nature neuroscience

The dorsomedial striatum encodes net expected return, critical for energizing performance vigor

Alice Y Wang[1,4], Keiji Miura[1–3] & Naoshige Uchida[1]

Decision making requires an actor to not only steer behavior toward specific goals but also determine the optimal vigor of performance. Current research and models have largely focused on the former problem of how actions are directed while overlooking the latter problem of how they are energized. Here we designed a self-paced decision-making paradigm, which showed that rats' performance vigor globally fluctuates with the net value of their options, suggesting that they maintain long-term estimates of the value of their current state. Lesions of the dorsomedial striatum (DMS) and, to a lesser degree, in the ventral striatum impaired such state-dependent modulation of vigor, rendering vigor to depend more exclusively on the outcomes of immediately preceding trials. The lesions, however, spared choice biases. Neuronal recordings showed that the DMS is enriched in net value–coding neurons. In sum, the DMS encodes one's net expected return, which drives the general motivation to perform.

Given stock options A and B, which one do you choose? Considerable progress has been made in the field of decision making and neuroeconomics on this problem of 'action selection'[1,2]. Optimal behavioral selection, however, not only depends on the ability to choose which action to perform but also the appropriate vigor to perform. For example, it may be wise to flexibly adjust one's motivation to invest according to the overall state of the stock market, such as investing with lower frequency during an economic crisis.

The importance of properly regulating response vigor becomes apparent when one considers costs associated with performing an action. A rapid response may increase the rate of obtaining rewards but may also increase energetic costs. Conversely, a slow response may be energetically efficient, yet it delays all future rewards. The cost of delaying future rewards critically depends on the net expected future reward given the current state of the animal. It is thus proposed that selection of response vigor should depend on the average or net expected reward (or 'state value'), whereas action selection depends on values specific to individual options (and the relative value between them)[3–5]. This idea echoes two aspects of motivation proposed in classic animal psychology: the motivation to steer toward making a specific action (the 'directing' effect or action-specific motivation) and the motivation to generally 'arouse' or speed up all prepotent actions in a nonspecific manner (the 'energizing' effect or action-general motivation)[3,6,7]. It should be noted that the directing effect may also speed up actions toward particular goals, but the energizing effect acts diffusely on a wider set of actions. Experimentally, many classical studies in animal psychology have shown that response vigor is modulated by the rate of reward, providing some limited support for the energizing effects of average reward rate[8]. More recently, it has been shown that manipulations of the size or probability of rewards affect choice direction and latencies in various choice tasks, which highlights motivation's directing effects[9–11]. However, whether response vigor is indeed regulated by average or net expected reward, that is, whether motivation energizes behavior in a global manner, remains controversial.

It is believed that the basal ganglia have important roles in action selection[1,2,12,13]. Some studies on individuals with Parkinson's disease and lesion studies using animal models, however, have suggested that the basal ganglia also have a prominent role in the regulation of response vigor[14]. Mounting evidence suggests that specific areas of the striatum encode specific types of values and regulate distinct aspects of value-dependent behavior[15–17]. Historically, the striatum, particularly the ventral striatum, has been linked to motivation[18,19], although other studies implicate the role of dorsal striatum in motivation[20–22]. However, previous studies have not separated the directing versus energizing aspects of motivation; therefore whether these processes can be mapped onto specific parts of the striatum remains unknown.

To address these questions, we designed a task that allows us to study both the directing and energizing aspects of behavioral regulation. We first examined whether response vigor is indeed modulated by net expected future rewards and how this process is separable from the directing effects that are specific to individual actions. Second, using lesions, we examined which part of the striatum is involved in the regulation of net value–dependent response vigor. Finally, we recorded the activity of single neurons in DMS and ventral striatum. The results demonstrate a critical role of the DMS in net value–dependent regulation of response vigor.

RESULTS

Self-paced decision task

We designed a self-paced, two-choice behavioral paradigm, where a rat self-initiated a trial by poking its snout into a central odor port. After a rat poked its snout into the odor port (trial initiation),

[1]Center for Brain Science, Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts, USA. [2]Precursory Research for Embryonic Science and Technology (PRESTO), Japan Science and Technology Agency, Saitama, Japan. [3]Graduate School of Information Sciences, Tohoku University, Sendai, Japan. [4]Present address: Department of Neurobiology, Harvard Medical School, Boston, Massachusetts, USA. Correspondence should be addressed to N.U. (uchida@mcb.harvard.edu).

Received 20 September 2012; accepted 9 March 2013; published online 14 April 2013; doi:10.1038/nn.3377

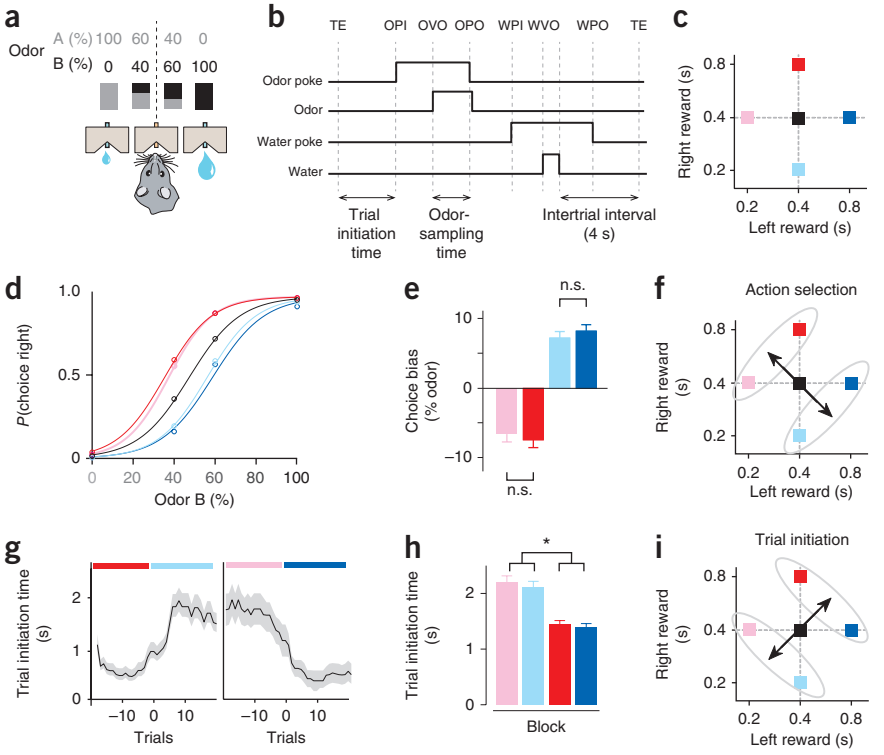


Figure 1 Design of a self-paced decision-making task: dissociating the effects of relative and net value on choice and performance vigor. (**a**) Schematic of the odor-mixture categorization task. Rats were trained to respond to the left or right reward port depending on the dominant component of a binary mixture of odors (A and B, grey and black rectangles, respectively). Dashed line indicates the stimulus category boundary. Concurrently, the sizes of reward at the left and right reward ports were manipulated in blocks of 40–60 trials. (**b**) Task epochs. Rats initiated a trial by making a nose poke into the central odor port. TE, trial end; OPI, odor poke in (entry); OVO, odor valve on (opening); OPO, odor poke out (exit); WPI, water poke in; WVO, water valve on (opening); WPO, water poke out (exit). (**c**) Design of block-wise reward manipulations, with each square denoting a block of trials (arbitrarily defined by indicated colors): reward values for left and right ports varied across blocks. Reward size was manipulated by changing the duration of the water valve being open. Dotted lines indicate the intermediate level of reward size. (**d**) Psychometric functions for blocks ($n = 15$ rats). The data were fitted with a logistic function. (**e**) Magnitude of shift in psychometric curves in bias blocks compared to no-bias blocks (red versus pink, $t_{28} = 0.5$, $P > 0.05$; blue versus cyan, $t_{28} = 0.8$, $P > 0.05$; t test, $n = 15$ rats). Error bars, s.e.m. (n.s., not significant; $n = 15$ rats). (**f**) Blocks in which rats show a rightward bias are circled in the upper-left quadrant and blocks in which rats show a leftward bias are circled in the lower-right quadrant. Arrow indicates the direction in which choice biases are modulated in the value space. (**g**) Trial initiation time across blocks (blocks are indicated by bars at the top of graphs) ($n = 15$, mean ± s.e.m., smoothed over 5 trials). Gradual shifts preceding a block switch (trial = 0) are due to smoothing of data. (**h**) Trial initiation time (mean ± s.e.m. of the median trial initiation time per block, colored as in **c**). $^*P < 0.05$ (*t*-test). (**i**) Blocks in which rats show fast initiation time are circled in the upper-right quadrant and blocks in which rats show a rightward bias are circled in the lower-left quadrant. Arrow indicates the direction in which choice biases are modulated in the value space. All colors correspond to the design outlined in **c**.

a randomly selected odor cue was presented after a variable delay (300–500 ms; **Fig. 1a**,**b**)[23]. We used pure odors as well as binary mixtures of odors of various ratios, where the dominant component in the mixture determined which reward port delivered water. To examine the effects of the rat's option values on behavior, we systematically varied the amount of water delivered at the left and right ports in blocks of 40–60 trials (below referred to as 'blocks'; **Fig. 1c**)[16,24].

Consistent with previous studies[16,24], the rat's choices were biased by the relative value of the options (**Fig. 1d–f**). For example, rats chose the right port more frequently in blocks where the right port delivered relatively more water, as shown in the shifts in psychometric curves (**Fig. 1d**,**e**). This bias toward the more valuable side was invariant to changes in the net value of the options (**Fig. 1e**). Choice biases existed primarily for mixed-odor but not pure-odor trials, indicating that rats did not make their selection until after odor sampling (**Fig. 1d**). Thus, action selection depended on the relative value of the rats' options.

Latency to trial initiation depends on the net value

We next examined the rats' response vigor by measuring the latency to initiate a trial after completing a previous trial. Overall, rats exhibited shorter latency in blocks of high net value compared with blocks of low net value (**Fig. 1g–i**; *t*-test, $n = 15$ rats, $t_{58} = 5.0$, $P < 0.05$, **Supplementary Figs. 1** and **2**). This change in initiation time consisted primarily of the slowing down of movement from the reward port to odor port, brief pauses near the port (as reflected by the increase in trials with 5–10 s initiation time) as well as a relatively minor contribution of disengagement from the task (grooming or resting outside the task, as reflected by the increase in trials with long (>10 s) initiation time; **Supplementary Fig. 1c**). On average, alternations in initiation time occurred within 5 trials after a change in net value (time constant: 4.9 ± 1.3 trials, mean ± s.e.m, $n = 15$ rats; **Fig. 1g**). These results suggested that trial initiation time depended on the net value of the rat's choices (**Fig. 1i**).

Past models predict that in states of high overall return, motivation should generally increase, whereas in states of low net value, motivation should generally decrease[3,4]. However, the above results do not demonstrate that response vigor is block- or state-dependent. Under our task design, rats may simply initiate slower immediately after receiving small rewards and faster immediately after receiving large rewards. To examine whether the rats' response vigor is globally energized by the block, we next examined initiation time after left and right reward trials separately.

If the motivation to initiate trials is driven exclusively by the immediately preceding trial's reward size, initiation time after left reward trials should be the same if the left reward size is identical regardless of the right reward size. However, we found that rats were significantly faster at trial initiation after left reward when the right reward size was larger (**Fig. 2a**, *t*-test, $t_{12} = 2.4$, $P < 0.01$, $n = 7$ sessions in one rat). Moreover, latency after left trials in the block where left delivers 0.4 s and right delivers 0.2 s of water (**Fig. 2a**, cyan) was no faster than in the block where left delivers 0.2 s and right delivers 0.4 s of water (**Fig. 2a**, pink) (*t*-test, $t_{12} = 0.7$, $P > 0.05$, $n = 7$ sessions in one rat). We saw a similar pattern for initiation time after all right reward trials (**Fig. 2b**), and we observed similar patterns in other rats. These results suggest that the motivation to initiate trials cannot be explained simply by the rat responding to reward size in the immediately preceding trial. Rather, the overall value of the block determined the rat's motivation to perform.

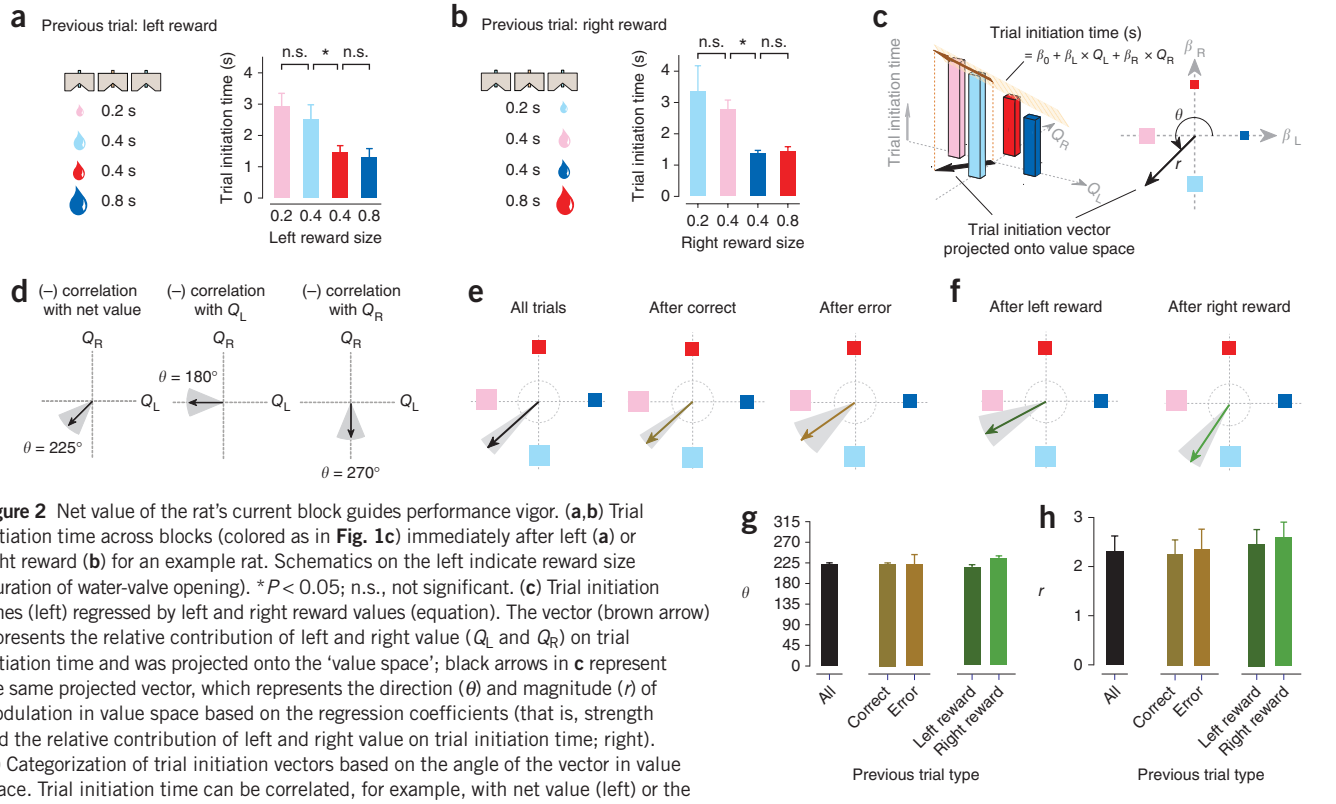


Figure 2 Net value of the rat's current block guides performance vigor. (**a**,**b**) Trial initiation time across blocks (colored as in **Fig. 1c**) immediately after left (**a**) or right reward (**b**) for an example rat. Schematics on the left indicate reward size (duration of water-valve opening). $^*P < 0.05$; n.s., not significant. (**c**) Trial initiation times (left) regressed by left and right reward values (equation). The vector (brown arrow) represents the relative contribution of left and right value (Q_L and Q_R) on trial initiation time and was projected onto the 'value space'; black arrows in **c** represent the same projected vector, which represents the direction (θ) and magnitude (r) of modulation in value space based on the regression coefficients (that is, strength and the relative contribution of left and right value on trial initiation time; right). (**d**) Categorization of trial initiation vectors based on the angle of the vector in value space. Trial initiation time can be correlated, for example, with net value (left) or the value of one of the options (middle and right). Negative correlation is indicated (–). (**e**,**f**) Example vectors for a single rat calculated separately for trial initiation times immediately after correct, error or pooled trials (**e**), or after left or right reward trials (**f**). Gray shading indicates the 95% confidence interval of the vector angle obtained by bootstrap method. (**g**,**h**) Population summary of the angles (θ; **g**) and the amplitudes (r, **h**) of vectors (error bars, s.e.m.; $n = 15$ rats).

To more succinctly quantify changes in initiation time across blocks (that is, how initiation time varies with left and right value), we represented trial initiation time with a single vector projected onto 'value space' with a polar angle (θ) and amplitude (r) (**Fig. 2c**). We regressed initiation time with value of left choice and right choice in each block (Q_L and Q_R, respectively):

$$\text{Trial initiation time (s)} = \beta_0 + \beta_L \times Q_L + \beta_R \times Q_R \quad (1)$$

We used these regression coefficients to project a single vector onto value space (**Fig. 2c**). The vector conveys two variables: θ (in polar coordinates) reveals how trial initiation time varies across blocks (that is, the relative contribution of Q_L and Q_R), and r is the strength of the modulation (**Fig. 2c**). Essentially, the vector illustrates the strength and the relative contribution of left and right value on trial initiation time. For example, a negative correlation with net value (the larger the value of left and right reward, the shorter the initiation time) will be projected as a vector directed toward the lower-left quadrant of value space, or $\theta = 225° \pm 22.5°$ (**Fig. 2d**). If initiation time negatively correlates with left value only, the vector would be horizontal, or $\theta = 180° \pm 22.5°$ (**Fig. 2d**). If initiation time negatively correlates with right value only, the vector would be vertical, or $\theta = 270° \pm 22.5°$ (**Fig. 2d**). For more stringent classification, we used a 95% confidence interval on θ calculated by bootstrapping.

Using the example rat shown in **Figure 2a**, we obtained a vector of $\theta = 222°$ (95% confidence interval lower bound = 213° and upper bound = 231°) and $r = 3.3$ (**Fig. 2e**). As θ fell within 22.5° of 225°, and the confidence interval did not cross 180° or 270°, we considered the rat's trial initiation time as net value–dependent. This analysis allowed us to summarize initiation times following different trial outcomes, such as correct, error, left reward or right reward trials, separately (n = 1,019, 244, 512, and 507 trials, respectively; **Fig. 2e**,**f**).

Using this analysis, we found that on average, trial-initiation time was net value–dependent, regardless of previous trial type (**Fig. 2g**,**h**, $n = 15$ rats; correct trials, $\theta = 223° \pm 2.3$, $r = 2.3 \pm 0.3$ error trials, $\theta = 222° \pm 21$, $r = 2.4 \pm 0.4$; left trials, $\theta = 212° \pm 6.2$, $r = 2.4 \pm 0.3$; and right trials, $\theta = 233° \pm 4.9$, $r = 2.5 \pm 0.3$). Thus, response vigor depended predominantly on the net value of both options, more so than on reward size of the immediately preceding trial. However, we observed that θ_R (vector angle representing initiation time immediately after right reward) was slightly larger than θ_L (vector angle representing initiation time after left reward) in most of the cases (**Fig. 3a**,**b**; paired t-test, $t_{14} = 2.9$, $P < 0.05$). This suggested that although initiation time primarily depends on the net value of the block (as most vectors were near 225°), reward value for the immediately preceding trial may still have an influence.

To quantify how the outcomes of multiple previous trials affected trial initiation time, we regressed initiation time by previous reward outcomes, where T is the current trial (for example, Q_{T-1} is the reward outcome of one previous trial):

$$\text{Trial initiation time (s)} = \beta_0 + \beta_{T-1} \times Q_{T-1} + \beta_{T-2} \times Q_{T-2} + \beta_{T-5} \times Q_{T-5} \quad (2)$$

We obtained the regression coefficients and fitted them to an exponential to obtain a decay constant, τ (**Fig. 3c**). The result showed that initiation time depends on integration over multiple previous trials ($\tau = 4.6 \pm 1.5$ trials, $n = 15$ rats). The difference between the

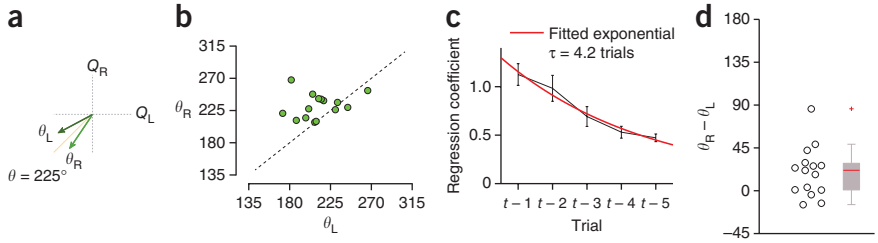


Figure 3 Performance vigor depends on the integration of outcomes of multiple previous trials. (**a**) Schematic showing how $\theta_R - \theta_L$ was computed. (**b**) Scatter plot of θ_R versus θ_L. The dashed line indicates identity. (**c**) Effect of multiple previous trial outcomes on trial initiation time for the population. Black line shows regression coefficients (mean ± s.e.m., $n = 15$ rats); red, fitted exponential curve (time constant, τ). *t*, current trial; $t - 1$, immediately preceding trial and so on. (**d**) Difference in angle between vectors representing trial initiation time immediately after left reward and after right reward trials (each point is the $\theta_R - \theta_L$ for one rat). The vertical line indicates the maximum and minimum values of non-outliers; points are considered as outliers if they are larger than $b + 1.5(b - a)$ or smaller than $a - 1.5(b - a)$, where *a* and *b* are the 25th and 75th percentiles, respectively.

angles, $\theta_R - \theta_L$, was very small (**Fig. 3d**; median, 21.3°; mean, 19.6°; **Supplementary Fig. 3**), and was not significantly different from the $\theta_R - \theta_L$ for trial-shuffled controls ($P > 0.05$, bootstrap test), suggesting that the effect of the outcome of the immediately preceding trial was small. The analysis based on the time constant (τ) complements the vector analysis in that although the effects of various trial types (for example, left, right, correct and error) are masked in the process of computing τ, we could obtain the number of trials back in time on which initiation time depends.

In all, these results demonstrate that the overall value of the block energized initiation of trials: rats exploit the task during states of high value, as the opportunity cost of wasted time is high, as predicted by normative model of motivation[4].

DMS is critical for net value–dependence of response vigor

The above behavioral paradigm allowed us to dissociate the effect of values on choice and response vigor. We next made selective lesions in rats' DMS or ventral striatum to test whether either region is involved in choice bias or response vigor. After training the rats, we injected ibotenic acid, an excitotoxic agent, into the anterior DMS or ventral striatum bilaterally, which caused lesions in a relatively large portion of each area (~2 mm in diameter) (**Fig. 4** and **Supplementary Fig. 4**). In control rats, we injected saline ('sham lesion'). Then we examined the behavior of ventral striatum–lesion, DMS-lesion and sham-lesion rats ($n = 5$ rats per condition; 3,365.6 ± 115 trials in 7.86 ± 0.09 sessions per rat). None of the rats exhibited impaired choice biases (**Fig. 4b**). Ventral striatum–lesion rats showed a trend for larger choice biases (**Fig. 4b**), but the shifts were not statistically significant (**Fig. 4c**). For mixed-odor trials only, ventral striatum–lesion rats had significantly more errors toward the relatively better choice (**Fig. 4d**; sham lesion, 37.9%; DMS lesion, 38.6%; ventral striatum, 45.1% error trials toward larger choice for mixed-odor trials; *t*-test between sham lesion and ventral striatum lesion, $n = 5$ rats per condition, $t_8 = 3.1$, $P < 0.05$). This higher error rate for difficult trials was not due to inadequate odor-sampling duration (**Fig. 4e**; *t*-test, all $P > 0.05$; sham lesion versus DMS lesion pure odors, $t_8 = 0.2$ and mixed odors, $t_8 = 0.6$; sham lesion versus ventral striatum lesion pure odors, $t_8 = 0.4$ and mixed odors, $t_8 = 0.8$).

Next, we examined the effect of lesions on trial initiation time. Although the mean initiation time was normal (**Fig. 5a**, *t*-test comparing high and low net value blocks, $n = 5$ rats per condition, all $P < 0.05$; sham, $t_{18} = 2.6$; DMS, $t_{18} = 3.0$; ventral striatum, $t_{18} = 4.3$), we observed the effect of lesions when trial initiation time was analyzed separately for different trial types, using the vector representation method (**Fig. 5b–d** and **Supplementary Fig. 5**). Trial initiation vectors for correct, error and all trials were dependent on net value (that is, within 22.5 degrees of 225° in all example rats, $n = 3{,}413$, 3,041, 3,776 total trials for sham lesion, DMS lesion and ventral striatum lesion, respectively). In contrast, vectors obtained for left and right reward trials separately revealed notable differences across lesion conditions (**Fig. 5c–f**). In the sham-treated rats, trial initiation vectors for both left and right trials were within ± 22.5° of 225°, indicating net value–dependent regulation of response vigor (**Fig. 5c**, sham lesion $\theta_L = 216°$ and $\theta_R = 225°$; **Supplementary Fig. 5b**, population $\theta_L = 219° \pm 5.8$ and $\theta_R = 236° \pm 7°$). For the DMS-lesion rat, however, initiation times after left trials depended on left value alone whereas initiation after right trials depended on right value alone (**Fig. 5c–f**, DMS lesion $\theta_R = 190°$ and $\theta_R = 289°$; **Supplementary Fig. 5b**, population $\theta_L = 189° \pm 4.5°$ and $\theta_R = 275° \pm 6.8°$). The DMS-lesion rat's trial initiation vectors for left reward and right reward trials were nearly orthogonal to one another (**Fig. 5c**, DMS lesion $\theta_R - \theta_L = 99°$; **Fig. 5e**, DMS lesion $\theta_R - \theta_L = 87° \pm 8.5°$) unlike the case for sham-lesion rats (**Fig. 5e**, $\theta_R - \theta_L = 20.1° \pm 5.3°$; DMS lesion versus sham lesion, $P < 0.05$, *t*-test; $t_8 = 6.1$), indicating that the rat's motivation depended primarily on the value of the immediately preceding trial. We observed a less dramatic but significant effect in the ventral striatum–lesion condition (**Fig. 5c–f**, ventral striatum lesion $\theta_L = 207°$ and $\theta_R = 248°$; **Supplementary Fig. 5b**, population $\theta_L = 203° \pm 3.7°$ and $\theta_R = 243° \pm 2.9°$). Vectors for left and right trials were more widely separated in ventral striatum–lesion rats (**Fig. 5c**, ventral striatum example $\theta_R - \theta_L = 41°$; **Fig. 5e**, ventral striatum population $\theta_R - \theta_L = 40.7° \pm 5.4°$) than for the sham rat (ventral striatum versus sham condition: *t*-test, $t_8 = 2.8$, $P < 0.05$), yet not orthogonal as in the DMS condition (ventral striatum versus DMS lesion: *t*-test, $t_8 = 5.1$, $P < 0.05$). Lastly, we examined the effects of multiple preceding trial outcomes on initiation time and found that the influence of the immediately preceding trial was significantly stronger for the DMS-lesion condition compared to the sham-lesion condition (**Fig. 5f**, $\tau = 1.2 \pm 0.4$ trials for DMS lesion; DMS lesion versus sham lesion: *t*-test, $t_8 = 2.7$, $P < 0.05$). We observed a similar but nonsignificant tendency for the ventral striatum lesion (**Fig. 5f**, $\tau = 2.6 \pm 0.9$ trials for ventral striatum; ventral striatum versus sham, *t*-test, $t_8 = 1.7$, $P > 0.05$). In total, these results demonstrate that the DMS had a larger role than the ventral striatum on integrating the total value of one's options, which was critical for promoting net value–dependent modulation of response vigor.

DMS is enriched with neural representations of net value

To examine the neural mechanisms that may underlie the importance of DMS in net value–dependent modulation of trial initiation, we monitored the neural activity in the DMS and ventral striatum (**Fig. 6a**,**b**) of another set of rats as they performed the behavioral paradigm (**Supplementary Fig. 6**). We recorded 522 neurons, with 364 from DMS and 158 from ventral striatum while rats performed the task ($n = 4$ rats; 91 ± 37.4 neurons and 39.5 ± 56.9 neurons in DMS and ventral striatum per rat; mean ± s.d.; total neurons recorded from each rat: 96, 154, 209 and 63, all from the left hemisphere).

First, we examined each neuron's time of maximum firing by pooling all trial types and found that as a population, the intertrial epoch was a

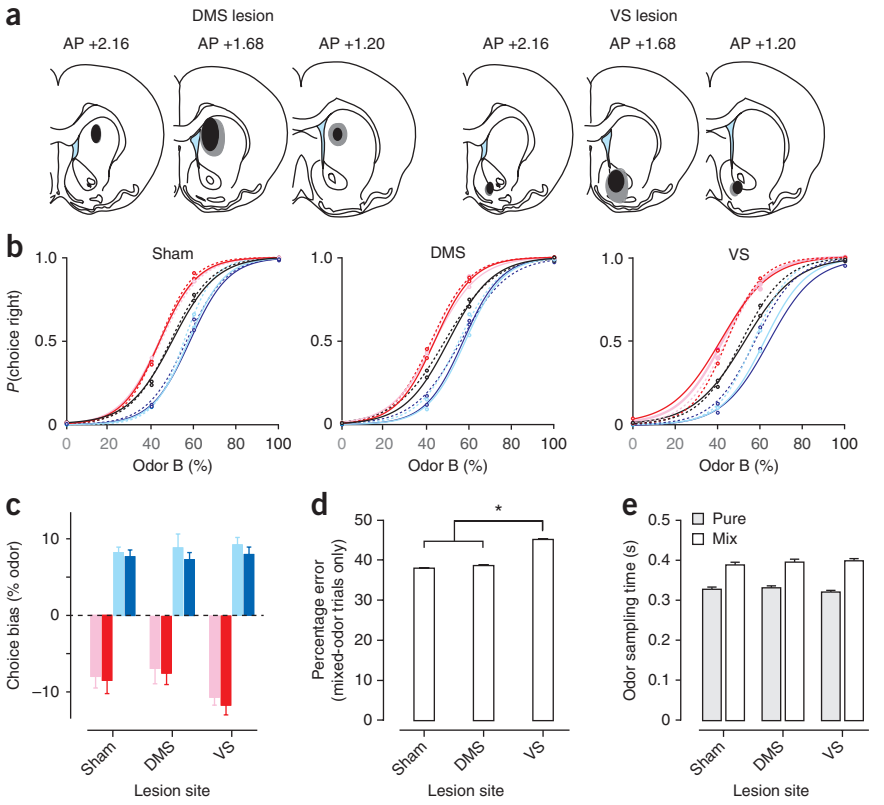


Figure 4 Lesions of DMS and ventral striatum (VS) have little effect on action selection. (**a**) Locations of selective lesions of the anterior DMS and VS. Gray area, maximum extent of lesion. Black area, minimum extent of lesion. AP, distance from bregma in the anterior-posterior axis. (**b**) Psychometric curves for indicated lesion conditions. Dashed lines indicate pre-lesion performance; solid lines indicate post-lesion performance (average performance over 15 rats). The data were fitted with a logistic function. (**c**) Quantification of the shifts in psychometric curves for all reward-bias blocks (colored as in **Fig. 1c**). Sham versus VS lesion for cyan and blue blocks, $t_{18} = 0.6$ and sham versus VS lesion for pink and red blocks, $t_{18} = 1.8$; $^*P < 0.05$ (*t*-test; $n = 5$ rats per lesion condition). We obtained the inflection points of psychometric curves for each block and rat. Leftward or rightward shifts in psychometric curves were calculated by subtracting the inflection points of reward-bias blocks (colored curves) from blocks with no reward bias (black curves in **b**) (mean ± s.e.m., $n = 15$ rats). (**d**) Performance accuracy for mixed-odor trials (mean ± s.e.m., $n = 15$ rats). $^*P < 0.05$. (**e**) Odor sampling time for pure- and mixed-odor trials (mean ± s.e.m., $n = 15$ rats).

time of high neural activity (30% or 152 of 522 cells had maximal firing during this intertrial interval; **Fig. 6a**). Note that peri-event time histograms (PETHs) show activity aligned to the median timing of the epochs. As intertrial intervals were especially variable, the peak timing of neurons appears especially smeared during this period. Based on this result and the effects of lesions on trial initiation, in the following neural analysis we focused on the epoch immediately before trial initiation (0–300 ms before odor poke-in). During this preinitiation epoch, rats have just prepared to start a new trial but have not yet made their decision. Moreover, this period is less contaminated by movement activities that are unrelated to the task such as grooming or resting, as rats are prepared to perform the task. Finally, owing to the blocked task structure, rats already know the expected values of their options during this epoch (except for first 5–10 trials of each block), enabling us to examine value-related activity.

To quantify how the neural responses during the preinitiation epoch are modulated by block-wise changes in value, we projected each neuron onto the same value space as our

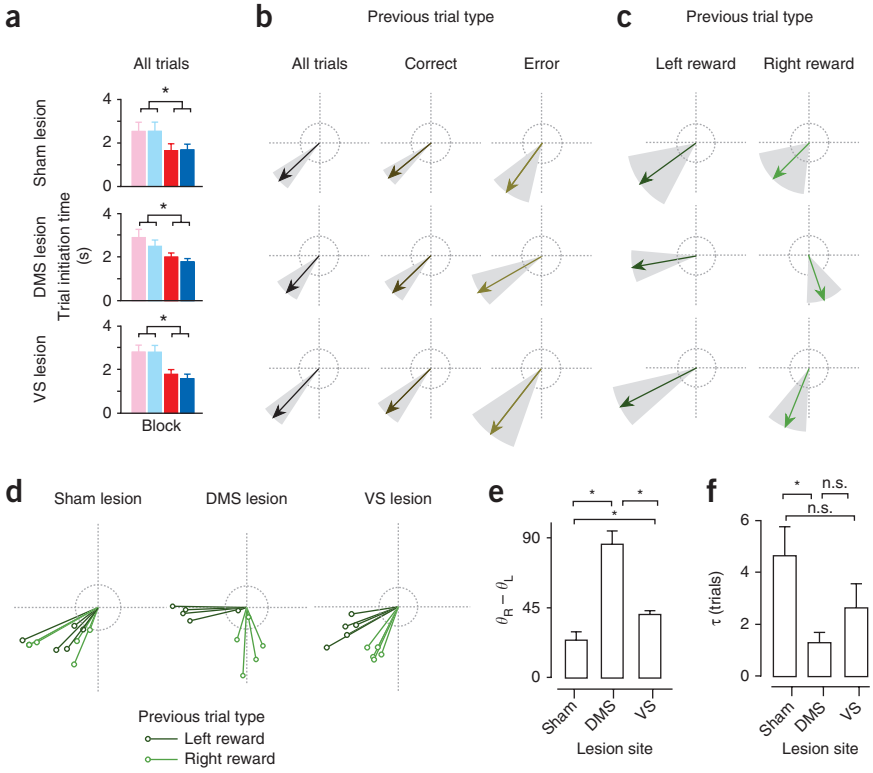


Figure 5 Lesions in DMS impaired net value-dependent modulation of vigor. (**a**) Trial initiation time across blocks (pooled for all trials) in indicated lesion conditions (mean ± s.e.m., $n = 5$ rats for each condition). VS, ventral striatum. $^*P < 0.05$. (**b**,**c**) Trial initiation vectors for one rat separated by previous trial outcome (correct, error, left reward and right reward). Gray, 95% confidence interval. Dotted rings represent a scale of vector length ($r = 1$). (**d**) Trial initiation vectors for all rats, separated by previous left and right reward outcome ($n = 5$ rats for each condition). (**e**) Difference in polar angle between left and right trial vectors of each rat (s.e.m., n.s., not significant and $^*P < 0.05$, *t*-test, $n = 5$ rats for each condition). (**f**) Effects of multiple previous trial outcomes on trial initiation time. Time constants are derived from exponential curves fitted to coefficients of multiple regression (mean ± s.e.m.; n.s., not significant and $^*P < 0.05$, $n = 5$ rats for each condition).

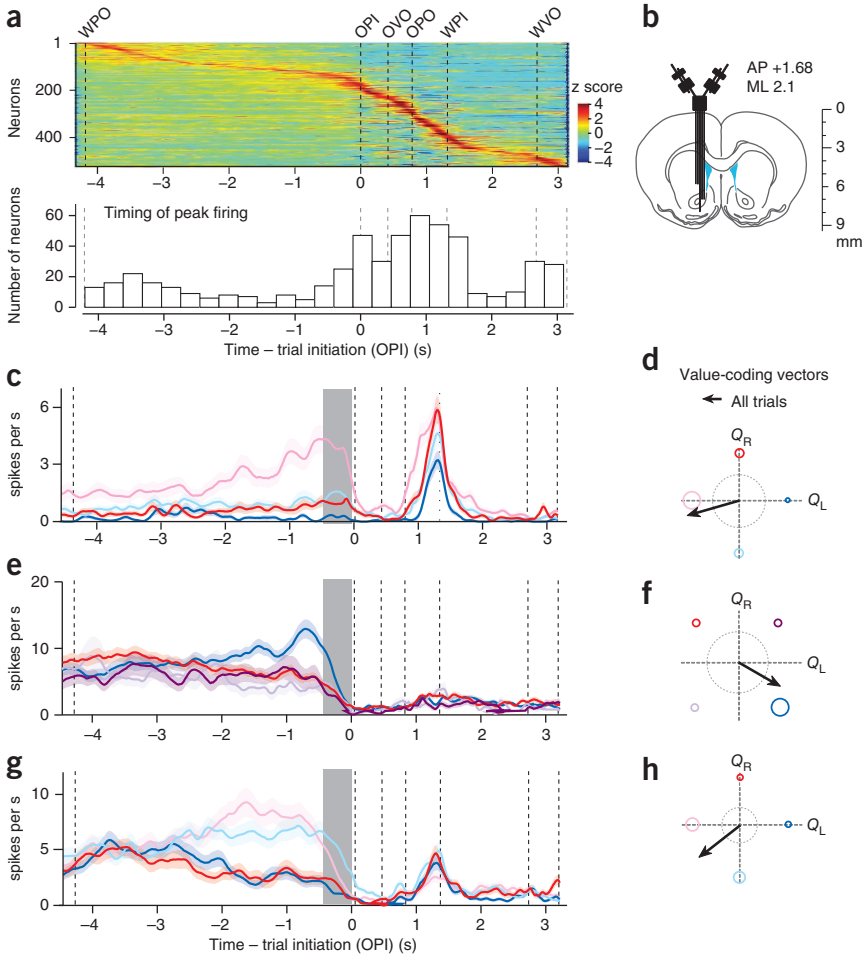


Figure 6 Value-coding neurons in the rat striatum. (**a**) PETHs of all neurons recorded (top); neurons are sorted by time of peak firing (*z*-scored). Time of peak firing (bottom, $n = 522$ neurons). OPI, odor poke-in (entry); OVO, odor valve on (opening); OPO, odor poke-out (exit); WPI, water poke-in; WVO, water valve on (opening); and WPO, water poke-out (exit). (**b**) Recording configuration. AP, distance (mm) from bregma in the anterior-posterior axis (positive values indicate anterior from bregma). ML, distance (mm) from the midline. (**c**–**h**) PETHs of single neurons (solid line indicates mean firing rate, shaded region indicates s.e.m.; colors refer to blocks as in **Fig. 1c** and **Supplementary Fig. 6**) and vector representation of firing rate modulation during pre-trial initiation period (PETHs in **c**,**e**,**g** correspond to vectors in **d**,**f**,**h**, respectively). Dotted rings represent a scale of vector length ($r = 4$).

behavioral analysis (**Fig. 6c**–**h**). We regressed firing rate with left and right value (Q_L and Q_R, respectively).

$$\text{Firing rate (spikes s}^{-1}) = \beta_0 + \beta_L \times Q_L + \beta_R \times Q_R \quad (3)$$

We systematically mapped neural responses to their respective decision-making processes: neurons encoding left value alone point toward 0° or 180°; those encoding right value point toward 90° or 270°. Such neurons encode values in a 'menu-invariant' manner ('absolute value'). Next, neurons encoding relative values ($Q_R - Q_L$) are modulated toward 135° or 315°, and neurons encoding net values ($Q_L + Q_R$) are modulated along 45° or 225°. Value-coding neurons are defined as those that exhibited significant modulation in any direction in this value space ($P < 0.01$, F-test).

We then examined the population of neurons whose activity was significantly modulated by value during pretrial initiation in the DMS and ventral striatum (**Fig. 7a** and **Supplementary Fig. 7**). The activity of 31% (113/364) of DMS neurons and 22.8% (36/158) of ventral striatum neurons were significantly modulated by value ($P < 0.01$, F-test in regression analysis). Although the DMS contained a higher proportion of value-coding neurons than the ventral striatum, the difference was not significant (χ^2 test, $\chi^2_1 = 3.41$, $P = 0.055$). However, the distribution of the types of value-coding neurons in the two regions differed. In the ventral striatum,

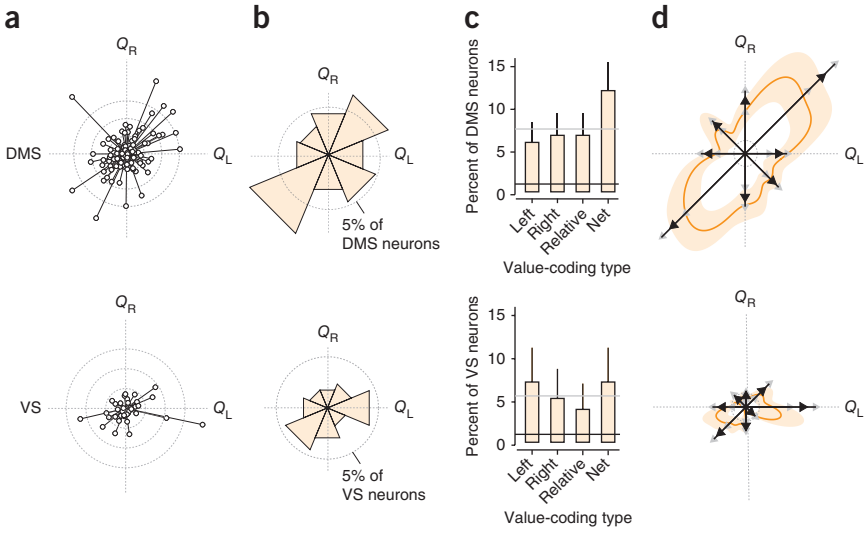


Figure 7 DMS, rather than ventral striatum (VS), predominantly encodes net value. (**a**–**c**) DMS value coding neurons (top). VS value coding neurons (bottom). Population data (**a**): vector representation of all value coding neurons ($P < 0.01$, $n = 149$ of 522 neurons, F-test in regression). Each circle represents one neuron. Inner dotted ring: $r = 8$; middle ring, $r = 16$; and outer ring, $r = 24$. Circular histogram depicting the proportion of neurons in each value-coding category (**b**). Dotted ring, 5%. Distribution of neurons encoding value of one action (left and right), relative value and net value (**c**). Mean ± 95% confidence interval (bootstrap). (**d**) Smoothed vector representations (thick orange line). Population vectors (**Fig. 7a**) were smoothed using a Von Mises distribution (kappa, 20). Mean ± 95% confidence interval (bootstrap; orange light shading). Vectors, resultant vectors of each value-coding type (black arrowheads). Gray arrowheads represent the boundaries of the 95% confidence interval.

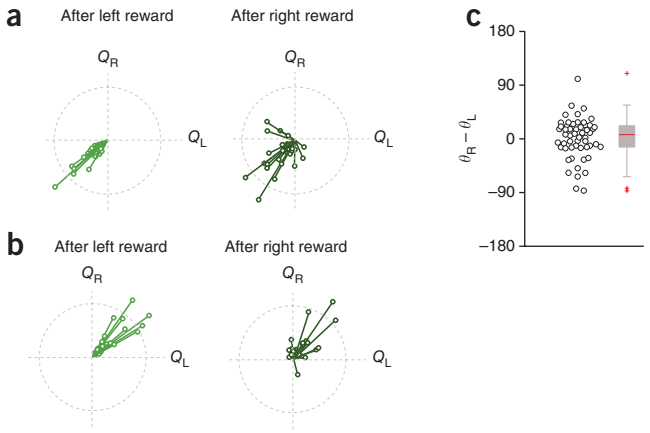


Figure 8 The outcome of immediately preceding trial does not significantly affect the activity of net value-coding neurons. (**a**,**b**) Vectors representing direction of value modulation plotted separately for trials after left and right reward for neurons with negative (**a**) or positive (**b**) correlation with net value. (**c**) Difference in angle between vectors for firing rate after left and right reward trials ($\theta_R - \theta_L$). Each circle represents one value-coding neuron ($n = 54$ neurons; left). In the box plot (right), the central mark indicates the median, and the edges of the box are 25th and 75th percentiles. Vertical line indicates the maximum and minimum values of non-outliers. Error bars indicate the maximum and minimum values of non-outliers. Points are considered as outliers if they are larger than $b + 1.5(b - a)$ or smaller than $a - 1.5(b - a)$, where a and b are the 25th and 75th percentiles, respectively.

the proportion of absolute and net value–coding neurons were significantly represented above chance ($P < 0.01$, binomial test), but relative value-coding neurons were not (**Fig. 7b**; $P > 0.0125$, Bonferroni correction for proportion of relative value-coding neurons). Furthermore, the distribution of value-coding categories (left absolute, right absolute, relative and net value) was not significantly different from a uniform distribution (χ^2 goodness of fit test against uniform, $\chi^2_3 = 2.0$, $P = 0.57$). In the DMS, all categories of neuron types were represented above chance ($P < 0.0125$, binomial test). Moreover, the distribution was significantly different from a uniform distribution (χ^2 goodness of fit test against a uniform distribution, $\chi^2_3 = 12.0$, $P = 0.0077$). The nonuniformity was due to the predominance of a single category: net value–coding neurons (**Fig. 7c**). Net value–coding neurons formed the only category that significantly deviated from what is expected from a uniform distribution ($P = 0.0028$, binomial test). In total, net value–coding neurons were most dominantly represented in the DMS, although all coding types were present.

An advantage of the polar coordinate method is that in addition to examining the number of neurons per category (classified by θ), we can also take into account the strength of representation (amplitude, r, of the vectors). Thus, we computed resultant vectors for each category of neurons with their 95% confidence boundaries (bootstrap) and projected them onto the polar coordinate (**Fig. 7d**). This analysis also supported that the DMS predominantly represents net value.

Finally, we examined whether the activity of net value–coding neurons was significantly affected by the immediately preceding trial's outcome. The vector analysis indicated that the majority of neurons (>80%) showed net value–dependent modulation of firing regardless of previous trial's outcome (**Fig. 8a**,**b**). For the population, the angle between the two vectors representing firing rates immediately after left reward and after right reward trials ($\theta_L - \theta_R$) were not significantly different from zero (**Fig. 8c**, paired t-test, $t_{53} = 0.9$, $P > 0.05$). Moreover, we regressed firing rate with net value and previous trial outcome and found that data for 94% (51/54) of the neurons were significant for net value whereas 13% (7/54) were significant for both net value and previous trial's outcome but only 3.7% (2/54) were significant for previous trial's outcome alone.

DISCUSSION

We teased apart two orthogonal components of decision making: how to choose between alternate actions and how to choose the vigor applied to actions. First, we demonstrated that response vigor fluctuates globally with the net value of the rat's options (or the average expected reward rate). Second, lesions in the DMS strongly diminished the global effects of net value on motivation, rendering rats' response vigor to depend on immediately preceding trial outcomes rather than on the net value. We observed a weaker effect in the ventral striatum. In contrast, action selection depended on the relative value of available options and was not affected by lesions of DMS or ventral striatum. Finally, the DMS was more enriched with net value–coding neurons than the ventral striatum. Together, these results demonstrated a critical role of the DMS in net value–dependent regulation of response vigor.

In decades past, measurement of response vigor in behaviors such as lever-press tasks and key-pecking tasks have been central[6,7,25], but action selection was not well measured. In recent years, the opposite problem has emerged, where the primary focus has been on action selection, often neglecting intertrial intervals. Although there have been a few studies that examined how incentive values of goals modulate reaction times of actions directed toward specific goals, such questions differ in that they do not address how long-term estimates of the rat's state can have global effects on response vigor[9–11,26]. The present study is one of the first to demonstrate that the total value of the rat's options (that is, the value of the rat's state) globally influences vigor. Such global considerations for choosing vigor have been previously hypothesized to be an optimal strategy for appropriately exploiting a task only when it is worthwhile and slowing down or relaxing when it is not. The present study paves a way toward dissociating and examining both aspects of decision making in a unified task.

The role of dorsomedial striatum in energizing motivation

The ventral striatum has long been linked to motivation[18,19] because of its role in intracranial self-stimulation[27], drug addiction[28] and effort-related decision making[29], and its anatomical connections with limbic structures[30]. In contrast, dorsal striatum is less known to be involved in motivation, but studies using dopamine-deficient mice have shown that restoration of dopamine into the dorsal striatum can sufficiently rescue rats' motivation to engage in reward-oriented behavior[20]. Studies in humans have also demonstrated the role of the dorsal striatum in motivation[9,31,32]. Further, DMS lesions impair reaction time and initiation latency in rats[22,33,34], electrical stimulation of primate caudate nucleus influences reaction time toward immediate goals[10], and several physiology and lesion studies have implicated the DMS in flexible, reward-oriented behavior[32–36]. Despite these results, none of the studies explicitly separated the directing and energizing effects of motivation. Our results suggest that the DMS may normally be critical for integrating the values of the rat's options, or computing net value, and regulating response vigor. In support of this, our electrophysiology results show that net-value representation is enriched in the DMS.

Our results do not exclude the possibility that the ventral striatum was involved in the regulation of response vigor in other behavioral contexts. It should be noted that in our task, the rate of reward depends on rats' actions, that is, the task is instrumental (rather than Pavlovian).

It is possible that there are distinct control mechanisms for response vigor in Pavlovian versus instrumental contexts.

Recent studies in rats and humans have indicated that DMS is involved in goal-directed behavior (or model-based learning)[37]. A hallmark of goal-directed behavior, but not of habitual behavior, is its sensitivity to the changes of the outcome values associated with specific actions (action-outcome associations). For example, if the outcome value of one of two potential actions is reduced, the performance of the devalued, but not the non-devalued action decreases[38,39]. It has been shown that lesions of DMS render rats insensitive to such devaluations, suggesting that DMS supports behavior based on associations between specific actions and outcomes. In contrast, our result showed that DMS lesions impaired the response vigor's dependency on the net value (values general to potential actions) but spared choice biases that depend on the values associated with specific actions. We note that previous studies[38] had highlighted the importance of the more posterior region of the DMS than the region we studied here. It is therefore possible that the anterior and posterior DMS underlie two distinct aspects of goal-directed behavior: anterior DMS in behavior dependent on action-general values (net values) and posterior DMS in behavior dependent on action-specific values.

Here we focused on the value of options in the task. However, one can argue that rats are choosing between performing the task we presented and other activities outside the task (such as grooming or resting). It is thus possible that the change in trial initiation time is due to the changes in the value of performing the task relative to the value of performing other activities. Alternatively, the normative theory of motivation would predict that the 'vigor' of grooming or other activities outside the task is also slowed down during the period of low net value, similar to trial initiation time in the task[4]. Measurement of the vigor of outside activities (for example, testing whether rats groom rapidly during high net value) will allow us to tease apart these possibilities.

A popular model of decision making describes a hierarchical architecture, where values of individual actions (absolute values) are represented in the DMS, and these representations are read out and used by downstream areas to compute the relative values needed for action selection[1]. Although this hypothesis comes from the finding that neurons in DMS (including the anterior- to mid-caudate) encode values of specific actions (absolute value)[15,16], our results show that net-value representation in the DMS is more enriched than has been previously thought. One difference between the present and previous studies is task epochs used for the analyses: we focused on the pre-trial initiation period whereas previous studies focused on the period just preceding action selection. However, this difference alone may not account for the discrepancy. First, our analysis using pre-action selection epochs did not support the dominance of absolute value representation (**Supplementary Fig. 8**). Second, and more importantly, the analysis methods of past studies had biases in the classification of their neuronal responses (**Supplementary Figs. 8** and **9**). Our analysis using polar representations is not prone to these biases.

Diversity of value coding

Our data show that value-coding neurons in the striatum are distributed rather continuously in the polar coordinate (**Fig. 7a**). Rather than simply forming distinct categories of value-coding types, the population seems to encode diverse linear combinations of value. Future studies will be needed to understand how this diversity is generated. One possibility is that relative and net-coding activities are indeed secondarily derived from absolute value representations. It is also possible that net- and relative-value representations in the DMS do not depend on reading out absolute-value representations, and are explicitly or directly represented. For example, tonic dopamine levels may convey net-value information to DMS neurons[4]. It will be crucial to examine how cortical afferents, dopaminergic and other neurotransmitter systems modulate this diversity of value-modulated responses.

Choosing the general pace of performance in conjunction with what specific action to take is vital for behavioral regulation. By providing a theoretical framework, a behavioral paradigm, and analytical tools, our study promotes a more inclusive understanding of decision making. Applying these approaches to future experiments in different brain regions will further our understanding of how the brain regulates value representation and goal-oriented behavior.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank J. Assad, K. Blum, O. Hikosaka, D. Lee, M. Livingstone, J. Maunsell and M. Meister for their comments on the manuscript; R. Born and the members of the Uchida laboratory for discussions; and S. Kim for support in behavioral experiments. This work was supported by a fellowship from National Science Foundation (A.Y.W.), Precursory Research for Embryonic Science and Technology (K.M.), a Smith Family New Investigator Award, the Alfred Sloan Foundation, the Milton Fund and the Startup Fund from Harvard University (N.U.).

AUTHOR CONTRIBUTIONS

A.Y.W. wrote the manuscript, performed the experiments, performed the analyses and was involved in the experimental design. K.M. performed the simulations. N.U. was involved in preparing the manuscript and experimental design.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Kable, J.W. & Glimcher, P.W. The neurobiology of decision: consensus and controversy. *Neuron* **63**, 733–745 (2009).
2. Lee, D., Seo, H. & Jung, M.W. Neural basis of reinforcement learning and decision making. *Annu. Rev. Neurosci.* **35**, 287–308 (2012).
3. Niv, Y., Joel, D. & Dayan, P. A normative perspective on motivation. *Trends Cogn. Sci.* **10**, 375–381 (2006).
4. Niv, Y., Daw, N.D., Joel, D. & Dayan, P. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl.)* **191**, 507–520 (2007).
5. Guitart-Masip, M., Beierholm, U.R., Dolan, R., Duzel, E. & Dayan, P. Vigor in the face of fluctuating rates of reward: an experimental examination. *J. Cogn. Neurosci.* **23**, 3933–3938 (2011).
6. Hebb, D.O. Drives and the C.N.S. (conceptual nervous system). *Psychol. Rev.* **62**, 243–254 (1955).
7. Hull, C.L. *Principles of Behavior: An Introduction to Behavior Theory* (D. Appleton-Century Company, Inc., Oxford, England, 1943).
8. Ferster, C.B. & Skinner, B.F. *Schedules of Reinforcement* (D. Appleton-Century Company, Inc., Oxford, England, 1957).
9. Nakamura, K. & Hikosaka, O. Role of dopamine in the primate caudate nucleus in reward modulation of saccades. *J. Neurosci.* **26**, 5360–5369 (2006).
10. Ding, L. & Gold, J.I. Separate, causal roles of the caudate in saccadic choice and execution in a perceptual decision task. *Neuron* **75**, 865–874 (2012).
11. Roesch, M.R. & Olson, C.R. Neuronal activity related to reward value and motivation in primate frontal cortex. *Science* **304**, 307–310 (2004).
12. Gurney, K., Prescott, T.J. & Redgrave, P. A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biol. Cybern.* **84**, 401–410 (2001).
13. Hikosaka, O., Nakamura, K. & Nakahara, H. Basal ganglia orient eyes to reward. *J. Neurophysiol.* **95**, 567–584 (2006).
14. Turner, R.S. & Desmurget, M. Basal ganglia contributions to motor control: a vigorous tutor. *Curr. Opin. Neurobiol.* **20**, 704–716 (2010).
15. Lau, B. & Glimcher, P.W. Value representations in the primate striatum during matching behavior. *Neuron* **58**, 451–463 (2008).
16. Samejima, K., Ueda, Y., Doya, K. & Kimura, M. Representation of action-specific reward values in the striatum. *Science* **310**, 1337–1340 (2005).

17. Cai, X., Kim, S. & Lee, D. Heterogeneous coding of temporally discounted values in the dorsal and ventral striatum during intertemporal choice. *Neuron* **69**, 170–182 (2011).
18. Kelley, A.E. Ventral striatal control of appetitive motivation: role in ingestive behavior and reward-related learning. *Neurosci. Biobehav. Rev.* **27**, 765–776 (2004).
19. Mogenson, G.J., Jones, D.L. & Yim, C.Y. From motivation to action: functional interface between the limbic system and the motor system. *Prog. Neurobiol.* **14**, 69–97 (1980).
20. Palmiter, R.D. Dopamine signaling in the dorsal striatum is essential for motivated behaviors: lessons from dopamine-deficient mice. *Ann. NY Acad. Sci.* **1129**, 35–46 (2008).
21. Amalric, M. & Koob, G.F. Depletion of dopamine in the caudate nucleus but not in nucleus accumbens impairs reaction-time performance in rats. *J. Neurosci.* **7**, 2129–2134 (1987).
22. Hauber, W. & Schmidt, W.J. Differential effects of lesions of the dorsomedial and dorsolateral caudate-putamen on reaction time performance in rats. *Behav. Brain Res.* **60**, 211–215 (1994).
23. Uchida, N. & Mainen, Z.F. Speed and accuracy of olfactory discrimination in the rat. *Nat. Neurosci.* **6**, 1224–1229 (2003).
24. Rorie, A.E., Gao, J., McClelland, J.L. & Newsome, W.T. Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP) of the macaque monkey. *PLoS ONE* **5**, e9308 (2010).
25. Skinner, B.F. *The Behavior of Organisms: An Experimental Analysis* (D. Appleton-Century Company, Inc., Oxford, England, 1938).
26. Tai, L.H., Lee, A.M., Benavidez, N., Bonci, A. & Wilbrecht, L. Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. *Nat. Neurosci.* **15**, 1281–1289 (2012).
27. Mogenson, G.J., Takigawa, M., Robertson, A. & Wu, M. Self-stimulation of the nucleus accumbens and ventral tegmental area of Tsai attenuated by microinjections of spiroperidol into the nucleus accumbens. *Brain Res.* **171**, 247–259 (1979).
28. Imperato, A. & Di Chiara, G. Preferential stimulation of dopamine release in the nucleus accumbens of freely moving rats by ethanol. *J. Pharmacol. Exp. Ther.* **239**, 219–228 (1986).
29. Salamone, J.D., Correa, M., Farrar, A.M., Nunes, E.J. & Pardo, M. Dopamine, behavioral economics, and effort. *Front. Behav. Neurosci.* **3**, 13 (2009).
30. Newman, R. & Winans, S.S. An experimental study of the ventral striatum of the golden hamster. I. Neuronal connections of the nucleus accumbens. *J. Comp. Neurol.* **191**, 167–192 (1980).
31. Delgado, M.R., Stenger, V.A. & Fiez, J.A. Motivation-dependent responses in the human caudate nucleus. *Cereb. Cortex* **14**, 1022–1030 (2004).
32. Volkow, N.D. *et al.* "Nonhedonic" food motivation in humans involves dopamine in the dorsal striatum and methylphenidate amplifies this effect. *Synapse* **44**, 175–180 (2002).
33. Brown, V.J. & Robbins, T.W. Elementary processes of response selection mediated by distinct regions of the striatum. *J. Neurosci.* **9**, 3760–3765 (1989).
34. Bailey, K.R. & Mair, R.G. The role of striatum in initiation and execution of learned action sequences in rats. *J. Neurosci.* **26**, 1016–1025 (2006).
35. Thorn, C.A., Atallah, H., Howe, M. & Graybiel, A.M. Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. *Neuron* **66**, 781–795 (2010).
36. Kimchi, E.Y. & Laubach, M. Dynamic encoding of action selection by the medial striatum. *J. Neurosci.* **29**, 3148–3159 (2009).
37. Balleine, B.W. & O'Doherty, J.P. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* **35**, 48–69 (2010).
38. Yin, H.H., Knowlton, B.J. & Balleine, B.W. Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. *Eur. J. Neurosci.* **22**, 505–512 (2005).
39. Yin, H.H., Ostlund, S.B., Knowlton, B.J. & Balleine, B.W. The role of the dorsomedial striatum in instrumental conditioning. *Eur. J. Neurosci.* **22**, 513–523 (2005).

ONLINE METHODS

Experiments. All procedures involving animals were carried out in accordance with US National Institutes of Health standards and approved by the Harvard University Institutional Animal Care and Use Committee. All values were represented by the mean ± standard error unless otherwise noted. All hypothesis testing was two-sided, unpaired tests unless otherwise noted.

Behavior. Fifteen male Long-Evans hooded rats (250–300 g) were trained to perform an odor-discrimination task for water reward[23]. Rats were pair-housed under a 12-h light-dark cycle during the training period (and were housed individually later for electrophysiology and lesion experiments). All experiments were performed during the rat's dark cycle. Rats self-initiated each trial by introducing their snout into a central port, which triggered odor delivery. Valid odor pokes were restricted to trials where rats delivered a nose poke that lasted at least 20 ms and delivered a single nose poke into the odor port. Multiple successive pokes, such as two pokes in a row, aborted the trial and triggered a 4-s intertrial interval. Valid odor pokes also must occur outside the intertrial interval. After a variable delay, drawn from a uniform random distribution of 0.3–0.5 s, a binary mixture of two pure odorants, caproic acid and 1-hexanol, was delivered at one of four concentration ratios (100/0, 60/40, 40/6, 0/100) in pseudorandom order within a session. After a variable odor sampling time, rats responded by withdrawing from the central port, which terminated the delivery of odor and moved to the left or right water port. Choices were rewarded according to the dominant component of the mixture, that is, at the left port for mixtures A/B < 50/50 and at the right port for A/B > 50/50. The next 'trial start' time commences 4 s after closure of the water valve. The rat's trial initiation time was defined as the latency it took for the rat to poke back into the odor port after this trial start time. Blocks were randomly interleaved within a session, which contained 5–9 blocks.

Odor pokes during the inter-trial interval. The task had a forced 4-s time out between trials, and the analyses focused on valid odor pokes only. However, it is possible that rats poke during this intertrial interval. However, very few odor pokes occurred before the end of the trial (<20%, odor pokes before time 0 s; **Supplementary Fig. 1**), and our conclusions were not significantly affected by the intertrial interval.

Movement time. Under vigor theory, one may predict that all components of actions will speed up during times of high net value[4]. We analyzed movement time (time from odor port exit to reward port) across blocks in individual rats and in the population. Although some exhibited showed faster movement time in high-net-value blocks (**Supplementary Fig. 2a**, *t*-test between high and low value blocks, $t_{28} = 2.8$, $P < 0.05$, $n = 7$ sessions per block type for one rat) this effect was very small and not consistent across the population (**Supplementary Fig. 2b**, *t*-test between high and low value blocks, $t_{58} = 0.5$, $P > 0.05$, $n = 15$ rats per block type). For the example rat shown, movement time was fastest at 0.28 s in high-value blocks and slowest at 0.31 s in low-value blocks (**Supplementary Fig. 2a**). Thus, the difference in time between high- and low-value blocks was on the order of tens of milliseconds, possibly because the rat was already moving near maximum speed in high-value blocks. Nevertheless, there was a trend for the population to speed up movement time in high-value blocks.

Comparing τ to the difference between θ_R and θ_L. A small difference in θ_R and θ_L corresponds to less dependence of trial initiation time on immediately preceding trials and should correspond to larger τ. Conversely, a large difference between θ_R and θ_L (~90°) should correspond to smaller τ. To understand the quantitative relationship between τ and $\theta_R - \theta_L$, we performed the following simulation. Using a given value of τ, we predicted trial initiation times based on rat's actual reward history. We then obtained θ_R and θ_L, using the predicted trial initiation times for right reward and left reward trials (**Supplementary Fig. 3**). This simulation predicted that there is indeed a negative relationship between τ and $\theta_R - \theta_L$, and that $\theta_R - \theta_L$ and τ obtained from the data fall within this prediction (**Supplementary Fig. 3**, τ and $\theta_R - \theta_L$, mean and standard error). This result demonstrates that our observations using trial initiation vectors and time constants (regression with reward history) are quantitatively consistent.

Lesions. After 6 weeks of training on the reward-manipulation task, rats were randomly selected to be in one of three conditions: DMS lesion, ventral striatum lesion and sham lesion. The experimenter was not blinded to group allocation. Rats were individually housed after surgery. Rats were lesioned with ibotenic acid (250 nl, 10 mg/ml) in either the DMS (AP 1.68, ML 2.0 and DV 4.5) or ventral striatum (AP 1.68, ML 1.5 and DV 7.4). Sham-lesion rats were bilaterally infused with saline of the same volume. We used 5 rats per category, as it was the minimum needed to obtain statistical significance for our analyses. After 1 week of recovery, rats were water-deprived. Behavioral performance for the first 7–8 d after recovery was used for the behavioral analysis. All rats ($n = 15$) were perfused and stained with Nissl as described before[40]. Some sections were used for neuron-specific labeling, where NeuN was used as primary antibody (1:200, Millipore) and Alexa Fluor 568 (1:500, Invitrogen) used as the secondary antibody[41] (**Supplementary Fig. 4**).

Electrophysiology. Electrophysiological experiments were performed as described before[40]. Briefly, rats were implanted with custom-made microdrives in the left, anteromedial striatum (1.7 mm anterior to bregma and 2.1 mm lateral to the midline). Extracellular recordings were obtained with 12 independently movable tetrodes using the Cheetah system (Neuralynx), and single units were isolated by manually clustering spike features with MClust (A. D. Redish). Cells were recorded at various depths between 3.5 mm and 9 mm ventral to the surface of the skull. The boundary between dorsal and ventral striatum was 6 mm deep. The depth of each cell was reconstructed by calculating the number of turns made on each tetrode screw (each turn = 0.32 mm) and confirmed using the final length of each tetrode (through histological examination and measuring the length of the tetrodes after removal of the drive). We recorded 522 neurons, with 364 neurons from DMS and 158 neurons from ventral striatum. The sample size is comparable to similar studies in the field and was sufficient for the statistical analyses used in this study.

Behavior data analysis. We developed a method to represent, using a single vector, how behavior and neural activity were modulated by changes across different reward blocks. The first and last 30 trials per session were excluded from the analysis for examining trial initiation, to examine steady-state behavior that is independent of satiety. The first 10 trials after each block transition were eliminated to exclude the effects of learning. To obtain a behavioral vector, we regressed each rat's trial initiation time by the reward amounts of the left and right water ports, which varied across blocks. The values of the left and right choices, Q_L and Q_R, were defined by reward amounts (water valve duration; we confirmed the linear relationship between the valve durations and the delivered reward amounts).

$$\text{Trial initiation time (s)} = \beta_0 + \beta_L \times Q_L + \beta_R \times Q_R$$

We used *F*-test to see whether behavior was significantly modulated by the option values (Q_L and Q_R). The *F*-test tests whether a proposed regression model as a whole fit the data significantly better than a simpler model (trial initiation = β_0)[42].

The amplitude (r) of the behavior vector was calculated by taking the square root of the sum of the square of the coefficients:

$$r = \sqrt{(\beta_L)^2 + (\beta_R)^2}$$

the polar angle (θ) was calculated by taking the four-quadrant arc-tangent of the coefficients:

$$\theta = \tan^{-1}\left(\frac{\beta_R}{\beta_L}\right)$$

To determine how trial initiation time was modulated by reward size, we divided the polar plot into 8 segments of 45°. θ values of the behavior vectors falling between −22.5° and +22.5° represents a positive correlation between trial initiation and size of left reward; behavior vectors whose θ values fall between 22.5° and 67.5° represents a positive correlation between trial initiation and net value, θ value between 67.5° and 112.5° represents a positive correlation with right reward, θ value between 112.5° and 157.5° represents positive correlation with right > left, θ value between 157.5° and 202.5° represents negative correlation with left reward and so on.

To determine whether the two vectors for left and right trials have significantly different angles, we shuffled left and right trials so that we effectively ignore

doi:10.1038/nn.3377

choice direction. We then obtain trial initiation vectors for the shuffled trials and determined whether their angle difference is significantly different from that derived from the original, unshuffled data.

Neuronal data analysis. Our analysis focused on the pre-trial initiation period, the 0–300 ms window before odor poke in. The first 10 trials of every block were not used to eliminate the effect of learning. Next, we regressed each neuron's firing rate by the reward amounts, which varied across blocks.

$$\text{Firing rate} = \beta_0 + \beta_L \times Q_L + \beta_R \times Q_R$$

We used *F*-test to select for neurons ($P < 0.01$). The *F*-test tests whether a proposed regression model as a whole fit the data significantly better than a simpler model (firing rate = β_0). Because this fitting is invariant to the choice of axes, *F*-test is also invariant to the choice of independent variables.

The amplitude (r) of each neuron was calculated by taking the square root of the sum of the coefficients:

$$r = \sqrt{(\beta_L)^2 + (\beta_R)^2}$$

the polar angle (θ) was calculated by taking the four-quadrant arc-tangent of the coefficients:

$$\theta = \tan^{-1}\left(\frac{\beta_R}{\beta_L}\right)$$

Similar to the behavioral vector classification, we divided the polar plot into eight segments of 45° to classify the neural responses. Neurons whose θ values fall between −22.5° and +22.5° were classified as left-positive value–coding, neurons whose θ fall between 22.5° and 67.5° were positive-state value–coding, neurons whose θ fall between 67.5° and 112.5° were right-positive value–coding, neurons whose θ fall between 112.5° and 157.5° were right-preferring relative value-coding and so on.

We used a more stringent method to classify the responses of our population. By bootstrapping, we obtained a 95% confidence interval on each angle. Here only neurons whose confidence intervals crossed exactly one classification boundary can be included (each boundary being 0, 45, 90, 135, 180, 225, 270 and 315). With the more stringent classification method, 80/522 neurons were classified as value-coding. Net value–coding neurons were in the majority, comprising 33.8% of the 80 value-coding neurons.

PETHs in **Figure 6** show the firing rates of each block condition between odor port entry and water port entry. As the timing between each task event is variable, for visualization purposes, the neural activities were aligned to all events, and 'time-warped,' so that firing rate (spikes/s) is preserved, but the time windows between epochs are kept constant. The timing of each epoch (odor poke-in, odor valve on and so on) was averaged across all trials. Firing rates between epochs were obtained for every trial and then averaged for the PETH display. Time-warped firing rates were used only for displaying PETHs and not for the analyses below.

We selected all neurons that had their peak firing activity between water poke-out and odor poke-in and obtained 181 neurons in total (127/364 in DMS and 54/158 in ventral striatum). We performed the above regression analysis and vector projection using a 250-ms window around the time of the each neuron's time of peak activity. With this analysis, the dominant neural response in the striatum was net value–coding (**Supplementary Fig. 7b**). DMS was net value–coding in both the positive and negative direction whereas the ventral striatum was primarily negative net value–coding (**Supplementary Fig. 7b**).

Comparing our proposed vector analysis with previous studies' analyses. Our finding that the DMS encodes net value seemingly contradicts with the current framework that the caudate (or DMS) primarily encodes the absolute values of choices, or 'action values'[15,16]. Previous studies have used different animal models (primates versus rodents), task structures (varying reward value by temporal discounting or reward probability versus varying reward size) and analysis epochs (after stimulus onset instead of during before trial initiation). However, the most critical differences may be in our analysis methods. In studies that suggested that DMS primarily encodes absolute action values, firing rates were regressed by the independent variables, Q_L and Q_R, representing the 'action value' of left and right, respectively. After running multiple regression (or Mann-Whitney *U*-test), neurons were classified to be absolute value–coding if the coefficients for either Q_L or Q_R, but not both, were significantly different from zero[15–17]. Yet neurons were relative (or net) value–coding if coefficients for both Q_L and Q_R were significantly different from zero (**Supplementary Fig. 8a**, conventional method type 1). Conversely, other studies that examined the role of ventral striatum and other areas in net value–coding applied the opposite regression: $Q_R - Q_L$ and $Q_L + Q_R$ as independent variables, where neurons were considered relative value– or net value–coding if either $Q_R - Q_L$ or $Q_L + Q_R$ were significant, respectively (conventional method type 2)[17,43,44].

We first performed a simulation (**Supplementary Fig. 8**) to test whether the two different methods can correctly capture a neural population that is uniformly involved in all aspects of value coding, that is, uniformly distributed in the polar coordinate representation (an 'extreme' case). To match numbers obtained in past striatal studies, we assumed about 36% of the population was statistically significant at the 5% level (for more general cases, **Supplementary Fig. 9**). Applying conventional method 1, 32% of the population are classified as absolute value–coding, but only 4% are relative value– or net value–coding. Therefore, the vast majority (89%) of value coding neurons was classified as absolutevalue–coding, or participate in valuation. Conversely, by applying conventional method 2 (**Supplementary Fig. 9**), 89% of significant neurons were relative value– or net value–coding, whereas the rest were absolute-value coding. Therefore, neither choice of axes for the regression was fair at classification for a uniform distribution. We visualized these systematic biases in classification in **Supplementary Figure 8a**,**b**, by plotting a randomly generated and uniformly distributed neural population distributed in two-dimensional space onto each methods' regression plane (neurons falling into different sectors of the critical regions are color-coded and classified accordingly). By replacing the Cartesian coordinate system with a polar coordinate system, so that each neuron's activity is represented by a polar angle (θ) and amplitude (r), equal numbers of neurons were classified into the four categories (**Supplementary Fig. 8c**). Therefore, we believe that analyzing neural responses with our polar method is fairer at classifying changes related to the absolute, relative and net values of the rat's options.

We obtained similar biases when we applied past regression methods onto our own data set. Similarly, we found that with conventional method type 1 (regression with Q_L and Q_R), 74.6% of all significant striatal neurons were classified as absolute value–coding (**Supplementary Fig. 8a**). On the contrary, using conventional method type 2, 79.4% of significant neurons were relative or state coding (**Supplementary Fig. 8b**). Lastly, we examined another time epoch (pre-odor valve on) for the analysis and obtained similar biases with conventional method 1 and 2 (**Supplementary Fig. 8**).

Therefore, discrepancies between our and previous studies may stem from differences in regression methods.

Simulations: results of categorization by both conventional methods are sensitive to the strengths of neural responses. In conventional regression methods, the relative proportions of different value-coding categories (absolute value– and relative/net value–coding neurons) are sensitive to the total proportion of statistically significant neurons. To illustrate this, we simulate neural populations that are uniformly distributed in two-dimensional space with various standard deviations. When there are no signals in the population (**Supplementary Fig. 9a**, a normal distribution with s.d. = 1), only chance levels of neurons are selected with a criterion of 5%. In this case, most neurons are significant for a single independent variable (single-positive) and very few are significant for both independent variables (double-positive). In contrast, in distributions with high s.d. and thus larger proportions of significant neurons, (**Supplementary Fig. 9d**, s.d. = 6.0), most of these significant neurons are significant for both independent variables (double-positive) whereas a smaller fraction are significant for only one independent variable (single-positive). Thus, the proportions of neurons classified as single-positive or double positive are sensitive to the s.d. of the distributions, or the amount of signal in a population. (Note that with conventional method type 1, single-positive types are absolute value–coding whereas double-positive types are relative/state–coding; the opposite holds for conventional method type 2).

By defining p as the probability of significance for each independent variable, then the probabilities of single-positive, double-positive and total positive (single-positive plus double-positive) types are given as follows (**Supplementary Fig. 9e**):

$$\text{Single-positive: } 2p - 2p^2$$

$$\text{Double positive: } p^2$$

$$\text{Total positive: } 2p - p^2$$

Therefore, the relative frequency of single-positive points among total positive points is given (**Supplementary Fig. 9f**):

$$\text{Single-positive/total: } \frac{2-2p}{2-p}$$

Therefore when $P = 0.20$ (20% of the neurons are significantly modulated by either variable), the distribution is about 1.53 s.d., so 88.9% of significant neurons are classified as single-positive whereas only 11.1% are classified as double-positive. (In this case, about 36% of neurons in total are deemed 'significant', which is close to the percentage obtained in the present and previous studies).

Contrary to the conventional methods, our method (the polar method) is invariant to overall proportions of significant neurons.

Our proposed method is invariant to the choice of axes for regression. We applied Q_L and Q_R as independent variables for multiple regression. With our proposed method of categorization, 32, 32, 30 and 55 neurons out of 522 in total were left absolute, right absolute, relative, and net value coding, respectively. When we applied $Q_L + Q_R$ and $Q_R - Q_L$ as independent variables, we achieved exactly the same results (to categorize the neurons, all categories are simply shifted +45°). This confirms that our method is invariant to the choice of independent variables (axes).

Value-coding cells and their anatomical cell types. We examined how the different types of value-coding neurons may be mapped onto various types of neurons. We plotted the firing rate of each neuron against its spike width and color coded each neuron by its value-coding type (**Supplementary Fig. 10**).

40. Cury, K.M. & Uchida, N. Robust odor coding via inhalation-coupled transient activity in the mammalian olfactory bulb. *Neuron* **68**, 570–585 (2010).
41. Jongen-Relo, A.L. & Feldon, J. Specific neuronal protein: a new tool for histological evaluation of excitotoxic lesions. *Physiol. Behav.* **76**, 449–456 (2002).
42. Draper, N.R. & Smith, H. *Applied Regression Analysis* (John Wiley & Sons, Inc., New York, 1966).
43. Ito, M. & Doya, K. Validation of decision-making models and analysis of decision variables in the rat basal ganglia. *J. Neurosci.* **29**, 9861–9874 (2009).
44. Seo, H. & Lee, D. Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J. Neurosci.* **27**, 8366–8377 (2007).

SUPPLEMENTARY INFORMATION

The dorsomedial striatum encodes net expected return, critical for energizing performance vigor

Alice Y. Wang[1], Keiji Miura[1,2,3], Naoshige Uchida[1]

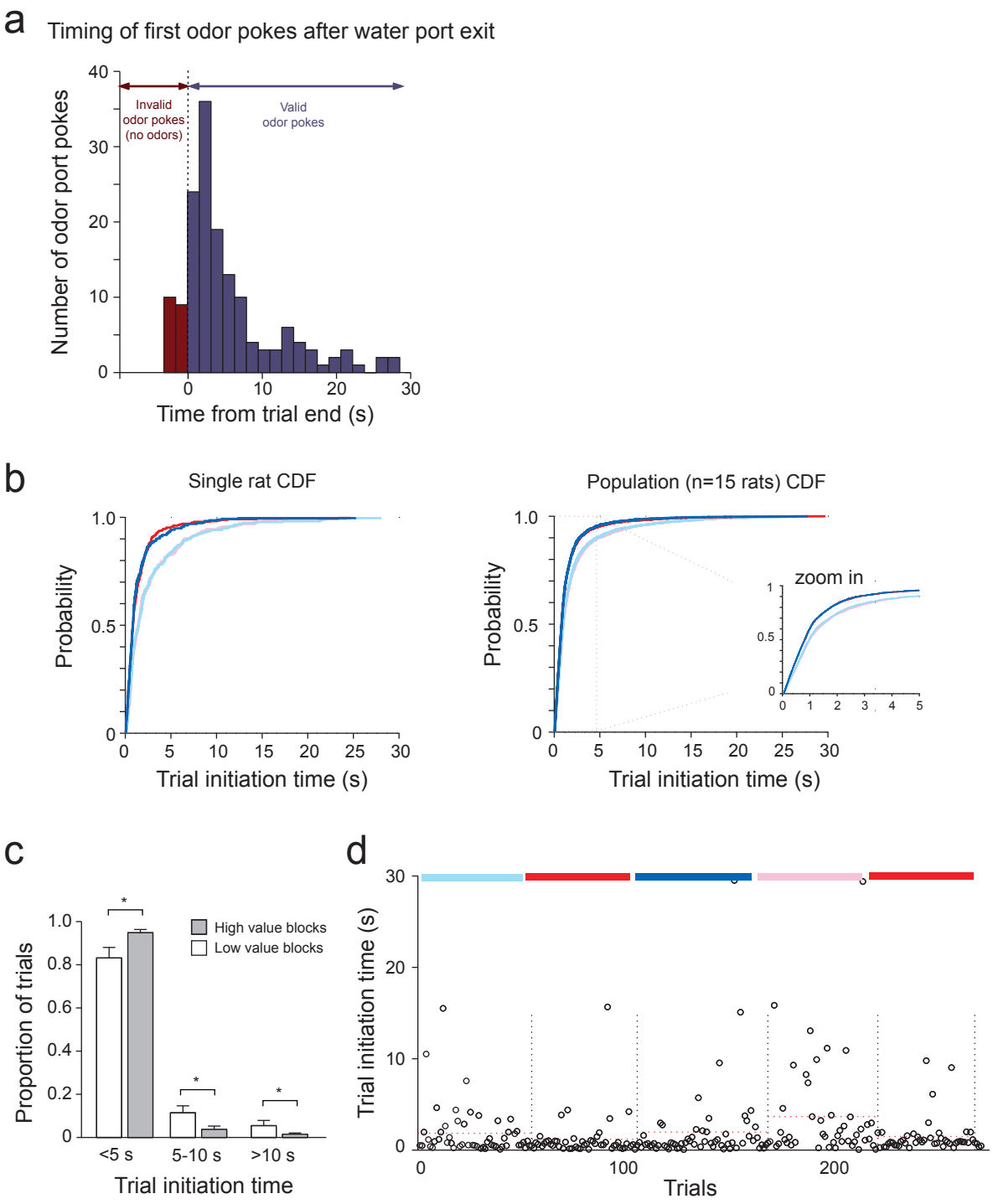
[1] Center for Brain Science, Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, USA

[2] PRESTO, JST, Saitama, Japan

[3]Graduate School of Information Sciences, Tohoku University, Sendai, Japan

Correspondence should be addressed to N.U. (uchida@mcb.harvard.edu).

Supplementary Figure 1-10.



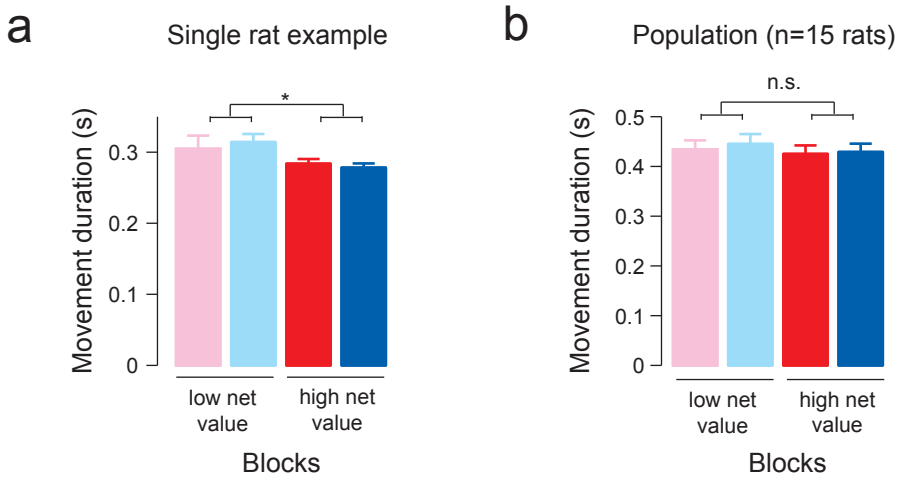
Supplementary Fig. 1| Distribution of trial initiation times

(a) Histogram of the timing of first nose pokes after water port exit relative to trial end time. Few nose pokes occurred prematurely, before the end of the inter-trial interval (red bars).In the main analysis, trial initiation times were calculated only for valid nose pokes (blue bars). However, including premature nose pokes does not affect the results presented in the paper.

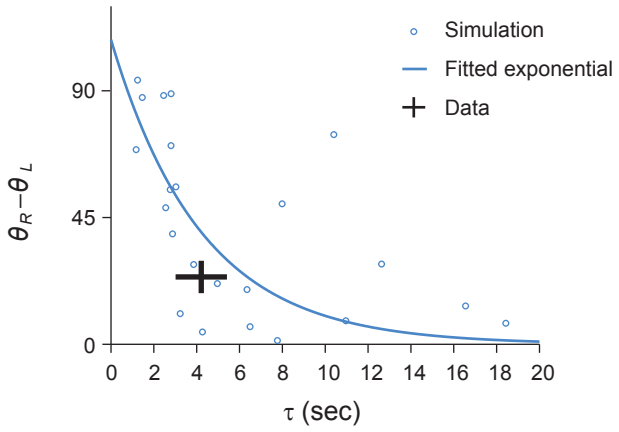
(b) Cumulative distribution of trial initiation times in different blocks (blocks are color-coded according to Fig. 1c or supplementary Fig. 1c) for a single rat (left) and the population (right, n=15 rats).

(c) Trial initiation time binned into short (< 5seconds), medium (between 5 and 10 seconds) and long (> 10 seconds) for high and low net value blocks (* denotes $P<0.01$, paired t-test; t_6=5.5, 5.1, and 4.4, respectively; n=7 sessions in 1 rat).

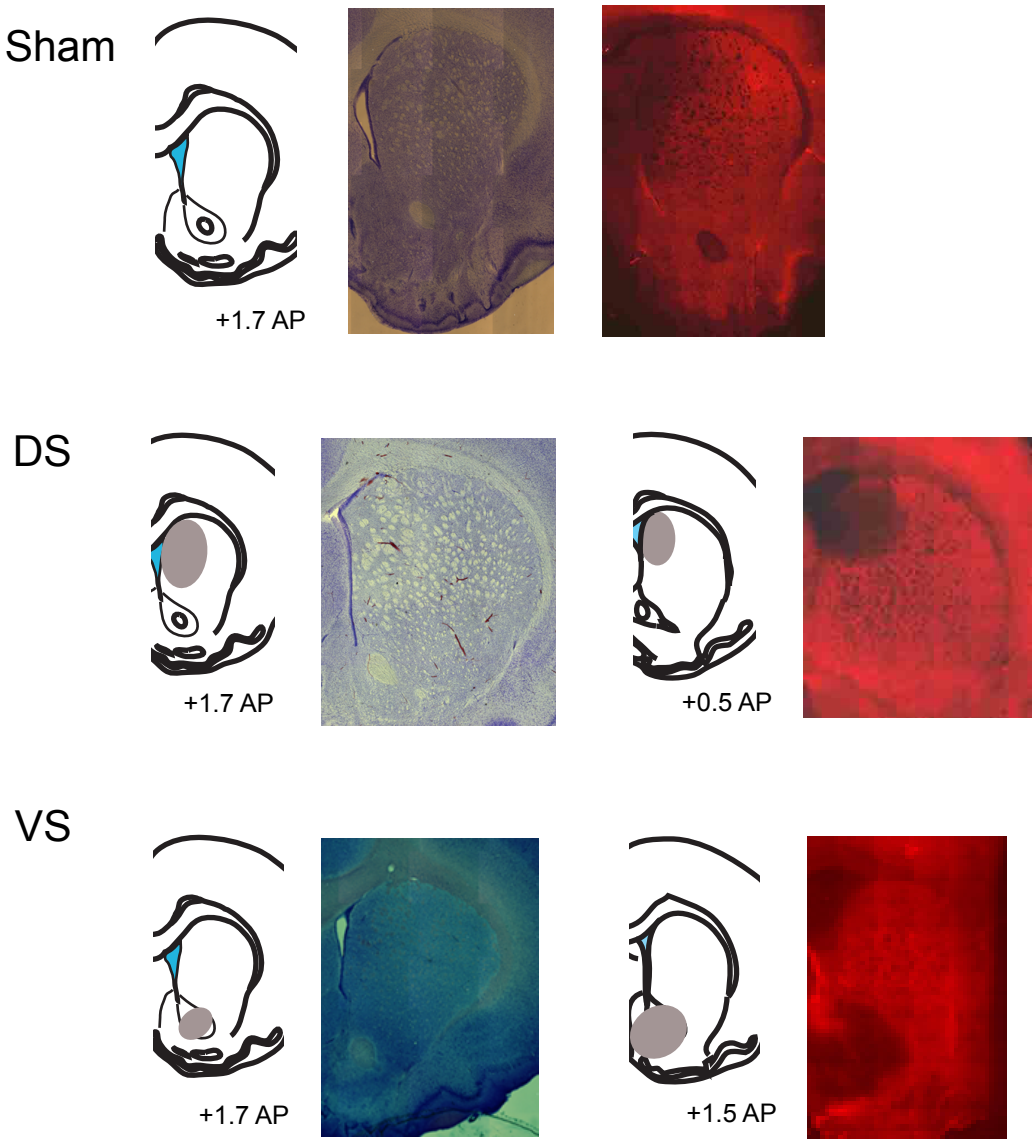
(d) Raw trace of trial initiation time across a behavioral session.



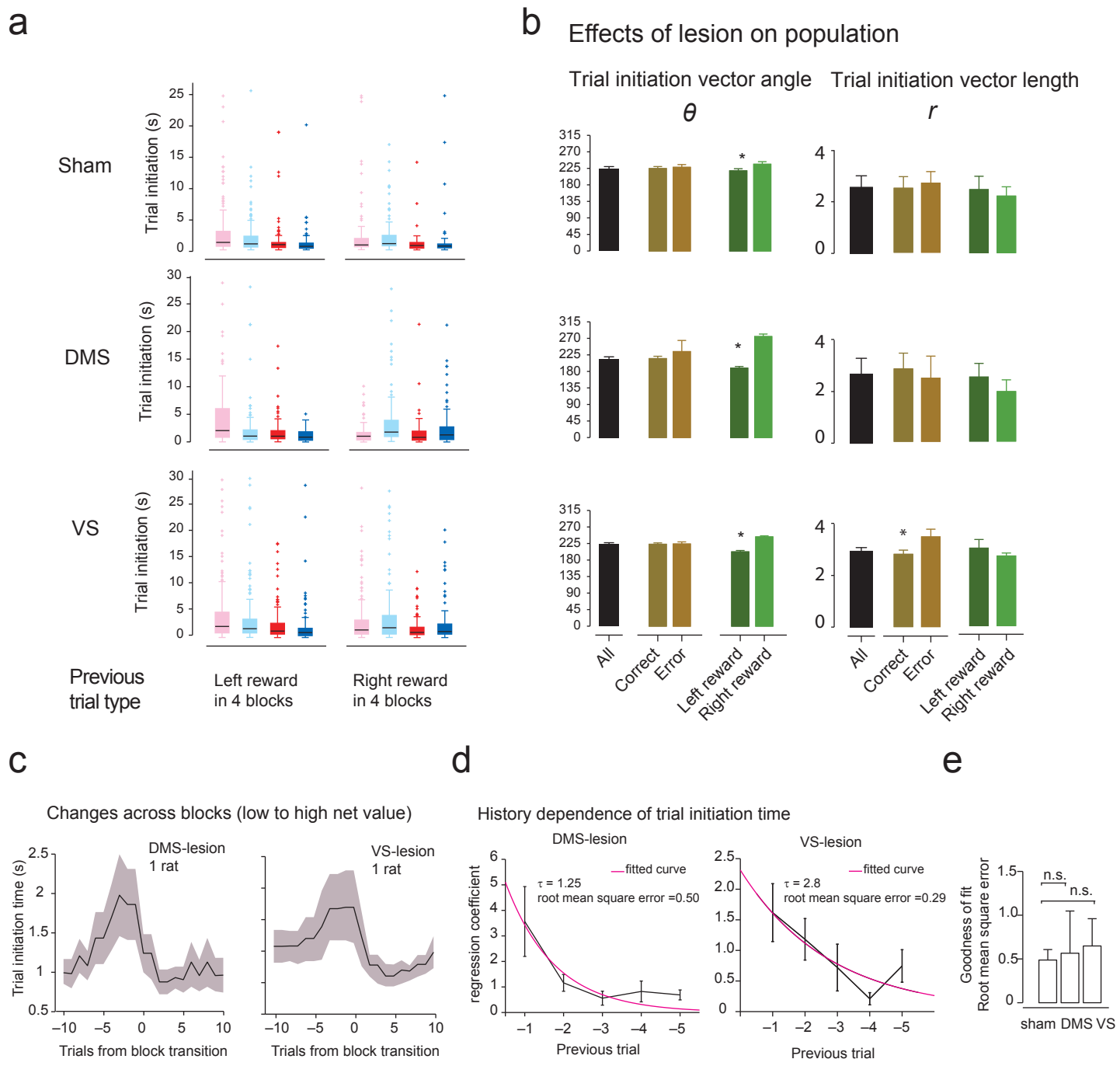
Supplementary Fig. 2| Movement duration across blocks
(a) Movement duration (time from odor port exit to reward port) in different blocks (blocks are color-coded according to **Fig. 1c**) for a single rat.
(b) Movement duration for the population (n=15 rats).



Supplementary Fig. 3| Simulation to determine the relationship between tau and $\theta_R - \theta_L$.
Black circles: $\theta_R - \theta_L$ obtained from simulation using a given time constant (τ). Blue: fitted exponential curve. Black cross: actual time constant(τ) and $\theta_R - \theta_L$ (mean ±s.e.m.).



Supplementary Fig. 4| Histological evidence of lesions
Top panels: Coronal depiction of approximate stereotaxic location of sham brain sections (left). Tissue is stained with cresyl violet (middle panel) and NeuN (rignt panel), a neuron-specific antibody that is visualized here using Alexa568 as the secondary antibody.
Middle panels: Coronal depiction of DMS lesions using cresyl violet staining (left two panels) and anti-NeuN labeling (right two panels).
Bottom panels: Coronal depiction of VS lesions using cresyl violet (left two panels) and anti-NeuN labeling (right two panels).



Supplementary Fig. 5| Effects of lesion on trial initiation time

(a) Boxplots of trial initiation time after left or after right reward trials in example sham, DMS, and VS-lesioned animals. The central mark indicates the median, and the edges of the box are 25th and 75th percentiles.

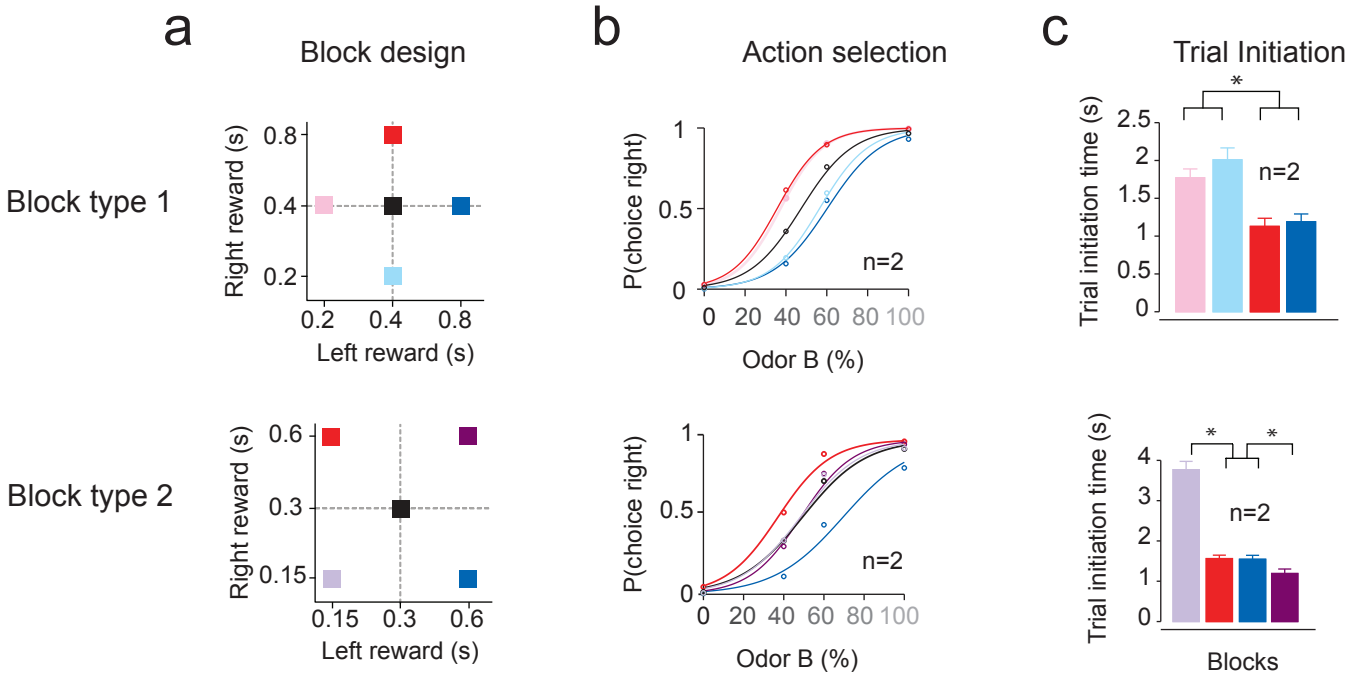
(b) Polar angles of vectors representing trial initiation after specific trial types (mean ±s.e.m., n = 5 rats for each group). **P*<0.05.

Amplitudes of vectors representing trial initiation after specific trial types (mean ±s.e.m., n = 5 rats for each group). **P*<0.05.

(c) Trial initiation time (mean±s.e.m.) across block changes from low net value to high net value, where 0 is the first trial after a block transition (left, n=1 DMS-lesioned; right, n=1 VS-lesioned).

(d) Effect of multiple previous trials on trial initiation. Black: regression coefficients (mean ± s.d., n=5 DMS -lesioned rats and n=5 VS-lesioned rats), Red: fitted exponential curve.

(e) Goodness of fit (root mean square error) of the exponential curves across conditions (n=5 rats per condition, mean ± s.d.).

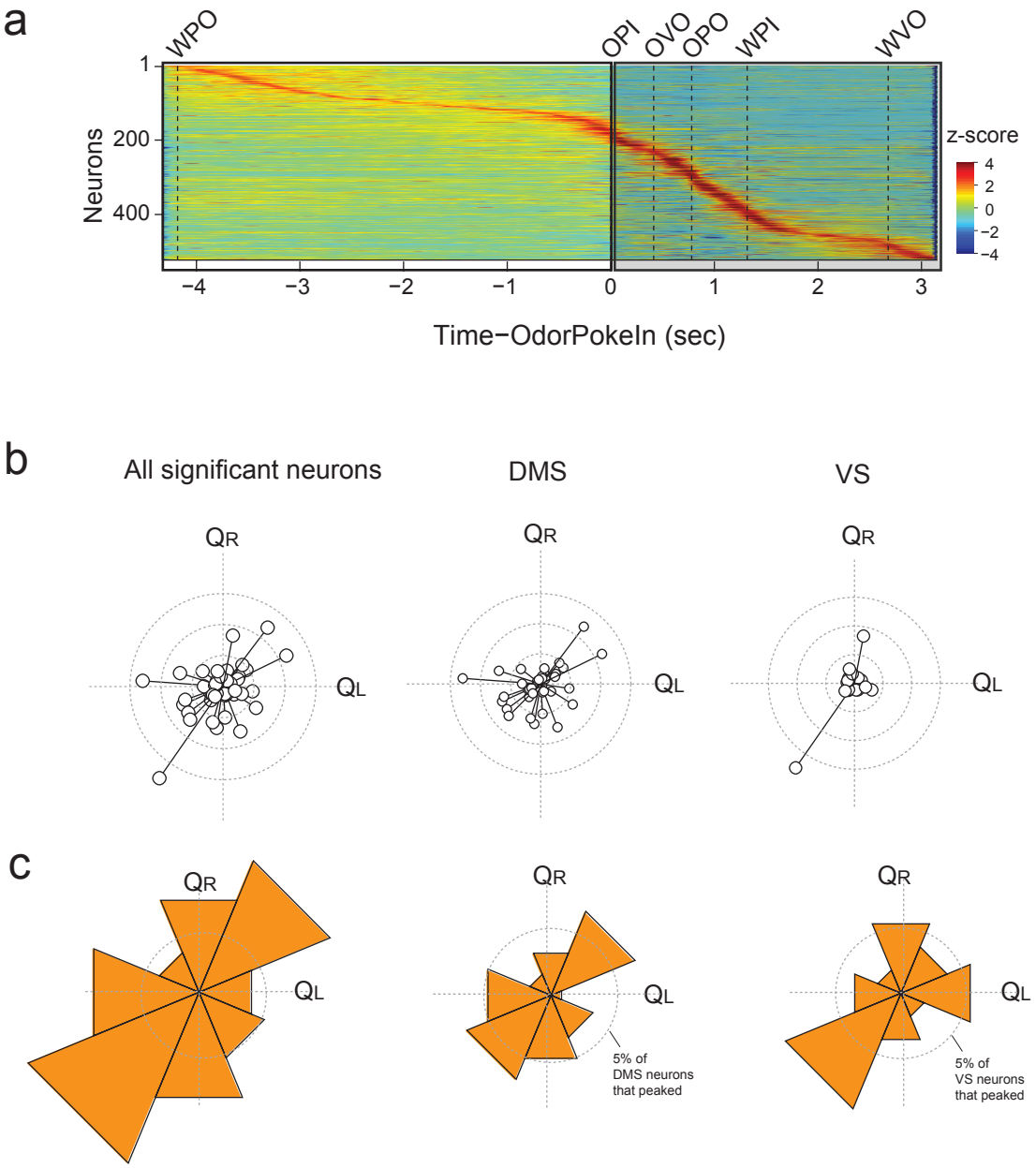


Supplementary Fig.6| Behavior during neuronal recording

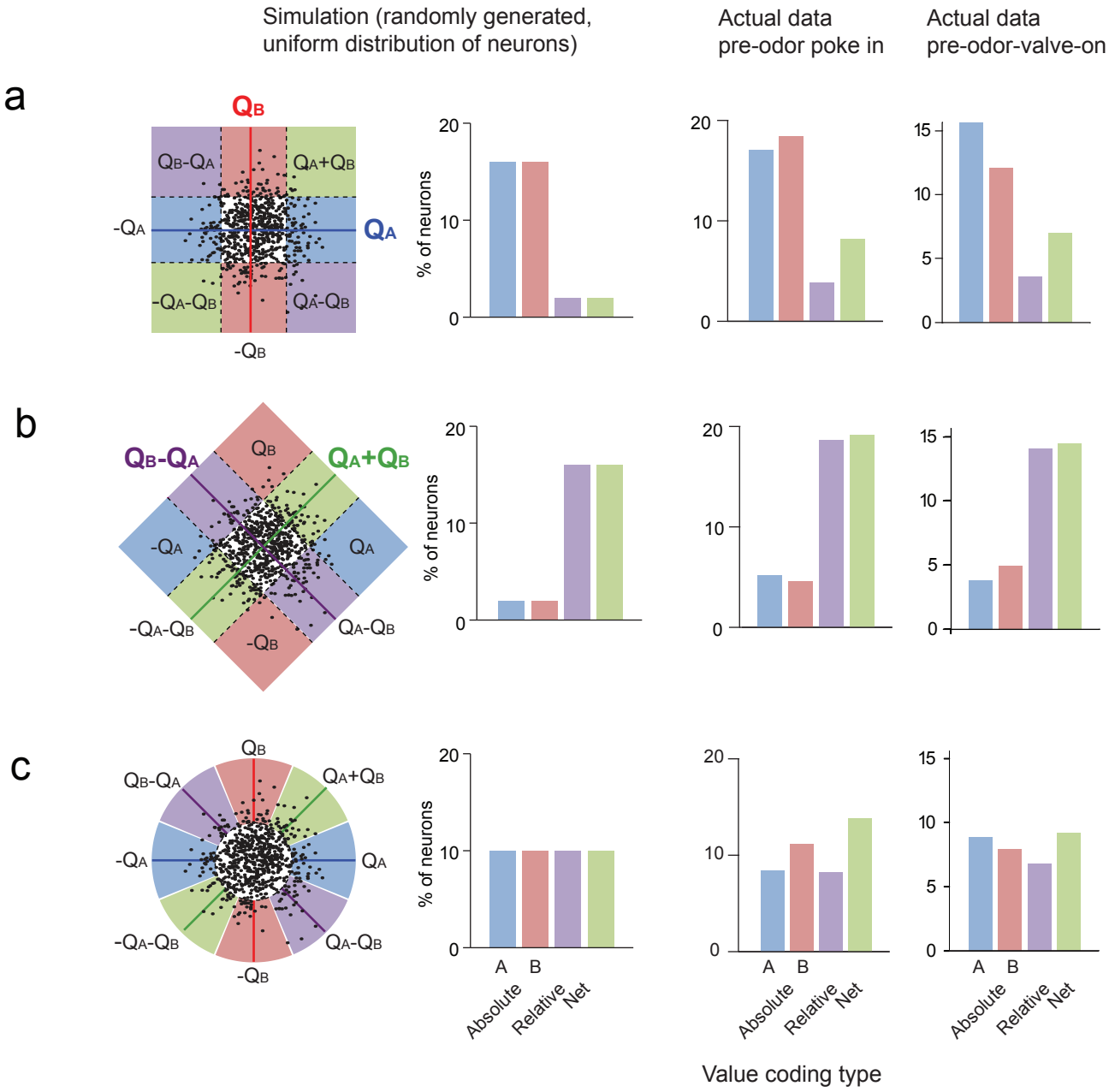
(a) Block-wise manipulation of left and right reward value.Two types of block manipulations (top and bottom) were used.

(b) Psychometric curves. Relative-values biased the choices in both task types (n = 2 rats each).

(c) Trial initiation time. Net value modulated trial initiation time in both task types (mean±s.e.m., n = 2 rats each, unpaired two-sided t-test, $P<0.05$*).



Supplementary Fig.7| Analysis using the peak time window for each neuron.
(a) Neurons that peak between WPO and OPI. Same as in Fig. 7a but only neurons that peak their firing between water-poke-out (WPO) and odor-poke-in (OPI) were used for the following analyses (n=181 neurons).
(b) Vectors of value coding neurons. Inner ring, r=16; outer ring, r=24.
(c) Proportion of neurons per value-coding category. Ring, 5% of neurons.



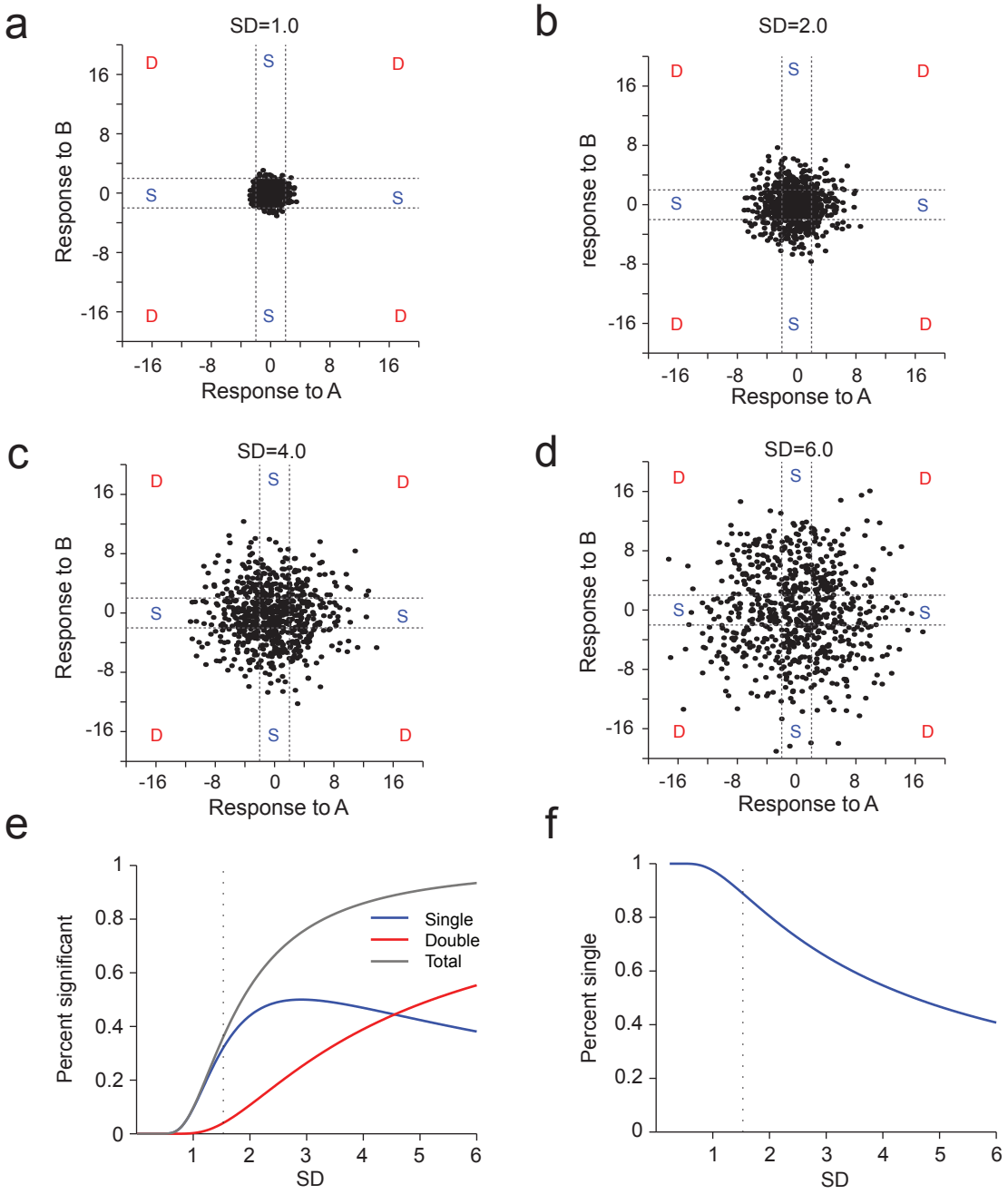
Supplementary Fig.8| Demonstration of classification biases with conventional regression methods using simulation and actual data

(a) Conventional method type 1: multiple regression with the independent variables, Q_A (X-axis, blue) and Q_B (Y-axis, red), representing values of A and B, respectively.

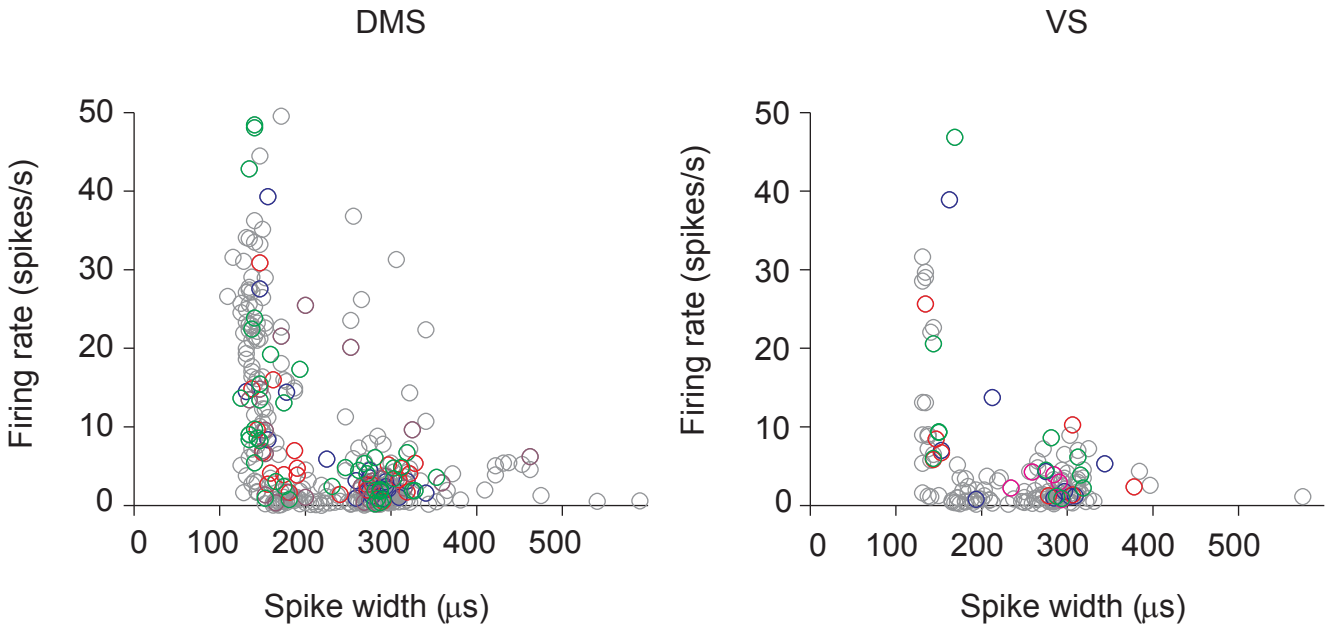
(b) Conventional method type 2: multiple regression with the independent variables, Q_B-Q_A (purple) and Q_A+Q_B (green) (tilted axes), representing relative and net values, respectively.

(c) Proposed method: polar representation to avoid bias in classification.

(a-c) Left panels: Black dots represent a randomly generated population of neurons that is uniformly distributed in two dimensional space (assuming 36% of neurons are significant). These simulated neurons are projected onto their respective regression planes, where dotted lines represent the critical boundary ($P = 0.05$) and critical regions are color-coded. Middle left panels: Distribution of simulated absolute, relative, and net value coding types. Middle right panels: Distribution of value coding types on actual data (time epoch, pre-odor-poke-in). Right panels: Distribution of value coding types on actual data at an alternative time window (time epoch, pre-odor-valve-on).



Supplementary Fig.9| Conventional methods are sensitive to the proportion of significant neurons
(a-d) Simulation of how uniformly distributed response populations at various standard deviations (SD) are classified with conventional methods. Each black dot represents a neuron. S, single positive (one independent variable is significant). D, double positive (two independent variables are significant). Dotted lines represent the critical boundary, *P*=0.05.
(e) Curves illustrates how the percent of neurons classified as single-positive (S) and double positive (D) varies as a function of the SD of the distribution.
(f) Curve illustrates the relative frequency of single-positive (S) neurons.
(e-f) Dotted vertical line represents when the criterion is 1.53 SD (equivalent to having ~36% of the population significant).



Supplementary Fig. 10| Value coding neurons and their anatomical cell types
Scatter plot of the spike width and firing rate of every recorded neuron (522 total). Neurons that are color-coded are value-coding and those in grey are non-value coding. Blue: left absolute value; red: right absolute value; purple: relative value; green: net value.Spike width is defined by the interval between the peak and trough of spike waveforms. There are at least two clusters: narrow spike waveform with relatively high firing rates (narrow spikes <200 us, 199 neurons) and wide spike waveforms with relatively low firing rates (wide spikes >200 us, 296 neurons). The latter likely corresponds to medium spiny neurons and the former, tonically active (TAN) neurons and/or fast spiking GABAergic interneurons (Kawaguchi, *J. Neurosci.*, 1993; Gage et al., *Neuron*, 2010). The proportion of value coding neurons among narrow-spiking and wide-spiking neurons were not statistically different (χ^2 test, $P = 0.13$; $\chi^2_1 = 2.3$; 25%, or 50/199 narrow-spiking neurons were value coding and 31% or 93/296 wide-spiking neurons were value coding). Among the narrow-spiking value coding neurons, there was a nonuniform distribution of value coding types, where net value coding neurons comprised the largest proportion (χ^2 , goodness of fit for a uniform distribution, $P = 0.0045$; 8/199, 16/199, 10/199, and 26/199 were left absolute, right absolute, relative, and net value coding, respectively). Among wide-spiking value coding neurons, net value coding types were also the most common, although the distribution of value coding types was not significantly different from uniform (χ^2 , goodness of fit for a uniform distribution, $P = 0.1772$; $\chi^2_3 = 4.9$; 22/296, 15/296, 17/296, and 28/296 fell into the respective categories indicated above). Finally, the proportions of absolute, relative, and net value coding neurons between narrow- and wide-spiking neurons were not statistically different from each other (χ^2 test, $P = 0.07$, 0.29, 0.55, and 0.37, for left absolute, right absolute, relative, and state coding neurons, respectively).

Kawaguchi, Y. Physiological, morphological, and histochemical characterization of three classes of interneurons in rat neostriatum. The Journal of Neuroscience13, 4908-4923 (1993)

Gage, G. J., Stoetzner, C. R., Wiltschko, A. B. & Berke, J. D. Selective Activation of Striatal Fast-Spiking Interneurons during Choice Execution. Neuron67, 466-479 (2010)