Astrocytes in the external globus pallidus coordinate flexibility of action strategy

Sa-Ik Hong
Mayo Clinic College of Medicine and Science
Seungwoo Kang
Mayo Clinic College of Medicine and Science
Minryung Song
Korea Advanced Institute of Science and Technology
Minsu Yang
Korea Advanced Institute of Science and Technology https://orcid.org/0000-0001-7685-1280
Matthew Baker
Mayo Clinic
Shinwoo Kang
Mayo Clinic College of Medicine and Science
Jeyeon Lee
Mayo Clinic
Sang Wan Lee
Korea Advanced Institute of Science and Technology
Doo-Sup Choi (Choids@mayo.edu)
Mayo Clinic College of Medicine and Science https://orcid.org/0000-0002-6796-9938

Article

Keywords: external globus pallidus, astrocyte, goal-directed behavior, habit, reward-seeking behavior

Posted Date: September 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-845991/v1

Abstract

The external globus pallidus (GPe) is an integrative hub and gateway for behavioral flexibility in reward-related behaviors. However, it remains unknown whether enriched astrocytes in the GPe guide behavioral flexibility. Here, we trained mice to exhibit goal-directed and habitual reward-seeking behaviors using the behavior tasks with effort- and time-based reward delivery, respectively. Then, we examined the temporal dynamics of GPe astrocytes during goal-directed and habitual learning. Overall, GPe astrocytes were substantially silenced during habitual learning compared to goal-directed learning. In the timescale of action events, GPe astrocyte activities were increased immediately after termination of reward-taking behavior before the following action. However, during habitual learning, the increase of astrocyte activity was not evident. Moreover, support vector machine (SVM) analysis demonstrated that GPe astrocytes dynamics predicted whether mice perform goal-directed or habitual behaviors. Interestingly, chemogenetic activation of GPe astrocytes, which dampened GPe neuronal firings and habitual behaviors, exhibting goal-directed behaviors. Strikingly, brief and repeated attentional stimulations recapitulated the effect of chemogenetic activation of GPe in intervening the habitual reward-seeking behaviors with increased GPe astrocyte activities. Our findings reveal a novel insight that increasing GPe astrocytic activities attenuates habitual behavior and improves behavioral flexibility, which may provide a potential therapeutic target for decision-making-related disorders, such as obsessive-compulsive disorder and addiction.

Introduction

Upon repetition of goal-directed behavior, brain circuits routinize the information processing as a habit, which is influenced less by executive functions and cost-effective in procuring the desired outcome serving an energy-saving autopilot system [1]. The ability to coordinate goal-directed and habitual controls can determine behavioral flexibility. However, a maladaptive habit with recurring and detrimental outcomes often develops with decision-making disorders, such as obsessive-compulsive disorder (OCD), substance and alcohol use disorders [2, 3].

In the striatopallidal circuits, the external globus pallidus (GPe) has been considered an integrative hub for the behavioral flexibility in reward-related behaviors because it coordinates the top-down neurotransmission from the two distinctive dorsal striatum regions, dorsomedial and dorsolateral striatum (DMS and DLS), the central neural substrates for goal-directed and habitual reward-seeking behaviors, respectively [4–7]. Indeed, the GPe redistributes the dorsal striatal inhibitory GABAergic tone and signals through the fast-spiking prototypical and slow-spiking arkypallidal neurons in the GPe to the downstream brain regions [8–10]. Interestingly, recent studies demonstrate that the GPe harbors abundant astrocytes, which is distinguishable to the neighboring brain regions, and its astrocytes directly modulate the neuronal activities in the GPe, leading to the consequent behavioral changes [11, 12]. However, the role of GPe astrocytes in goal-directed and habitual action strategies remains to be determined.

We utilized the two operant behavior tasks contingent upon effort- or time-based reward (10 μl of 20% sucrose solution) delivery to establish goal-directed and habitual action selections. Then, we employed *in vivo* calcium imaging to probe temporal dynamics of GPe astrocytes during goal-directed and habitual learnings. We validated whether GPe astrocyte dynamics contain information about the subject's behavior task type using the machine learning approach. Lastly, we determined whether the promotion of the calcium signaling in GPe astrocytes adapted by habitual learning attenuates the habitual reward-seeking behaviors.

Results

Learning with loose A-O contingency drives habitual seeking behavior. Goal-directed behaviors and habits can be quickly and efficiently developed by the random ratio (RR) task with effort-based reward delivery and the random interval (RI) task with time-based reward delivery [4]. We used sucrose solution as a reward since it has an intense rewarding effect in mice, especially in operant conditioning [13–15]. Upon nose-poke in an active hole, 10 μl of 20% sucrose solution was immediately delivered in the magazine through an automated syringe as we previously described [16–18].

Firstly, we trained mice to acquire action (nose-poke)-outcome (reward in the magazine) contingency in an operant chamber through 4 days of FR1 (fixed ratio 1) conditioning. The two randomly assigned groups showed similar nose-poking behaviors [**Supplementary Fig. 1a**, two-way RM ANOVA, nose-poke rate, $F(1,13) = 0.092$, $p = 0.767$; **Supplementary Fig. 1b**, two-way RM ANOVA, discrimination ratio, $F(1,13) = 0.030$, $p = 0.865$; **Supplementary Fig. 1c**, session duration, two-way RM ANOVA, $F(1,13) = 0.878$, $p = 0.366$; **Supplementary Fig. 1d**, latency to magazine, two-way RM ANOVA, $F(1,13) = 1.843$, $p = 0.198$; $n = 8$/RR, $n = 7$/RI groups] and the repeated FR1 tasks facilitated operant behaviors [**Supplementary Fig. 1d, 1**st vs. 4th session, two-way RM ANOVA, $F(1.923,25) = 33.84$, $p = 0.016$ for RR, $p = 0.002$ for RI, $n = 8$/RR, $n = 7$/RI groups].

After mice completed the FR1 conditioning, mice were subjected to the RR and RI schedules (Fig. 1a-c). At the last session of each RR or RI schedule, mice behaved with higher action-outcome (A-O) contingency during the RR task compared to the RI task (Fig. 1d, Mann-Whitney test, $p = 0.009$) and this RR task resulted in goal-directed behavior (Fig. 1e, Wilcoxon test, $p = 0.016$, $n = 8$), but the RI task developed habitual seeking behavior (Fig. 1e, two-tailed Wilcoxon test, $p = 0.672$, n = 7). Besides, A-O contingencies during the RR and RI tasks were negatively correlated with the habit index of the evaluation test (Fig. 1f, Spearman correlation coefficient $r = -0.600$, $p = 0.020$, $n = 15$). Thus, these results indicate that goal-directed or habitual learning may adjust an action strategy to procure the reward depending on the RR and RI schedules.

GPe astrocytes show distinct dynamics during learnings for goal-directed behavior and habit. The GPe is the integrative hub for goal-directed and habitual action selections and highly contains GFAP proteins (**Supplementary Fig. 2a**), indicating an abundance of astrocytes [11, 12]. Thus, we sought to characterize temporal astrocytic dynamics in the GPe during the goal-directed and habitual learnings. We used mice

with GFAP promotor-driven GCaMP6s expression. We recorded intracellular calcium imaging at the last session of the RR and RI tasks, where goal-directed behaviors and habits were assumably estimated to be sufficiently learned (**Supplementary Fig. 2b**). As expected, mice's operant performances highly represented the behavioral task rules at both task conditions (**Supplementary Fig. 2c-f, n** = 282/RR, *n* = 66/RI groups).

In operant conditioning, we analyzed a total of 6 behavior events: whether mice perform a nose-poke that gives a reward (rNP) or a nose-poke without a reward (nrNP), whether mice entered the magazine with the reward (M_{entry} R+) or without the reward (M_{entry} R-), and whether mice exit magazine after reward consumption (M_{exit} R+) or without reward experience (M_{exit} R-) (Fig. 2a). Then, we collected mice's behavior and fiber-photometry data from the pro-8 seconds and the post-8 seconds (total of 16 s, 480 frames) around the behavioral event. All data sets from one mouse were averaged and expressed as n = 1. Also, to analyze nose-poking behavior across time in the session, a bin was created with 2 seconds (60 frames) as one block (Fig. 2a).

Firstly, we confirmed that mice trained with the RR20 or RI120 task performed distinct reward-seeking and -taking behaviors at the last session of RR and RI schedules (**Supplementary Fig. 3a-f** top for reward-seeking behavior; **Supplementary Fig. 3a-f** bottom for reward-taking behavior; n = 6/group). Moreover, the RR task showed a higher seeking-taking association compared to the RI task in all the behavioral events except for magazine entry without reward (**Supplementary Fig. 3g**) as well as throughout the session (**Supplementary Fig. 3h**). Besides, the seeking-taking association showed positive correlation with A-O contingency (**Supplementary Fig. 3i**, Spearman correlation coefficient $r = 0.811$, $p = 0.002$, $n = 12$). This indicates that mice may seek the reward to acquire the reward in learning for goal-directed behavior, whereas mice may seek the reward regardless of the presence of the reward in habit learning.

Based on these distinct behavioral performances between the goal-directed behavior and habit learnings, we examined GPe astrocyte dynamics surrounding the tasks and behavioral events. Throughout the 6 behavior events, mice showed an increase in GPe astrocyte dynamics during the RR20 task compared to the RI120 task [Fig. 2b (left) for rewarded nose-poke, $F(1,10) = 8.565$, $p = 0.015$; Fig. 2b (right) for non-rewarded nose-poke, $F(1,10) = 9.136$, $p = 0.013$; Fig. 2e (left) for magazine entry with reward, $F(1,10) = 12.16$, $p = 0.006$; Fig. 2e (right) for magazine entry without reward, $F(1,10) = 29.21$, $p < 0.001$; Fig. 2h (left) for magazine exit with reward, $F(1,10) = 7.396$, $p = 0.022$; Fig. 2h (right) for magazine exit without reward, $F(1,10) = 24.18$, $p = 0.001$; two-way RM ANOVA, n = 6/group]. However, comparison of GPe astrocyte dynamics between before and after the rewarded (Fig. 2c, Wilcoxon test, $p = 0.156$ for RR, $p = 0.219$ for RI, $n = 6$/group) and non-rewarded nose-poking behaviors (Fig. 2c, Wilcoxon test, $p = 0.063$ for RR, $p = 0.313$ for RI, n = 6/group) showed non-significance. Interestingly, the changes in GPe astrocyte dynamics appeared at the initiation and termination of reward-taking behavior. GPe astrocyte dynamics were reduced when mice entered the magazine [Fig. 2f, Wilcoxon test, $p = 0.031$ for RR M_{entry} (R+), RR M_{entry} (R-), and RI M_{entry} (R-)]. GPe astrocytic dynamics were largely changed more when mice experienced reward [Fig. 2g, two-way ANOVA, $F(1,21) = 14.82$, $p = 0.001$, $n = 6$/group], whereas GPe astrocyte dynamics

were highly increased when mice exit from the magazine [Fig. 2i, Wilcoxon test, $p = 0.031$ for RR M_{exit} (R+), RI M_{exit} (R+), and RR M_{exit} (R-), $n = 6$/group; Fig. 2j, two-way ANOVA, $F(1,21) = 33.40$, $p < 0.001$, n = 6/group]. Moreover, the changes of GPe astrocyte dynamics with reward experience were greater than those without reward experience when mice showed magazine entry [Fig. 2g, two-way ANOVA, $F(1,21) = 14.82$, $p = 0.001$, $n = 6$/group] and exit [Fig. 2j, two-way ANOVA, $F(1,21) = 33.40$, $p < 0.001$; $n = 6$/group] but not for nose-poke [Fig. 2d, two-way ANOVA, $F(1,21) = 1.759$, $p = 0.199$; $n = 6$/group]. These results show that GPe astrocyte dynamics may be closely related to reward-taking behaviors and habit learning, with the reductions in A-O contingency and association between reward-seeking and -taking behaviors resulting in silenced GPe astrocyte dynamics.

Predictable goal-directed action selection by using GPe astrocyte dynamics. Next, we employed a support vector machine (SVM), an optimal neural network with minimal risk of overfitting [19], to identify whether the astrocytic activity in the GPe can predict which type of action selection mice performed. To accommodate the temporal dynamics of GPe astrocyte, as opposed to the average value of GPe astrocytic activities, we utilized all the trials of the RR and RI tasks. In addition, the time window was set to 1 second before and after the behavioral events to focus on task-driven changes in event-related GPe astrocyte dynamics. Overall, we were able to successfully differentiate temporal patterns of GPe astrocyte between the RR and RI tasks for all the behavioral events (Fig. 3a, **Supplementary Table 1**). Especially, GPe astrocyte activity patterns at the magazine exit highly predicted the types of learning tasks [Fig. 3b; *posthoc* Dunn's test after Kruskal-Wallis test; $p < 0.001$ for accuracy, sensitivity, and specificity (NP vs. M_{exit}, M_{entry} vs. M_{exit}); $n = 1600$/comparison]. These results suggest that GPe astrocyte activities are a good predictor of action selection underlying goal-directed learning.

Chemogenetic activation of GPe astrocytes reduced habitual seeking behaviors. Since we observed that habit learning might reduce the Ca^{2+} signals in the GPe astrocytes, parallelly with the reduction in A-O contingency, we sought to examine whether enhancing the astrocyte activity in the GPe could attenuate the habitual reward-seeking behaviors. We used chemogenetics to manipulate cellular activity in a cell-type-specific manner.

First, we injected the GFAP promoter-dependent hM3Dq-expressing virus into the GPe of mice and implanted optic cannula into the GPe (Fig. 4a). For comparison, we prepared the control group where GFAP promoter-driven mCherry-expressing virus was delivered (Fig. 4b) so that we performed three controls (vehicle in mCherry mice, C21 in mCherry mice, and the vehicle in hM3Dq mice) for *in vivo* calcium imaging and behavioral experiments. Four to five weeks after the virus delivery, we validated whether C21 treatment increases intracellular calcium ion levels in GPe astrocytes using *in vivo* calcium imaging. C21 (3 mg/kg, *i.p.*) increased the average fluorescence intensity in hM3Dq mice, but not in mCherry mice (Fig. 4b, c, Mann-Whitney test, $p = 0.008$ for hM3Dq mice, $p = 0.540$ for mCherry mice, $n = 5$/group). Furthermore, we assessed behavior 30 min after systemic injection of either vehicle or C21. C21 did not alter locomotor activity in the open field [Fig. 4d, two-way ANOVA, $F(1,11) = 0.775$, $p = 0.397$, $n = 7$/hM3Dq, $n = 6$/mCherry group], although GPe governs motor-related behaviors [20, 21].

In operant conditioning, we developed a habit and goal-directed behavior by using RI and RR schedules, respectively. And then, we performed the evaluation tests (Fig. 4e). As we expected, the RR schedule yielded a reduction of the habitual behavior index (100% of the index indicates intact habit) (Fig. 4f, Mann-Whitney test, $p < 0.001$, $n = 14$/RR, $n = 30$/RI groups). C21 treatment reduced habit index after the RI task in GPe hM3Dq-expressing mice compared to vehicle treatment, but not after the RR task (Fig. 4g, Wilcoxon test, $p = 0.079$ for RR, $p = 0.029$ for RI, $n = 14$/RR, $n = 21$/RI groups). Consistently, C21 treatment in hM3Dq mice exhibited a significant difference of nose-poking behaviors between the valued and devalued states similar to the behaviors in vehicle-treated hM3Dq mice after the RR schedule (Fig. 4h, Wilcoxon test, $p < 0.001$ for VEH and C21 groups, $n = 14$/group). In contrast, C21 treatment in hM3Dq mice showed outcome value-dependent flexible changes in nose-poking behaviors, which were not observed in vehicle-treated hM3Dq mice after the RI schedule (Fig. 4i, Wilcoxon test, $p = 0.663$ for VEH groups, $p = 0.026$ for C21 groups, $n = 21$/group). Such drastic changes were observed specific to hM3Dq mice whereas mCherry mice with vehicle or C21 treatment did not display after the RI task (**Supplementary Fig. 4a**, Wilcoxon test, $p = 0.945$; **Supplementary Fig. 4b**, Wilcoxon test, $p = 0.969$ for VEH groups, $p = 0.445$ for C21 groups; $n = 8$/groups). Therefore, these results indicate that GPe astrocyte activation may dampen habitual seeking behavior.

Activity increase in GPe astrocytes reduces GPe neuronal activities. Previous studies have demonstrated that astrocytic Gq-GPCRs induce calcium influx, promoting the direct astrocyte-neuron interaction and results in temporal activity changes in the surrounded neurons and their circuits [22–24]. Interestingly, previous research reported that the neuronal responses induced by the astrocytic activation depend on their neuronal characteristics [18, 25, 26]. Thus, we examined how the astrocyte activation via modified Gq-pathway GPCRs (chemogenetic hM3Dq) affects neighbor GPe neurons. To preserve the cellular structure of the recorded neurons and surrounding astrocytes expressing chemogenetic GPCRs, we performed *in vivo* electrophysiological recordings with the mice expressing GFAP promoter-driven hM3Dq in the GPe (Fig. 5a). The expression of hM3Dq did not affect the spontaneous firing rates of GPe neurons (**Supplementary Fig. 5a**, Mann-Whitney test, $p = 0.387$; **Supplementary Fig. 5b**, Kolmogorov-Smirnov test, $p = 0.566$; $n = 49$/mCherry, $n = 46$/hM3Dq groups).

Importantly, systemic application of C21 (3 mg/kg, *i.p.*) reduced the spontaneous firing rates of GPe neurons of mice expressing astrocytic hM3Dq, compared to those without hM3Dq expression (Fig. 5b-c, two-way ANOVA, $F(1,93) = 43.13$, $p = 0.075$; Fig. 5d, Kolmogorov-Smirnov test, $p < 0.001$; $n = 49$/mCherry, $n = 46$/hM3Dq groups) at the time point we assessed animal behaviors. This reduction of firing was generally observed independent of the basal firing frequencies of the neurons (Fig. 5e, Spearman correlation, $p = 0.372$; Fig. 5f, Spearman correlation, $p = 0.494$; $n = 49$/mCherry, $n = 46$/hM3Dq groups), unlike the significant positive correlations between firing rates before and after C21 application in both groups (**Supplementary Fig. 6a**, Spearman correlation, $p < 0.001$; **Supplementary Fig. 6b**, Spearman correlation, $p < 0.001$; $n = 49$/mCherry, $n = 46$/hM3Dq groups). These results suggest that the astrocytic activation reduce the overall GPe neuronal activities, regardless of their neuronal characteristics.

Attentional stimulation reduces habitual seeking behaviors. Vigilance is a state of readiness to respond to certain state changes [27]. Since habit is often associated with insensitivity to environmental changes [4], we hypothesized that attentional stimulation may awake the cognitive states, which were implicitly silenced while learning the optimal behavior routine in the RI tasks, incorporating time-based reward delivery.

To check the relationship between vigilance with attentional stimulation and GPe astrocyte activities, we recorded *in vivo* Ca^{2+} imaging in GPe astrocytes and neurons (Fig. 6a). First, we dropped the ball in a parabola style to touch the mouse face to induce the vigilance state artificially (Fig. 6a). We employed GFAP-Cre/GCaMP6s mice for GPe astrocytes and viral hSyn-GCaMP6s mice for neurons. Retrograde viral hSyn-GCaMP6s-tdTomato injected into the dorsal striatum was expressed in the GPe (Fig. 6b, n = 32/group). Here, we observed that attentional stimulation enhanced intracellular Ca^{2+} levels in GPe astrocytes and neurons (Fig. 6c, Wilcoxon test, $p < 0.001$, $n = 32$/group).

Next, to confirm that attentional stimulation does not integrate any side effects can intervene the operant behavior, we conducted brief stimulation for 4 min and evaluated whether attentional stimulation alters locomotion or induces anxiety-like behavior (Fig. 6d). We concluded that attentional stimulation did neither change locomotor activities (Fig. 6e, two-way ANOVA, $F(1,11) = 0.713$, $p = 0.417$; Fig. 6f, Mann-Whitney test, $p = 0.628$; $n = 7$/Ctrl, $n = 6$/Stim groups), nor induce anxiety-like behavior (Fig. 6g, Mann-Whitney test, $p = 0.836$, $n = 7$/Ctrl, $n = 6$/Stim groups).

After the operant conditioning with the RI task, 8 times of attentional stimulation for 4 min before the extinction test reduced habit index compared to the no stimulation group (Fig. 6h, i, Wilcoxon test, $p < 0.001$, $n = 9$/groups). In addition, the stimulation group showed the reduction of nose-poke rate to devalued state, which was not observed in the no stimulation group (Fig. 6j, Wilcoxon test, $p = 0.410$ for Ctrl groups, $p = 0.020$ for Stim groups, $n = 9$/groups). These results demonstrated that attentional stimulation might reduce habitual reward-seeking behavior and manifest goal-directed behavior by reactivating the cognitive states related to vigilance.

Discussion

In this study, we demonstrated that activation of GPe astrocyte activities attenuates habitual reward-seeking behaviors. Using Cre-dependent astrocyte-specific calcium dynamics, we showed that GPe astrocyte activities are correlated with critical behavioral events in an operant chamber. Also, our prediction model elucidated that the pattern of GPe astrocyte activities can predict whether mice are exhibiting goal-directed or habitual learning. We verified that activation of GPe astrocytes dampens the GPe neuronal activities. Furthermore, both chemogenetic activation and attentional stimulation increase GPe astrocyte activities and facilitate the transition from habitual to goal-directed behaviors.

Although the GPe is classically considered a component of the motor-suppressing indirect pathway [28–30], it is also attributed to non-motor functions such as addiction [31], reward learning and prediction [32–35], and sleep [36, 37]. Remarkably, the GPe is an essential nucleus for action selection within basal ganglia circuits

[29, 38, 39]. Of note, agreeing with our finding, several previous studies confirmed that astrocytes are highly enriched in the GPe [11, 40–42], implying a crucial role of astrocytes in the GPe. Especially, considering that at least three different types of GABAergic neurons, denoted arkypallidal (Arky)-GPe, LIM homeobox 6 (Lhx 6)-GPe and parvalbumin (PV)-GPe neurons, are having unique feedforward or feedback circuits [8–10], the role of GPe astrocytes is critical for accurate regulation of behavioral flexibility. While previous studies were focused on GPe neurons or the GPe astrocytes on movement function or disorders, our current study reveals the role of GPe astrocytes in reward-seeking behaviors. Notably, we found that GPe astrocyte activities are sensitively increased upon magazine exit when mice were exhibiting goal-directed reward-seeking (random ratio schedule, RR). By contrast, the GPe astrocyte activities are silenced during habitual reward-seeking (random interval schedule, RI) behaviors. Since mice may need to decide the next action (going back to nose-poke, re-try to visit the magazine, or spending time without specific reward-related action), especially when they failed in procuring the expected reward, increased GPe astrocyte activities may reflect the action selection activities in goal-directed behaviors to maximize the reward collection through updating action-outcome contingency.

On the other hand, dampened GPe astrocyte activities during habitual reward-seeking may indicate that mice have no clear direction to choose the next action. Our findings show that activation of GPe astrocytes decreases GPe neuronal activities, suggesting that GPe neuronal activities are decreased during goal-directed and increased during habitual behavior. Our previous study demonstrated that enhanced astrocyte activities in the dorsomedial striatum (DMS) reduced GPe neuronal activities and abolished habitual reward-seeking behaviors similar to this study [18]. Based on direct and indirect BG circuits, decreased GPe neuronal activities will yield increased excitatory glutamatergic STN activities on GPi, which may result in potentiating GPi activities [4, 43, 44]. Also, our finding suggests that goal-directed behaviors or behavioral flexibility require inhibition of continuous momentum of behaviors or tendency to resist to form a stimulus-response association (**Supplementary Fig. 7**). Although our study focuses on the GPe, which receives GABAergic inputs from the dorsal striatum, both the nucleus accumbens core and shell are critical for goal-directed and habitual learning and behaviors [45, 46]. Therefore, additional researches are required to delineate the circuit-specific regulation of reward-seeking behaviors.

The GPe integrates divergent neural circuits for information processing to predict and perform motor and reward-related actions [35]. The GPe receives not only GABAergic input from the dorsal striatum, especially from dopamine D2 receptor or adenosine A2A receptor-expressing medium spiny neurons (MSNs), but also glutamatergic input from the subthalamic nucleus (STN). Also, PV-GPe neurons mainly project to STN and parafascicular nucleus (PF) of the thalamus, while Lhx6-GPe neurons project to STN and substantia nigra pars compacta (SNc). Interestingly, Arky-GPe neurons uniquely project back to the dorsal striatum [8, 9]. Considering these complex input and output circuits of the GPe, the abundant GPe astrocytes may serve as fine moderators. Previously, we showed that astrocyte-driven adenosine signaling regulates the MSNs in the DMS and habitual reward-seeking behaviors [18]. On the other hand, GPe neurons and astrocytes are expressing several receptors including glutamate, dopamine, and

adenosine receptors [47–49]. Thus, further molecular studies will reveal the precise role of astrocytes in moderating the GPe circuit-dependent activities.

As previously described [17, 18], we utilized sucrose as a reward. After fasting, mice are highly motivated to seek out sucrose, which facilitates the operant learning [14]. Although sucrose is not fully representing the addictive drugs [50], sucrose has common rewarding and reinforcing properties as substance abuse, especially in rodents [15, 51]. Thus, our research is highly relevant to addiction since the hallmark of addiction is an impaired ability to stop or control the substance or non-substance activities, which is attributed to the progressive dominance of habits over goal-directed behaviors [52, 53]. Habits are a state of insensitive to outcome devaluation or degradation of action-outcome contingency [54], as we measured the difference between valued and devalued nose-poking behaviors during the extinction period. In this assessment, the deficit of inhibition of unnecessary or counter-productive behavior is mainly attributed to maladaptive habitual behavior as a behavioral inflexibility, which can develop to or overlap with compulsive reward-seeking behaviors when reinforcement of reward becomes insensitive to costs or negative consequences [2, 55]. In this striatum-focused perspective, a transition from ventral (reinforcement and motivation), to dorsomedial (goal-directed), to dorsolateral (habitual) striatum [2, 55] implies progressive development of addictive behaviors[56]. As we observed in our study, the dampened GPe astrocyte activities during habit formation with random interval schedules may potentiate GPe neuronal activities and thereby weaken the glutamatergic STN neurons to GPi, which shift to disinhibition of GPi inputs to the thalamus as strengthening the direct MSNs pathway. Notably, the uncontrollable habits and deficit of behavioral flexibility are viewed as compulsions [6, 57, 58]. However, several experimental and theoretical studies disagree with the progressive transition from goal-directed to habitual-reward seeking behaviors and emphasize the excessive goal-directed behaviors as addiction [56, 59, 60]. Importantly, imbalance of goal-directed and habits are traits, not the addiction itself, but leading to a more visible addictive component such as decision-making [61]. Impairment of behavioral flexibility in regard to balancing goal-directed and habits are implicated in obsessive-compulsive disorder (OCD) [62, 63] or ritual behaviors [58] as well. Then, how do goal-directed or habitual reward-seeking behaviors influence the choice depending on contexts or cues? What are the other nuclei or circuits being able to modify or determine the decision-making process engaging infralimbic cortex [64–66] or orbitofrontal cortex [67–69]? For these, future studies will provide more precise answers. At least, however, our present study demonstrated that GPe astrocyte activities intervene in the critical transition between goal-directed and habitual reward-seeking behaviors.

We employed the support vector machine (SVM) to generate prediction models based on six different behavioral components [3 behavioral events (nose-poke, magazine entry, and magazine exit) x 2 conditions (R + and R-)] and correlated to real-time GPe astrocyte activities measured by fiber-photometry with GCaMP6s. SVM is one of the most widely used machine learning algorithms for classification and regression because their predictions do not suffer from overfitting in the structural risk minimization sense [19, 70]. Our SVM model helped us elucidate the role of GPe astrocyte encoding behavioral types

(goal-directed and habitual behaviors). To our understanding, this is the first report dissecting the behavioral components (goal-directed and habitual) aligned with temporal and spatial cell-specific activities (GPe astrocytes). In this study, we manually labeled all the behavioral components in the recorded video to compare them with GPe astrocyte activities. Recently, automated video tracking systems and various software allow detecting the behavioral components [71] for more detailed behavioral components [72, 73]. Importantly, machine-learning algorithms can make more precise predictions if they benefit from data generated through dissection of behavioral components with high temporal resolutions and cell-specific neuronal and non-neuronal activities using endomicroscopy or fiber-photometry.

Goal-directed behavior requires attention and vigilance to modify their actions based on the outcome while habit demands less attention [74–76]. After confirming that chemogenetic activation of GPe astrocytes abolished the habitual reward-seeking behaviors, we also found that attentional stimulation can increase GPe astrocytes and dampen the habitual reward-seeking. In rats, sustained temporal attention prevents habit formation in an operant chamber [77]. Since vigilance is commonly defined as sustained attention or tonic alertness [27, 78, 79], brief intervention to increase vigilance may be effective to disrupt an automatic or inattentive routine of repeating the same behaviors. For visual and spatial attention as we employed, the basal ganglia are one of the central regions involved in vigilance and attention [80]. Not surprisingly, other brain regions are attributed to visual attention including the visual cortex, prefrontal cortex, and thalamus [81]. Thus, further studies are warranted to clarify whether our attention stimulation approach changes the astrocytes and neuronal activities in other brain regions.

In summary, our study demonstrated that GPe astrocyte activities are reduced during habit learning. Chemogenetic and attention stimulation can prevent habitual reward-seeking, at least partly through activation of GPe astrocyte. This novel finding may help to treat maladaptive habit-related disorders such as addiction and obsessive-compulsive disorders.

Declarations

Conflict of Interest

D.S.C. is a scientific advisory board member to Peptron Inc., and the Peptron had no role in the preparation, review, or approval of the manuscript; nor the decision to submit the manuscript for publication. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgment

We thank all the laboratory members for their helpful discussion and comments. We thank Ms. DeAnn Frederixon for her editing and proofreading. The main Figure 7 and Supplementary Figure 7 were created with BioRender.com. This research was supported by the Samuel C. Johnson for Genomics of Addiction Program at Mayo Clinic, the Ulm Foundation, and the National Institute of Health (AA018779, AA029258,

AG072898, AA027773). This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2019M3E5D2A01066267, Development of metacognitive AI for rapid learning), and Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2019-0-01371, Development of brain-inspired AI with human-like intelligence).

Author Contributions

S.I.H, S.W.L., and D.S.C. designed the study. S.I.H. performed all the behavioral, chemogenetic experiments and analyzed all the data. S.K. (Seungwoo Kang) performed electrophysiological experiments and analyzed the electrophysiological data. M.S., M.A.Y., J.L., S.W.L. performed machine-learning analysis. M.B. and S.K. (Shinwoo Kang) assisted S.I.H. for behavioral experiments and data analysis. All the authors wrote the manuscript.

References

1. Balleine, B.W. & Dickinson, A. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* **37**, 407-419 (1998).

2. Koob, G.F. & Volkow, N.D. Neurocircuitry of addiction. *Neuropsychopharmacology* **35**, 217-238 (2010).

3. Robbins, T.W., Vaghi, M.M. & Banca, P. Obsessive-Compulsive Disorder: Puzzles and Prospects. *Neuron* **102**, 27-47 (2019).

4. Yin, H.H. & Knowlton, B.J. The role of the basal ganglia in habit formation. *Nat Rev Neurosci* **7**, 464-476 (2006).

5. Lovinger, D.M. & Gremel, C.M. A Circuit-Based Information Approach to Substance Abuse Research. *Trends Neurosci* **44**, 122-135 (2021).

6. Lipton, D.M., Gonzales, B.J. & Citri, A. Dorsal Striatal Circuits for Habits, Compulsions and Addictions. *Front Syst Neurosci* **13**, 28 (2019).

7. Mendelsohn, A.I. Creatures of Habit: The Neuroscience of Habit and Purposeful Behavior. *Biol Psychiatry* **85**, e49-e51 (2019).

8. Gittis, A.H.*, et al.* New roles for the external globus pallidus in basal ganglia circuits and behavior. *J Neurosci* **34**, 15178-15183 (2014).

9. Mastro, K.J., Bouchard, R.S., Holt, H.A. & Gittis, A.H. Transgenic mouse lines subdivide external segment of the globus pallidus (GPe) neurons and reveal distinct GPe output pathways. *J Neurosci* **34**, 2087-2099 (2014).

10. Abrahao, K.P. & Lovinger, D.M. Classification of GABAergic neuron subtypes from the globus pallidus using wild-type and transgenic mice. *J Physiol* **596**, 4219-4235 (2018).

11. Cui, Q.*, et al.* Blunted mGluR Activation Disinhibits Striatopallidal Transmission in Parkinsonian Mice. *Cell Rep* **17**, 2431-2444 (2016).

12. Tatsumi, K.*, et al.* Voluntary Exercise Induces Astrocytic Structural Plasticity in the Globus Pallidus. *Front Cell Neurosci* **10**, 165 (2016).

13. Berridge, K.C. & Kringelbach, M.L. Affective neuroscience of pleasure: reward in humans and animals. *Psychopharmacology (Berl)* **199**, 457-480 (2008).

14. Sieburg, M.C.*, et al.* Reward Devaluation Attenuates Cue-Evoked Sucrose Seeking and Is Associated with the Elimination of Excitability Differences between Ensemble and Non-ensemble Neurons in the Nucleus Accumbens. *eNeuro* **6** (2019).

15. Winterdahl, M.*, et al.* Sucrose intake lowers mu-opioid and dopamine D2/3 receptor availability in porcine brain. *Sci Rep* **9**, 16918 (2019).

16. Hong, S.I., Bullert, A., Baker, M. & Choi, D.S. Astrocytic equilibrative nucleoside transporter type 1 upregulations in the dorsomedial and dorsolateral striatum distinctly coordinate goal-directed and habitual ethanol-seeking behaviors in mice. *Eur J Neurosci* **52**, 3110-3123 (2020).

17. Hong, S.I., Kang, S., Chen, J.F. & Choi, D.S. Indirect Medium Spiny Neurons in the Dorsomedial Striatum Regulate Ethanol-Containing Conditioned Reward Seeking. *J Neurosci* **39**, 7206-7217 (2019).

18. Kang, S.W.*, et al.* Activation of Astrocytes in the Dorsomedial Striatum Facilitates Transition from Habitual to Goal-Directed Reward-Seeking Behavior *Biol Psychiatry* **88**, 797-808 (2020).

19. Vapnik, V. *Statistical Learning Theory* (John-Wiley & Sons, Inc, 1998).

20. Lilascharoen, V.*, et al.* Divergent pallidal pathways underlying distinct Parkinsonian behavioral deficits. *Nat Neurosci* **24**, 504-515 (2021).

21. Aristieta, A. & Gittis, A. Distinct globus pallidus circuits regulate motor and cognitive functions. *Trends Neurosci* (2021).

22. Adamsky, A.*, et al.* Astrocytic Activation Generates De Novo Neuronal Potentiation and Memory Enhancement. *Cell* **174**, 59-71 e14 (2018).

23. Cao, X.*, et al.* Astrocyte-derived ATP modulates depressive-like behaviors. *Nat Med* **19**, 773-777 (2013).

24. Scofield, M.D. & Kalivas, P.W. Astrocytic dysfunction and addiction: consequences of impaired glutamate homeostasis. *Neuroscientist* **20**, 610-622 (2014).

25. Cavaccini, A., Durkee, C., Kofuji, P., Tonini, R. & Araque, A. Astrocyte Signaling Gates Long-Term Depression at Corticostriatal Synapses of the Direct Pathway. *J Neurosci* **40**, 5757-5768 (2020).

26. Erickson, E.K.*, et al.* Cortical astrocytes regulate ethanol consumption and intoxication in mice. *Neuropsychopharmacology* **46**, 500-508 (2021).

27. Oken, B.S., Salinsky, M.C. & Elsas, S.M. Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clin Neurophysiol* **117**, 1885-1901 (2006).

28. Albin, R.L., Young, A.B. & Penney, J.B. The functional anatomy of basal ganglia disorders. *Trends Neurosci* **12**, 366-375 (1989).

29. Smith, Y., Bevan, M.D., Shink, E. & Bolam, J.P. Microcircuitry of the direct and indirect pathways of the basal ganglia. *Neuroscience* **86**, 353-387 (1998).

30. Hegeman, D.J., Hong, E.S., Hernandez, V.M. & Chan, C.S. The external globus pallidus: progress and perspectives. *Eur J Neurosci* **43**, 1239-1265 (2016).

31. Beier, K.T.*, et al.* Rabies screen reveals GPe control of cocaine-triggered plasticity. *Nature* **549**, 345-350 (2017).

32. Schroll, H.*, et al.* Differential contributions of the globus pallidus and ventral thalamus to stimulus-response learning in humans. *Neuroimage* **122**, 233-245 (2015).

33. Gdowski, M.J., Miller, L.E., Parrish, T., Nenonene, E.K. & Houk, J.C. Context dependency in the globus pallidus internal segment during targeted arm movements. *J Neurophysiol* **85**, 998-1004 (2001).

34. Adler, A.*, et al.* Temporal convergence of dynamic cell assemblies in the striato-pallidal network. *J Neurosci* **32**, 2473-2484 (2012).

35. Arkadir, D., Morris, G., Vaadia, E. & Bergman, H. Independent coding of movement direction and reward prediction by single pallidal neurons. *J Neurosci* **24**, 10047-10056 (2004).

36. Qiu, M.H., Chen, M.C., Wu, J., Nelson, D. & Lu, J. Deep brain stimulation in the globus pallidus externa promotes sleep. *Neuroscience* **322**, 115-120 (2016).

37. Qiu, M.H., Yao, Q.L., Vetrivelan, R., Chen, M.C. & Lu, J. Nigrostriatal Dopamine Acting on Globus Pallidus Regulates Sleep. *Cereb Cortex* **26**, 1430-1439 (2016).

38. Bogacz, R., Martin Moraud, E., Abdi, A., Magill, P.J. & Baufreton, J. Properties of Neurons in External Globus Pallidus Can Support Optimal Action Selection. *PLoS Comput Biol* **12**, e1005004 (2016).

39. Kita, H. Globus pallidus external segment. *Prog Brain Res* **160**, 111-133 (2007).

40. Dervan, A.G.*, et al.* Astroglial plasticity and glutamate function in a chronic mouse model of Parkinson's disease. *Exp Neurol* **190**, 145-156 (2004).

41. Lange, H., Thorner, G., Hopf, A. & Schroder, K.F. Morphometric studies of the neuropathological changes in choreatic diseases. *J Neurol Sci* **28**, 401-425 (1976).

42. Salvesen, L.*, et al.* Changes in total cell numbers of the basal ganglia in patients with multiple system atrophy - A stereological study. *Neurobiol Dis* **74**, 104-113 (2015).

43. Knowlton, B.J., Mangels, J.A. & Squire, L.R. A neostriatal habit learning system in humans. *Science* **273**, 1399-1402 (1996).

44. Yin, H.H. & Knowlton, B.J. Contributions of striatal subregions to place and response learning. *Learn Mem* **11**, 459-463 (2004).

45. Mannella, F., Gurney, K. & Baldassarre, G. The nucleus accumbens as a nexus between values and goals in goal-directed behavior: a review and a new hypothesis. *Front Behav Neurosci* **7**, 135 (2013).

46. Penner, M.R. & Mizumori, S.J. Neural systems analysis of decision making during goal-directed navigation. *Prog Neurobiol* **96**, 96-135 (2012).

47. Araque, A.*, et al.* Gliotransmitters travel in time and space. *Neuron* **81**, 728-739 (2014).

48. Lovinger, D.M. Neurotransmitter roles in synaptic modulation, plasticity and learning in the dorsal striatum. *Neuropharmacology* **58**, 951-961 (2010).

49. Xin, W. & Bonci, A. Functional Astrocyte Heterogeneity and Implications for Their Role in Shaping Neurotransmission. *Front Cell Neurosci* **12**, 141 (2018).

50. Bobadilla, A.C.*, et al.* Cocaine and sucrose rewards recruit different seeking ensembles in the nucleus accumbens core. *Mol Psychiatry* **25**, 3150-3163 (2020).

51. Leblond, M., Fan, D., Brynildsen, J.K. & Yin, H.H. Motivational state and reward content determine choice behavior under risk in mice. *PLoS One* **6**, e25342 (2011).

52. Everitt, B.J. & Robbins, T.W. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat. Neurosci.* **8**, 1481-1489 (2005).

53. Everitt, B.J. & Robbins, T.W. Drug Addiction: Updating Actions to Habits to Compulsions Ten Years On. *Annu Rev Psychol* **67**, 23-50 (2016).

54. Robbins, T.W. & Costa, R.M. Habits. *Curr Biol* **27**, R1200-R1206 (2017).

55. Everitt, B.J. & Robbins, T.W. From the ventral to the dorsal striatum: Devolving views of their roles in drug addiction. *Neuroscience and biobehavioral reviews* (2013).

56. Hong, S.I., Kang, S., Baker, M. & Choi, D.S. Astrocyte-Neuron Interaction in the Dorsal Striatum-Pallidal Circuits and Alcohol-Seeking Behaviors. *Neuropharmacology* **In Press** (2021).

57. Luscher, C., Robbins, T.W. & Everitt, B.J. The transition to compulsion in addiction. *Nat Rev Neurosci* **21**, 247-263 (2020).

58. Graybiel, A.M. Habits, rituals, and the evaluative brain. *Annu Rev Neurosci* **31**, 359-387 (2008).

59. Hogarth, L. Addiction is driven by excessive goal-directed drug choice under negative affect: translational critique of habit and compulsion theory. *Neuropsychopharmacology* **45**, 720-735 (2020).

60. Singer, B.F., Fadanelli, M., Kawa, A.B. & Robinson, T.E. Are Cocaine-Seeking "Habits" Necessary for the Development of Addiction-Like Behavior in Rats? *J Neurosci* **38**, 60-73 (2018).

61. Veatch, L.M. Disruptions in sleep time and sleep architecture in a mouse model of repeated ethanol withdrawal. *Alcohol Clin Exp Res* **30**, 1214-1222 (2006).

62. Gillan, C.M.*, et al.* Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *Am J Psychiatry* **168**, 718-726 (2011).

63. Graybiel, A.M. & Rauch, S.L. Toward a neurobiology of obsessive-compulsive disorder. *Neuron* **28**, 343-347 (2000).

64. Hitchcott, P.K., Quinn, J.J. & Taylor, J.R. Bidirectional modulation of goal-directed actions by prefrontal cortical dopamine. *Cereb Cortex* **17**, 2820-2827 (2007).

65. Smith, K.S. & Graybiel, A.M. A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron* **79**, 361-374 (2013).

66. Smith, K.S., Virkud, A., Deisseroth, K. & Graybiel, A.M. Reversible online control of habitual behavior by optogenetic perturbation of medial prefrontal cortex. *Proc Natl Acad Sci U S A* **109**, 18932-18937 (2012).

67. Gremel, C.M. & Costa, R.M. Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nat Commun* **4**, 2264 (2013).

68. Volkow, N.D. & Fowler, J.S. Addiction, a disease of compulsion and drive: involvement of the orbitofrontal cortex. *Cereb Cortex* **10**, 318-325 (2000).

69. Voon, V.*, et al.* Disorders of compulsivity: a common bias towards learning habits. *Mol Psychiatry* **20**, 345-352 (2015).

70. Orru, G., Pettersson-Yeo, W., Marquand, A.F., Sartori, G. & Mechelli, A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev* **36**, 1140-1152 (2012).

71. Geuther, B.Q., *et al.* Robust mouse tracking in complex environments using neural networks. *Commun Biol* **2**, 124 (2019).

72. Markowitz, J.E. & Datta, S.R. The striatum specifies the statistics of behavior. *Neuropsychopharmacology* **45**, 222-223 (2020).

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

74. Franz, E.A. The allocation of attention to learning of goal-directed actions: a cognitive neuroscience framework focusing on the Basal Ganglia. *Front Psychol* **3**, 535 (2012).

75. Gasbarri, A., Pompili, A., Packard, M.G. & Tomaz, C. Habit learning and memory in mammals: behavioral and neural characteristics. *Neurobiol Learn Mem* **114**, 198-208 (2014).

76. Lisman, J. & Sternberg, E.J. Habit and nonhabit systems for unconscious and conscious behavior: implications for multitasking. *J Cogn Neurosci* **25**, 273-283 (2013).

77. Lin, Z., *et al.* Sustaining temporal attention prevents habit expression during operant learning in rats. *Sci Rep* **10**, 10303 (2020).

78. Nuechterlein, K.H., Parasuraman, R. & Jiang, Q. Visual sustained attention: image degradation produces rapid sensitivity decrement over time. *Science* **220**, 327-329 (1983).

79. Parasuraman, R. Memory load and event rate control sensitivity decrements in sustained attention. *Science* **205**, 924-927 (1979).

80. Wang, L. & Krauzlis, R.J. Involvement of Striatal Direct Pathway in Visual Spatial Attention in Mice. *Curr Biol* **30**, 4739-4744 e4735 (2020).

81. Luo, T.Z. & Maunsell, J.H.R. Attention can be subdivided into neurobiological components corresponding to distinct behavioral effects. *Proc Natl Acad Sci U S A* **116**, 26187-26194 (2019).

82. Quiroga, R.Q., Nadasdy, Z. & Ben-Shaul, Y. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput* **16**, 1661-1687 (2004).

83. Derusso, A.L., *et al.* Instrumental uncertainty as a determinant of behavior under interval schedules of reinforcement. *Front Integr Neurosci* **4** (2010).

Methods

Animals. All experimental procedures were approved by the Mayo Clinic Institutional Animal Care and Use Committee and performed following NIH guidelines. C57BL/6J mice and bi-transgenic GFAP-Cre (Stock No. 024098) /GCaMP6s (Stock No. 028866) mice were purchased from Jackson Laboratory (Bar Harbor,

ME). Mice were housed in standard Plexiglas cages. The colony room was maintained at a constant temperature (24 ± 1°C) and humidity (60 ± 2%) with a 12 h light/dark cycle (lights on at 07:00 A.M.). We used 8- to 10-week-old male mice for all experiments. Mice were allowed *ad libitum* access to food and water. For the operant conditioning tests, mice were food restricted to 85% of their baseline weight, at which time they were maintained for the duration of experimental procedures.

Microsurgery and chemogenetics. Mice were anesthetized with isoflurane (1.5% in oxygen gas) with the VetFlo™ vaporizer with a single-channel anesthesia stand (Kent Scientific Corporation, Torrington, CT) and placed on the digital stereotaxic alignment system (Model 1900; David Kopf Instruments, Tujunga, CA). Hair was trimmed and the skull was exposed using an 8-gauge electrosurgical skin cutter (KLS Martin, Jacksonville, FL). The skull was leveled using a dual tilt measurement tool. Holes were drilled in the skull at the appropriate stereotaxic coordinates. For chemogenetics, viruses were infused unilaterally to the GPe (AP -0.46 mm, ML +2.0 mm, DV -3.2 mm from bregma) at 100 nl/min for 3 min through a 33-gauge injection needle (cat #NF33BV; World Precision Instruments) using a micro-syringe pump (Model UMP3; World Precision Instruments). For *in vivo* Ca^{2+} imaging of GPe neurons, the virus was infused unilaterally to the DLS (AP +0.6 mm, ML +2.4 mm, DV -3.1 mm from bregma) at 100 nl/min for 5 min through a 33-gauge injection needle using a micro-syringe pump. The injection needle remained in place for an additional 6 min following the end of the injection. Following stereotaxic surgery, we injected buprenorphine sustained-release LAB (1 mg/kg, s.c.; ZooPharm, Laramie, WY, USA) to alleviate post-surgery pain. Mice were utilized for the experiments 4-5 wk after the virus injection. We injected viruses at following titers: AAV5-GFAP-mCherry, 1.6 X 10^{13} GC/ml (Vector Biolabs, Malvern, PA); AAV5-GFAP-hM3Dq-mCherry, 1.2 X 10^{13} GC/ml; AAVr-hSyn1-GCamP6s-P2A-nls-dTomato, 9.0 X 10^{12} GC/ml. AAV-GFAP-hM3Dq-mCherry was a gift from Bryan Roth (Addgene viral prep #50478-AAV5; http://n2t.net/addgene:50478; RRID: Addgene_50478). AAV-hSyn1-GCaMP6s-P2A-nls-dTomato was a gift from Jonathan Ting (Addgene viral prep #51084-AAVrg; http://n2t.net/addgene:51084; RRID: Addgene_51084). We purchased compound 21 from Hello Bio (Princeton, NJ).

Behavioral experiments. 1) *Operant conditioning.* The operant chamber consisted of an active hole, an inactive hole, a magazine, a house light, a speaker, and a cue light at each nose port. Reward (20% sucrose solution) was presented to the liquid receptacle in the magazine once per reward signal by a syringe pump. The nose-poking behavior of mice was followed by two contradicting results in some cases. One is the case of rewarded nose-poke. When the mice did a nose-poke in the active hole, the chamber presented tone, light from the nose port, and one reward from the magazine. On the other hand, in the case of non-rewarded nose-poke, the chamber did not present anything, no matter how nose-poke the mice did (**Fig. 1a**). The operant conditioning schedule is as follows. On the first day, magazine training was performed for 30 minutes a day to learn the space of the magazine for the mice. After that, the fixed ratio 1 (FR1) was performed for 4 sessions and 60 minutes, and if mice obtained 60 rewards, the session was terminated regardless of the remaining time. For rapid learning of operant behavior, 10 ul of 20% sucrose was placed in the active hole as bait before the start of the FR1 session. Moreover, in two consecutive sessions, when the average latency time from nose-poke to the magazine was less than 2

seconds, no bait was placed in the active hole from the next session. In the last session, the fourth session of the FR1 task, all mice performed operant behavior without prior bait presentation. Until the fourth session of the FR1 task, if the average latency time from nose-poke to the magazine was more than 2 s, we planned to discard that mice data. Nevertheless, no mouse was excluded. Afterward, a random ratio (RR) schedule was performed to develop goal-directed behaviors, and a random interval (RI) schedule was performed to form habits. The RR schedule consisted of RR2 (1 session), RR5 (2 sessions), RR10 (3 sessions), and RR20 (2 sessions). Each task was performed for 30 minutes, and as in the FR1 task, when mice obtained 60 rewards, the session was terminated regardless of the remaining time. The RI schedule consisted of RI30 (1 session), RI60 (3 sessions), and RI120 (4 sessions). Each task was performed for 30 minutes, and there was no case that the session was terminated because mice obtained 60 rewards. The rules for FR1, RR, and RI tasks were as follows. In the FR1 task, only a rewarded nose-poke exists since all nose-pokes presented one tone, cue light, and reward. In the RR task, multiple nose-pokes were required to obtain one reward, and only the last nose-poke was a rewarded nose-poke when a certain number of nose-pokes was reached, and all others were non-rewarded nose-pokes. The RI task requires a certain amount of time to flow between two rewarded nose-pokes. So, the nose-poke over time becomes a non-rewarded nose-poke. For example, RR20 is programmed to obtain 1 reward only after performing an average of 20 nose-pokes after a rewarded nose-poke, and RI120 is a rewarded nose-poke and must perform a nose-poke after an average of 120 s (**Fig. 1a**). The timepoints of nose-poke and magazine entry, session duration, latency time from nose-poke to magazine approach, time spent in the magazine, and the number of nose-poke and magazine entry were recorded using the Med-PC-IV software and time resolution was 10 ms. Discrimination ratio (%) was calculated as the number of nose-poke for active hole divided into total numbers of nose-pokes.

2) Evaluation test. According to sensation-specific satiety theory, the evaluation test was performed to determine the goal-directed behaviors and habits of the mice through the change of the reward outcome value. The evaluation test consists of manipulating the reward value for 60 min in the home cage, and then the extinction test in the operant chamber for 10 min. The reward outcome value was manipulated as follows: the valued state that was provided an unlimited food chow to make the outcome of nose-poke valuable, and the devalued state that was provided an unlimited amount of 20% sucrose solution as same as the reward to devaluate the outcome value. Furthermore, the evaluation test was performed after the last sessions of FR1, RR20, and RI120. Habit index (%) = 100 - |difference in the number of nose-poke between the valued and devalued states| / total nose-pokes in the valued and devalued states x 100. If the habit index is 100%, it indicates the mouse may have an intact habit. The smaller the habit index is, the more goal-directed behavior is.

3) Open-field test. The open-field test was conducted in chambers (Med Associates, St Albans, VT). Locomotor activities in mice were recorded for 30 min and the first 10 min were used to measure anxiety-like behavior. We recorded mice behaviors 30 min after drug treatment (vehicle, C21) and immediately after the attentional stimulation period. Distance traveled was measured using beam breaks. Time spent in open zone (%) = time spent in open zone / total time x 100.

Attentional stimulation. We dropped a ball in a parabola style to touch the mouse face to give mice attentional stimulation. The distance from the central axis to the ball was 10 cm and potential energy was 0.059 Joules. We applied this stimulation once every 30 s for a total of 8 times for 4 min before the open-field test and extinction test. For *in vivo* Ca^{2+} imaging, we gave mice this attentional stimulation once every 30 s for a total of 6-8 times.

Immunofluorescence. Brains were fixed with 4% paraformaldehyde (Sigma-Aldrich, St. Louis, MO) and transferred to 30% sucrose (Sigma-Aldrich) in phosphate-buffered saline at 4°C for 72 hr. Brains were then frozen in dry ice and sectioned at 40 μm using a microtome (Leica Corp., Bannockburn, IL). Brain slices were stored at −20°C in a cryoprotectant solution containing 30% sucrose (Sigma-Aldrich) and 30% ethylene glycol (Sigma-Aldrich) in phosphate-buffered saline. Sections were incubated in 0.2% Triton X-100 (Sigma-Aldrich), 5% bovine serum albumin in phosphate-buffered saline for 1 h followed by incubation with the primary antibody in 5% bovine serum albumin overnight at 4°C. Primary antibodies that were used in the present study included mouse anti-GFAP Alexa Fluor 488 [1:100, monoclonal immunoglobulin G1 (IgG1), #53-9892-82; Thermo Fisher, Waltham, MA, USA] antibody. After three times of washing with phosphate-buffered saline, the sections were mounted onto a glass slide coated with gelatin and cover-slipped with a VECTASHIELD® antifade mounting medium with DAPI (4',6-diamidino-2-phenylindole) (Vector Laboratories, Burlingame, CA). Images were obtained using an LSM 700 laser scanning confocal microscope (Carl Zeiss, Heidelberg, Germany) using a 10x lens (**Fig. 2a** and **3a**) and a 63x water-immersion lens (**Fig. 3a**).

***In vivo* calcium imaging with fiber-photometry.** We recorded the cellular calcium transients in real-time *in vivo* by fiber photometry as described previously [18]. Briefly, we implanted an optic cannula (200/240 μm diameter, 200 μm end fiber) and fixed it into the GPe (AP -0.46 mm, ML +2.0 mm, DV -3.0 mm from bregma) of the cell-type specifically GCaMP6s expressing mice by stereotaxic surgery. The implanted fiber was linked to a patch cord, allowing the light-emitting diode intensity at the interface of the fiber tip to raged in 60 μW consistently. These output signals were projected onto a photodetector by the same optical fiber, passed through a GFP filter. Data analysis was performed using the CineLyzer software (Ver. 4.4, Plexon) and light intensities were measured as relative fluorescence change ($\Delta F/F_0$). $\Delta F/F_0$ (%) was calculated relative to the signal (F_0) averaged in the 3 s preceding the time alignment. The camera for observing the mice's performance and the camera for observing the fluorescence change of GCaMP6s were synchronized and recorded at 30 frames per second. In operant conditioning, we recorded Ca^{2+} imaging in the GPe of mice during the last sessions of RR20 and RI120. For attentional stimulation, we measured cellular calcium transient for 3-4 min. To compare calcium transients between before and after the behavioral event, the $\Delta F/F_0$ data was averaged over the 2 seconds before or after the behavioral event. Furthermore, to compare the difference between calcium transients before and after the behavioral event, we subtracted averaged $\Delta F/F_0$ data before the behavioral event from averaged $\Delta F/F_0$ data after the behavioral event.

***In vivo* electrophysiological recording.** The experiments were performed as described previously [18]. Briefly, mice, 3-4 weeks following virus injections, were anesthetized by intraperitoneal injection of urethane (1.5 g/kg, Sigma-Aldrich) [21] and placed horizontally on a stereotaxic frame (RWD Lifesciences, San Diego, CA). Respiratory rate and pedal withdrawal reflex during anesthesia were constantly monitored and the physiological body temperature was maintained using a small animal feedback-controlled warming pad (Kent Scientific Corporation, Torrington, CT). After the scalp incision, small burr holes were drilled for the insertion of high impedance microelectrodes (Cambridge NeuroTech, Cambridge, UK). The reference wire (Ag/AgCl, 0.03" in diameter, A-M systems) was placed in the contralateral parietal cortex. Electrophysiological signals were digitized at 30 kHz and band-pass filtered from 300 to 3000 Hz (RHS 2000, Intan technologies, Los Angeles, CA). The analysis was performed by a custom-written code in MATLAB (R2019a; The MathWorks). Spike sorting was performed offline by superparamagnetic clustering of wavelet coefficients (Wave_Clus toolbox) [82]. An amplitude threshold of 4 medians of the absolute value of the signal was used for spike detection in each channel. The sorting result was then complemented by visual inspection of waveforms, variability, and interspike-interval (ISI) distributions. Spiking frequency was calculated in 1 s bins.

Prediction modeling. To investigate whether astrocyte activity can predict behavioral tasks (RR or RI) or reward experience (R+ or R-), we used a support vector machine (SVM) and 4-fold cross-validation. Average astrocyte activity during 1 s periods immediately before and after the behavior of interest (namely NP, M_{entry}, and M_{exit}) was used as input. Data from different mice were normalized by dividing by F_0 and then pooled together (n=6). The number of labels was 2 (rewarded or non-rewarded). To balance the dataset (3035 times of rNP and 174 times of nrNP in the RR task, 64 times of rNP and 1045 times of nrNP in the RI task; 120 times of either M_{entry} or M_{exit} (R+) and 170 times of either M_{entry} or M_{exit} (R-) in the RR task, 58 times of either M_{entry} or M_{exit} (R+) and 549 times of either M_{entry} or M_{exit} (R-) in the RI task), we performed random under sampling. Random under sampling was repeated 400 times for each analysis and an averaged performance of the SVM was calculated. We used MATLAB (version R2020a) for SVM analyses.

Data analysis. All data are represented as mean ± SEM and were analyzed by two-tailed Mann Whitney U test, two-tailed Wilcoxon matched-pairs signed-rank test, Kolmogorov–Smirnov test, Kruskal-Wallis test followed by Dunn's multiple comparisons test, two-way analysis of variance (ANOVA) followed by Tukey's multiple comparisons tests, two-way repeated-measures ANOVA followed by Bonferroni's multiple comparisons tests, Spearman correlation, simple linear regression, normality test with Shapiro-Wilk test using Prism 9.0 (GraphPad Software, San Diego, CA). For the statistical significance was set at $p < 0.05$. Detailed statistical data with exact p values are listed in Supplemental Table S1.

To display operant behavior patterns before and after the behavioral events, we collected reward-seeking and -taking behaviors during the last sessions of the RR20 and RI120 schedules. For reward-seeking behavior, we calculated the accumulated numbers of nose-poke in every 2 s (a bin). Moreover, for reward-taking behavior, we gave scores (0 if mice stayed at the magazine, otherwise 1 if mice stayed outside of

the magazine) as the probability of reward-taking behavior in each frame (30 frames per second). For the action-outcome (A-O) contingency, we created two behavioral data arrays from the last sessions of the RR20 and RI120 tasks: the action array for accumulated numbers of rewarded and non-rewarded nose-pokes and the outcome array for accumulated numbers of reward outcomes in every 60 s period [83]. Then, we calculated the Spearman *r* correlation coefficient between two arrays. The A-O contingency (%) = |Spearman *r* correlation coefficient| x 100. All the *r* values (from 8 mice) in the RR group were positive numbers. Three of seven mice in the RI group showed negative numbers of the *r* values.

For seeking-taking association, we calculated two types of the association at the last session of the RR20 and RI120 tasks: behavioral event-dependent seeking-taking association (**supplementary Fig. 3g**) and the seeking-taking association throughout the session (**supplementary Fig. 3h**). We created two arrays: the reward-seeking behavior array which is the accumulated numbers of total nose-poke and the reward-taking behavior array which is the averaged probability of taking behavior in every 2 s (60 frames) or 60 s (1800 frames) period. Every 2 s of the bin was for the behavioral event-dependent association, and every 60 s of the bin was for seeking-taking association throughout the session. Then, we calculated the Spearman *r* correlation coefficient. The seeking-taking association (%) = |Spearman *r* correlation coefficient| x 100.

Data availability. All data are available from the authors upon reasonable request.

Code availability. All code used in this manuscript is available at https://github.com/brain-machine-intelligence/Hong_2021.

Figures

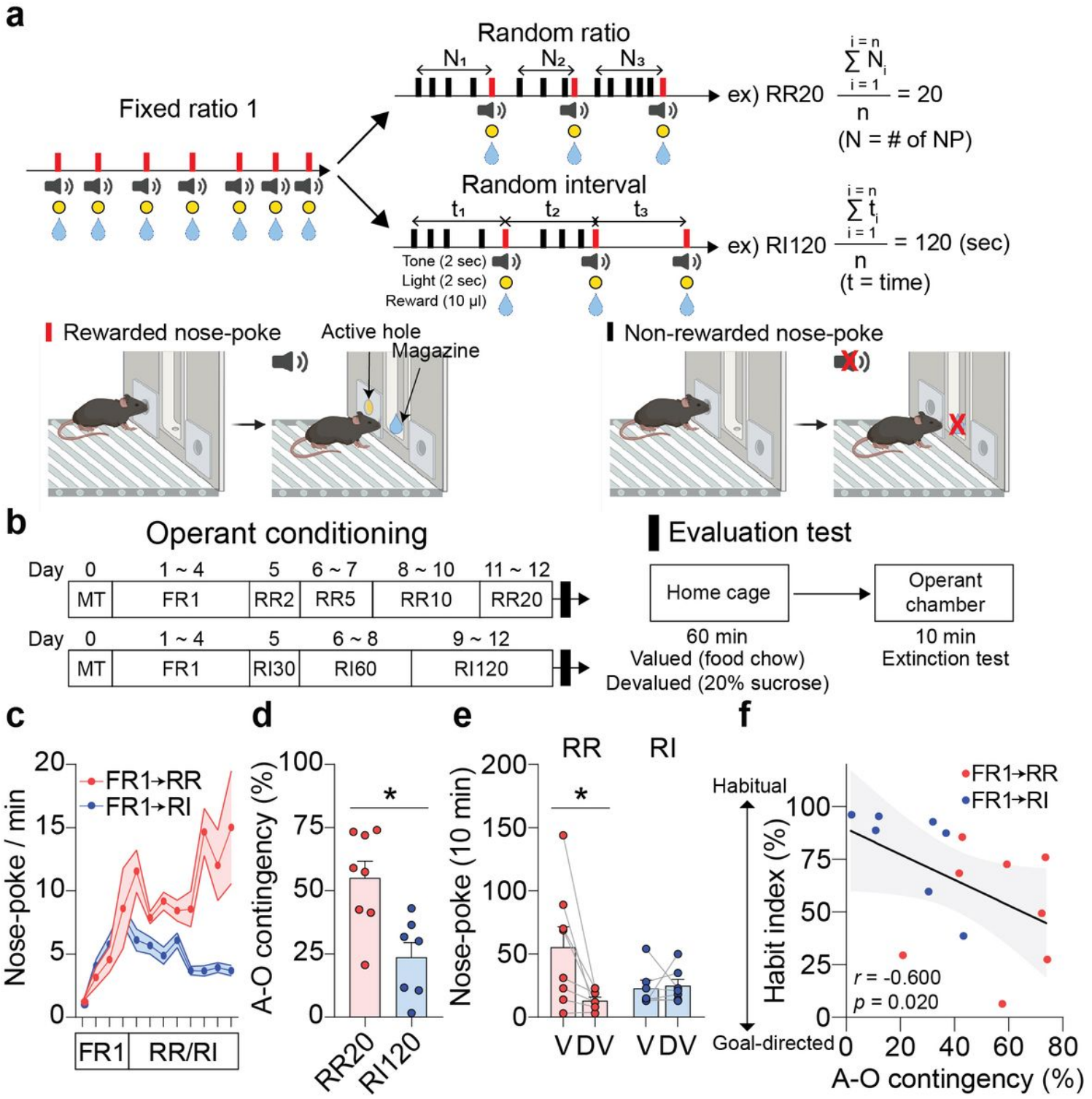

Figure 1

Reduced A-O contingency after habit learning. a, Task schemes for operant conditioning. b, Experimental schedules for operant conditioning and evaluation test. c, Nose-poking behaviors from the FR1 to RR20 or RI120 task. d, Action-outcome (A-O) contingency in the last sessions of RR20 and RI120 tasks. e, Nose-poke changes in the outcome evaluation tests after the RR or RI task. f, Correlation between A-O contingency and habit index. Data represented mean ± SEM; f, shadow indicated 95% confidence bands

of the best-fit line; c-f, n = 8/RR group, n = 7/RI group. *p < 0.05 comparing each group. d, two-tailed Mann-Whitney test; e, Wilcoxon matched-pairs signed-rank test; f, Spearman correlation analysis. A-O, action-outcome; DV, devalued; FR1, fixed ratio 1; MT, magazine training; RI, random interval; RR, random ratio; V, valued.

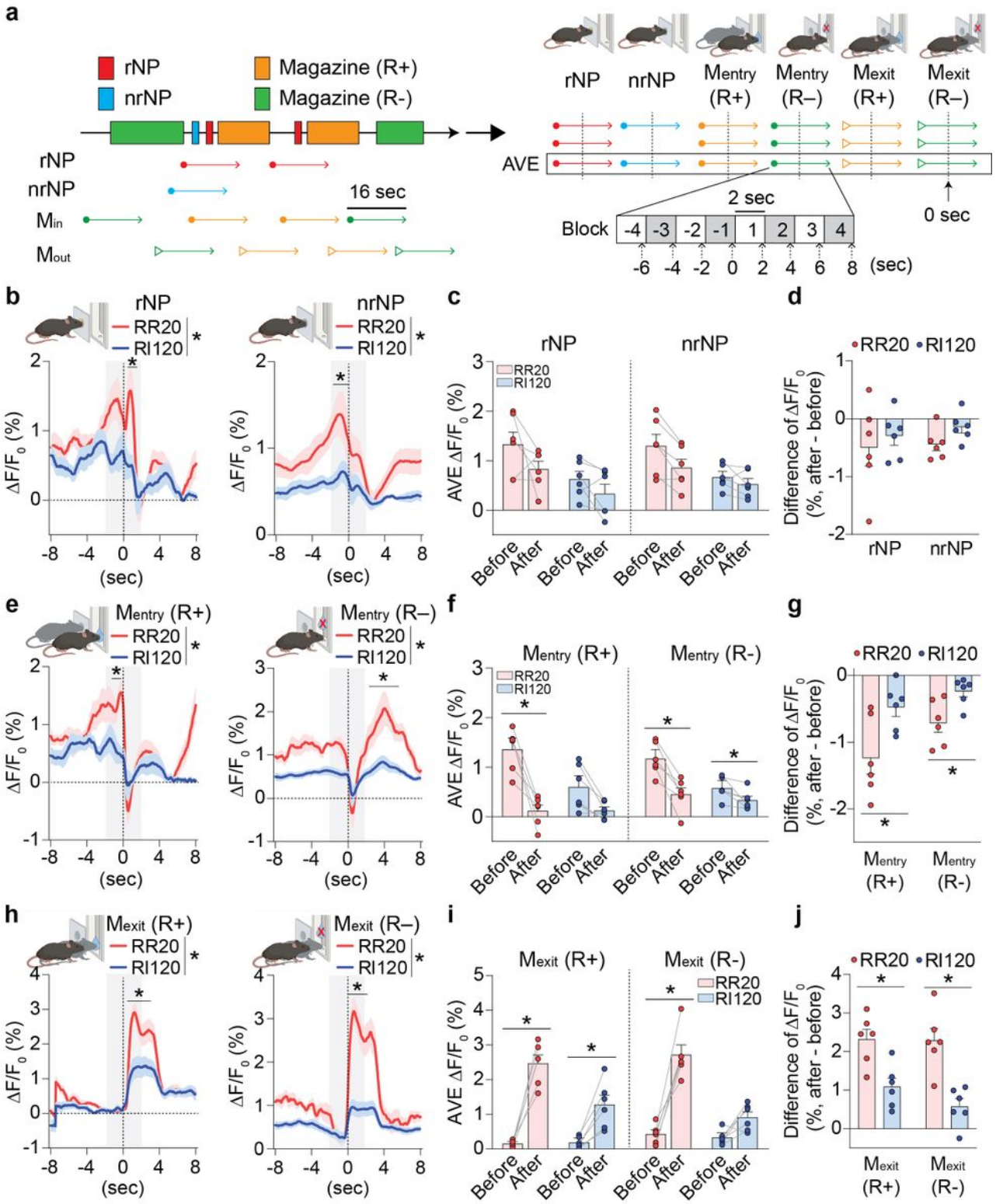

Figure 2

Astrocyte dynamics in the GPe during goal-directed and habitual reinforcement learning. a, Data collection method. We collected behavior and fiber-photometry data from 8 s before each behavior event to 8 s after each behavior event. Fluorescence changes around (b) rewarded nose-poke and non-rewarded nose-poke, (e) magazine entry with and without reward, (h) magazine exit with and without reward during the RR and RI tasks. Comparison of the average fluorescence change in the previous 2 s (before) and the next 2 s (after) based on behavioral events (c, nose-poke; f, magazine entry; i, magazine exit). Difference of fluorescence changes between before and after (d) rewarded nose-poke or non-rewarded nose-poke, (g) magazine entry with or without reward, (j) magazine exit with or without reward. Data represented mean ± SEM; n = 6/group. *p < 0.05 comparing each group. b, e, and h, Two-way repeated-measures ANOVA followed by Bonferroni's multiple comparisons tests; c, f, and i, two-tailed Wilcoxon matched-pairs signed-rank test; d, g, and j, two-way ANOVA followed by Tukey's multiple comparisons tests. Mentry, magazine entry; Mexit, magazine exit; nrNP, non-rewarded nose-poke; RI, random interval; rNP, rewarded nose-poke; RR, random ratio; R+, with reward; R-, without reward.

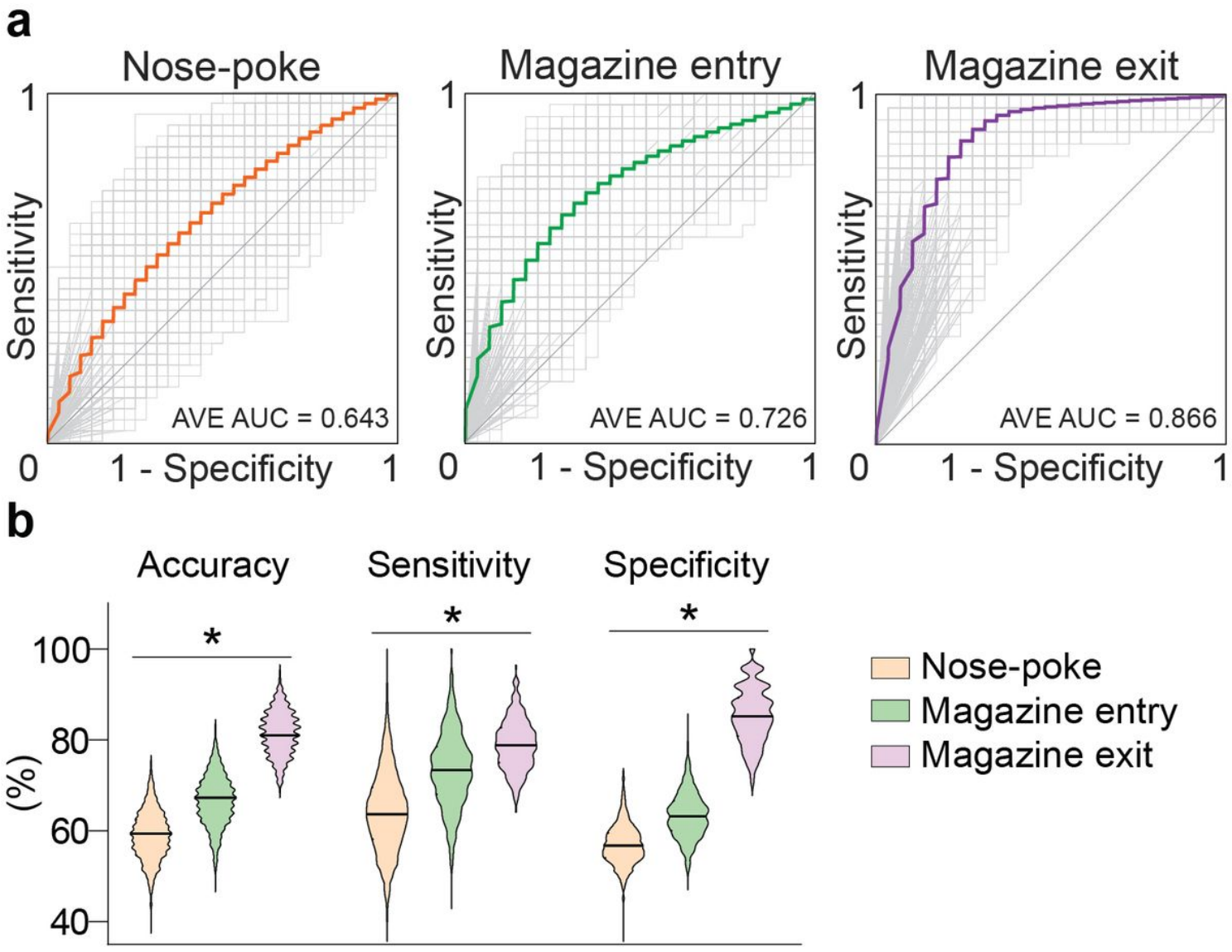

Figure 3

Predictable behavioral tasks by using GPe astrocyte dynamics of behavioral events. a, Averaged receiver operator characteristic (ROC) curves for prediction of behavioral tasks around behavioral events (nose-poke, magazine entry, and magazine exit). b, Comparison of accuracy, sensitivity, and specificity among the behavioral events. Data represented mean ± SEM; n = 1600/group. *p < 0.05 comparing nose-poke and magazine exit groups. Kruskal-Wallis test followed by Dunn's multiple comparisons test.

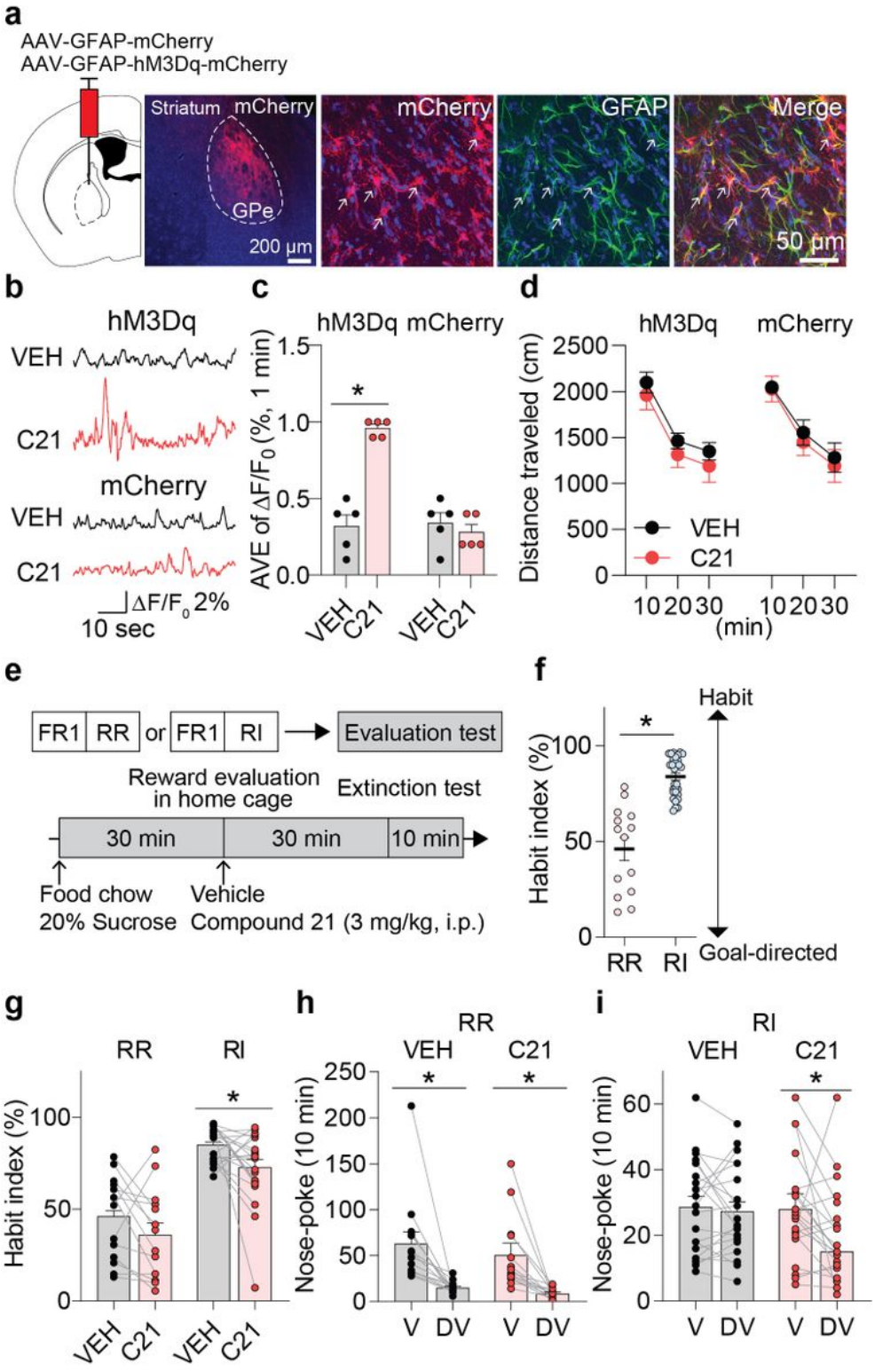

Figure 4

Effects of chemogenetic increase in GPe astrocyte activity on goal-directed behavior and habit. a, Histological validation of viral delivery of hM3Dq in GPe astrocytes. b and c, Functional validation of GPe astrocytic hM3Dq. d, No change in locomotor activity. e, Experimental scheme for operant conditioning. f, Habit index after RI or RR task. g, Habit index of the vehicle (VEH) or compound 21 (3 mg/kg, i.p.; C21) treatment group after RR or RI task. Nose-poke between valued and devalued states after h, RR or i, RI task. Data represented mean ± SEM; c, n = 5/group, d, n = 5/group, f-i, n = 14 for RR group, 21 for RI group. *p < 0.05 comparing each group. c and f, two-tailed Mann-Whitney test; g-i, two-tailed Wilcoxon matched-pairs signed-rank test. C21, compound 21; DV, devalued; FR1, fixed ratio 1; RR, random ratio; V, valued; VEH, vehicle.

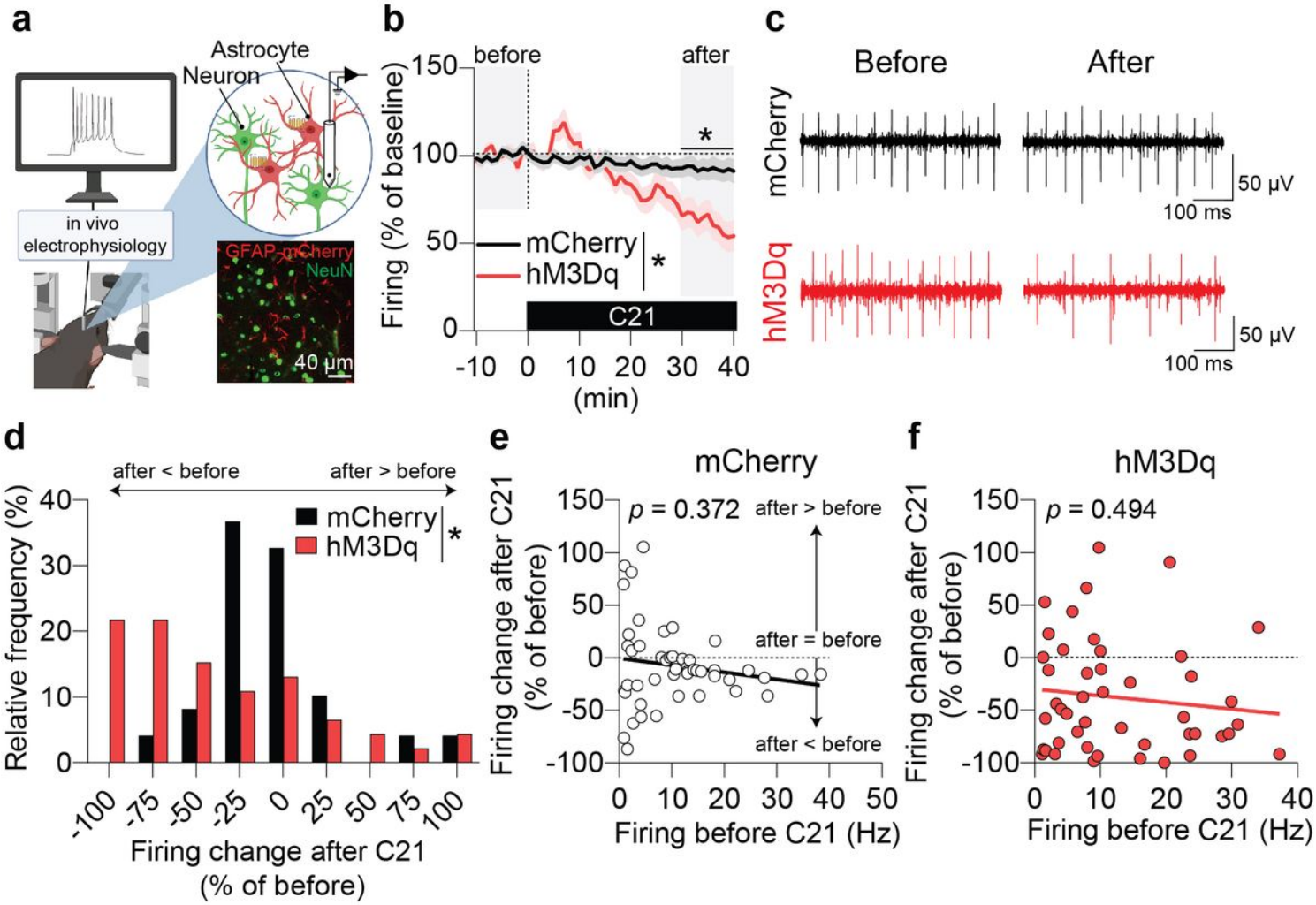

Figure 5

Effects of chemogenetic increase in GPe astrocyte activity on GPe neuronal activities. a, Experimental scheme, and protein expressions of GFAP promoter-driven mCherry and NeuN. b, Relative firing changes along with C21 treatment (3 mg/kg, i.p.) in mice who expressed non-hM3Dq (as the mCherry group) and hM3Dq in the GPe. c, Different firing patterns between before and after in the hM3Dq group. d, Different distributions of relative firing changes after C21 treatment between the mCherry and hM3Dq groups. Correlations between GPe neuronal firing before C21 treatment and relative firing changes after C21 treatment in (e) the mCherry and (f) hM3Dq groups. Data represented mean ± SEM; n = 49 from 5

mice/mCherry group, n = 46 from 5 mice/hM3Dq group. *p < 0.05 comparing each group. b, two-way repeated-measures ANOVA followed by Bonferroni's multiple-comparisons test; d, Kolmogorov–Smirnov test; e and f, Spearman correlation analysis. C21, compound 21.

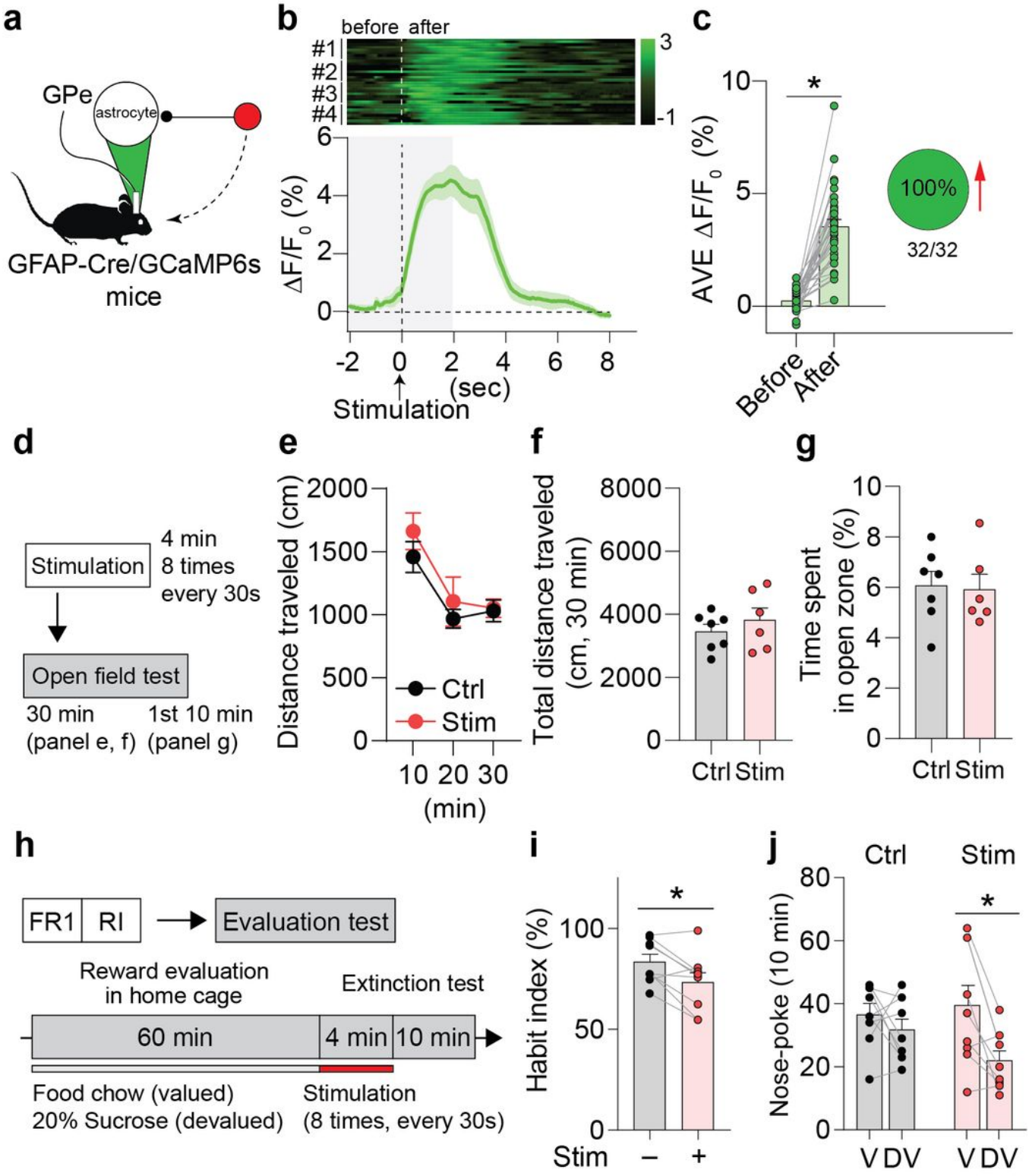

Figure 6

Effects of attentional stimulation on habits. a, Scheme for attentional stimulation. b, Activity changes of GPe astrocytes by attentional stimulation. c, Difference of ΔF/F0 between before and after (each 2 s)

attentional stimulation. d, Experimental scheme for an open field test with attentional stimulation. e and f, No change in locomotor activity for 30 min. g, No change in anxiety-like behavior. h, Experimental scheme for operant conditioning with attentional stimulation. i, Reduced habit index by attentional stimulation. j, Change in nose-poke between valued and devalued states by attentional stimulation. Data represented mean ± SEM; b and c, n = 32 from 4 mice/astrocyte group; e-g, n = 7/control group, n = 6/stimulation group; i and j, n = 10/group. c, i, and j, *p < 0.05 comparing each group. c, i, and j, two-tailed Wilcoxon matched-pairs signed-rank test, e, two-way repeated-measures ANOVA followed by Bonferroni's multiple-comparisons test, f and g, two-tailed Mann-Whitney test. Ctrl, control; FR1, fixed ratio 1; DV, devalued state; RI, random interval; Stim, stimulation; V, valued state.

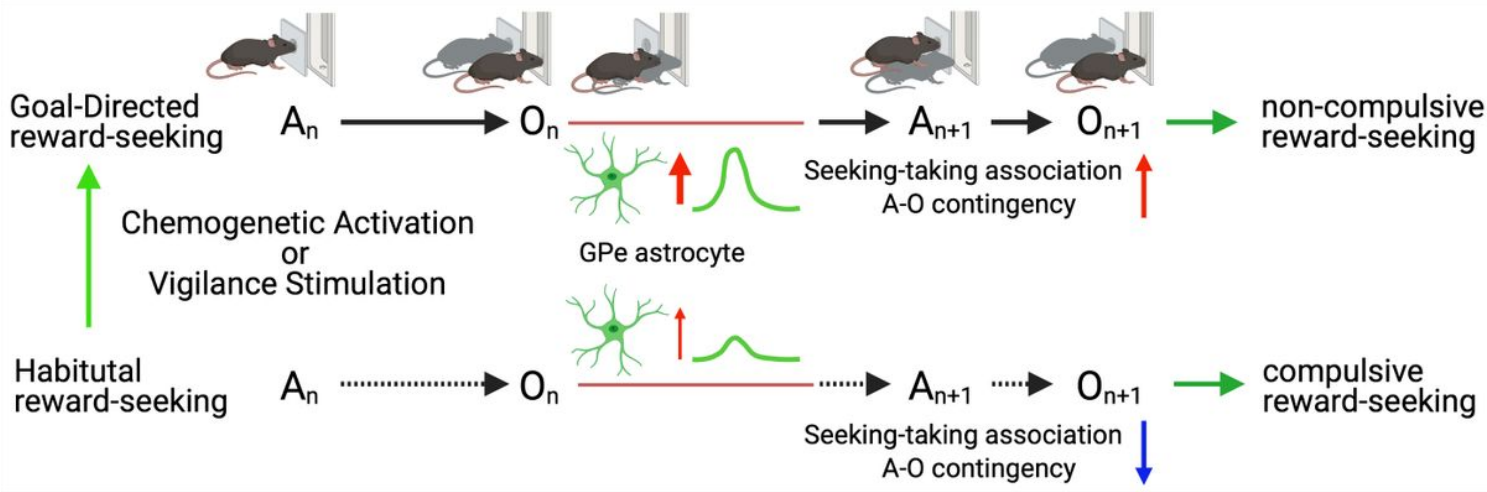

Figure 7

Schematics summarizing the results of this present study. With goal-directed behavior, action-outcome contingency is higher, and GPe astrocyte activity increases to plan the next action. Oppositely with the habit, action-outcome contingency is lower, and GPe astrocyte activity is reduced. Consequently, habit proceeds as compulsive reward-seeking. Thus, to prevent this, goal-directed behavior can be restored by increasing GPe astrocyte activity or giving vigilance stimulation, and as a result, non-compulsive reward-seeking can be performed to treat diseases related to addiction and maladaptive decision-making.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- GPeSupplFinal2.docx