risk of unphysiological synchronization of striatal neurons under specific optogenetic stimulation protocols. Indeed, striatal neurons are involved in action selection[29,62–65] and neurons promoting distinct goals are not activated synchronously with the same kinetics under physiological conditions throughout various behaviours[30,33,34]. DREADDS partially circumvent this issue as a gain of function approach in this study. However, one limitation of our DREADD approach is that their CNO ligand may be converted to clozapine[66], an antipsychotic with sedative effects potentially decreasing of behavioural impulsivity[67–69] but our mCherry controls excluded DREADD-independent effects.

Our results precise and extend the literature in the differential involvement of subparts of the striatum[70]. The role of DMS to sustain instrumental associations (action reward) has been widely proven before[6,7,13,16,71]. Here, we show that the DMS is not critical only to maintain the instrumental responding of mice, but also to determine their preferences under risk. We further provide a computational rationale (and iSPN verus dSPN distinction) for the effect of DMS on risky choice[72] in terms for sensitivity to reward variance and reward saturation. While NAc has also been implicated in risky decision-making, with a specific involvement of iSPNs as in our study[15], we found a lower NAc effect than the one of the DMS. This may be due to the predominantly instrumental nature of the task, as the NAc rather supports the acquisition of stimulus-outcome associations[73–76]. There was no effect of DLS manipulations in the IGT, consistent with the known dissociation of DMS and DLS[12,13,17] in respectively controlling goal-directed behavior and habits. Despite extensive training, mice did not develop rigid preferences

(i.e. most mice had balanced choice behavior, Figure 1) suggesting they did not develop habits (although we did not test choices in extinction). We also provide novel data on the major anatomical and functional distinction 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 in complementary ways to promote and oppose (or refine) action selection[33,63,77–79]. However, in the DMS, we found that DREADD-mediated activation of both iSPNs and dSPNs enhanced risky choice, yet through distinct computational mechanisms (i.e. uncertainty-sensitivity and reward saturation, respectively). This points at more complex interactions between pathways. The activity of iSPNs has been related to the encoding of non-rewarding events, or changes in reward value (prediction errors), leading to the updating of action value[36] and henceforth task-switching[79,80]. Our computational account is consistent with this view, as reward saturation (i.e. how subjective value saturates with reward magnitude) may arise from learning effects. As we could disentangle reward saturation from sensitivity to uncertainty in our multiple-choice setup, our results extend and explains the ones obtained in tasks involving binary choices. For instance, activating iSPNs may increase seemingly stochastic choices in serial choices in which the exploratory choice may also constitute a risky (high variance) choice[81] because reward saturation affects the perception of risk (Figure 4 and Supp. Figure 4). In line with above point, for DMS-dSPNs facilitation, a straightforward interpretation would be that animals displayed poorer decision-making, but we found that increased proneness to risk may be a better explanation. We also shed a new light on previous results obtained following manipulation of dopamine levels or receptors. DAT (dopamine transporter)- KD (knock down) mice display riskier choices than controls in the IGT[49], and also "safe" decision-makers in the mouse IGT presented lower levels of dopamine in the dorsal striatum than average mice. Our results suggest that dopamine effects may arise from distinct D1R- and D2R- mediated mechanisms, combining to promote riskier choices. 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[28–34,63]. 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 could also be determined[82]. Moreover, we could also check if, as our results suggest, the activity of dSPNs when mice are making a choice differs between the DMS, the DLS and the NAc. There are multiple questions to answer and using *in vivo* techniques we could confirm what our results have hinted by increasing neuronal excitability rather than directly activating targeted cells. These questions should be carefully investigated in future work.

Understanding and aiding individuals grappling with decision-related disorders, such as pathological gambling[83] and drug addiction[84], necessitates a shift in perspective. Rather than viewing impairments in decision-making through a purely normative lens, it is crucial to perceive them as discrepancies between anticipated outcomes and actual choices. To address this, we propose a framework in which decision-making "profiles" arise from evolutionary tradeoffs, implemented by differences in parameters from the deliberative machinery leading to goal-directed choices. Within this framework, biological factors, such as our manipulation

of striatal subpopulations, or social[42] factors, can induce shifts in the spectrum of potential strategies. These shifts are probably governed by "meta-learning" rules[85] which offer promising avenues for enhancing translational research efforts.

References

1. Dolan, R. J. & Dayan, P. Goals and Habits in the Brain. *Neuron* **80**, 312–325 (2013).
2. Rangel, A., Camerer, C. & Montague, P. R. A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* **9**, 545–556 (2008).
3. Graybiel, A. M. The basal ganglia and chunking of action repertoires. *Neurobiol. Learn. Mem.* **70**, 119–136 (1998).
4. Hikosaka, O. Neural systems for control of voluntary action--a hypothesis. *Adv. Biophys.* **35**, 81–102 (1998).
5. Redgrave, P., Prescott, T. J. & Gurney, K. The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience* **89**, 1009–1023 (1999).
6. Hart, G., Leung, B. K. & Balleine, B. W. Dorsal and ventral streams: the distinct role of striatal subregions in the acquisition and performance of goal-directed actions. *Neurobiol. Learn. Mem.* **108**, 104–118 (2014).
7. Kupferschmidt, D. A., Juczewski, K., Cui, G., Johnson, K. A. & Lovinger, D. M. Parallel, but Dissociable, Processing in Discrete Corticostriatal Inputs Encodes Skill Learning. *Neuron* **96**, 476-489.e5 (2017).
8. Mai, B., Sommer, S. & Hauber, W. Dopamine D1/D2 Receptor Activity in the Nucleus Accumbens Core But Not in the Nucleus Accumbens Shell and Orbitofrontal Cortex Modulates Risk-Based Decision Making. *Int. J. Neuropsychopharmacol.* **18**, pyv043 (2015).

9. Stopper, C. M., Khayambashi, S. & Floresco, S. B. Receptor-Specific Modulation of Risk-Based Decision Making by Nucleus Accumbens Dopamine. *Neuropsychopharmacology* **38**, 715–728 (2013).
10. Vandaele, Y. *et al.* Distinct recruitment of dorsomedial and dorsolateral striatum erodes with extended training. *eLife* **8**, e49536 (2019).
11. Yang, J.-H. & Liao, R.-M. Dissociable contribution of nucleus accumbens and dorsolateral striatum to the acquisition of risk choice behavior in the rat. *Neurobiol. Learn. Mem.* **126**, 67–77 (2015).
12. Yin, H. H., Knowlton, B. J. & Balleine, B. W. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.* **19**, 181–189 (2004).
13. Yin, H. H., Ostlund, S. B., Knowlton, B. J. & Balleine, B. W. The role of the dorsomedial striatum in instrumental conditioning. *Eur. J. Neurosci.* **22**, 513–523 (2005).
14. Yin, H. H. *et al.* Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. *Nat. Neurosci.* **12**, 333–341 (2009).
15. Zalocusky, K. A. *et al.* Nucleus accumbens D2R cells signal prior outcomes and control risky decision-making. *Nature* **531**, 642–646 (2016).
16. Balleine, B. W., Liljeholm, M. & Ostlund, S. B. The integrative function of the basal ganglia in instrumental conditioning. *Behav. Brain Res.* **199**, 43–52 (2009).
17. Yin, H. H. & Knowlton, B. J. The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.* **7**, 464–476 (2006).
18. Peak, J., Hart, G. & Balleine, B. W. From learning to action: the integration of dorsal striatal input and output pathways in instrumental conditioning. *Eur. J. Neurosci.* **49**, 658–671 (2019).

19. Bariselli, S., Fobbs, W. C., Creed, M. C. & Kravitz, A. V. A competitive model for striatal action selection. *Brain Res.* **1713**, 70–79 (2019).

20. Bateup, H. S. *et al.* Distinct subclasses of medium spiny neurons differentially regulate striatal motor behaviors. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 14845–14850 (2010).

21. Carvalho Poyraz, F. *et al.* 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. Durieux, P. F. *et al.* D2R striatopallidal neurons inhibit both locomotor and drug reward processes. *Nat. Neurosci.* **12**, 393–395 (2009).

23. Durieux, P. F., Schiffmann, S. N. & de Kerchove d'Exaerde, A. 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. Kravitz, A. V. *et al.* Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature* **466**, 622–626 (2010).

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

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

27. Gerfen, C. R. *et al.* D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. *Science* **250**, 1429–1432 (1990).

28. Barbera, G. *et al.* Spatially Compact Neural Clusters in the Dorsal Striatum Encode Locomotion Relevant Information. *Neuron* **92**, 202–213 (2016).

29. Cui, G. *et al.* Concurrent activation of striatal direct and indirect pathways during action initiation. *Nature* **494**, 238–242 (2013).

30. Klaus, A. *et al.* The Spatiotemporal Organization of the Striatum Encodes Action Space. *Neuron* **95**, 1171-1180.e7 (2017).

31. Markowitz, J. E. *et al.* The Striatum Organizes 3D Behavior via Moment-to-Moment Action Selection. *Cell* **174**, 44-58.e17 (2018).

32. Parker, J. G. *et al.* Diametric neural ensemble dynamics in parkinsonian and dyskinetic states. *Nature* **557**, 177–182 (2018).

33. Weglage, M. *et al.* Complete representation of action space and value in all dorsal striatal pathways. *Cell Rep.* **36**, 109437 (2021).

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

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

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

37. Fitoussi, A. *et al.* 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. Kurikawa, T., Haga, T., Handa, T., Harukuni, R. & Fukai, T. Neuronal stability in medial frontal cortex sets individual variability in decision-making. *Nat. Neurosci.* **21**, 1764–1773 (2018).

39. Rivalan, M., Coutureau, E., Fitoussi, A. & Dellu-Hagedorn, F. 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. Pittaras, E., Callebert, J., Chennaoui, M., Rabat, A. & Granon, S. Individual behavioral and neurochemical markers of unadapted decision-making processes in healthy inbred mice. *Brain Struct. Funct.* **221**, 4615–4629 (2016).
41. Dongelmans, M. *et al.* Chronic nicotine increases midbrain dopamine neuron activity and biases individual strategies towards reduced exploration in mice. *Nat. Commun.* **12**, 6945 (2021).
42. Torquet, N. *et al.* Social interactions impact on the dopaminergic system and drive individuality. *Nat. Commun.* **9**, 3081 (2018).
43. Belin, D., Mar, A. C., Dalley, J. W., Robbins, T. W. & Everitt, B. J. High impulsivity predicts the switch to compulsive cocaine-taking. *Science* **320**, 1352–1355 (2008).
44. Elton, A., Smith, C. T., Parrish, M. H. & Boettiger, C. A. Neural Systems Underlying Individual Differences in Intertemporal Decision-making. *J. Cogn. Neurosci.* **29**, 467–479 (2017).
45. Spoelder, M. *et al.* Individual differences in voluntary alcohol intake in rats: relationship with impulsivity, decision making and Pavlovian conditioned approach. *Psychopharmacology (Berl.)* **234**, 2177–2196 (2017).
46. Hynes, T. J. *et al.* Dopamine neurons gate the intersection of cocaine use, decision making, and impulsivity. *Addict. Biol.* **26**, e13022 (2021).
47. Ferland, J.-M. N. & Winstanley, C. A. Risk-preferring rats make worse decisions and show increased incubation of craving after cocaine self-administration. *Addict. Biol.* **22**, 991–1001 (2017).

48. SanMiguel, N. *et al.* Individual differences in the energizing effects of caffeine on effort-based decision-making tests in rats. *Pharmacol. Biochem. Behav.* **169**, 27–34 (2018).

49. Young, J. W., van Enkhuizen, J., Winstanley, C. A. & Geyer, M. A. 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. Rivalan, M., Ahmed, S. H. & Dellu-Hagedorn, F. Risk-prone individuals prefer the wrong options on a rat version of the Iowa Gambling Task. *Biol. Psychiatry* **66**, 743–749 (2009).

51. Bechara, A., Damasio, H., Tranel, D. & Damasio, A. R. Deciding advantageously before knowing the advantageous strategy. *Science* **275**, 1293–1295 (1997).

52. Cutler, A. & Breiman, L. Archetypal Analysis. *Technometrics* **36**, 338–347 (1994).

53. Hart, Y. *et al.* Inferring biological tasks using Pareto analysis of high-dimensional data. *Nat. Methods* **12**, 233–235 (2015).

54. Roth, B. L. DREADDs for Neuroscientists. *Neuron* **89**, 683–694 (2016).

55. Balleine, B. W. The Meaning of Behavior: Discriminating Reflex and Volition in the Brain. *Neuron* **104**, 47–62 (2019).

56. Budaev, S., Jørgensen, C., Mangel, M., Eliassen, S. & Giske, J. Decision-Making From the Animal Perspective: Bridging Ecology and Subjective Cognition. *Front. Ecol. Evol.* **7**, (2019).

57. Hultman, C. *et al.* Exploring decision-making strategies in the Iowa gambling task and rat gambling task. *Front. Behav. Neurosci.* **16**, (2022).

58. Gomez-Marin, A. & Ghazanfar, A. A. The Life of Behavior. *Neuron* **104**, 25–36 (2019).

59. Zeeb, F. D. & Winstanley, C. A. 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).

60. Cassotti, M., Aïte, A., Osmont, A., Houdé, O. & Borst, G. What have we learned about the processes involved in the Iowa Gambling Task from developmental studies? *Front. Psychol.* **5**, (2014).

61. Marder, E. & Goaillard, J.-M. Variability, compensation and homeostasis in neuron and network function. *Nat. Rev. Neurosci.* **7**, 563–574 (2006).

62. Shin, J. H., Song, M., Paik, S.-B. & Jung, M. W. 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).

63. Tecuapetla, F., Jin, X., Lima, S. Q. & Costa, R. M. Complementary Contributions of Striatal Projection Pathways to Action Initiation and Execution. *Cell* **166**, 703–715 (2016).

64. Isomura, Y. *et al.* Reward-Modulated Motor Information in Identified Striatum Neurons. *J. Neurosci.* **33**, 10209–10220 (2013).

65. Jin, X., Tecuapetla, F. & Costa, R. M. Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences. *Nat. Neurosci.* **17**, 423–430 (2014).

66. Manvich, D. F. *et al.* 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).

67. Haidary, H. A. & Padhy, R. K. Clozapine. in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2024).

68. Maurice-Gélinas, C., Deslauriers, J., Monpays, C., Sarret, P. & Grignon, S. 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).

69. Spivak, B., Shabash, E., Sheitman, B., Weizman, A. & Mester, R. The effects of clozapine versus haloperidol on measures of impulsive aggression and suicidality in chronic

schizophrenia patients: an open, nonrandomized, 6-month study. *J. Clin. Psychiatry* **64**, 755–760 (2003).

70. Cox, J. & Witten, I. B. Striatal circuits for reward learning and decision-making. *Nat. Rev. Neurosci.* **20**, 482–494 (2019).
71. Malvaez, M. Neural substrates of habit. *J. Neurosci. Res.* **98**, 986–997 (2020).
72. Schumacher, J. D., van Holstein, M., Bagrodia, V., Le Bouder, H. B. & Floresco, S. B. Dorsomedial striatal contributions to different forms of risk/reward decision making. *Neurobiol. Learn. Mem.* **178**, 107369 (2021).
73. Calipari, E. S. *et al.* In vivo imaging identifies temporal signature of D1 and D2 medium spiny neurons in cocaine reward. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 2726–2731 (2016).
74. Gmaz, J. M., Carmichael, J. E. & van der Meer, M. A. Persistent coding of outcome-predictive cue features in the rat nucleus accumbens. *eLife* **7**, e37275 (2018).
75. Salamone, J. D. & Correa, M. The mysterious motivational functions of mesolimbic dopamine. *Neuron* **76**, 470–485 (2012).
76. Yin, H. H., Ostlund, S. B. & Balleine, B. W. Reward-guided learning beyond dopamine in the nucleus accumbens: the integrative functions of cortico-basal ganglia networks. *Eur. J. Neurosci.* **28**, 1437–1448 (2008).
77. Li, Y. *et al.* Optogenetic Activation of Adenosine A2A Receptor Signaling in the Dorsomedial Striatopallidal Neurons Suppresses Goal-Directed Behavior. *Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol.* **41**, 1003–1013 (2016).
78. Macpherson, T. *et al.* Parallel and hierarchical neural mechanisms for adaptive and predictive behavioral control. *Neural Netw. Off. J. Int. Neural Netw. Soc.* **144**, 507–521 (2021).

79. Nonomura, S. *et al.* Monitoring and Updating of Action Selection for Goal-Directed Behavior through the Striatal Direct and Indirect Pathways. *Neuron* **99**, 1302-1314.e5 (2018).

80. Shin, J. H., Kim, D. & Jung, M. W. Differential coding of reward and movement information in the dorsomedial striatal direct and indirect pathways. *Nat. Commun.* **9**, 404 (2018).

81. Delevich, K. *et al.* Activation, but not inhibition, of the indirect pathway disrupts choice rejection in a freely moving, multiple-choice foraging task. *Cell Rep.* **40**, 111129 (2022).

82. Samejima, K., Ueda, Y., Doya, K. & Kimura, M. Representation of action-specific reward values in the striatum. *Science* **310**, 1337–1340 (2005).

83. van den Bos, R. *et al.* Cross-species approaches to pathological gambling: a review targeting sex differences, adolescent vulnerability and ecological validity of research tools. *Neurosci. Biobehav. Rev.* **37**, 2454–2471 (2013).

84. Stoops, W. W. & Kearns, D. N. Decision-making in addiction: Current knowledge, clinical implications and future directions. *Pharmacol. Biochem. Behav.* **164**, 1–3 (2018).

85. Schweighofer, N. & Doya, K. Meta-learning in reinforcement learning. *Neural Netw. Off. J. Int. Neural Netw. Soc.* **16**, 5–9 (2003).

86. Gong, S. *et al.* Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J. Neurosci. Off. J. Soc. Neurosci.* **27**, 9817–9823 (2007).

87. Jonas, P. & Burnashev, N. Molecular mechanisms controlling calcium entry through AMPA-type glutamate receptor channels. *Neuron* **15**, 987–990 (1995).

88. Jane, D. E., Lodge, D. & Collingridge, G. L. Kainate receptors: pharmacology, function and therapeutic potential. *Neuropharmacology* **56**, 90–113 (2009).

89. Kourrich, S., Rothwell, P. E., Klug, J. R. & Thomas, M. J. Cocaine experience controls bidirectional synaptic plasticity in the nucleus accumbens. *J. Neurosci. Off. J. Soc. Neurosci.* **27**, 7921–7928 (2007).

Acknowledgments: We thank for critical reading of the manuscript.. We thank Delphine Houtteman, Souad Laghmari, Perrine Hagué and Laetitia Cuvelier for technical assistance and mouse colonies in Brussels, LIMIF for assistance in imaging techniques. ECR was a research fellows 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 FWC, Fondation ULB, AXA Research Fund (Chaire AXA) to AKE, Agence Nationale de la Recherche (ANR-22-CE16-0013 LEARN) and 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 in vivo and ex vivo experiments : ECR and DR. Methodology and investigation for 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 no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

FIGURES AND LEGENDS

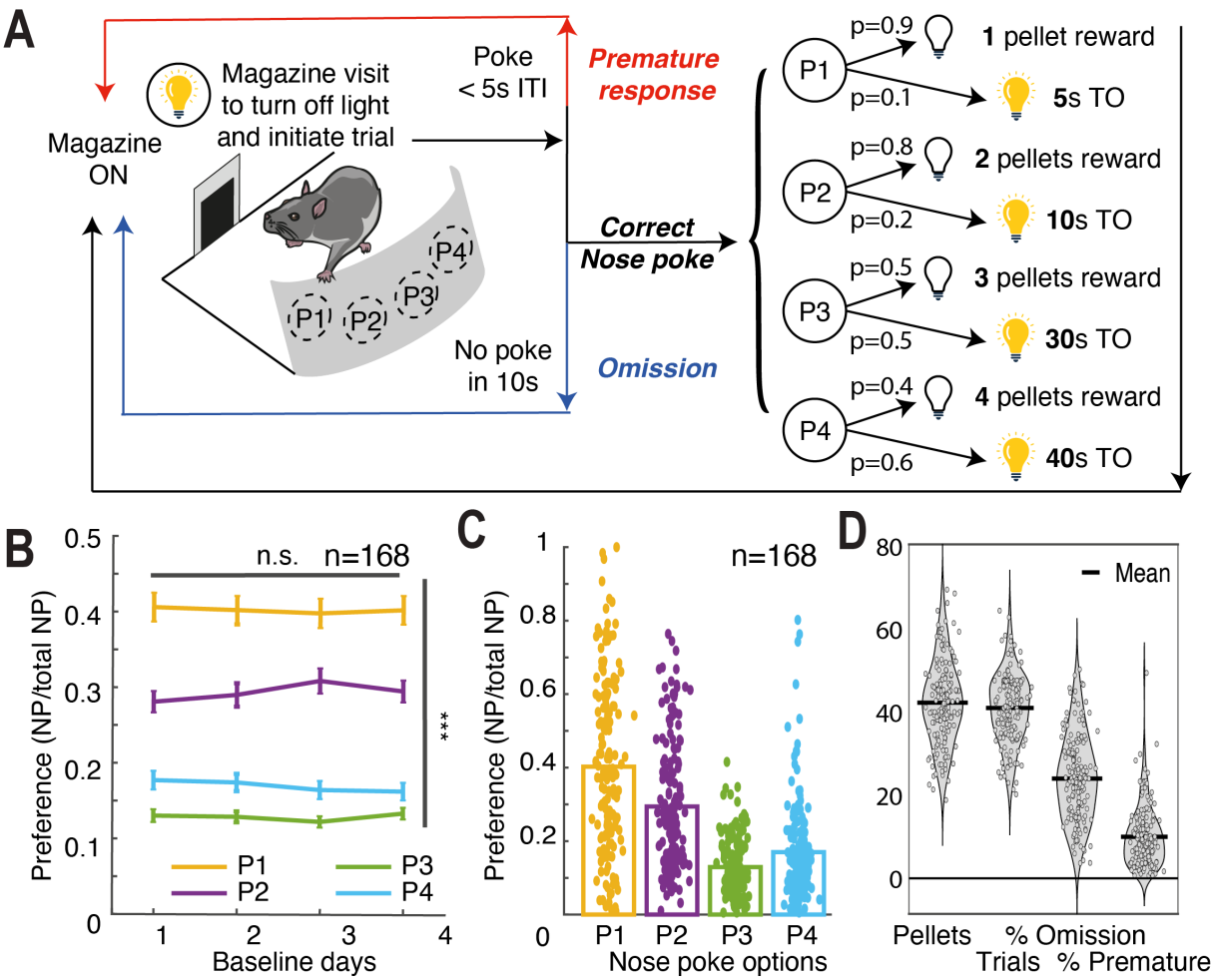


Figure 1

Figure 1. Mice exhibited different but stable preferences in a gambling task. **A.** Task schema: upon initiating a trial by a magazine entry, mice had to choose. **B.** Preferences for the four options against days of training, showing that mice learn the task and display a stable preference for safe options (repeated measure ANOVA: time: $F_{(3,668)}=92.2$, $p<1e^{-12}$; options $F_{(3,668)}=0.1$, n.s.). **C.** Preferences from all mice (all conditions) display substantial inter-individual variability. **D.** Other behavioral measures in the gambling task (number of trials, proportions of premature responses, omissions) also display substantial inter-individual variability, suggesting different choice strategies.

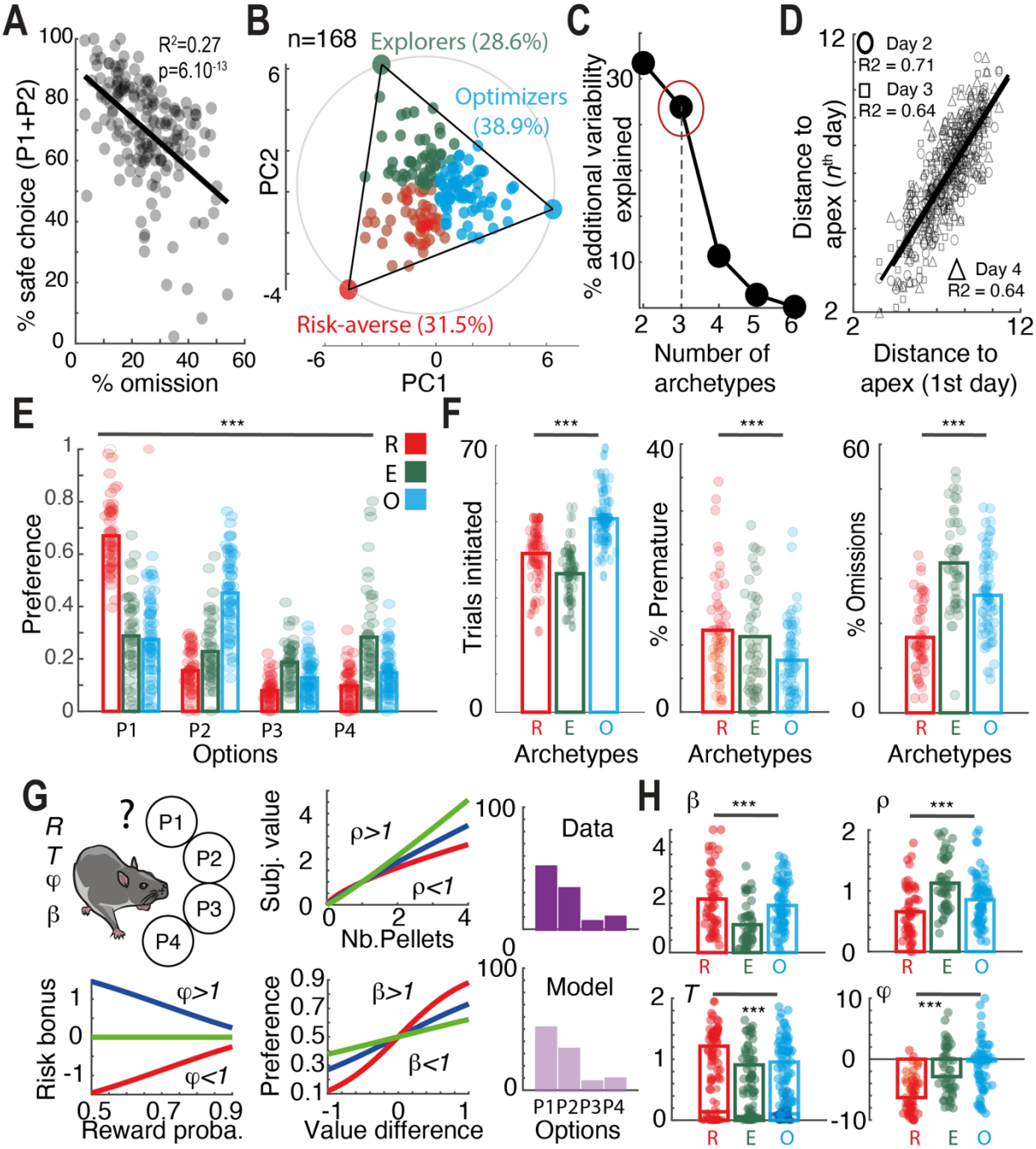


Figure 2

Figure 2. Variability in mice behavior can be described as distinct strategies arising from different sensitivities to reward, punishment, and uncertainty. **A.** Preference for the safe choice was correlated with the % of omission ($R^2 = 0.27$, $p=6.10^{-13}$), suggesting that the different choice strategies and motivational/attentional measures reflect underlying cognitive profiles. **B**. Archetypal analysis performed on 8 measures (number of trials, pellets obtained, % omission, % premature responses, 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 (PC) 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)). In this framework, an individual can be described as a convex

combination of the 3 archetypes. **C.** Variability (in the 8 measures x 168 mice) explained as a function of the number of archetypes, showing that 3 archetypes constitute a parsimonious description of the data. **D.** Distance to archetypes (R apex) 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), suggesting that behavior in the task arose from stable cognitive profiles. **E**. Preferences (2-ways ANOVA: $F_{(6,66)}$=115.86, p=1.10^{-99}) from mice assigned to their closest archetype indicate that P1 (certain, small reward, short TO) was preferred by Risk-averse (R) mice, and Optimizers (O) preferred P2 (larger, less certain reward, more advantageous relative to TO than P1) while Explorers (E) mice did not exhibit preferences. **F**. These choice strategies were associated with different in other task measures (Trials: 1-way ANOVA, $F_{(2,165)}$=69.82, p=1.10^{-22}; Premature: 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}, with Risk-averse mice displaying less omissions, Optimizers having earned more pellets and less premature responses, while Explorers mice made more omissions and gained less pellets. **G**. Left, 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 on how much they positively valued a pellet (ρ 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, φ parameter) and the precision in their decision process ("inverse temperature" β parameter). Right, example of mouse preference fit by the choice model. **H**. The choice model suggests that the strategies unraveled by the archetypal analysis arise from different sensitivities to task variables (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}): mice were in the Risk-averse archetype if their sensitivity to punishment was higher and their attitude toward risk more negative; Explorers mice mainly displayed lower β (low precision, or high noise, in the decision process); and Optimizers exhibited higher sensitivity to reward and their attitude towards risk was neutral.

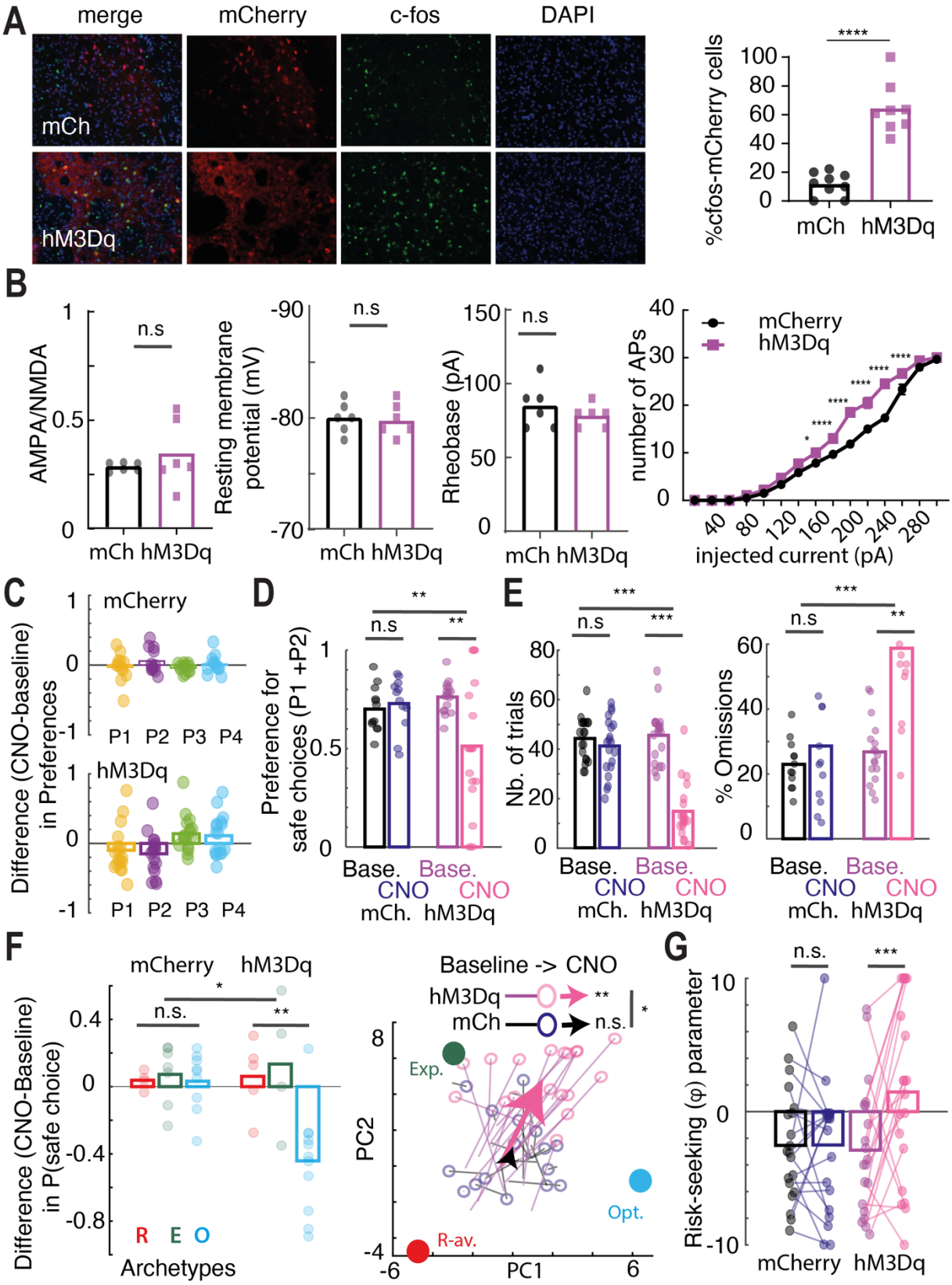

A
merge
mCherry
c-fos
DAPI
mCh
hM3Dq
%cfos-mCherry cells
mCh
hM3Dq
B
AMPA/NMDA
n.s
mCh hM3Dq
Resting membrane potential (mV)
Rheobase (pA)
number of APs
mCherry
hM3Dq
injected current (pA)
C
mCherry
hM3Dq
Difference (CNO-baseline) in Preferences
P1 P2 P3 P4
D
Preference for safe choices (P1 +P2)
Base. CNO
Base. CNO
mCh. hM3Dq
E
Nb. of trials
% Omissions
F
mCherry hM3Dq
Difference (CNO-Baseline) in P(safe choice)
R E O
Archetypes
Baseline -> CNO
hM3Dq
mCh
Exp.
Opt.
R-av.
PC1
PC2
G
Risk-seeking (φ) parameter
mCherry hM3Dq

Figure 3

Figure 3. DMS d-SPNs promote risk-seeking. **A.** Microscope fluorescence for control (up) and hM3Dq (bottom) individuals in D1-Cre mice injected in the dorsomedial striatum (DMS): DAPI (blue) indicate nuclei, mCherry (red) transfected neurons and C-fos (green) CNO-activated neurons. White arrowheads point at hM3Dq-expressing neuronal somas. C-fos expression was significantly higher in the hM3Dq group comparing to controls (*** t-test, $t_{(16)}$=8.5, p<0,0001, Unpaired t-test). Scale bars= 20µm. Graph represents mean±SEM. **B**. Electrophysiological recordings after CNO treatment during the IGT. From left to right: AMPA/NMDA ratio ($t_{(10)}$=0.93, n.s.), resting membrane potential ($t_{(10)}$=0.2, n.s.), and rheobase ($t_{(10)}$=0.96, n.s.) did not differ between control and hM3Dq-expressing mice after CNO treatment, but hM3Dq-expressing neurons had an increased intrinsic excitability compared to controls (f-I curve, 2-way ANOVA $F_{(14,140)}$ = 13, p<0.0001). **C.** Difference in all preferences for the four options between CNO and baseline, for mCherry controls and hM3Dq mice. **D**. Preferences of « safe » mice were grouped for safe (advantageous) options P1+P2 compared to unsafe (P3+P4) options, showing that CNO activation of DMS D1-expressing neurons transfected with hM3Dq decreased the preference for the safe option (Repeated-Measures-ANOVA, $F_{(1,38)}$=5.83, p = 0.02; CNO effect for mCherry group: n.s.; for Gq group: p=0.017). **E.** CNO treatment also decreased the number of trials (Repeated-Measures-ANOVA, $F_{(1,38)}$=100, p = 3.10^{-12}; CNO effect for mCherry group: n.s.; for Gq group: p=2.10^{-11}) and increased the % of omissions (Repeated-Measures-ANOVA, $F_{(1,38)}$=18.08, p = 1.10^{-4}; CNO effect for mCherry group: n.s.; for Gq group: p=2.10^{-6}), suggesting that the effect on choices reflect a wider cognitive effect. **F**. Left: data split by archetypes indicates that the decrease in safe choice was mainly due to an effect on Optimizers (2-way ANOVA: $F_{(1,34)}$=5.25, p = 0.01; R+E mice: n.s.; O mice: p=0.005). Right: ternary plot showing the global effect of CNO on strategies, with each point (black, mCherry controls, pink, hM3Dq animals) representing an animal on CNO treatment, with a connected line showing the displacement in the ACP space from baseline to CNO. The larger arrows show group averages (black, mCherry controls, pink, hM3Dq animals), with hM3Dq animals moving away from the R apex under CNO (Repeated-Measures-ANOVA, $F_{(1,38)}$=5.16, p = 0.03; CNO effect for mCherry group: n.s.; for Gq group: p=0.006). **G**. Data fit by the choice model suggest that the decreased choice of the safe option reflects an increase in risk-seeking upon CNO treatment (Repeated-Measures-ANOVA, $F_{(1,38)}$=4.84, p = 0.03; CNO effect for mCherry group: n.s.; for Gq group: p=0.01).

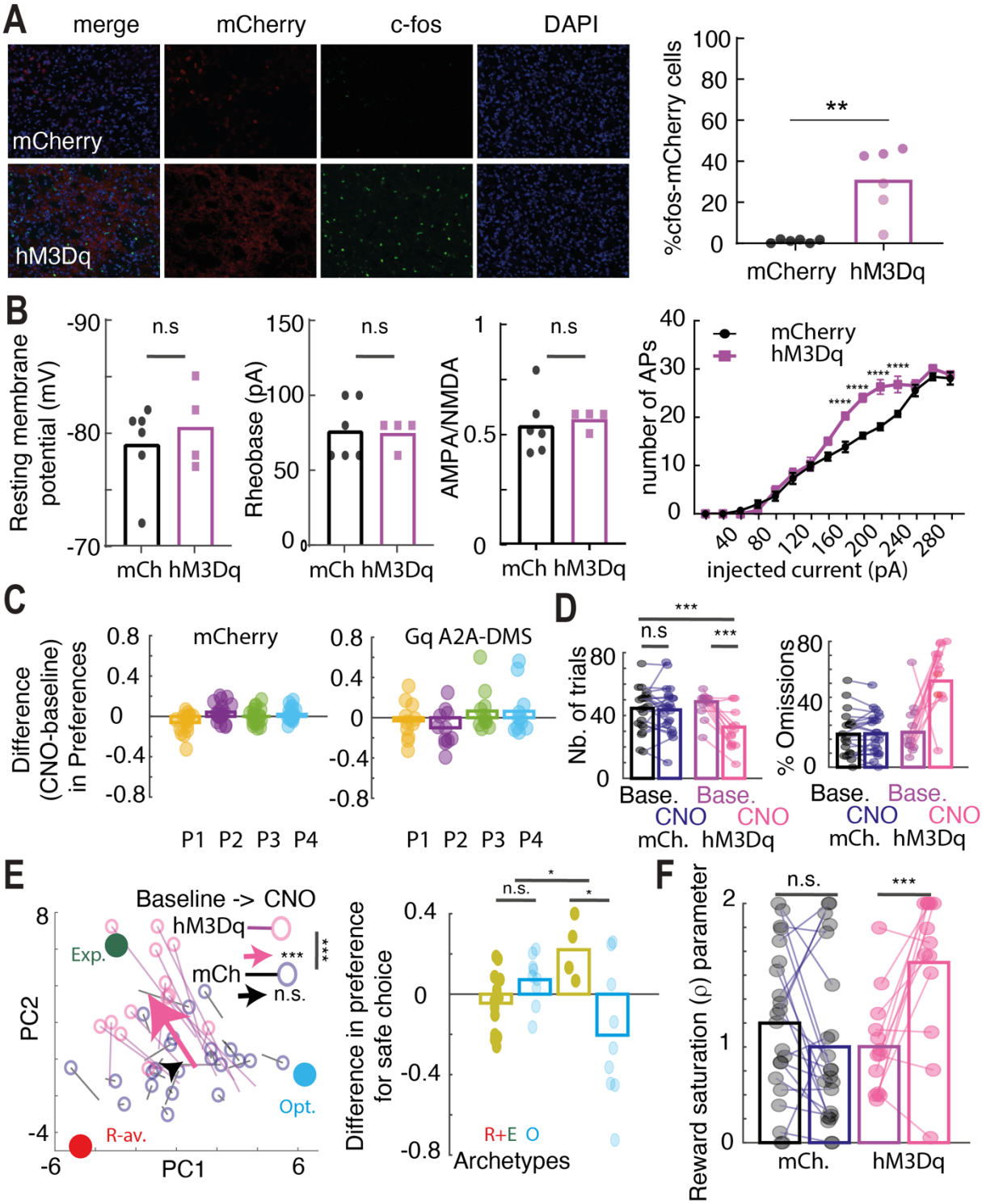


Figure 4

Figure 4. DMS iSPNs favor large gain. **A.** Microscope fluorescence for control (up) and hM3Dq (bottom) individuals in A2A-Cre mice injected in the dorsomedial striatum (DMS): DAPI (blue) indicate nuclei, mCherry (red) transfected neurons and C-fos (green) CNO-activated neurons. White arrowheads point at hM3Dq-expressing neuronal somas. C-fos expression was significantly higher in the hM3Dq group comparing to controls (***p<0,0001, U(6)=0, Mann-Whitney). Scale bars= 20µm. Graph represents mean±SEM. **B**. Electrophysiological recordings after CNO treatment during the IGT. From left to right:

resting membrane potential ($t_{(8)}$=0.39, n.s.), rheobase ($t_{(8)}$=0.63, n.s.) and AMPA/NMDA ratio ($t_{(8)}$=0.15, n.s.) did not differ between control and hM3Dq-expressing mice after CNO treatment, but hM3Dq-expressing neurons had an increased intrinsic excitability compared to controls (f-I curve, 2-way ANOVA $F_{(14,140)}$ = 13). **C.** Difference in all preferences for the four options between CNO and baseline, for mCherry controls and hM3Dq mice. **D.** CNO treatment also decreased the % of trials (Repeated-Measures-ANOVA, $F_{(1,33)}$=25.6, p = 8.10^{-7}; CNO effect for mCherry group: n.s.; for Gq group: p=1.10^{-4}) and increased the % of omissions (Repeated-Measures-ANOVA, $F_{(1,33)}$=25.08, p = 2.10^{-5}; CNO effect for mCherry group: n.s.; for Gq group: p=0.001), suggesting that the effect on choices reflect a wider cognitive effect. **E**. Left: ternary plot showing the global effect of CNO on strategies, with each point (black, mCherry controls, pink, hM3Dq animals) representing an animal on CNO treatment, with a connected line showing the displacement in the ACP space from baseline to CNO. The larger arrows show group averages (black, mCherry controls, pink, hM3Dq animals) with hM3Dq moving towards the E apex under CNO (Repeated-Measures-ANOVA, $F_{(1,33)}$=37.69, p = 6.10^{-7}; CNO effect for mCherry group: n.s.; for Gq group: p=2.10^{-4}). Right: data split by archetypes indicates that opposite effect of DMS A2A-expressing neurons on safe choice on Optimizers compared to Risk averse and Explorers mice (). **G**. Data fit by the choice model suggest that the decreased choice of the safe option reflects a decrease in reward saturation upon CNO treatment (Repeated-Measures-ANOVA, $F_{(1,33)}$=14.23, p = 6.10^{-4}; CNO effect for mCherry group: n.s.; for Gq group: p=0.003).

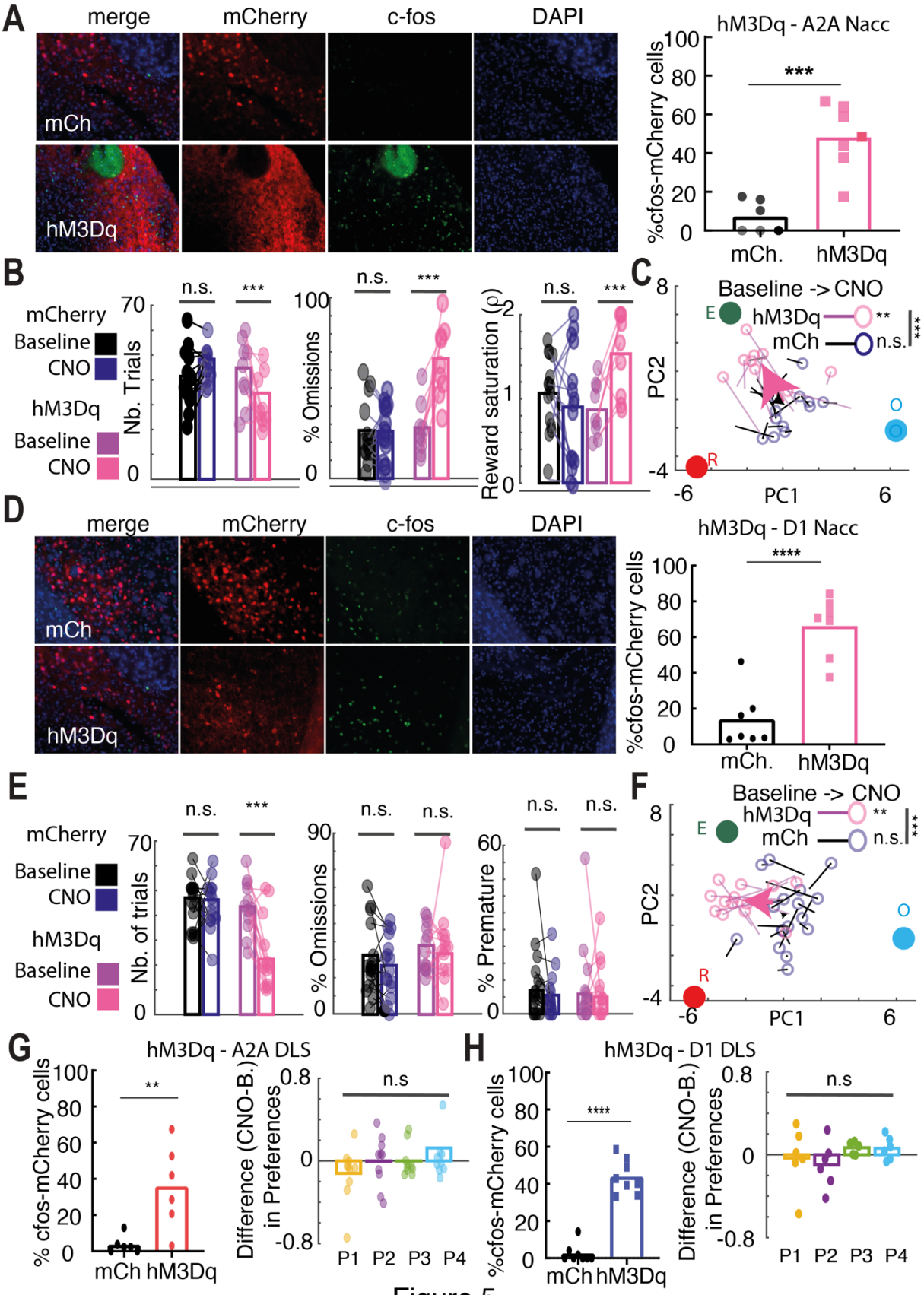
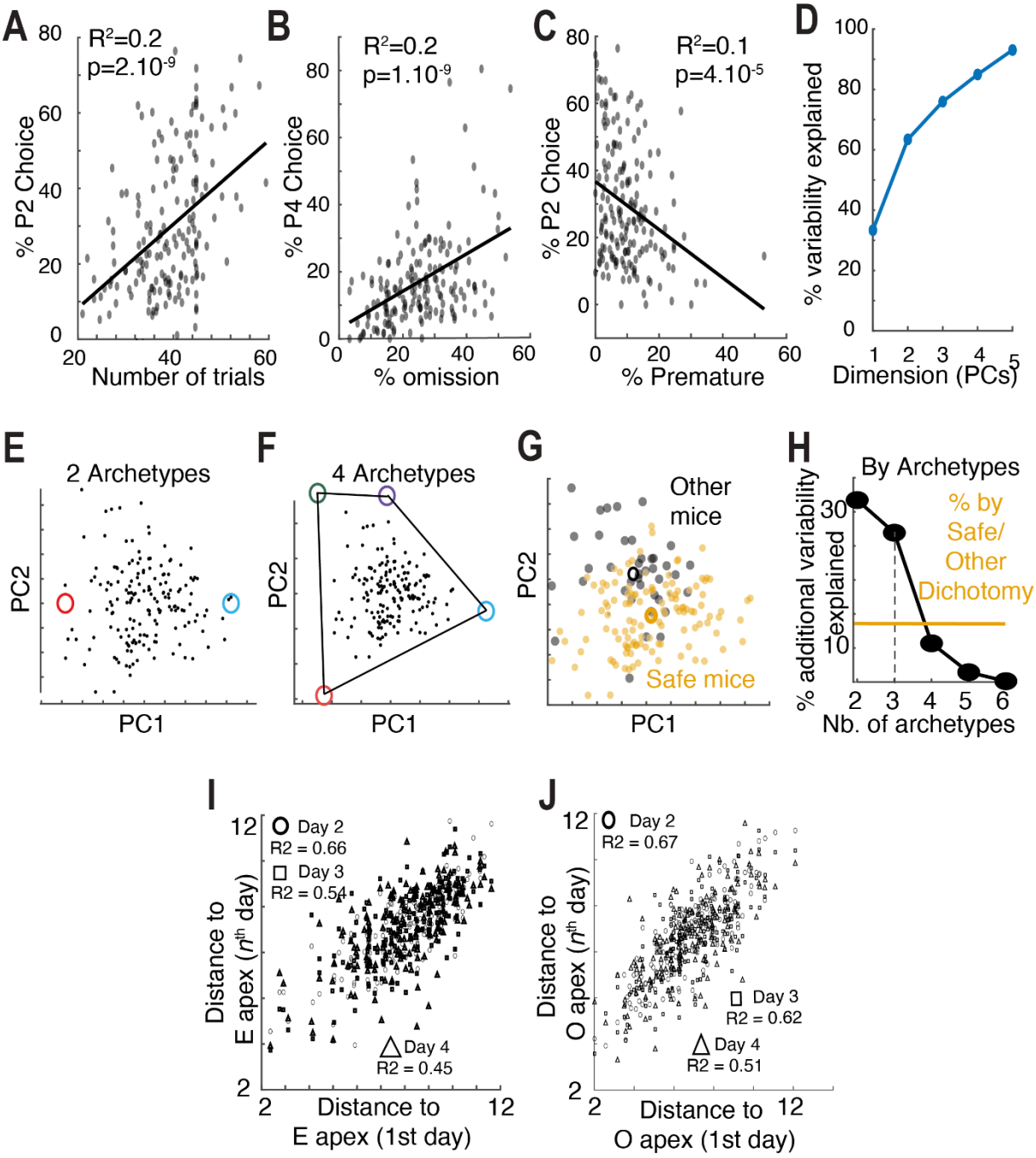

A
merge
mCherry
c-fos
DAPI
mCh
hM3Dq
hM3Dq - A2A Nacc
%cfos-mCherry cells
***
mCh.
hM3Dq
B
mCherry
Baseline
CNO
hM3Dq
Baseline
CNO
Nb. Trials
% Omissions
Reward saturation (ρ)
n.s.
***
C
Baseline -> CNO
hM3Dq
mCh
**
n.s.
***
PC2
PC1
E
O
R
D
hM3Dq - D1 Nacc
****
E
Nb. of trials
% Omissions
% Premature
F
G
hM3Dq - A2A DLS
% cfos-mCherry cells
Difference (CNO-B.) in Preferences
n.s
P1
P2
P3
P4
H
hM3Dq - D1 DLS

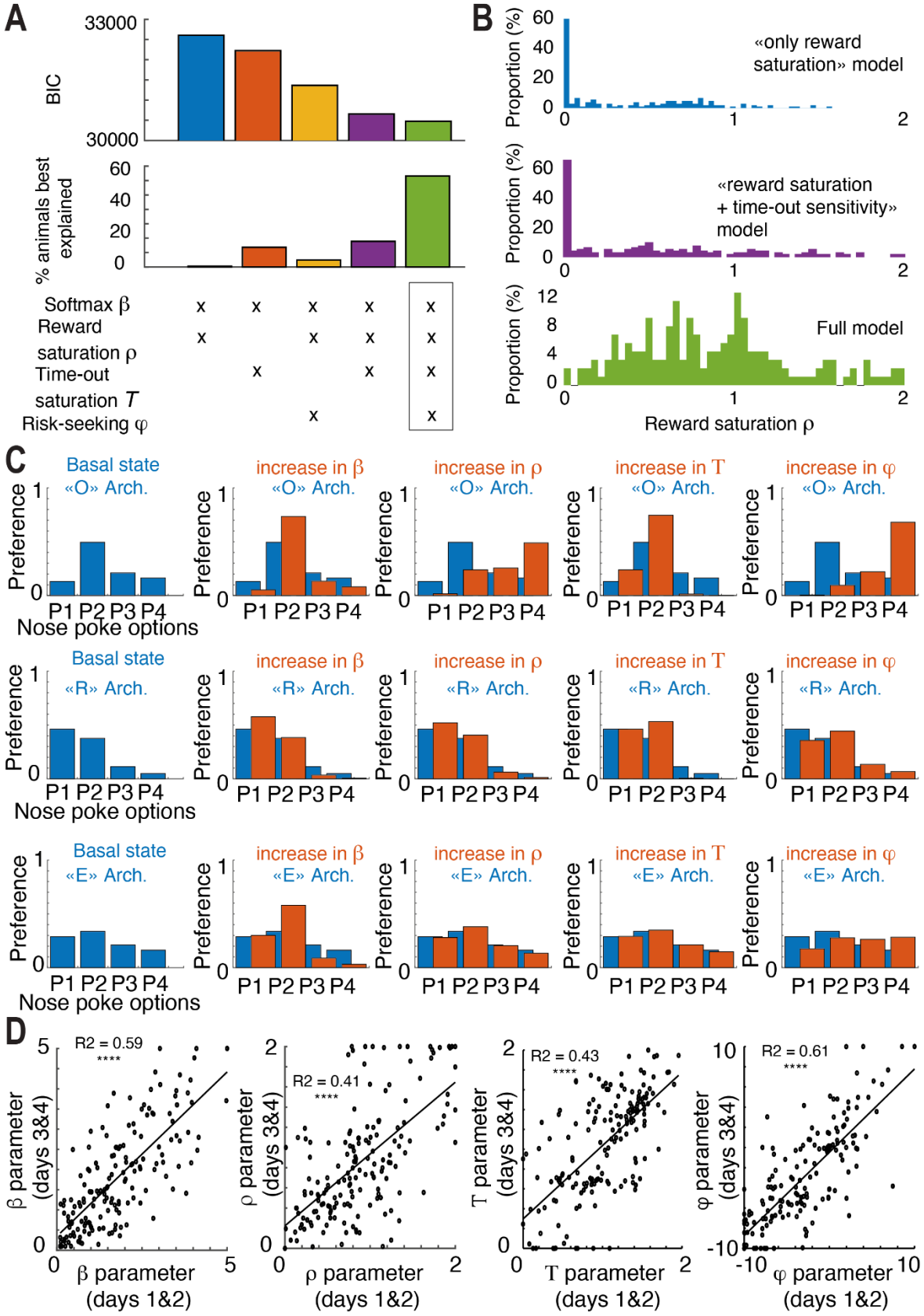
Figure 5

Figure 5. Lesser involvement of NAc and DLS i- and d-SPN in the mouse gambling task. **A.** Microscope fluorescence for control (up) and hM3Dq (bottom) individuals in A2A-Cre mice injected in the nucleus accumbens (Nacc): DAPI (blue) indicate nuclei, mCherry (red) transfected neurons and C-fos (green) CNO-activated neurons. White arrowheads point at hM3Dq-expressing neuronal somas. C-fos expression was significantly higher in the hM3Dq group comparing to controls (***p<0,0001, Unpaired t-test). Scale bars= 20µm. Graph represents mean±SEM. **B**. From left to right : CNO treatment decreased the number of trials initiated trials (Repeated-Measures-ANOVA, $F_{(1,22)}$=25.43, p = 5.10^{-5}; CNO effect for mCherry group: n.s.; for Gq group: p=0.001), increased the % of omissions (Repeated-Measures-ANOVA, $F_{(1,22)}$=25.6, p = 8.10^{-7}; CNO effect for mCherry group: n.s.; for Gq group: p=1.10^{-4}), and decreased reward saturation parameter from the model fit (Repeated-Measures-ANOVA, $F_{(1,22)}$=26.03, p = 0.03; CNO effect for mCherry group: n.s.; for Gq group: p=0.0076). **C**. Ternary plot showing the global effect of CNO on strategies, with each point (black, mCherry controls, pink, hM3Dq animals) representing an animal on CNO treatment, with a connected line showing the displacement in the ACP space from baseline to CNO. The larger arrows show group averages (black, mCherry controls, pink, hM3Dq animals) with hM3Dq moving towards the E apex under CNO (Repeated-Measures-ANOVA, $F_{(1,22)}$=33.22, p = 2.10^{-4}; CNO effect for mCherry group: n.s.; for Gq group: p=0.005). **D.** Microscope fluorescence for control (up) and hM3Dq (bottom) individuals in D1-Cre mice injected in the Nacc: DAPI (blue) indicate nuclei, mCherry (red) transfected neurons and C-fos (green) CNO-activated neurons. White arrowheads point at hM3Dq-expressing neuronal somas. C-fos expression was significantly higher in the hM3Dq group comparing to controls (***p<0,0001, Unpaired t-test). Scale bars= 20µm. Graph represents mean±SEM. **E**. CNO treatment only decreased the number of trials (Repeated-Measures-ANOVA, $F_{(1,28)}$=16.68, p = 3.10^{-4}; CNO effect for mCherry group: n.s.; for Gq group: p=4.10^{-4}), but neither the % of premature responses nor the % of omissions. **F**. ternary plot showing the global effect of CNO on strategies, with each point (black, mCherry controls, pink, hM3Dq animals) representing an animal on CNO treatment, with a connected line showing the displacement in the ACP space from baseline to CNO. The larger arrows show group averages (black, mCherry controls, pink, hM3Dq animals) with hM3Dq moving away from the O apex under CNO Repeated-Measures-ANOVA, $F_{(1,28)}$=13.53, p = 1.10^{-3}; CNO effect for mCherry group: n.s.; for Gq group: p=0.03). **G**. In DLS A2A-expressing neurons, c-fos expression was significantly higher in the hM3Dq group comparing to controls (***p<0,0001, Unpaired t-test). There was no difference in all preferences for the four options between CNO and baseline, for mCherry controls and hM3Dq mice. **H**. In DLS D1-expressing neurons, c-fos expression was significantly higher in the hM3Dq group comparing to controls (***p<0,0001, Unpaired t-test). There was no difference in all preferences for the four options between CNO and baseline, for mCherry controls and hM3Dq mice.

SUPPLEMENTARY FIGURES AND LEGENDS

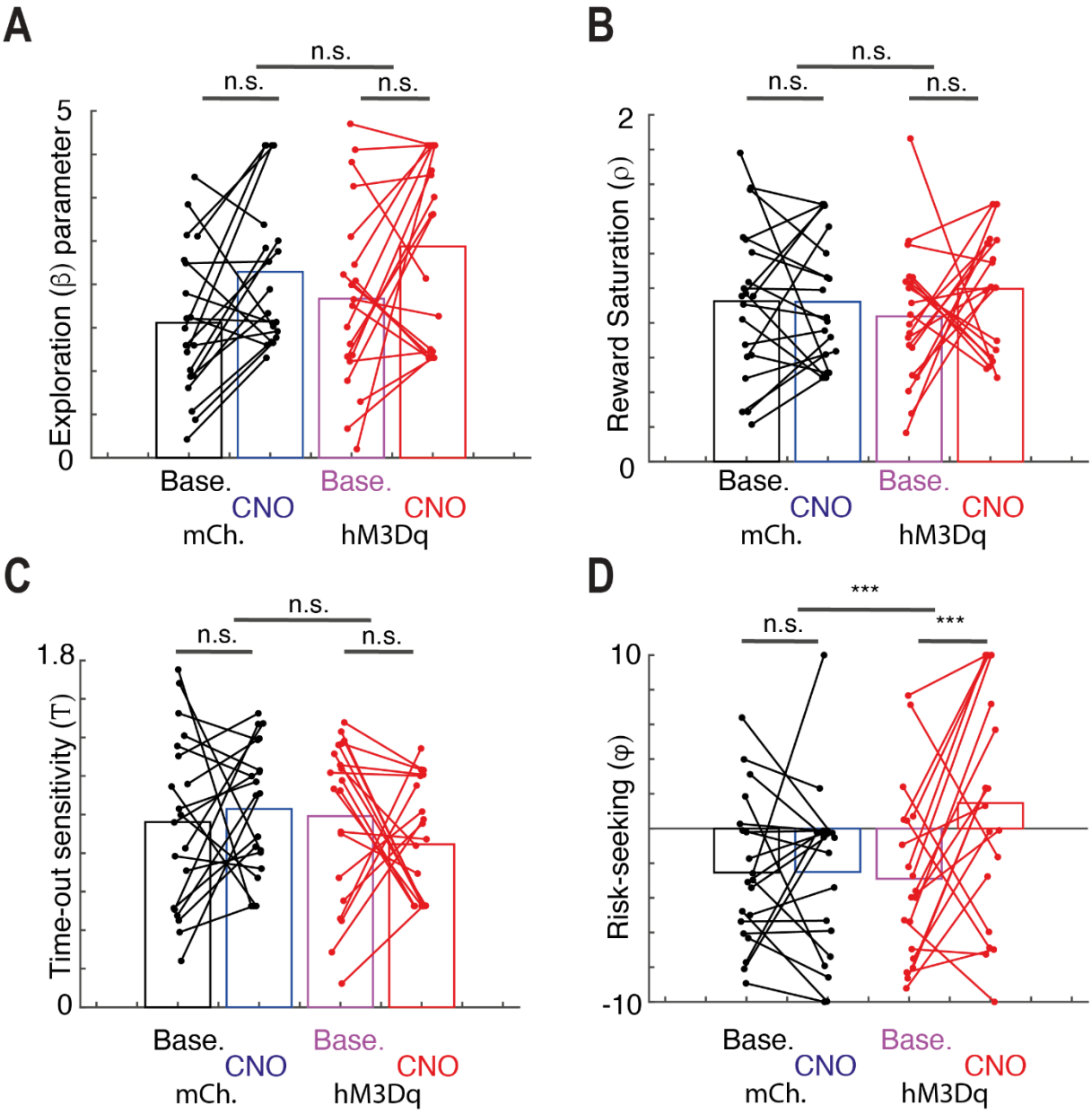


Supplementary Figure 1. Additional analyses related to Archetypes. (**A,B,C**). Other examples of correlations between behavioral measures: (**A**) preference for the P2 choice was positively correlated with the number of trials ($R^2 = 0.2$, $p=2.10^{-9}$), (**B**) Preference for the P4 choice was positively 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**) Projection in the PC (1st and 2nd) space of mice according to the safe/not safe classical dichotomy. (**H**) Comparison of the % of explained variance by archetypes (black, same as Fig. 2C) and by safe/not safe dichotomy (yellow). (**I,J**) Correlation over days between distances (same as Fig. 2D) to the E (**I**) and O (**J**) apices.

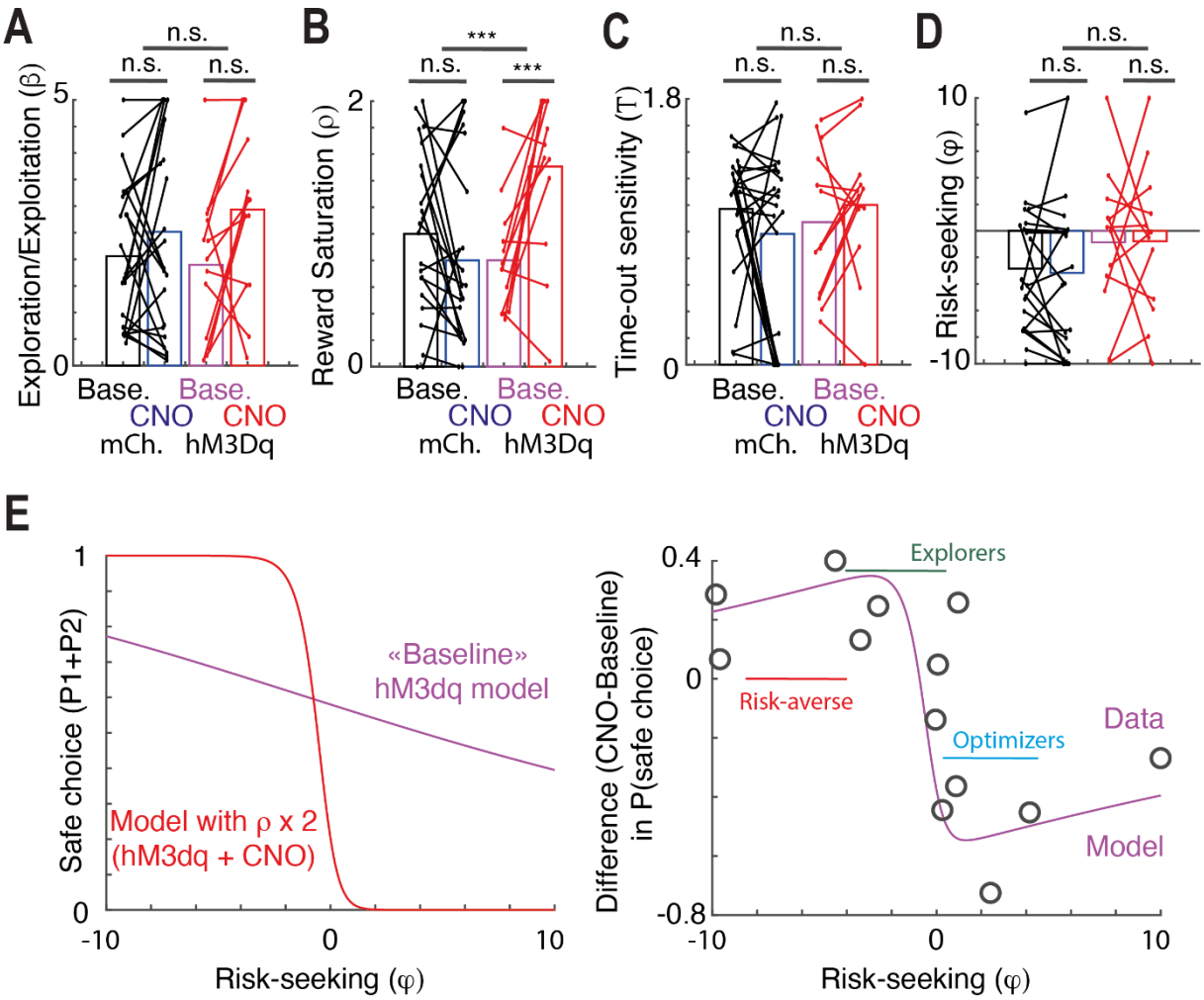


Supplementary Figure 2. Additional analyses related to the Reinforcement Learning (RL) model. **A**. Bayesian Information Criterion (top, model likelihood penalized for model complexity) and proportion of animals best explained (middle) for different versions of the

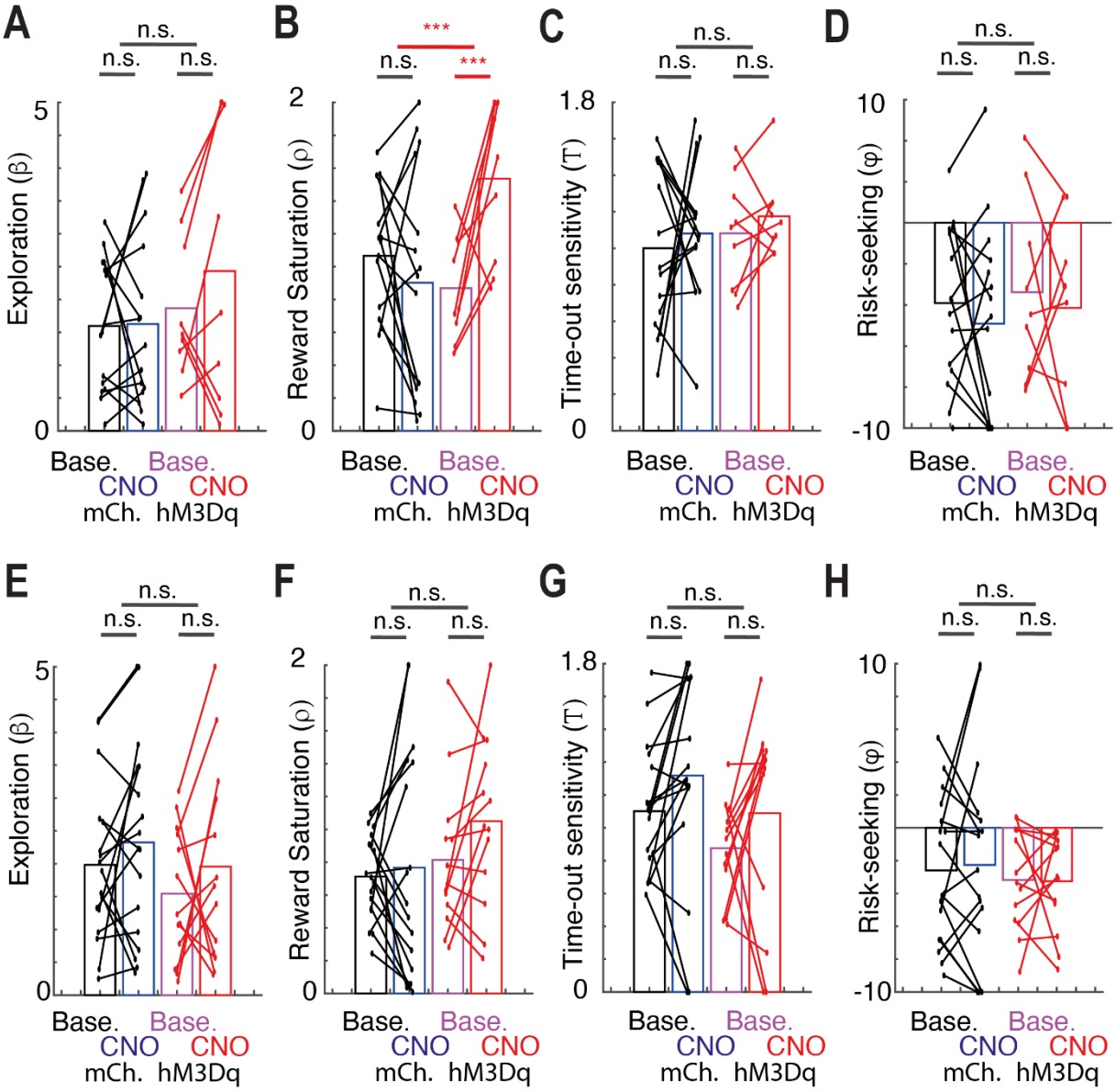
Reinforcement-Learning model (bottom), with the 4-parameters model showing smaller (i.e. better) BIC. Middle. **B**. Distribution of fitted values for reward saturation parameter in different models (increasing complexity from top to bottom) showing that simpler versions of the model fit the data with extreme, implausible values for the parameters (i.e. reward saturation is expected to lie to close to 1 or slightly below 1). **C.** Effects of parameter changes (columns: first column displays the basal model, second to fifth columns show the effects of changes in β, ρ, T, and φ) on the distribution of preferences for the 4 options (P1-P4) depending on the archetype (rows: O for Optimizers, R for Risk-averse, E for explorers). **D**. Correlations between model parameters fitted on consecutive days (from left to right: β, ρ, T, and φ)).



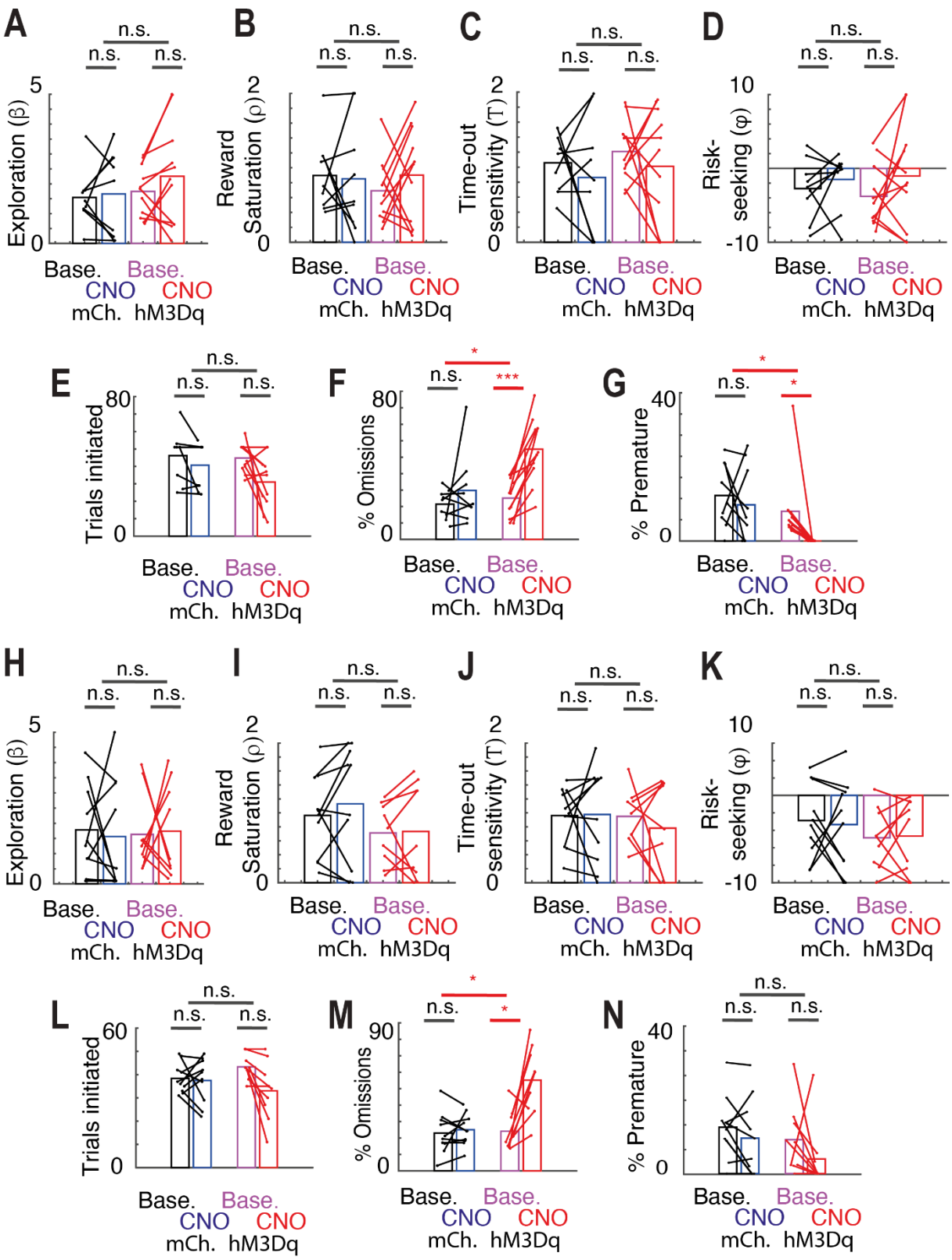
Supplementary Figure 3. Model fit for D1-DMS animals. (**A,B,C,D**) correspond to (explore/exploit ; reward saturation; time-out sensitivity; risk-seeking) parameters for DREADD (hM3Dq) animals and mCherry (mCh) controls, under baseline and CNO conditions. Only the risk-seeking was significantly different in the (hM3Dq, CNO) condition.



Supplementary Figure 4 Model fit for A2A-DMS animals. (**A,B,C,D**) correspond to (explore/exploit ; reward saturation; time-out sensitivity; risk-seeking) parameters 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**) Safe choices (P1 plus P2 preferences) as a function of φ in models with ρ=0.5 (”baseline”) and ρ=2 (”+CNO”) show that changes in safe choices induced by an increase in ρ depend non-linearly on φ . (**F**) Difference in safe choices as a function of φ predicted from an average model (black curve) and difference in safe choices as a function of φ in A2A-DMS animals (actual difference in safe choices plotted against animal fitted in the baseline condition).



Supplementary Figure 5. Model fit for Nac animals. (**A,B,C,D**) correspond to (explore/exploit ; reward saturation; time-out sensitivity; risk-seeking) parameters for A2A-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. (**E,F,G,H**) same parameters for D1-NAc - DREADD (hM3Dq) animals and mCherry (mCh) controls, under baseline and CNO conditions, with no significant effect of CNO.



Supplementary Figure 6. Model fit and behavioral measures for DLS animals. (**A to G**) correspond to (**A** explore/exploit ; **B** reward saturation; **C** time-out sensitivity; **D** risk-seeking)

parameters; number of trials (**E**), % of omissions (**F**), and % of premature responses (**G**) 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}) and % of premature responses ($F_{(1,18)}$=6.89, p = 0.02; $T_{(10)}$ = 2.76, p = 0.02) in A2A-DLS-hM3Dq animals. (**H to N**) 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).

METHODS

Animal care and use

All procedures were performed according to the Institutional Animal Care Committee guidelines and were approved by the Local Ethical Committee (Comité d'Ethique et de Bien-Être Animal du pôle santé de l'Université Libre de Bruxelles (ULB), Ref. No. 646 N). Mice were maintained on a 12-hour dark/light cycle (lights on at 8 pm) with *ad libitum* access to food and water. The room temperature was set to 22 ± 2 °C with constant humidity (40–60%). The behavioral tests were performed during the dark photoperiod. Both male and female transgenic mice (≥ 8 weeks) were used in all behavioral experiments.

Transgenic mouse generation

The genetic background of all transgenic mice used in this paper is C57Bl/6J. The mice were heterozygous and maintained by breeding with C57Bl/6 mice. Three transgenic mouse lines were used: A_{2A}-Cre[22], D_1-Cre (EY262; GENSAT)[86]. Simple transgenic A_{2A}-Cre or D_1-Cre mice were used for the virally mediated targeting of iSPNs or dSPNs, respectively.

Viral injections

Under Avertin anesthesia (2,2,2-tribromoethaol 1.25%, 2-methyl-2-butanol 0.78%; 20 µL/g, i.p.; Sigma Aldrich) male A_{2A}-Cre and D_1-Cre mice (≥ 8 weeks old), which allowed us to target A2A- and D1- expressing neurons, respectively, received two injections (at 100 nL/min) under stereotaxic control in DLS (AP +0,8 L ±2,42 DV -3,2) or DMS (AP +1,2 L ±1,33 DV -3,2) or the NAc (AP +1,95 L ±1,2 DV -4,85). of a Cre-dependent virus encoding hM3Dq (AAV5-hSyn1-DIO- hM3Dq-mCherry, Addgene) or mCherry alone as control (AAV5-hSyn1-DIO-mCherry , Addgene). Coordinates were relative to Bregma according to the Franklin and Paxinos atlas third edition. Volumes of injection were: DLS 0,4 µl, DMS 0,35 µl, Nac 0,2 µl, and these were delivered through a cannula connected to a Hamilton syringe (10 µL) placed in a syringe pump (KDS-310-PLUS, KDScientific). Cannulas were lowered into the brain and left in place for 10 min after infusion. A minimum 3 weeks elapsed between the stereotaxic injection to ensure optimal protein expression levels.

Mouse Gambling task

Behavioral testing was performed in 6 identical operant chambers (IMETRONIC, Pessac, France). Four circular holes were available in a curved wall (the central one was obstructed, only 4 holes were actually available), with LEDs at their rear and a food dispenser of 20 mg chocolate pellets (Purified Dustless Precision Pellets for rodents, Bioserv, New Jersey, USA) on the opposite wall. Infrared beams in the nose-poke holes and the magazine allowed the detection of entries to them. Additionally, a house light located in the ceiling of the chamber was used under specific conditions. These chambers were in individual soundproof closets, to avoid any external disturbance during training. Video cameras placed on the top of the chambers allowed tracking of every session. The chambers are connected to a computer and were controlled by the POLY software (IMETRONIC).

We used a rodent version of the Iowa Gambling Task (IGT) (or Rodent IGT, RGT) adapted from Young and colleagues (2011)[49]. Mice have to choose among 4 different options associated to a certain gain and a certain loss (Figure 1). To achieve that goal, mice must undergo several phases of training. Mice were placed on a reduced diet two days before training, aiming to reach 85-90% of their original weight to enhance performance, given that food motivation drives the IGT. Initially, they received 2.5 grams of food per mouse, adjusted the next day based on individual weight loss. During IGT training, food amounts remained stable but could be adjusted if mice fell out of the 85-90% weight range. Once the final training phase commenced, food quantities were fixed to avoid performance interference. The whole protocol lasted between 1 month and a half and 3 months, depending on the animal. Mice were trained daily at 9 am in reverse cycle.

*Magazine habituation***.** Mice were trained for 10 minutes per day during 3 consecutive days. During these sessions every light was off. 15 seconds after the start of the session, the pellet dispenser was activated to distribute a single pellet. The LED of the magazine was then lit. The light goes out when mice get the pellet, and the 15 seconds cycle would start again. The magazine would stay lit until retrieval of the pellet to allow association between the lit magazine and food.

Nose-poke habituation. 4 nose-pokes were available, and a lit nose-poke meant that it was an active nose-poke. This training is divided in two phases. In the initial stage lasting 4 days, they responded to one active nose-poke out of 4, chosen randomly in two daily sessions. If successful, 3 pellets were dispensed; otherwise, sessions could last 10 minutes. The first 2 days, pellets were placed in active nose-pokes to encourage exploration. In the second phase, all 4 nose-pokes remained active throughout the 30-minute sessions, delivering a single pellet upon visitation. Training persisted until mice achieved 40 responses in two consecutive days, with individual progress varying. Upon reaching this criterion, mice progressed to the next training phase.

*Forced-choice IGT***.** Different nose-pokes had specific rewards and punishments. Mice underwent 3 days of forced 30-minute sessions, where a random nose-poke was activated, leading to either a reward or punishment afterward.

Final IGT. Mice underwent 30-minute sessions with up to 100 trials. Sessions began with a lit magazine that mice had to approach. A 5-second interval preceded illumination of 4 nose-poke lights; premature responses triggered a time-out. After the interval, mice had 10 seconds to nose-poke, and omissions led to new trials. Correct responses extinguished lights and rewarded mice based on specific nose-pokes probabilities (P1, P2, P3, P4). Rewards involved pellet delivery, followed by an 8-second period before the next trial. Punishments initiated a time-out with flashing lights, after which mice could start a new trial. Perseverative responses were noted but not punished. Mice were trained until their preference stabilized over four consecutive days, establishing a baseline.

CNO treatment.

When mice reached a stable pattern of preference over 4 days of IGT (see below), we started a treatment with 0,9% NaCl for 2 days. All injections were applied 30 minutes before the start of the session. After receiving the CNO injection (1mg/kg), mice performed their last session of IGT. Immediately after the end of this session, mice were euthanized. A part of the group was

used to perform electrophysiological recordings. In this group there was a balanced representation of hM3Dq-expressing mice and control mice.

Immunohistochemistry

After the behavioral experiments were completed, mice were deeply anesthetized with avertin (2,2,2-tribromoethaol 1.25%, 2-methyl-2-butanol 0.78%; 20 µL/g, i.p.; Sigma Aldrich) and transcardially perfused with PBS followed by 4% paraformaldehyde in PBS. Brains were removed and postfixed overnight at 4°C. Then, 30-µm coronal slices containing the striatum were cut with a vibratome (VT1000 S; Leica) and stored in PBS. Sections are incubated overnight in a dilution (1/2000 in 1%NHS 0.1%PBST) of a rabbit C-fos primary antibody (SantaCruz). Next, slices were washed twice for 5 minutes with PBST and incubated 1,5 h in a dilution (1/200 in 1%NHS 0.1%PBST) of donkey anti-rabbit secondary antibody Alexa 647 (far red) (Jackson ImmunoResearch). The slices were then washed twice for 5 minutes with PBST. Finally, a DAPI nuclear staining, is performed (10 minutes in a 1/5000 DAPI solution in 0,01M PBS), washed 5min in 0.01M and mounted on glass slides and cover slipped with Fluoromount.

For GFP immunostaining, the protocol is the same with chicken anti-GFP primary antibody (1/2000, SantaCruz) and a goat anti-chicken secondary antibody Alexa 647 (1/400, Jackson ImmunoResearch).

Image acquisition was done with the Zeiss Axioimager Z1 on 20X magnification and processed with the AxioVision (Zeiss) software. Cell counting was made with the opensource software FIJI. Additionally, we also used the Axio Zoom.V16 (Zeiss) and its tiling tool, to take whole slice surface images.

Acute brain slice preparation. Recordings were made ex vivo, on brain slices that were kept alive. Animals were anesthetized with halothane and then euthanized by decapitation. Their brains were quickly collected, stuck to a vibratome plate and submerged in the vibratome container filled with cutting solution at 4°C. Brain slices were cut 220 µm thick and then transferred to an incubator chamber filled with artificial cerebrospinal fluid (aCSF) (table MM-1) at 34°C, for at least 60 minutes before recordings started. This resting period is required for recovery of proper neuronal metabolic activity. Both cutting solution and aCSF were constantly oxygenized, as they were supplied with a 95% O2 and 5% CO2 mixture to avoid hypoxia. After recovery, brain slices were transferred one by one to the recording chamber, which was constantly perfused with oxygenized aCSF at room temperature. The set-up was equipped with a camera and a Zeiss upright microscope (Axioskop 2FS Plus, Zeiss), that allowed first, to identify the different areas in the striatum in the slices (5x/0,15 EC PlanNEOFLUAR objective lens, Zeiss) and secondly, cells (63x water-immersion objective, Zeiss). Depending on the group of mice tested, we recorded either in the DLS, DMS or Nacc (according to the injection site of the group).

Recordings. The patch pipette was obtained with a vertical two-stage puller (PIP 5, HEKA), and during the recordings, it was filled with intracellular solution (table MM-1). The Ag/AgCl electrode in the pipette was connected to an EPC-10 (Heka) amplifier, whose signal was recorded with the Patchmaster software (Heka). Intrinsic excitability was studied using current

clamp. First, the resting membrane potential was measured (without applying any current). Afterwards, a negative amount of current was injected into the cell to maintain its membrane potential at a value of −80mV. Increasing intensity currents (increase of 10 pA) were then injected for depolarising the cell, and triggering action potentials. Rheobase is defined as the minimum amount of injected current that allows the induction of an action potential. Evoked Excitatory Post-Synaptic Current (EPSCs) were recorded with the internal solution supplemented with spermine (0,1 mM). Spermine blocks calcium-permeable AMPA-R at positive potentials, and thus, allowed us to distinguish calcium-permeable from non-calcium permeable AMPA-R currents[87]. Evoked EPSCs were isolated from GABAergic currents by application of gabazine (or SR-95531, 10 µm). We don't make any reference to kainate-R because of the difficulties to distinguish them from AMPA-R, as they cannot be easily separated pharmacologically[88]. Thus, from now on, we will use the shortcut AMPA-R currents to refer to EPSCs resulting from both AMPA-R and kainate-R. 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 the specific stimulation of cortico-striatal pathway. The duration of stimulation current pulses was constant and set at 0.2ms. 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). Slices were stimulated every 10 seconds. Evoked EPSCs were recorded successively at −70mV, 0mV and +40mV 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[89]. AMPA-R mediated EPSCs were obtained at a holding potential of −70mV where no NMDA-R conductance is observed, due to the magnesium block. A sum of AMPA-R and NMDA-R mediated EPSCs is evoked at a holding potential of +40mV. In MSNs, AMPA-R mediated currents were pharmacologically isolated by application of L689,560 (10µM) to the solution, a non-competitive NMDA-R antagonist. The NMDA-R current is obtained by the off-line subtraction of the two traces. eEPSCs were digitized at a frequency of 10 kHz and filtered on-line with a low-pass Bessel filter, using a cut-off frequency set at 3 kHz. 10 eEPSC traces were averaged for each condition.

Statistical analyses

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

Archetypal analysis. Computations were done using the ParTI routine in Matlab (ref Uri Halon). Briefly, given an n × 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 % of premature responses and the % of omissions), the archetypal analysis finds the matrix Z of k multidimensional archetypes (k is the number of archetypes). Z is obtained by minimizing || X-α ZT | | 2, with α the coefficients of the archetypes (αi,1..k ≥ 0 and ∑ αi,1..k = 1), and | |.||2 a matrix norm. The archetype is also a convex combination of the data points Z = XTδ, with δ ≥ 0 and their sum must be 169. The α-coefficient depicts the relative archetypal composition of a given observation. For k = 3 archetypes and an observation i, αi,1; αi,2; αi,3 ≥ 0 and αi,1 + αi,2 + αi,3 = 1. A ternary plot can then be used to visualize data (αi,1; αi,2; αi,2) are used to assign individual behavior to its nearest archetype (i.e., k max(αi,1; αi,2; αi,3)). αi,j are also used as variable to estimate population archetypal composition. Pure archetype corresponds to 1, the archetypal composition decreases linearly with increasing distance from the archetype, 0 correspond to points on the opposite side.

Reinforcement-learning model

Decision model. Decision-making models determined 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(sum(V_a)/(1 + exp(−ß(sum(V_i))), where β is an inverse temperature parameter reflecting the sensitivity of choice to the difference of decision variables (values) V_i. Large β corresponds to exploitation, i.e. choosing the option that seems the best so far, while small β corresponds to exploration (at β=0 all options are chosen equally).

Decision variable or value *V* of an option is modeled as *V* = *R* + *P* + *U*; i.e. 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 a model parameter how the subjective value of the option depends on the reward magnitude. ρ value close to 1 corresponds to linear subjective value (the subjective value equals the reward size). 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 the opportunity cost of time, i.e. the average reward to be lost if there is a time out: *P* = (1-*p)**(TimeOut duration)^(*T*)* *R;* where *T* scales the perception of the time-out duration. *T* value close to 1 corresponds to linear perception of time, and γ smaller (respectively, larger) than 1 corresponds to underweighting (respectively, overweighting) long durations. Risk is given by φ*sqrt(*p**(1-*p*) .* (*R*-*P*).^(2)) corresponding to the variance in reward, the value of which is scaled by the free parameter φ (negative φ corresponds to risk aversion and positive φ 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 choice c = c1..T, data likelihood is the product of their probability given by the softmax decision rule. We used the fmincon function in Matlab to perform the fits, with the constraints that β ∈]0,10], ϕ ∈ [-10,10] and *Rs* ∈]0,2] and *Rp* ∈]0,2].