Role of rodent secondary motor cortex in value-based action selection

Jung Hoon Sul[1], **Suhyun Jo**[1,2], **Daeyeol Lee**[3], and **Min Whan Jung**[1,2]

[1] Neuroscience Laboratory, Institute for Medical Sciences, Ajou University School of Medicine, Suwon 443-721, Korea

[2] Neuroscience Graduate Program, Ajou University School of Medicine, Suwon 443-721, Korea

[3] Department of Neurobiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA

Abstract

Despite widespread neural activity related to reward values, signals related to upcoming choice have not been clearly identified in the rodent brain. Here, we examined neuronal activity in the lateral (AGl) and medial (AGm) agranular cortex, corresponding to the primary and secondary motor cortex, respectively, in rats performing a dynamic foraging task. Choice signals arose in the AGm before behavioral manifestation of the animal's choice earlier than in any other areas of the rat brain previously studied under free-choice conditions. The AGm also conveyed significant neural signals for decision value and chosen value. In contrast, upcoming choice signals arose later and value signals were weaker in the AGl. We also found that AGm lesions made the animal's choices less dependent on dynamically updated values. These results suggest that rodent secondary motor cortex might be uniquely involved in both representing and reading out value signals for flexible action selection.

Value-based decision making consists of two broad steps of valuation and selection. Previous studies have shown value-related neuronal activity in a number of different brain structures such as striatum[1–3], parietal cortex[4–6], anterior cingulate cortex (ACC)[7–8], orbitofrontal cortex (OFC)[8–10] and other parts of the prefrontal cortex (PFC)[8,11] in rats and monkeys. Also, after the outcome of the animal's choice is revealed, neuronal activity related to reward prediction error (RPE), namely the difference between the actual and predicted rewards, is found in multiple brain areas including midbrain dopamine regions[12], OFC[8] and striatum[2]. These results indicate that valuation of action outcomes engages multiple brain structures.

The neural system responsible for making a choice (selecting a single action out of multiple alternatives) based on values of expected outcomes has been more elusive. Although previous studies have found future choice-related neural signals in several brain areas such as the striatum[1], dorsolateral PFC[13], supplementary eye field (SEF)[14] and parietal cortex[6,15–16] in monkeys during a free-choice task, choice-related signals in multiple brain structures do not necessarily indicate that they are all involved in the final action selection process. Whereas valuation can be processed in parallel, choice must involve a process that selects a single action out of multiple alternatives that is to be executed by the motor system. Thus, although multiple brain regions might display upcoming choice signals in a given behavioral setting, it is likely that action selection takes place in a specific neural system and then the resulting choice signals propagate to other systems for the purpose of executing or evaluating the chosen action. Therefore, in order to identify the neural system responsible for final action selection, it would be important to compare relative time courses of choice signals across different brain regions and to examine effects of local lesion or inactivation on choice behavior of animals under the same behavioral setting.

Another important question is how different components of value-based decision making, namely value representation, action selection and action evaluation, overlap across different brain structures. Given that many brain areas conveying value signals tend to encode upcoming actions chosen by the animal (see above), it is likely that these functions are served by partially overlapping brain structures. However, the exact distribution of these functions remains unclear, in part because only a few studies have examined all of these signals in multiple brain areas under the same experimental setting (e.g., refs. 8,11,17).

We investigated these issues in the rat brain. In our previous studies with rats[2,8,18], value-related neural signals have been found in the lateral OFC, medial PFC (dorsal ACC, prelimbic cortex and infralimbic cortex), dorsal striatum, and ventral striatum. By contrast, clear signals related to upcoming choice were found in none of these brain structures. We therefore hypothesized that upcoming choice signals would be found in a motor structure. The first aim of this study was to test this prediction, which was supported by the data. The second aim was to test whether upcoming choice signals co-exist with value signals at the final stage of decision making. Our results showed that the secondary motor cortex of the rat processes both action selection and value signals.

RESULTS

Choice behavior and movement trajectory

Three rats performed a dynamic foraging task[19] (Fig. 1a), choosing freely between two goals that delivered a fixed amount of water reward with different probabilities. Although reward probabilities were constant within a block of 35–45 trials (four blocks per session), water was delivered stochastically in each trial and no sensory stimuli signalled block transitions so that the animals could detect changes in reward probabilities only by trial and error. Nevertheless, the animals quickly detected changes in relative reward probabilities after a block transition and biased their choices toward the goal with a higher reward probability, which was well described by a reinforcement learning (RL) model (Fig. 1b). The animals obtained rewards in 61.7±4.5% (mean±SD) of trials, which is significantly

lower than the amount of rewards expected for the RL model with optimal parameters ($72.1 \pm 2.3\%$; t-test, $p < 0.001$). A logistic regression analysis showed that past rewards up to two previous trials biased the animals to repeat the same goal choice, with the reward from the most recent trial having the greatest effect. There was also a tendency for the animals to alternate their goal choices (Fig. 1d).

A trial was divided into the delay, go, approach to reward, reward consumption and return stages as in our previous studies[2,8] (Fig. 1a). The animals showed a stereotyped pattern of movements on the maze (Fig. 1c). The animal's movement trajectory during the early delay stage reflected the previous goal choice (initial 0.78 ± 0.16 s of the delay stage), but this was not the case during subsequent behavioral stages. The onset of the approach stage was when the animal's movement trajectory began to diverge depending on the animal's upcoming goal choice (i.e., first behavioral manifestation of the animal's goal choice; Supplementary Fig. 1 online). The analysis of the animal's movement trajectory revealed, as in our previous study (see supplemental Fig. S6 of ref. 2), that the movement trajectory did not vary with the animal's future goal choice in any of the behavioral stages before the onset of the approach stage.

Neural signals for choice

Single units were recorded simultaneously from rostral parts of the lateral and medial agranular cortex (AGl and AGm, respectively; Fig. 2) in the right hemisphere of the three rats. A total of 227 and 411 well-isolated single units (500 spikes during each recording session) were recorded from the AGl and AGm, respectively. We first analyzed neural activity related to the animal's choice, its outcome and their interaction in the current and two preceding trials using a multiple regression analysis. Fig. 3a shows the fractions of neurons that significantly (according to eq. 1, see Online Methods) modulated their activity according to these behavioral variables in non-overlapping 500-ms time windows across different behavioral stages. Neural signals for the animal's upcoming choice in the current trial [$C(t)$] were low during the delay stage, but began to increase during the go stage first in the AGm and then in the AGl (Fig. 3a). By analyzing neural activity with a higher temporal resolution, we found that the onset of the upcoming choice signal (see Online Methods for its definition) was approximately 500 and 150 ms before the onset of the approach stage in the AGm and AGl, respectively, and that the fraction of choice-encoding neurons was significantly larger in the AGm both before and after the onset of the approach stage (Fig. 3b). A similar pattern was observed when the time course of the upcoming choice signal was examined by decoding the animal's goal choice from population activity (Fig. 3c). The distribution of the choice signal onset for individual neurons was also different between the AGm and AGl (Fig. 3d,e). Among a total of 95 AGm neurons showing significant choice-related activity during the 500-ms time period before the approach stage onset, 48 (50.5%) and 47 (49.5%) discharged at higher rates in the left and right goal-choice trials, respectively, which is not significantly different (χ^2-test, $p=0.918$). The corresponding numbers of AGl neurons were 16 (55.2%) and 13 (44.8%), respectively, which is not significantly different, either ($p=0.732$).

Large fractions of neurons encoded signals for the animal's chosen action [$C(t)$] after the animal revealed its choice (approach, reward and return stages; Fig. 3a). It should be noted that sensory and motor events associated with left and right goal choices were different during these subsequent behavioral stages. Neural signals for already chosen actions might reflect the animal's choice or sensorimotor features (such as movement direction) that were different for the left vs. right sides of the maze. Many neurons encoded reward [i.e., choice outcome, $R(t)$] or chosen action×reward interaction [$X(t)$] after the outcome of the animal's choice was revealed (reward stage; Fig. 3a). These signals, especially chosen action signals, persisted until the next trial so that low, but significant levels of previous choice [$C(t–1)$], previous reward [$R(t–1)$], and previous interaction [$X(t–1)$] signals were observed (Fig. 3a). Similar results were obtained when the analysis was based on the magnitudes of regression coefficients rather than the fractions of significant neurons (Supplementary Fig. 2a online). Example neurons that modulated their activity according to the animal's choice or reward in the previous trial are shown in Supplementary Fig. 3 online.

Neural signals for decision value

Neural signals related to valuation process were examined by relating neural activity to action values that were estimated with a model-based RL algorithm[19] (Supplementary Note online). The RL algorithm predicted the animal's actual choices well (Fig. 1b), suggesting that the animal's subjective values for alternative actions were reliably estimated by this model. We examined neural signals for decision value (ΔQ), which was defined as the difference between the left and right action values ($Q_L–Q_R$), and chosen value (Q_c), which was the value of chosen action in a given trial[8] (eq. 2). Decision value and chosen value would be useful in deciding which goal to choose and evaluating the value of chosen action, respectively. Decision value signals fluctuated around the chance level in the AGl. The AGm conveyed significant, but still weak decision value signals so that the difference in the strength of decision value signals was not large between the AGm and AGl (Fig. 4a). Given that decision value signals are also weak in the OFC, medial PFC and striatum in rats[2,8], this finding suggests that decision value signals might be only weakly represented as persistent activity in multiple areas of the rat brain. An example AGm neuron that modulated its activity according to decision value is shown in Supplementary Fig. 3 online.

We also examined neural signals for action value by replacing decision value with the left and right action values in the regression (eq. 2). Unlike decision value signals, action value signals were below chance level during the delay stage in the AGm (Fig. 4b). There were only 14 AGl neurons (out of 227, 6.2%; binomial test, $p=0.248$) and 24 AGm neurons (out of 411, 5.8%; binomial test, $p=0.246$) that significantly modulated their activity according to the value of at least one action (Q_L or Q_R; $p < 0.025$, alpha=0.05 was corrected for multiple comparisons) during the last 1 s of the delay stage. The fraction of action value-encoding neurons during the last 1 s of the delay stage (5.8%) was significantly lower than that of decision value-encoding neurons (11.2%; χ^2-test, $p=0.006$) in the AGm.

Neural signals for chosen value

Chosen value signals arose after the animal revealed its choice (approach stage) and then slowly decayed during the reward stage in the AGm (Fig. 4). Chosen value signals in the

AGl were significantly weaker than in the AGm. During the 1-s period centered at the onset of the reward stage, 66 AGm neurons (16.1%; binomial test, $p < 0.001$) and 17 AGl neurons (7.5%; binomial test, $p=0.065$) modulated their activity according to chosen value, which were significantly different (χ^2-test, $p=0.002$). Similar results were obtained when the analysis was based on the magnitudes of regression coefficients (supplementary Fig. 2b online). For some neurons in the AGm and AGl, chosen value signals changed significantly depending on the animal's chosen action; 43 (10.5%) AGm and 24 (10.6%) AGl neurons showed significant chosen action×chosen value interaction (Supplementary Fig. 4 online).

To examine whether chosen value signals are combined with other signals necessary to evaluate chosen action, we examined neural signals for the animal's chosen action, reward, and chosen value around the time of reward stage onset at a higher temporal resolution (100-ms time steps). All three signals were robustly present in the AGm during the early reward stage (Fig. 5a). Of 36 AGm neurons (out of 411, 8.8%) that concurrently conveyed signals for reward and chosen value during the first 1 s of the reward stage, 26 (72.2%) also conveyed neural signals for the animal's chosen action (Fig. 5b). Thus, some AGm neurons conveyed all the signals necessary to compute RPE and update the value of chosen action. It should be noted that the likelihood of false negative (type II error) increases when multiple tests are applied conjunctively; the expected number of AGm neurons that encode both reward and chosen value by chance is only 1.0 among 411 neurons. Further analyses suggested that these signals are combined to compute RPE as well as to update the value of chosen action in the AGm (Supplementary Fig. 5).

Effects of AGm lesions

Coding of both choice and value signals in the AGm suggests that this area might play a critical role in value-based action selection. We further tested this possibility by examining behavioral effects of bilateral lesions of the rostral AGm (Fig. 6a). AGm lesions induced no significant changes in the animal's running speed (Fig. 6f) or choice bias (Fig. 6g). However, both short-lasting and long-lasting changes were observed in several related measures of the animal's choice behavior (see Supplementary Table 1 online for the results of statistical analyses). The animal's tendency to repeat the same goal choice after failing to obtain a reward (lose-stay) increased transiently during 3 to 5 days following lesions (Fig. 6b,d,e). By contrast, a decrease in the probability of choosing the higher arming-probability goal persisted for 10 days (Fig. 6c,h).

Long-lasting changes were also observed in several RL model-related measures. The inverse temperature (β) of the RL model decreased significantly following lesions, whereas there was no significant change in the learning rate (α; Fig. 6i,j). In addition, the lesioned animals were more likely to choose the goal that was associated with a lower action value than the alternative (Fig. 6k). In particular, they were more likely to repeat the goal choice that was associated with a lower action value (Fig. 6l), whereas the likelihood to switch their choices to a lower-value goal was unaffected (Fig. 6m). Finally, the goodness of fit for the SP model (negative log-likelihood normalized by the number of trials) and the accuracy of model prediction for the animal's choices (% correct) were significantly reduced following the

lesions (Fig. 6n,o). No significant changes in any of these parameters were observed following sham lesions except for a reduced proportion of lose-stay trials (Fig. 6b–o).

DISCUSSION

We examined neuronal activity related to choice and value in two regions of the rodent motor cortex. Neural signals for the animal's upcoming goal choice arose in the rostral AGm several hundred milliseconds before behavioral manifestation of the animal's goal choice. The AGm also conveyed significant decision value and chosen value signals before and after choice was made, respectively. By contrast, these neural signals arose later and less frequently in the AGl. Behaviorally, selective AGm lesions altered animal's choice behavior so that it is less dependent on values estimated from the animal's experience. Taken together, these results suggest that the AGm, but not the AGl, is involved in both valuation and selection of animal's voluntary actions.

Future-choice signals in the rat brain

Our previous investigations using free-choice tasks did not find clear preparatory signals related to action selection in the medial PFC (ACC, prelimbic cortex and infralimbic cortex), lateral OFC, dorsal striatum, or ventral striatum in rats[2,8] (Fig. 7a). Choice behavior of the animals in these studies was well accounted for by RL models, suggesting that the failure to find clear choice signals was not due to random and impulsive choices of the animals. Although a previous study demonstrated upcoming action selection signals in the superior colliculus of rats, the task used in that study was not a free-choice, but an odor discrimination task, and the earliest onset of choice-related activity was only ~300 ms before movement initiation[20]. Thus, upcoming choice signals found in the rostral AGm appear to be the earliest unambiguous action selection signals discovered so far in the rat brain under a free-choice condition. It is notable that the rostral AGm sends direct projections to the superior colliculus[21], raising the possibility that action selection signals observed in the superior colliculus might originate in the AGm. Disruptive effects of AGm lesions on the animal's choice behavior also support this possibility. Because our results were obtained from freely moving animals, we cannot completely rule out the possibility that the advanced action selection signals in the AGm were due to subtle, undetected variations in the animal's behavior. Nevertheless, the physiological and behavioral results obtained in the present study provide converging evidence for a primary role of the rostral AGm in action selection during a dynamic foraging task.

AGm and supplementary motor area

The AGl is regarded as a major part of the primary motor cortex because it receives strong projections from the ventrolateral, but not mediodorsal nucleus of the thalamus[22], sends direct projections to the spinal cord[23–24], and evokes movements when stimulated with low intensity electrical pulses[23,25–26]. On the other hand, the AGm has been proposed to be homologous to the premotor area (PMA), supplementary motor area (SMA), and/or frontal/ supplementary eye field in primates[21,23,25–29]. The rostral part of the AGm, in particular, has frequently been proposed to be homologous to the SMA in primates[23,25–27] because it receives inputs from both ventrolateral and mediodorsal nucleus of the thalamus[22,30], sends

strong projections to the primary motor cortex (i.e., AGl)[21–22], contains a second motor representation of the body[26,31], and, while sending direct projections to the spinal cord[24], does not readily evoke movements when stimulated electrically[23,25–26], all of which are characteristics of the SMA in primates[32]. Thus, although the homology between rat and primate motor cortical areas is not entirely clear[33], the rodent rostral AGm and primate SMA appear to share multiple anatomical and functional characteristics.

It is notable that substantial evidence highlights a critical role for the primate SMA in action selection under free-choice conditions[32,34]. SMA lesions often cause an inability to generate spontaneous movements of the contralateral limb, and sometimes cause the alien-limb syndrome in which the contralateral arm produces unwanted spontaneous movements such as grasping nearby objects. Physiological studies have shown discharges of SMA neurons and readiness potentials over the SMA before movement onset, and brain imaging studies have found activity in the SMA and pre-SMA during a free-choice task[32]. Similarly, neural signals for upcoming saccades during a free-choice task arose earlier in the SEF than in the frontal eye field and lateral intraparietal cortex[14]. Collectively, these and our results suggest that the supplementary motor regions (SMA, pre-SMA and SEF) in monkeys and the rostral AGm in rats might be part of the neural system where the future action is selected and propagated to downstream motor structures, such as the primary motor cortex and superior colliculus, for execution under free-choice conditions. Considering that different neural systems might be in charge of final action selection under different behavioral conditions[34–35], it will be important for future studies to compare relative time courses of neural signals for upcoming action across different parts of the rodent brain during different behavioral tasks.

Convergence of choice and value signals in the AGm

Our results indicate the involvement of the rostral AGm not only in action selection, but also in valuation, which is consistent with the finding that AGm activity is modulated by expected reward[36]. The SMA and PMA in monkeys also convey upcoming choice-[32,37] as well as expected reward-related signals[38–40]. Similarly, neurons in the primate superior colliculus convey both choice- and value-related signals[41]. Thus, co-existence of choice and value signals in motor structures might be a common feature across different species. However, the findings that expected outcome-dependent signals were stronger in the posterior parts of the lateral frontal cortex[38] and similarly modulated by reward and penalty unlike in the OFC[39] led to the suggestion that such signals might reflect motivation-regulated motor preparation rather than subjective values of specific actions. In our task, the level of motivation before the animal's choice is presumably similar across trials, because the sum of reward probabilities was held constant. Therefore, it is difficult to explain decision value signals, which were found well before upcoming choice signals, in terms of motivation-related movement preparation. Further analyses also showed that chosen value signals are unlikely to represent motivation-dependent motor preparation (Supplementary Fig. 6 online). These results suggest that value-related neural signals observed in the rostral AGm represent action values rather than movement preparation signals that are modulated by motivation.

Role of AGm in value-based action selection

The results obtained from the lesion experiments are also consistent with neurophysiological findings. The animals showed no obvious motor deficits following AGm lesions, indicating that the AGm is not involved in direct control of motor output. However, the lesioned animals were less likely to switch their choices after failing to obtain a reward (lose-switch) and to choose the goal with a higher reward probability, suggesting a deficit in the process of modifying future choices adaptively according to previous choice outcomes. When a RL model was applied to the choice data, learning rate did not change with AGm lesions, suggesting that the animal's ability to update values was not completely lost, which is consistent with widespread value signals found in multiple areas of the rat brain[2,8]. On the other hand, the lesioned animals showed increased randomness in action selection (i.e., low inverse temperature), indicating that their action selection became less dependent on values. Similarly, the AGm-lesioned rats tended to repeat a goal choice that was associated with a lower action value than the alternative, indicating that AGm-lesioned animals failed to adjust action selection according to altered action values.

Interestingly, the effect of AGm lesions on lose-switch was temporary, whereas the effects of lesions on other parameters lasted for the entire post-lesion sessions, suggesting that action selection functions of the AGm might be partially taken over by other brain structures. Perhaps relatively simple choice behavior such as lose-switch can be resumed by other brain structures such as the basal ganglia[42], whereas more elaborate value-based action selection requires the AGm. Collectively, our results suggest that the AGm plays a critical role in flexible action selection based on internally represented values.

Relationship with other brain structures

The valuation process in the AGm might not be independent of other brain structures. In rats, lesions to either the OFC or dorsal striatum impair reversal learning, and lesions to the OFC also impair reinforcer devaluation[43–44]. Thus, the AGm is unlikely to operate as a stand-alone module to serve valuation and choice functions. We found that the strength of decision value signals increases as one moves from the OFC to mPFC and AGm, whereas the opposite pattern was observed for the strength of action value signals (Fig. 7b). This raises the possibility that absolute value signals are transformed into relative value signals in the structures of the rat brain that are involved in action selection, which might also be the case in monkeys[45]. The AGm might require afferent action value signals from such brain structures as the OFC and dorsal striatum in order to compute decision value signals.

ONLINE METHODS

Behavioral task

Animals were trained in a free binary choice task on a modified figure 8-shaped maze (Fig. 1a) as in our previous studies[2,8,19]. The animal had to navigate from the central stem to one of the two goal sites to obtain water reward (30 μl) in each trial. The reward was controlled by a concurrent variable-ratio/variable-ratio reinforcement schedule, in which each choice contributed to the ratio requirement of both goals[19]. Thus, in each trial, reward could be available at neither, either, or both goals. If a particular goal was baited with reward, it

remained available, although it did not accumulate, in the subsequent trials until the animal visited there (referred to as 'dual assignment with hold' task)[47–48]. The animals performed four blocks of trials in each recording session. The number of trials in each block was 35 plus a random number drawn from a geometric mean of 5 and truncated at 45. Reward probability of a goal was constant within a block of trials, but changed across blocks without any sensory cues. The animals therefore had to detect changes in relative reward probabilities by trial and error. The following four combinations of reward probabilities were used in each session: 0.72:0.12, 0.63:0.21, 0.21:0.63 and 0.12:0.72. The sequence of reward probabilities was determined randomly with the constraint that the option with the higher-reward probability always changed its location at the beginning of a new block. The animal's head position was monitored by tracking an array of light-emitting diodes mounted on the headstage at 60 Hz. The experimental protocol was approved by the Institutional Animal Care and Use Committee of the Ajou University School of Medicine.

Behavioral stages

Each trial consisted of the delay, go, approach to reward, reward consumption, and return stages (Fig. 1a) as in our previous studies[2,8]. A trial began when the animal returned from a goal site to the central stem via the lateral alley and broke the central photobeam (Fig. 1a). A 2-s delay was imposed at the outset of each trial by raising the central connecting bridge (delay stage). The bridge was lowered at the end of the delay stage allowing the animal to move forward (go stage). The approach stage was the time period during which the animal turned and ran toward either goal site on the upper alley. The onset of the approach stage was determined separately for each behavioral session as the time when the left-right positions became significantly different for the left- and right-choice trials (*t*-test, $p < 0.05$) for the first time near the upper branching point (see Supplementary Fig. 1 online)[2,8]. Thus, the onset of the approach stage marked the first behavioral manifestation of the animal's goal choice at the branching point. The reward stage began when the animal broke a photobeam that was placed 6 cm in front of the water delivery nozzle, which triggered an immediate delivery of water in rewarded trials. The return stage began when the animal broke a photobeam that was placed 11 cm away from the water delivery nozzle, and ended when the animal returned to the central stem and broke the central photobeam (i.e., beginning of the next trial). The mean durations of the five behavioral stages were (mean ±SD) 2.00±0.00 (delay), 0.70±0.18 (go), 0.83±0.22 (approach), 6.13±1.50 (reward), and 1.97±0.23 s (return stage). The mean durations of the reward stage for the rewarded and unrewarded trials were 8.56±1.95 and 2.23±0.89 s, respectively. Thus, the animals stayed longer in the reward area in rewarded than unrewarded trials, although the animals licked the water-delivery nozzle in most unrewarded trials as in rewarded trials. Therefore, we cannot exclude the possibility that movement-dependent neuronal activity might have contributed to reward-related neuronal activity during the reward stage.

Logistic regression analysis

Effects of previous choices and their outcomes on the animal's goal choice were estimated using the following logistic regression model[2,48]:

$$\log\left(\frac{p_L(i)}{p_R(i)}\right) = \sum_{j=1}^{10} \gamma_j^r (R_L(i-j) - R_R(i-j)) + \sum_{j=1}^{10} \gamma_j^c (C_L(i-j) - C_R(i-j)) + \gamma_0,$$

where $P_L(i)$ [or $P_R(i)$] is the probability of selecting the left (or right) goal in the i-th trial. The variables $R_L(i)$ (or $R_R(i)$) and $C_L(i)$ (or $C_R(i)$) are reward delivery at the left (or right) goal (0 or 1) and the left (or right) goal choice (0 or 1) in the i-th trial, respectively. The coefficients γ_j^r and γ_j^c denote the effect of past rewards and choices, respectively, and γ_0 is a bias term. The numbers of total trials used in the regression for the three animals were 2,070, 1,907 and 2,112.

Reinforcement learning model

Our previous study[19] showed that the animal's choice behavior in the current task is better explained by a model-based than a model-free RL algorithm. We therefore employed a model-based RL algorithm that takes into consideration that the reward probability of the unchosen target increases as a function of the number of consecutive alternative choices[19] ('stacked probability' or SP algorithm; see Supplementary Note online). The SP algorithm was indeed superior in explaining animal's choice behavior to a model-free RL algorithm that changed only the value of chosen (but not unchosen) action (see Supplementary Table 2 online). However, results from the analyses of neural data were similar regardless of which RL algorithm was used (see Supplementary Fig. 7 online). Parameters α (learning rate) and β (inverse temperature) were estimated for the entire dataset from each animal using a maximum likelihood procedure[2,8]. The values of α and β for three animals were 0.58 and 2.02, 0.55 and 2.04, and 0.32 and 2.59, respectively. To compare the animal's performance to an optimal learner, we determined the optimal values for the parameters α and β of the SP algorithm by selecting the combination that yielded the maximal amount of rewards (average of 10 simulations) among 400 pairs of α (0.05 to 1.0 in steps of 0.05) and β (0.25 to 5.0 in steps of 0.25) values.

Unit recording

Two sets of six tetrodes were implanted in the right AGm (2.7 – 3.2 mm A and 1.0 mm L to bregma) and right AGl (2.7 – 3.2 mm A and 2.7 mm L to bregma; Fig. 2) of three well-trained rats under deep sodium pentobarbital anaesthesia (50 mg/kg). After at least one week of surgery recovery time, tetrodes were gradually lowered to obtain isolated action potentials from single neurons. Once the recording began, tetrodes were advanced for a maximum of 75 μm per day. Unit signals were amplified ×10,000, filtered between 0.6–6 KHz, digitized at 32 KHz and stored on a personal computer using a Cheetah data acquisition system (Bozemann, MT, USA). Unit signals were also recorded with the animals placed on a pedestal before and after each experimental session to examine stability of recorded unit signals. When recordings were completed, small marking lesions were made by passing an electrolytic current (50 μA, 30 s, cathodal) through one channel of each tetrode and recording locations were verified histologically[49].

Isolation and classification of units

Single units were isolated by manually clustering spike waveform parameters (MClust 3.4, Redish; see Supplementary Fig. 8a online). Recorded units were classified into two categories of broad-spiking neurons (putative pyramidal cells) and narrow-spiking neurons (putative interneurons; see Supplementary Fig. 8b online). The majority of the analyzed units were broad-spiking neurons (AGm: n=350, 85.2%; AGl: n=175, 77.1%). Although both types of neurons were included in the analyses, essentially the same results were obtained when narrow-spiking neurons were excluded from the analyses (data not shown).

Multiple regression analysis

Modulation of neuronal activity according to the animal's choice or its outcome was analyzed using the following regression model:

$$S(t){=}a_0{+}a_1C(t){+}a_2C(t{-}1){+}a_3C(t{-}2){+}a_4R(t){+}a_5R(t{-}1){+}a_6R(t{-}2){+}a_7X(t){+}a_8X(t{-}1){+}a_9X(t{-}2){+}\varepsilon(t),$$

(eq. 1)

where $S(t)$ indicates spike discharge rate, $C(t)$, $R(t)$ and $X(t)$ represent the animal's choice (left or right; dummy variable, -1 or 1), its outcome (reward or no reward; dummy variable, -1 or 1) and their interaction (dummy variable, -1 or 1), respectively, in trial t, $\varepsilon(t)$ is the error term, and $a_0 \sim a_9$ are the regression coefficients.

Neuronal activity related to values was examined using the following regression model as in our previous study[8]:

$$S(t){=}a_0{+}a_1C(t){+}a_2R(t){+}a_3X(t){+}a_4\Delta Q(t){+}a_5Q_c(t){+}a_6C(t{-}1){+}a_7R(t{-}1){+}A(t){+}\varepsilon(t),$$

(ep. 2)

where $\Delta Q(t)$ and $Q_c(t)$ denote decision value and chosen value that were estimated using the SP algorithm (or Rescorla-Wagner rule for the analyses shown in Supplementary Fig. 7 online), respectively. A slow drift in the firing rates can potentially inflate the estimate of value-related signals[8]. To control for this, the above model included a set of autoregressive terms, indicated by $A(t)$ that consisted of spike discharge rates during the same epoch in the previous three trials as the following:

$$A(t){=}a_{10}S(t-1){+}a_{11}S(t-2){+}a_{12}S(t-3),$$

where $a_{10} \sim a_{12}$ are regression coefficients.

Population decoding of goal choice

We examined how well activity of a simultaneously recorded neuronal ensemble predicted animal's goal choice using a template matching procedure with leave-one-out cross-validation[50] (Fig. 3c). The analysis was applied to the activity of neuronal ensembles containing 5 simultaneously recorded neurons after matching the distributions of neuronal ensemble size for the two regions by randomly dropping neurons from oversized AGm ensembles [size of analyzed ensembles, 5–21 neurons; 10.1±5.0 (mean±SD)]. We then calculated the percentage of trials neuronal ensemble activity during a 100-ms sliding time

window (50-ms time steps) correctly predicted the animal's goal choice (% correct decoding).

Determination of choice signal onset

The latency of upcoming choice signals was determined for the entire neural population as well as for individual neurons. For the neural population, we examined the fraction of neurons encoding the current choice signal using a multiple regression analysis (eq. 1) at a fine temporal resolution (a moving window of 100 ms that was advanced in 50-ms time steps; Fig 3b). The onset of the upcoming choice signal was determined as the first time point where the fraction of upcoming choice-encoding neurons exceeded and remained significantly higher than chance level (binomial test, alpha=0.05) for a minimum of 250 ms (5 bins) within 1 s time periods before and after the approach onset[8]. For individual neurons, we examined whether neuronal activity associated with the left- vs. right-choice trials during a 100-ms sliding time window (50-ms time steps) was significantly different based on a *t*-test (alpha=0.05). The onset of the upcoming choice signal was determined as the first time point where the left- vs. right-choice-associated neuronal activity was significantly different for a minimum of 150 ms (3 bins) within 1 s time periods before and after the approach onset (Fig. 3e).

Test of AGm lesion effects

Twelve additional animals were trained in the same task for 20 days as described in our previous study[19]. After testing the animals for additional 10 days to establish pre-lesion baseline performance (148–178 trials per session), 0.1 M quinolinic acid (lesion group, n=6 animals) or 0.9 % saline (vehicle; sham lesion group, n=6 animals) was infused into five locations of the rostral AGm (2.7 mm A and 1 mm L to bregma, 0.2 μl; 3.2 A and 1 L, 0.2 μl; 3.7 A and 1.2 L, 0.2 μl; 4.2 A and 1 L, 0.15 μl; 4.2 A and 3 L, 0.15 μl) in each hemisphere under deep sodium pentobarbital anaesthesia (50 mg/kg). Following one week of recovery from surgery, the animals were tested again in the same task (148–173 trials per session) for 10 days. The animals were then overdosed with sodium pentobarbital and their brains were processed according to a standard histological procedure as previously described[49]. The extents of lesions were determined based on light microscopic examinations of histological sections (40 μm thick) that were stained with cresyl violet. One animal was discarded from the lesion group because of incomplete lesions (a very small lesion was detected on only one side of the brain) and one sham animal died during surgery, leaving five animals for each group.

Statistical analysis

Statistical significance of a regression coefficient was determined using a *t*-test, and statistical significance of the fraction of neurons for a given variable with a binomial test. Effects of AGm lesions on behavioral parameters were tested with two-way repeated measure ANOVA followed by paired *t*-tests. A *p* value < 0.05 was used as the criterion for a significant statistical difference unless noted otherwise. Data are expressed as mean±SEM unless noted otherwise.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Samejima K, Ueda Y, Doya K, Kimura M. Representation of action-specific reward values in the striatum. Science. 2005; 310:1337–1340. [PubMed: 16311337]

2. Kim H, Sul JH, Huh N, Lee D, Jung MW. Role of striatum in updating values of chosen actions. J Neurosci. 2009; 29:14701–14712. [PubMed: 19940165]

3. Lau B, Glimcher PW. Value representations in the primate striatum during matching behavior. Neuron. 2008; 58:451–463. [PubMed: 18466754]

4. Platt ML, Glimcher PW. Neural correlates of decision variables in parietal cortex. Nature. 1999; 400:233–238. [PubMed: 10421364]

5. Sugrue LP, Corrado GS, Newsome WT. Matching behavior and the representation of value in the parietal cortex. Science. 2004; 304:1782–1787. [PubMed: 15205529]

6. Seo H, Barraclough DJ, Lee D. Lateral intraparietal cortex and reinforcement learning during a mixed-strategy game. J Neurosci. 2009; 29:7278–7289. [PubMed: 19494150]

7. Seo H, Lee D. Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. J Neurosci. 2007; 27:8366–8377. [PubMed: 17670983]

8. Sul JH, Kim H, Huh N, Lee D, Jung MW. Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. Neuron. 2010; 66:449–460. [PubMed: 20471357]

9. Padoa-Schioppa C, Assad JA. Neurons in the orbitofrontal cortex encode economic value. Nature. 2006; 441:223–226. [PubMed: 16633341]

10. Roesch MR, Taylor AR, Schoenbaum G. Encoding of time-discounted rewards in orbitofrontal cortex is independent of value representation. Neuron. 2006; 51:509–520. [PubMed: 16908415]

11. Seo H, Lee D. Behavioral and neural changes after gains and losses of conditioned reinforcers. J Neurosci. 2009; 29:3627–3641. [PubMed: 19295166]

12. Schultz W. Predictive reward signal of dopamine neurons. J Neurophysiol. 1998; 80:1–27. [PubMed: 9658025]

13. Barraclough DJ, Conroy ML, Lee D. Prefrontal cortex and decision making in a mixed-strategy game. Nat Neurosci. 2004; 7:404–410. [PubMed: 15004564]

14. Coe B, Tomihara K, Matsuzawa M, Hikosaka O. Visual and anticipatory bias in three cortical eye fields of the monkey during an adaptive decision-making task. J Neurosci. 2002; 22:5081–5090. [PubMed: 12077203]

15. Pesaran B, Nelson MJ, Andersen RA. Free choice activates a decision circuit between frontal and parietal cortex. Nature. 2008; 453:406–409. [PubMed: 18418380]

16. Louie K, Glimcher PW. Separating value from choice: delay discounting activity in the lateral intraparietal area. J Neurosci. 2010; 30:5498–5507. [PubMed: 20410103]

17. Kennerley SW, Dahmubed AF, Lara AH, Wallis JD. Neurons in the frontal lobe encode the value of multiple decision variables. J Cogn Neurosci. 2009; 21:1162–1178. [PubMed: 18752411]

18. Kim YB, et al. Encoding of action history in the rat ventral striatum. J Neurophysiol. 2007; 98:3548–3556. [PubMed: 17942629]

19. Huh N, Jo S, Kim H, Sul JH, Jung MW. Model-based reinforcement learning under concurrent schedules of reinforcement in rodents. Learn Mem. 2009; 16:315–323. [PubMed: 19403794]

20. Felsen G, Mainen ZF. Neural substrates of sensory-guided locomotor decisions in the rat superior colliculus. Neuron. 2008; 60:137–148. [PubMed: 18940594]

21. Reep RL, Corwin JV, Hashimoto A, Watson RT. Efferent connections of the rostral portion of medial agranular cortex in rats. Brain Res Bull. 1987; 19:203–221. [PubMed: 2822206]

22. Donoghue JP, Parham C. Afferent connections of the lateral agranular field of the rat motor cortex. J Comp Neurol. 1983; 217:390–404. [PubMed: 6886060]

23. Donoghue JP, Wise SP. The motor cortex of the rat: cytoarchitecture and microstimulation mapping. J Comp Neurol. 1982; 212:76–88. [PubMed: 6294151]

24. Miller MW. The origin of corticospinal projection neurons in rat. Exp Brain Res. 1987; 67:339–351. [PubMed: 3622693]

25. Sanderson KJ, Welker W, Shambes GM. Reevaluation of motor cortex and of sensorimotor overlap in cerebral cortex of albino rats. Brain Res. 1984; 292:251–260. [PubMed: 6692158]

26. Neafsey EJ, et al. The organization of the rat motor cortex: a microstimulation mapping study. Brain Res. 1986; 396:77–96. [PubMed: 3708387]

27. Reep RL, Goodwin GS, Corwin JV. Topographic organization in the corticocortical connections of medial agranular cortex in rats. J Comp Neurol. 1990; 294:262–280. [PubMed: 2332532]

28. Passingham RE, Myers C, Rawlins N, Lightfoot V, Fearn S. Premotor cortex in the rat. Behav Neurosci. 1988; 102:101–109. [PubMed: 3355650]

29. Akintunde A, Buxton DF. Differential sites of origin and collateralization of corticospinal neurons in the rat: a multiple fluorescent retrograde tracer study. Brain Res. 1992; 575:86–92. [PubMed: 1504786]

30. Reep RL, Corwin JV, Hashimoto A, Watson RT. Afferent connections of medial precentral cortex in the rat. Neurosci Lett. 1984; 44:247–252. [PubMed: 6728294]

31. Wang Y, Kurata K. Quantitative analyses of thalamic and cortical origins of neurons projecting to the rostral and caudal forelimb motor areas in the cerebral cortex of rats. Brain Res. 1998; 781:135–147.

32. Nachev P, Kennard C, Husain M. Functional role of the supplementary and pre-supplementary motor areas. Nat Rev Neurosci. 2008; 9:856–869. [PubMed: 18843271]

33. Uylings HB, van Eden CG. Qualitative and quantitative comparison of the prefrontal cortex in rat and in primates, including humans. Prog Brain Res. 1990; 85:31–62. [PubMed: 2094901]

34. Tanji J. The supplementary motor area in the cerebral cortex. Neurosci Res. 1994; 19:251–268. [PubMed: 8058203]

35. Buschman TJ, Miller EK. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science. 2007; 315:1860–1862. [PubMed: 17395832]

36. Kargo WJ, Szatmary B, Nitz DA. Adaptation of prefrontal cortical firing patterns and their fidelity to changes in action-reward contingencies. J Neurosci. 2007; 27:3548–3559. [PubMed: 17392471]

37. Hoshi E, Tanji J. Distinctions between dorsal and ventral premotor areas: anatomical connectivity and functional properties. Curr Opin Neurobiol. 2007; 17:234–242. [PubMed: 17317152]

38. Roesch MR, Olson CR. Impact of expected reward on neuronal activity in prefrontal cortex, frontal and supplementary eye fields and premotor cortex. J Neurophysiol. 2003; 90:1766–1789. [PubMed: 12801905]

39. Roesch MR, Olson CR. Neuronal activity related to reward value and motivation in primate frontal cortex. Science. 2004; 304:307–310. [PubMed: 15073380]

40. Sohn JW, Lee D. Order-dependent modulation of directional signals in the supplementary and presupplementary motor areas. J Neurosci. 2007; 27:13655–13666. [PubMed: 18077677]

41. Thevarajah D, Mikulic A, Dorris MC. Role of the superior colliculus in choosing mixed-strategy saccades. J Neurosci. 2009; 29:1998–2008. [PubMed: 19228954]

42. Balleine BW, Delgado MR, Hikosaka O. The role of the dorsal striatum in reward and decision-making. J Neurosci. 2007; 27:8161–8165. [PubMed: 17670959]

43. Schoenbaum G, Roesch MR, Stalnaker TA. Orbitofrontal cortex, decision-making and drug addiction. Trends Neurosci. 2006; 29:116–124. [PubMed: 16406092]

44. Ragozzino ME. The contribution of the medial prefrontal cortex, orbitofrontal cortex, and dorsomedial striatum to behavioral flexibility. Ann N Y Acad Sci. 2007; 1121:355–375. [PubMed: 17698989]

45. Kable JW, Glimcher PW. The neurobiology of decision: consensus and controversy. Neuron. 2009; 63:733–745. [PubMed: 19778504]

46. Paxinos, G.; Watson, C. The Rat Brain in Stereotaxic Coordinates. Academic Press; San Diego: 1998.

47. Staddon JE, Hinson JM, Kram R. Optimal choice. J Exp Anal Behav. 1981; 35:397–412. [PubMed: 16812224]

48. Lau B, Glimcher PW. Dynamic response-by-response models of matching behavior in rhesus monkeys. J Exp Anal Behav. 2005; 84:555–579. [PubMed: 16596980]

49. Baeg EH, et al. Fast spiking and regular spiking neural correlates of fear conditioning in the medial prefrontal cortex of the rat. Cereb Cortex. 2001; 11:441–451. [PubMed: 11313296]

50. Baeg EH, et al. Dynamics of population code for working memory in the prefrontal cortex. Neuron. 2003; 40:177–188. [PubMed: 14527442]
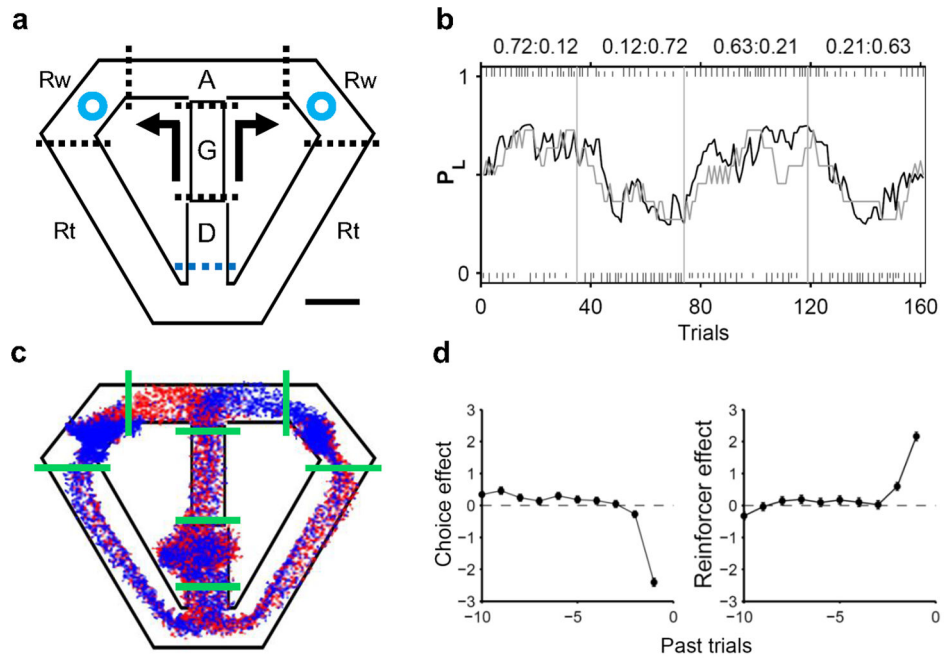
Figure 1.
Behavioral task and choice behavior. (**a**) Rats were allowed to choose freely between two locations (blue circles) that delivered water with different probabilities. The task was divided into five stages: delay (*D*), go (*G*), approach to reward (*A*), reward (*Rw*) and return (*Rt*) stages. Dotted lines indicate approximate stage transition points. Arrows indicate movement directions. Scale bar: 10 cm. (**b**) Choice behavior of an animal during an example behavioral session. The gray line shows the probability for the animal to choose the left goal (P_L; moving average of 10 trials). The black line indicates the probability of choosing the left goal in the RL model. Tick marks denote trial-by-trial choices of the animal (upper, left choice; lower, right choice; long, rewarded; short, unrewarded), and vertical lines denote block transitions. Reward probabilities for each block are indicated on top. (**c**) Movement trajectories during an example session. Each dot indicates the animal's head position (sampling frequency: 60 Hz). Green solid lines indicate stage transitions. Beginning from the reward onset, blue and red traces indicate trials associated with the left and right upcoming choice of the animal, respectively. Trials were decimated (3 to1) to enhance visibility. (**d**) Effects of past choices and their outcomes on the current goal choice are shown for one animal. The ordinate shows the coefficients of a logistic regression model. Positive coefficients indicate increased likelihood for the animal to repeat rewarded past choices (reinforcer effect) or past goal choices irrespective of their outcomes (choice effect). Error bars, SEM.
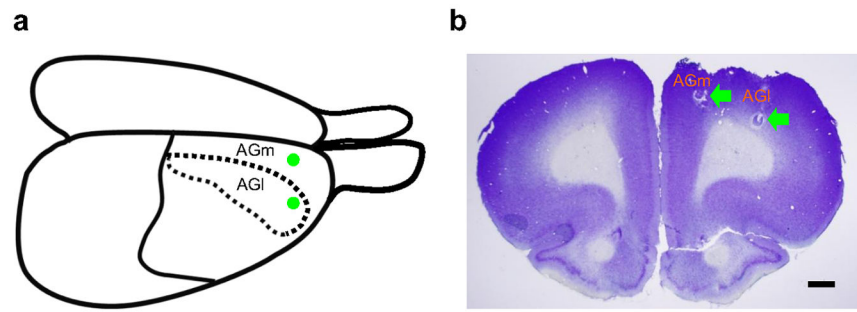
Figure 2.
Recording sites. Single units were recorded from the rostral AGm and AGl. (**a**) Dorsolateral view of the rat brain. Green circles indicate approximate electrode implantation sites. (**b**) Coronal section of the brain stained with cresyl violet. Green arrows indicate marking lesions. Scale bar: 1 mm.
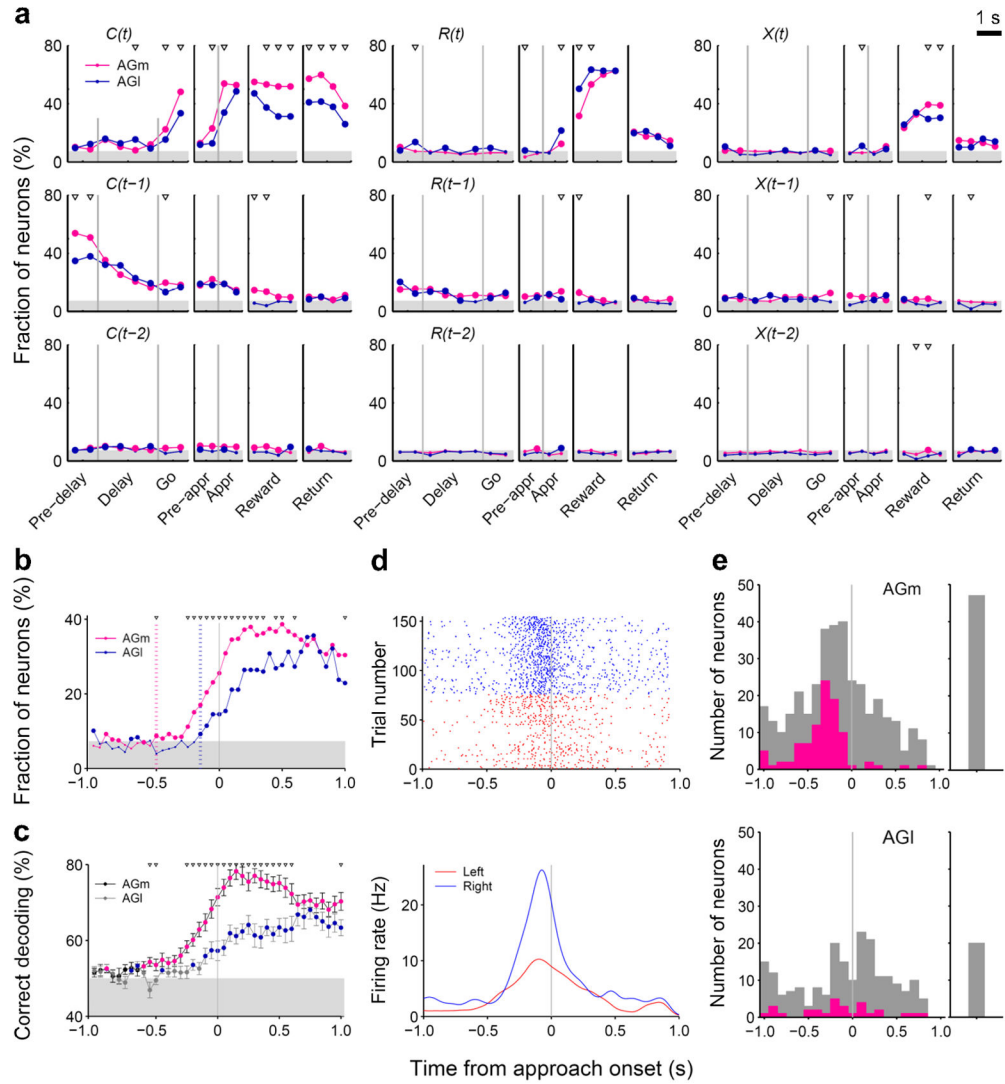
Figure 3.
Neural signals for the animal's choice and reward. (**a**) Fractions of neurons that significantly modulated their activity according to the animal's choice (*C*), reward (*R*), or their interaction (*X*) in the current (*t*) and previous trials (*t–1* and *t–2*) in non-overlapping 500-ms time windows (Pre-delay: last 1 s of the return stage; Pre-appr: last 1 s of the go stage; Appr: approach stage; eq. 1). The vertical lines indicate the onset of a behavioral stage. Large circles indicate that the fraction is above chance level (binomial test, $p < 0.05$), and triangles indicate significant differences across regions (χ^2-test, $p < 0.05$). The shading indicates the mean of the minimum fractions significantly above chance which is slightly different for the AGm and AGl. (**b**) Fractions of current choice [*C(t)*]-encoding neurons are plotted in a 100-ms time window advanced in 50-ms time steps. The red and blue dotted lines indicate the onsets of the current choice signals for the AGm and AGl, respectively (see Online Methods). (**c**) Neuronal population decoding of the animal's goal choice (% correct decoding) (100-ms window, 50-ms time steps). Red (AGm) and blue (AGl) symbols indicate significant differences from the chance level (50 %; *t*-test, $p < 0.05$). Triangles

indicate significant differences across regions (t-test, $p < 0.05$). Error bars, SEM. (**d**) An example AGm neuron that modulated its activity according to the animal's upcoming goal choice. Both spike raster (top) and spike density functions (bottom; Gaussian kernel with σ=50 ms) are shown. (**e**) Distribution of the choice signal onset for individual neurons (see Online Methods). Red indicates those neurons that significantly modulated their activity according to the animal's upcoming choice during the 500 ms time window before the approach onset (eq. 1). Gray indicates the remaining neurons. Those neurons that were determined not to have a choice onset during 1 s time periods before and after the approach onset are plotted separately on the right.
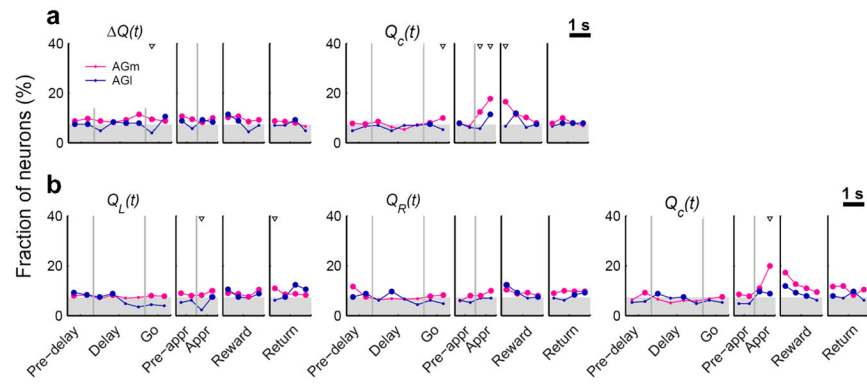
Figure 4.
Neural signals for values. (**a**) Fractions of neurons that significantly modulated their activity according to decision value (Q) or chosen value (Q_c) in non-overlapping 500-ms time windows (eq. 2). (**b**) Fractions of neurons that significantly modulated their activity according to the left or right action value [$Q_L(t)$ or $Q_R(t)$, respectively]. The same regression analysis (eq. 2) was performed with decision value replaced with the left and right action values. Same format as in Fig. 3a.
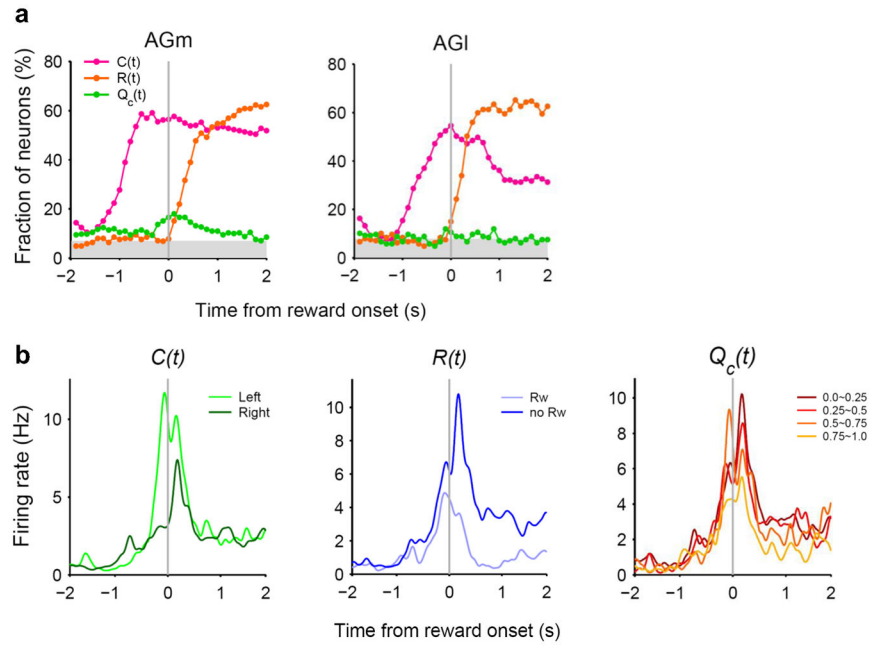
Figure 5.
Convergence of neural signals for chosen action, reward, and chosen value in the AGm. (**a**) Fractions of neurons that significantly modulated their activity according to chosen action, reward, or chosen value are shown in a 500-ms time window advanced in 100-ms time steps around the time of reward stage onset. (**b**) An example AGm neuron that significantly modulated its activity according to chosen action, reward, and chosen value in the reward stage. Spike density functions (Gaussian kernel, σ=50 ms) are shown separately for different choices (left vs. right), different choice outcomes (rewarded vs. unrewarded), or different levels of chosen value.
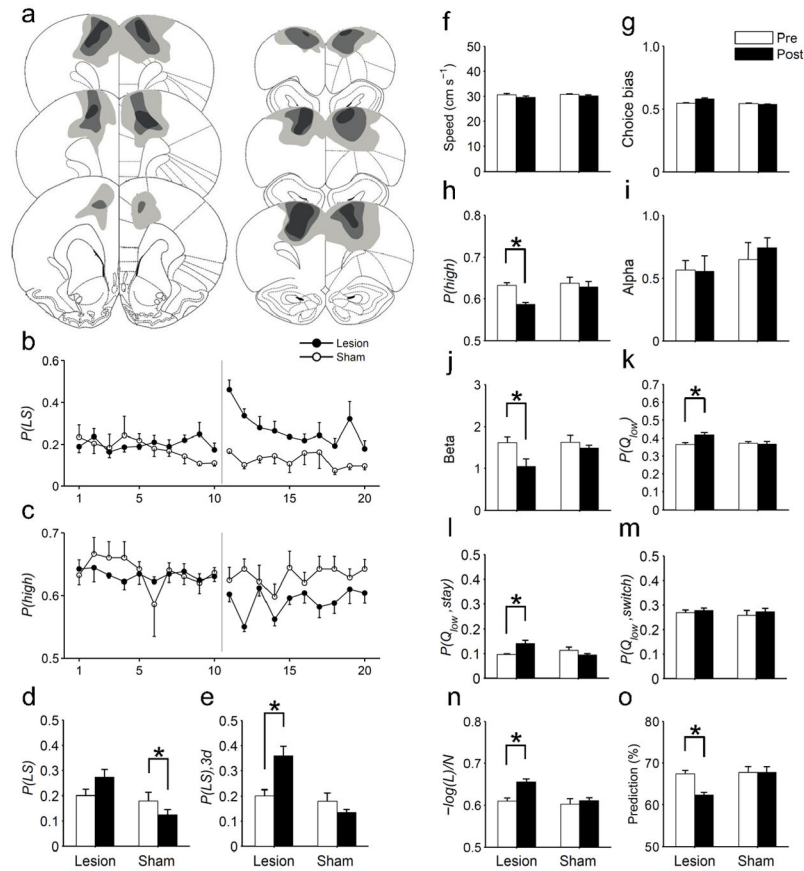
Figure 6.
Effects of AGm lesions. (**a**) The extent of AGm lesions. The diagram shows maximal (light gray), representative (medium gray) and minimal (dark gray) lesion extents across five animals. Adapted from ref. 46, with permission from Elsevier. (**b,c**) Time courses for the probabilities of lose-stay [*P(LS)*] and selecting the higher arming-probability goal in each block [*P(high)*]. (**d**) *P(LS)* averaged for 10 days before and after lesions. (**e**) *P(LS)* for the initial 3 days following lesions [*P(LS), 3d*] was compared with *P(LS)* during the pre-lesion period (10 days). (**f**) Mean running speed across go, approach and return stages before and after AGm or sham lesions. (**g**) Choice bias (fraction of the choice for a preferential goal in a given session). (**h**) *P(high)* before and after lesions. (**I,j**) Learning rate (α) and inverse temperature (β) of the RL model. (**k**) The probability of choosing the lower action-value goal [*P(Q_{low})*]. (**l,m**) *P(Q_{low})* is shown separately for staying [*P(Q_{low}, stay)*] and switching [*P(Q_{low}, switch)*] trials. (n) Normalized negative log-likelihood [$-log(L)/N$] of the SP model. (**o**) Prediction of the animal's choices by the SP model (% correct). The data in (**d–o**) are averaged values during the pre-lesion and post-lesion periods (10 days each) except for *P(LS), 3d* (**e**). Asterisks indicate significant differences (paired t-test, $p < 0.05$). Error bars, SEM.
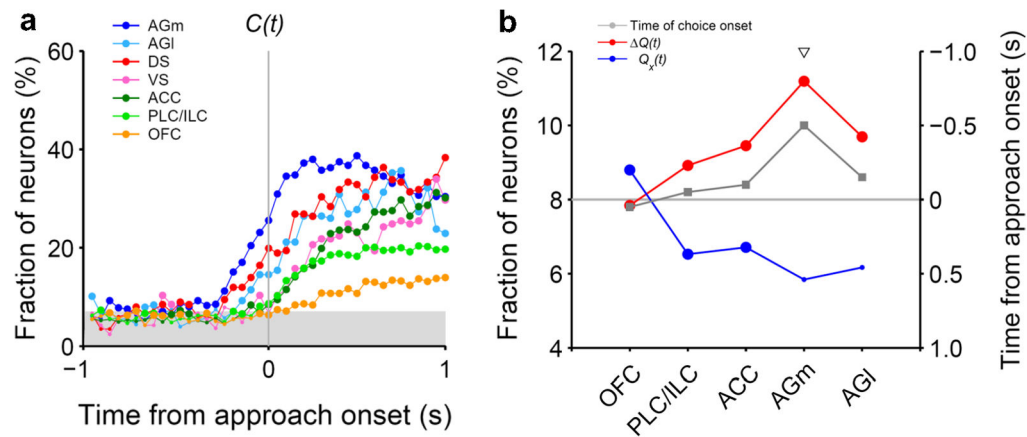
Figure 7.
Regional variations in neural signals related to valuation and choice in rodent frontal cortex and striatum. (**a**) Time courses of neural signals for upcoming action selection. (**b**) Fractions of neurons encoding at least one action value [$Q_x(t)$, blue; corrected for multiple comparisons] or decision value [$Q(t)$, red] during the last 1 s of the delay stage (left axis). The onset time of the upcoming choice signal is also indicated for each region in gray (right axis). Large (or small) circles denote significant (or insignificant) fractions (binomial test, alpha=0.05). The triangle indicates a significant difference between the fraction of action value and decision value-encoding neurons (χ^2-test, $p < 0.05$). PLC/ILC, prelimbic cortex/infralimbic cortex; DS, dorsal striatum; VS, ventral striatum.