Hierarchical encoding of reward and effort across the cortex and basal ganglia during cost-benefit decision making

Oliver Härmson[1,2,7], Isaac Grennan[1,7], Brook Perry[1], Robert Toth[1], Colin G. McNamara[1,5], Hayriye Cagnan[1,6], Sanjay G. Manohar[2,3], Mark E. Walton[2,4,7], Andrew Sharott[1,7]

[1]Medical Research Council Brain Network Dynamics Unit, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom
[2]Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom
[3]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom
[4]Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, United Kingdom
[5]Current address: APC Microbiome Ireland, University College Cork, Cork, Ireland
[6]Current address: Department of Bioengineering, Imperial College London, London, United Kingdom.
[7]These authors contributed equally to this work

To whom correspondence should be addressed:
Oliver Härmson: oliver.harmson@bndu.ox.ac.uk
Andrew Sharott: andrew.sharott@bndu.ox.ac.uk

Abstract

Adaptive value-guided decision-making requires weighing up the costs and benefits of pursuing an available opportunity. Though neurons across frontal cortical-basal ganglia circuits have been repeatedly shown to represent decision-related parameters, it is unclear whether and how this information is co-ordinated. To address this question, we performed large-scale single unit recordings across medial/orbital frontal cortex and basal ganglia as rats decided whether varying reward payoffs outweighed the associated physical effort costs. We found that single neurons across the circuit could encode different combinations of the canonical decision variables (reward, effort and choice). Co-active cell assemblies - ensembles of neurons that repeatedly co-activated within short time windows (<25ms) within and across structures - were able to provide representations of the same decision variables through the synchronisation of individual neurons with different coding properties. Together, these findings demonstrate a hierarchical encoding structure for cost-benefit computations, where individual neurons with diverse encoding properties are coordinated into larger, lower dimensional spaces within and across brain regions to that can signal decision parameters on the millisecond timescale.

Introduction

Deciding whether to invest effort in the pursuit of potential rewards is a challenge that organisms continually face. Diverse information must be utilised to calculate and compare energetic cost of future physical and/or cognitive effort and the probability and magnitude of prospective reward. In line with the richness of these computations, a wide range of areas across frontal cortical-basal ganglia circuit have been implicated in shaping such cost-benefit decisions [1-9]. Single-units across this network code decision-relevant variables such as reward probability and effort cost [10-20]. In primates, individual neurons in frontal cortices can integrate reward- and effort-related information [21]. It is unclear whether such integration is also a property of basal ganglia neurons, which have more direct anatomical pathways to controlling action [22, 23]. Direct comparison of cortical and basal ganglia encoding will be necessary to fully define the level of specialisation of different populations of neurons.

A more complex question is how encoding at the level of single neurons is *coordinated* across such a vast network of brain regions. A common proposition is that single neuron responses reflect and contribute to modular processing across specialised brain areas that arrive at a behavioural output through serial information transmission [24-26], akin to that described for sensory pathways. In sensorimotor tasks, information from cortex is transferred reliably and topographically to connected areas of striatum [27]. However, this need not be the case for cost-benefit calculations, whereby information presented over different timescales needs to be remembered and then utilised at the point of choice and action. Such computations could utilise the ability of cortico-basal ganglia-thalamic loops to *maintain* representations of decision parameters [28]. Indeed, maintaining representations is likely to be necessary to support the level of parallel, distributed processing suggested to be necessary to facilitate adaptive decision-making [29-31]. However, it is unclear how such parallel processing frameworks manifest at the level of populations of single neurons across cortical and subcortical networks.

One mechanism to achieve such distributed processing is the formation and expression of co-firing assemblies, whereby groups of neurons repeatedly fire action potentials within short-latency (10-30ms) of each other [32, 33]. These assemblies have been postulated to play a vital role in large scale representations and co-firing assembly patterns in the hippocampus have been repeatedly demonstrated to encode spatial location [34-36]. In addition, coding by short-timescale coincident spiking can be effectively read-out by downstream neural populations [32,

37]. Despite the potential of co-firing assemblies to resolve some of the current issues related to mechanisms underlying cost-benefit decision-making, it is not clear whether they are formed in cortico-basal ganglia circuits and, if so, whether they can encode relevant variables. Moreover, it is unclear whether such assemblies might be in a position to integrate information across single neurons with variable coding properties.

To address these questions, we performed large-scale recordings of single neurons across the rat cortico-basal ganglia network while subjects evaluated whether to accept or reject cost-benefit offers. Individual neurons across the entire network encoded variable combinations of canonical decision variables (reward, effort and choice), with a particularly pronounced, phasic and multiplexed signal in the dorsal medial striatum and subthalamic nucleus area. Co-firing assemblies were detected within and across all cortico-basal ganglia sites and could encode the three main decision parameters independently from the individual firing rates of the member neurons. Remarkably, these often included individual neurons whose firing rates signalled a parameter *different* to the tuning of the overall assembly, or neurons with no discernible rate-coding of the canonical decision variables at all. Co-firing assemblies thus coordinate individual neurons with wide-ranging coding properties and give rise to emergent representations that can mediate cost-benefit computations at the network level.

Results

Choice on a novel rat accept/reject cost-benefit choice paradigm is shaped by anticipated costs and benefits.

To elicit a trade-off between rewards and associated effort costs, we designed a novel cost-benefit task, inspired by foraging scenarios. Rats chose whether to pursue or forgo a reward available at the other end of a linear corridor based on its size and the set level of effort (Fig. 1a, online methods). Trials were initiated by poking and holding their noses in an initiation port ('pre-cue') while an auditory cue was presented ('offer'), which signalled the size of the available reward on that trial (4 levels: nil, small, medium, large; Fig. 1b). Once the port light illuminated ('Go cue'), rats could then leave the port ('action') and decide to either (i) pursue the offered reward by running to the opposite end of the corridor ('accept') or (ii) forgo the offered reward by refraining from entering the corridor ('reject'). In both cases, trials were followed by an inter trial interval (ITI), after which a new trial could be initiated. The effort cost of pursuing the reward was manipulated by placing different numbers of barriers in the corridor, fixed across blocks of ~20 trials (3 levels: low, moderate, high; Fig 1c).

Rats' choices were sensitive to both the benefits and costs of acting: animals were more likely to pursue reward when it was larger and when the effort cost was lower (n=12, non-implanted animals, Fig. 1c-e). This was reflected as a robust positive influence of reward ($\beta=1.15 \pm 0.10$, $p<.001$, binomial mixed effects' model) and a negative influence of effort ($\beta=-0.33 \pm 0.02$, $p<.001$) over choice (Fig. 1e, online methods). Despite the lack of an explicit cue to signal the effort level, it nonetheless had a negative influence on decisions throughout trial blocks (Fig. S1a). Rats also progressively accepted fewer offers over the course of a session as they gained more rewards ($\beta=-0.47 \pm 0.02$, $p<.001$), suggesting sensitivity to outcome value (Fig. 1e). This was corroborated by a follow-up experiment where rats (n=6) were given free access to lab chow for 1h prior to testing, which resulted in a significant reduction in willingness to accept any offer (Fig. S1b-c), indicating that choices on this task are not reflective of a fixed decision-making policy.

Sensitivity to reward and effort remained significant in a subset (n=4) of rats who were implanted for electrophysiological recordings in an adapted version of this paradigm (Fig. S1d-e, online methods). Reward offers also consistently influenced action initiation latency on successful trials (Fig. S1f) and premature nose-poke exit rates during the offer cue (Fig. S1j)

in both sets of animals. Rats' decisions on this paradigm are thus robustly sensitive to the offered reward magnitude, effort cost and outcome value across subtle variations of task parameters.

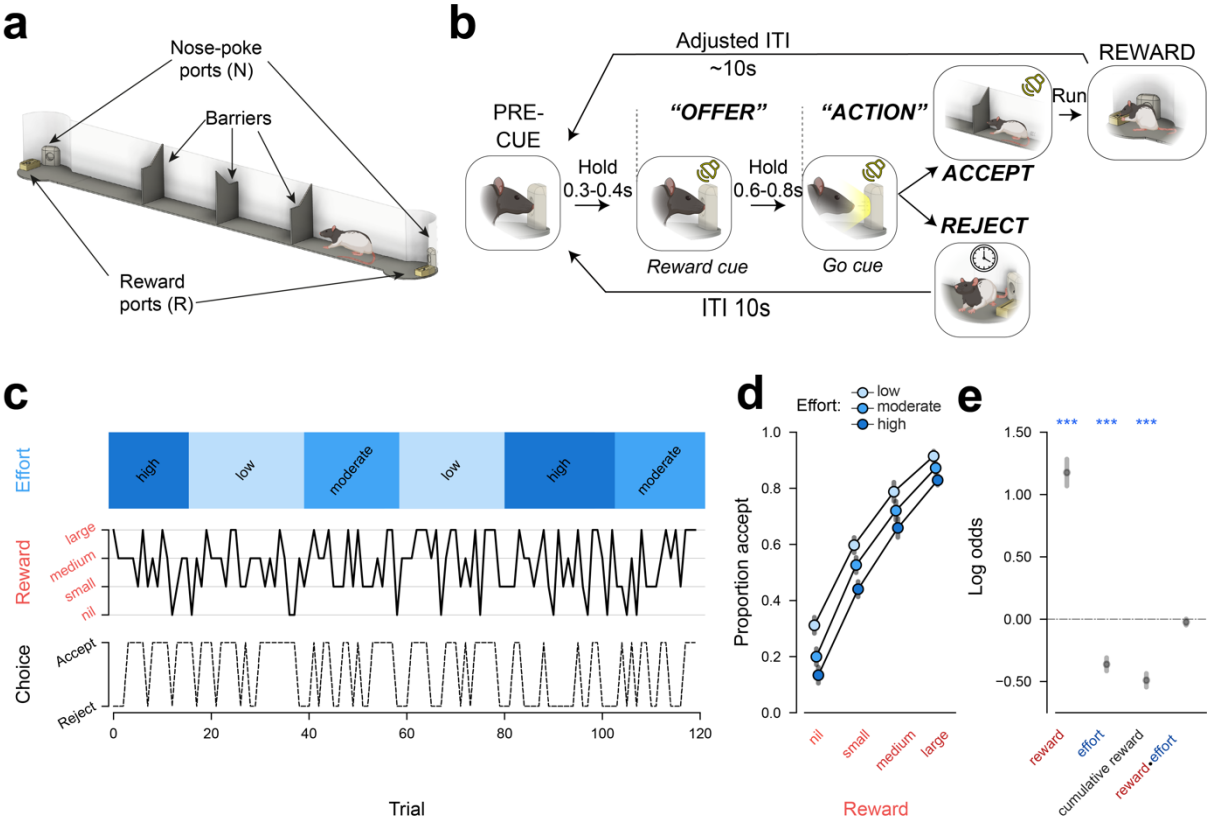
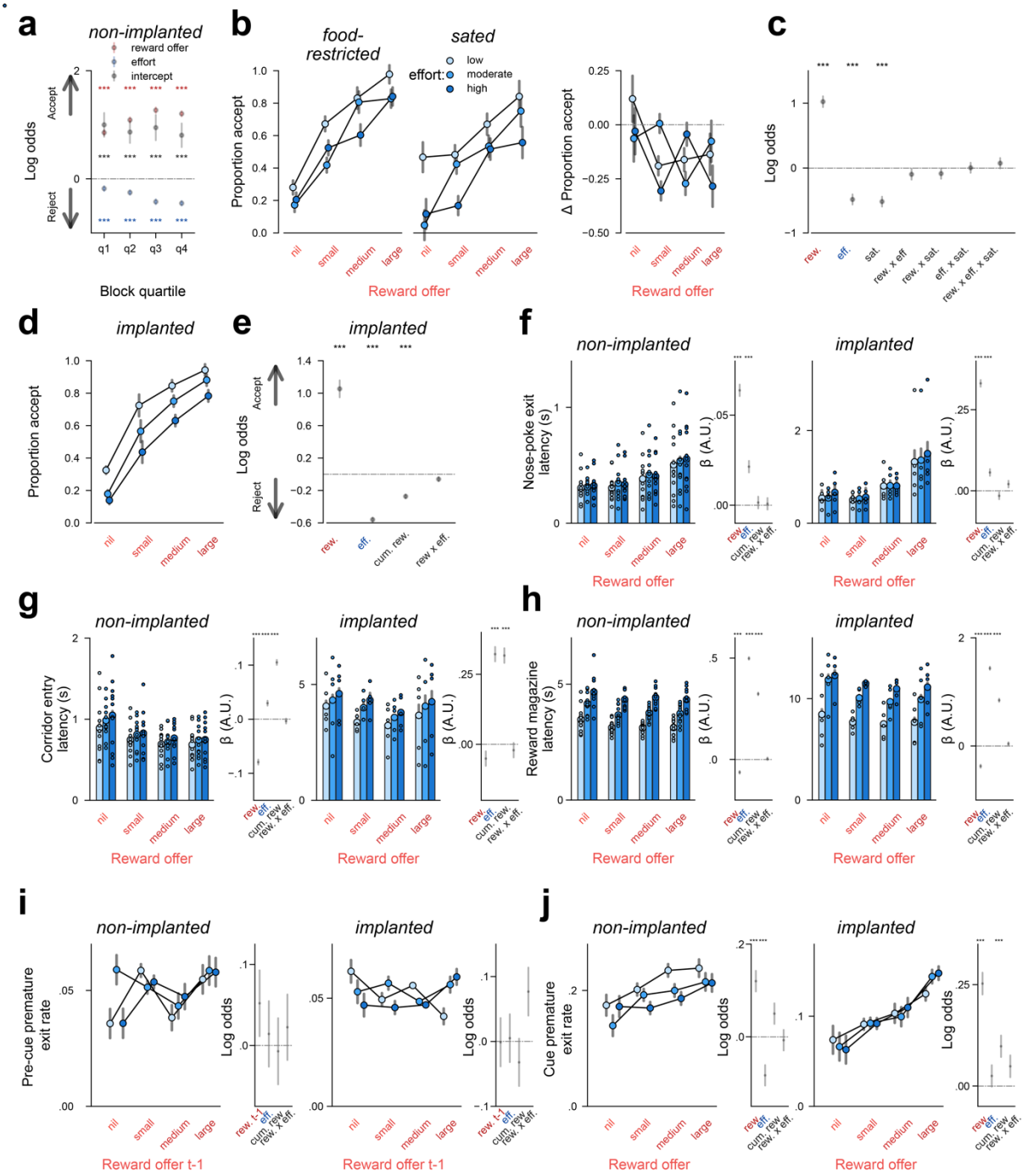


Figure 1. Choice on a novel rat accept/reject choice paradigm is shaped by the costs and benefits. a, Illustration of the experimental apparatus which consisted of a 130cm corridor with a variable number of barriers with circular arenas at either end, each of which contained a nose-poke port and a reward port. **b,** Rats chose between reward available at the far end of a linear corridor ('accept') vs opting out ('reject'). To ensure average reward rate was minimally affected by rats' accept/reject choices, ITIs were adjusted on 'accept' trials by subtracting the cohort median run time (online methods). Trials could always be initiated from the same arena where the last trial had finished (same arena on rejected trials; opposite side after an accepted trial). **c,** Reward/effort variation and choice in an example session. Reward magnitude (4 levels, middle panel) is signalled trial-by-trial by an auditory cue, whereas effort (3 levels, upper panel), is expressed in blocks as the number of barriers in the corridor. Choice is depicted as a dotted line on the bottom panel. **d**, Rates of accepting offers as a function of reward (x-axis) and effort (n=12). Coloured dots indicate mean ± SEM, adjusted for the within-subject variance across all levels of effort. **e,** weightings of task variables on behavioural choice from a binomial mixed effects model, depicted as mean ± SEM. Effect of reward x effort: β=-0.01 ± 0.02, p = .69. *** p < .001.



Supplemental Figure. 1 Choice on a rat accept/reject choice paradigm is shaped by the costs and benefits associated with the offer. a, weightings of reward and effort within quartiles of each effort block. Data are depicted as regression coefficients for reward (red) and effort (blue) ± SEM. ***, $p < .01$, corrected for family-wise error rate. **b,** Psychometric curves depicting rates of accepting offers as a function of reward (x-axis) and effort (black lines dotted with coloured scatter) from our within-subject outcome devaluation experiment (n=6). Data are depicted separately for rats' food-restricted and sated state on the left and right, respectively, with the difference (delta) scores between the sated and food-restricted proportions depicted on the right. **c,** weightings of experimental variables on behavioural choice from a binomial mixed effects' model fitted on data from the outcome devaluation experiment (n=6). **d,** Implanted rats' (n=4, sub-selected from the initial cohort of 12) rates of accepting offers as a function of reward and effort. **e,** weightings of task variables on implanted rats' choice from a binomial mixed effects' model. **f-h,** mean nose-poke exit (f), corridor entry (g), and reward magazine (h) latencies (left panels) and associated coefficient weightings from general linear mixed effects' models (right

panels, online methods). Nose-poke exit latencies were defined as the interval between cue onset and nose-poke exit, while corridor and reward magazine entry latencies were calculated as the time duration between nose-poke exit and closest corridor light-gate or reward magazine entry, respectively. Bars and larger scatters represent mean ± within-subject SEM, while small scatter dots represent per-subject median values. Colour-code is expressed as per panel **b**. **i-j,** mean pre-cue and cue premature nose-poke exit rate, respectively, and associated coefficient weightings from a binomial hierarchical model (online methods). Generalised linear mixed effects' models in i and j were fitted without nil reward trials. ***, $p < .01$, corrected for family-wise error rate using the Benjamini-Hochberg (BH) method.

Cost-benefit decision variable representations are asymmetrically distributed across the frontal cortical-basal ganglia network

Having established that choices on our task were concurrently sensitive both reward and effort, we first aimed to identify neural correlates of the underlying computations at the level of single neurons. Given the role of the frontal-basal ganglia network in motivating cost-benefit choice [4, 9, 17, 38], we targeted the medial/ventral orbitofrontal cortex (MO/VO), cingulate gyrus 1/2 (anterior cingulate cortex, ACC), dorsal medial striatum (DMS), ventral pallidum (VP) and subthalamic area (STA, encompassing the subthalamic nucleus, zona incerta, substantia nigra pars reticulata and compacta) (Fig 2a-d, Fig. S2a-e) using a bespoke driveable multielectrode implant to record individual neurons from well-trained rats (n=4) performing the cost-benefit task (online methods).

We first examined the propensity of cortical and/or basal ganglia neurons to encode single or multiple decision variables. We modelled each identified neuron's z-scored firing rate as the weighted sum of offered reward, effort and subsequent accept/reject choice (Fig. 2e, online methods) during epochs containing the *entire decision process* (from the offer cue onset to either corridor entry for accepted trials or the session median corridor entry for rejected trials). Significant proportions of units in each area linearly signalled each canonical decision variable (Fig. 2f, all $p < .001$, one-sided binomial test against a chance rate of 5% with Benjamini-Hochberg adjustment for multiple tests). Strikingly, however, reward and choice tuning was more frequent in the DMS and STA compared to other structures, (Fig. 2f, see legend for statistical comparisons). Moreover, the co-representation of decision parameters within single neurons was even more common in DMS and STA compared to the targeted frontal cortex (Fig. 2f, see legend for statistical comparisons). Together, these data suggest that single neuronal representations are highly distributed, but that there are also significant regional asymmetries, with particular specialisation in the DMS and STA.

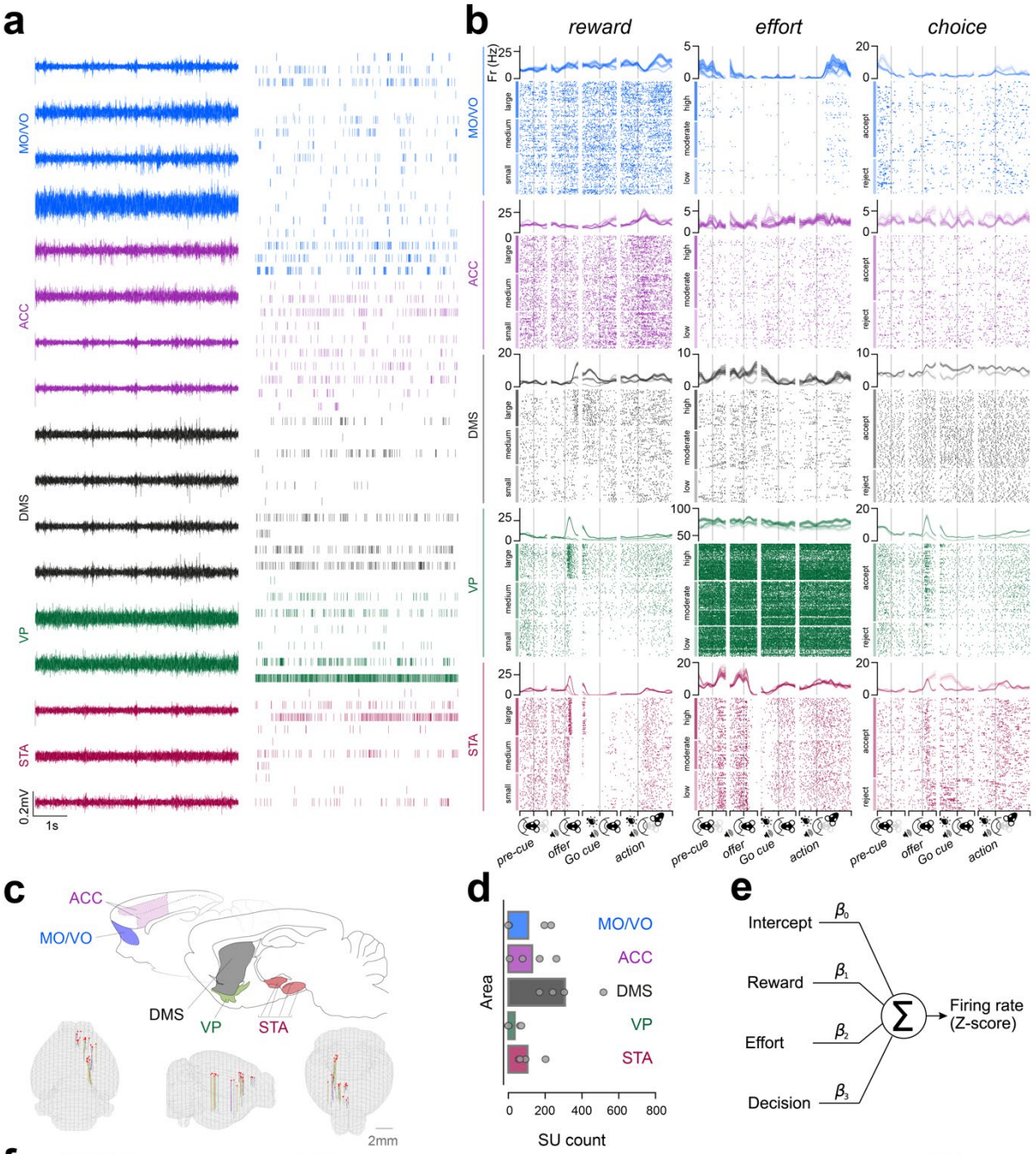
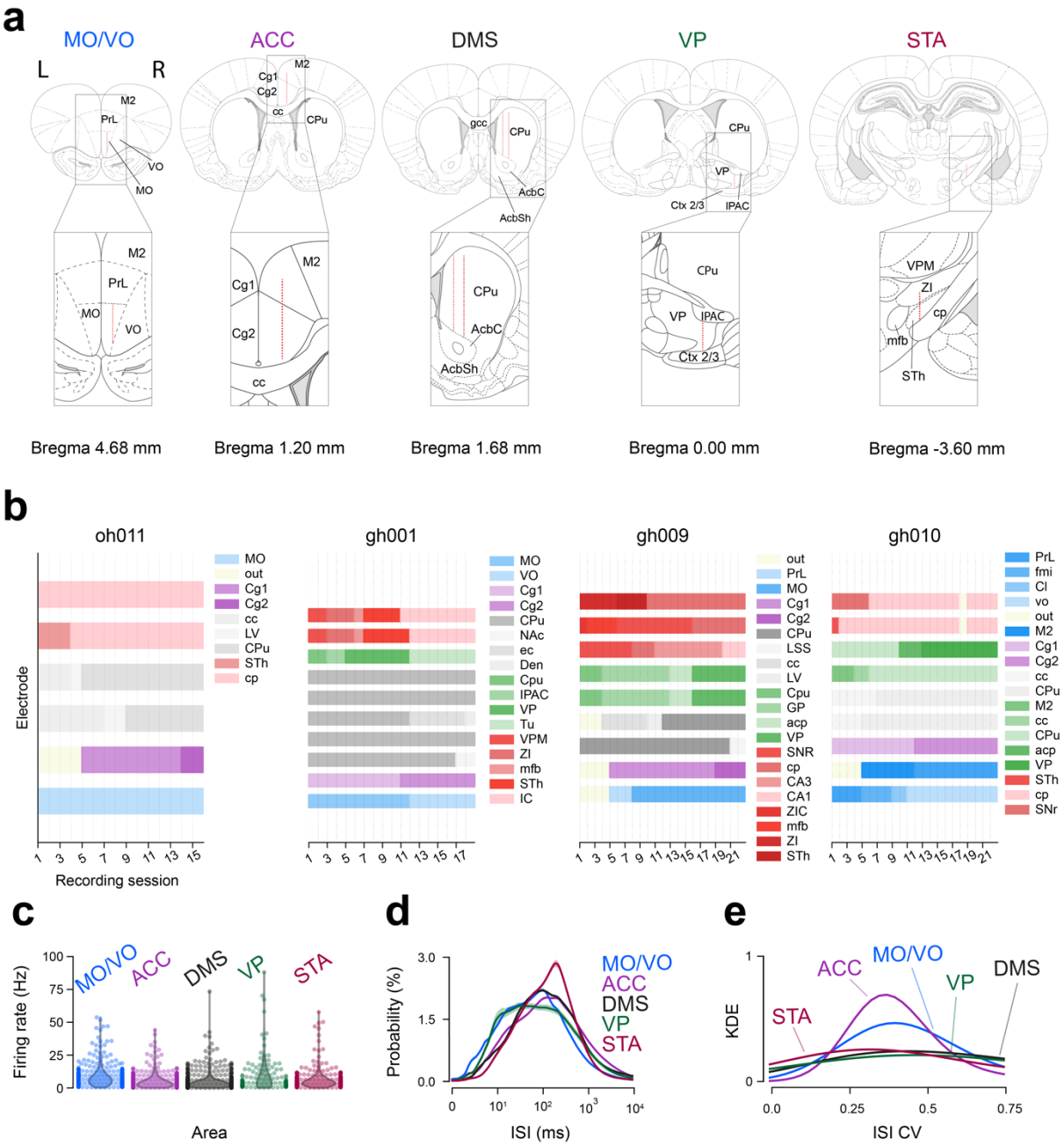

a
MO/VO
ACC
DMS
VP
STA
0.2mV
1s
b
reward
effort
choice
Fr (Hz)
large
medium
small
high
moderate
low
accept
reject
pre-cue
offer
Go cue
action
c
ACC
MO/VO
DMS
VP
STA
2mm
d
Area
MO/VO
ACC
DMS
VP
STA
0
200
400
600
800
SU count
e
Intercept
Reward
Effort
Decision
Firing rate
(Z-score)
f
MO/VO
(n=430)
26%
23%
12%
5%
12%
2%
7%
4%
15%
28%
ACC
(n=508)
21%
19%
13%
3%
12%
1%
4%
3%
12%
20%
DMS
(n=1196)
52%
26%
22%
10%
6%
5%
14%
5%
11%
35%
Reward
Effort
Choice
VP
(n=134)
28%
29%
13%
8%
17%
1%
5%
2%
9%
18%
STA
(n=410)
37%
24%
18%
6%
11%
2%
11%
4%
12%
29%

Figure 2. Single-neurons encode combinations of reward, effort, and choice across the frontal cortex and basal ganglia. a, High-pass filtered signal (left) and individual neuron spiking times (right, online methods) displayed from an example five-site simultaneous recording from one example subject. **b,** peri-event time histograms (PSTHs, top panels) and raster plots (bottom panels) of 15 individual example neurons (one per panel) that monotonically and significantly varied their firing rate as a function of levels of reward, effort and the rats' eventual choice (online methods). Data are time-locked to the beginning of the pre-cue, offer, Go cue and action epochs on the x axes, signified by vertical grey bars. PSTH data are displayed as averages across all other task conditions ± SEM (shaded areas). Data are displayed individually for each trial and grouped from top to bottom by levels of a decision variable, indicated on the left by coloured vertical bars. **c,** top: representations of targeted areas on sagittal sections adapted from [39]; bottom: overview of trode locations on 3D meshes registered to the Allen Common Coordinate Framework 3D space using SharpTrack [40], viewed from the back (left), right side (middle) and front (right). **d,** Overview of the average and per-subject single unit (SU) yield in each brain area. Means across subjects are depicted as bars and subject-by-subject cellular yields are marked as grey dots. We isolated on average 108 ± 107 (MO/VO, mean ± standard deviation), 129 ± 96 (ACC), 308 ± 130 (DMS) 34 ± 34 (VP) and 104 ± 58 (STA) single units per animal. **e,** Schematic of the multiple linear regression model used to analyse the contributions of decision variables on individual neuronal firing rates. **f,** Proportions of neurons that were tuned to canonical decision variables across a wide time window in each region (online methods), detected on the basis of significant coefficient of partial determination (CPD) values ($p < .050$). Pairwise comparisons for reward-coding neuron proportions: DMS vs all other structures: all $|z| > 5.13$, $p < .001$, two-sided z-test of proportions, adjusted for multiple comparisons using the BH procedure. STA vs all structures except VP: all $|z| > 3.37$, $p < .001$; STA vs VP: $|z| = 1.74$, $p < .050$. All other $|z| < 1.80$, $p > .050$. Effort-coding neuron proportion comparisons: ACC vs DMS $|z| = 3.22$, $p = .001$. ACC vs VP: $|z| = 2.69$, $p = .007$; all other $|z| < 1.91$, $p > 050$. Pairwise comparisons for choice-coding proportions: DMS vs other regions: all $|z| > 2.47$, $p < .013$; STA vs ACC and VP: both $|z| > 2.48$, $p < .012$. MO/VO vs ACC and VP: both $|z| > 2.31$, $p < .020$. All other $|z| < 0.61$, $p > .050$. Pairwise comparisons of the summed percentages of all multiplexing neuron types (intersecting Venn diagram areas) between regions: DMS vs all other regions: all $|z| > 3.91$, $p < .001$; STA vs MO/VO and ACC: both $|z| > 2.06$, $p < .020$; all other $|z| < 1.46$, $p > .072$, one-sided z-test of proportions without BH correction. MO/VO = medial/vental orbitofrontal cortex, ACC = anterior cingulate cortex, DMS = dorsal medial striatum, VP = ventral pallidum, STA = subthalamic area. Note that the for a given decision variable in DMS and STA, the proportion of neurons encoding that variable in additon to others is bigger that the proportion of neurons encoding that variable.



Supplemental Figure 2. Anatomic localisation and basic firing properties of single units recorded across five frontal-basal ganglia regions. a, planned stereotrode tracts. Stereotrode trajectories are depicted as dashed red lines superimposed on adapted atlas schematics on whole coronal sections (above) and zoomed insets (below; adapted from Paxinos and Watson 2006). **b,** Estimated targeting data are displayed from each identified electrode tract (bars) within each animal (panels) across all acquired recording sessions (x-axis). Electrode tract targets are colour-coded by the targeted area (blue - MO/VO; purple - ACC; grey - DMS; green - VP; red - STA). **c,** average firing rates of each identified neuron (dots) and kernel density estimates (KDEs) of the neuronal population (violin bodies) coloured by the targeted area. **d,** histogram of average inter-spike interval (ISI) values of neurons in each region. Values are depicted as the population mean probabilities (dark lines) ± SEM (shaded areas). Histograms values are smoothed with a gaussian kernel with a standard deviation of 1.0. **e,** KDEs of neuronal coefficients of ISI variation (lines, colour-coded by area). AcbC - nucleus accumbens core; AcbSh - nucleus accumbens shell; cp - cerebral peduncle; CA1/3 - field CA1/3 of the hippocampus; Cg1/Cg2 - cingulate gyrus 1/2, cc - corpus callosum; CPu - caudate/putamen; cp - cerebral peduncle; ec - external capsule; Den - dorsal endopiriform nucleus; fmi - forceps minor of the corpus callosum; M2 - secondary motor cortex; IPAC - internal nucleus of the

posterior limb of the anterior commissure; LSS - lateral stripe of the striatum; VPM - ventral posteromedial thalamic nucleus; ZI - zona incerta; mfb - median forebrain bundle; out - outside the brain; PrL - prelimbic cortex; STh - subthalamic nucleus; Tu - olfactory tubercle.

The dynamics of cost-benefit representations in cortex and basal ganglia are underpinned by diversity in temporal and directional properties.

Our finding that basal ganglia neurons were more specialised that those in the frontal cortex contrasts with previous studies suggesting that abstract cognitive variables (prospective reward value and effort cost) are preferentially signalled in higher-order regions and actions resolved in lower-order structures (Cisek 2012). However, such organisation could be more evident over finer timescales. We thus analysed the coefficients of partial determination (CPD) with finer temporal resolution for each canonical decision variable, neuron, and area (Fig. 3a, online methods). To characterise the temporal properties of the CPD signal, we further analysed (1) CPD-phasicness - an index of the dispersion of encoding in each area's single neuron population over time (online methods) - and (2) latency to maximal CPD.

Prospective reward was significantly encoded from the onset of the reward offer cue until the decision point *in all areas* (Fig. 3a, upper left). Strikingly, the reward signal in ACC, DMS and STA units was transient, while the VP and frontal cortical encoding was more dispersed (ACC, DMS and STA phasicness tests: all $|z| > 2.18$, $p < .029$, two-sided Mann-Whitney U test; Fig. 3a, upper left, inset). Strikingly, peak reward CPD was reached earlier in the STA compared to all other areas (Fig. 3a, bottom left; STA maximum CPD latency vs other areas: $|z| > 3.55$, $p < .001$). This was accompanied by commensurate recruitment of units to signal reward (Fig. 3b, left). Whereas the time course of reward coding was clearly linked to the onset of the offer cue, effort was signalled tonically (all CPD phasicness tests, $|z| < 1.34$, $p > .050$; Fig. 3a, upper middle) with similar density (Fig. 3a upper middle, inset; all $|z| < 1.08$, $p > .050$) and peak CPD latency (Fig. 3b lower middle; all $|z| < 1.60$, $p > .050$) in all structures. Choice, on the other hand, elicited phasic firing rate modulation in the MO/VO, ACC, and DMS (Fig. 3a, upper right; MO/VO, ACC and DMS CPD phasicness tests: $|z| > 3.47$, $p < .001$; all other $|z| < 0.93$, $p > .050$) with similar CPD phasicness across regions (all between-area $|z| < 1.72$; $p > .050$) and similar peak CPD latencies (Fig. 3b, bottom right; all between-area $|z| < 1.72$, $p > .050$) weighted towards the Go cue/action epochs.

We speculated that these phasic vs. distributed signals could be underpinned by groups of neurons that differ in their coding valence with respect to the decision parameters (Roesch and Olson 2004;

Kennerley et al. 2009). Visual inspection revealed that distributed reward coding in all areas was driven predominantly by negatively coding neurons, while phasic coding in the DMS and STA was mediated mostly by positively tuned single units (Fig. 3c, Fig. S3-4). In line with this, maximum reward cpd latencies were significantly higher in negative DMS and STA neurons compared to positive ones (comparison of population-median maximum cpd latencies: both $|z| > 5.31$, $p < .001$, Mann-Whitney U test). Such temporally disperse coding can be underpinned by either (1) persistent temporally disperse activity within single units or (2) tiling of task time by phasic signals of individual neurons (Guo et al. 2017; Runyan et al. 2017). Examination of the phasicness of individual positively vs. negatively reward-coding units revealed that the latter were significantly less phasic than the former in the DMS and STA (both $|z| > 3.41$, $p < .001$, Mann-Whitney U test), while phasicness did not differ between these populations in other regions or for other task variables (all other $|z| < 1.64$, $p > .050$).

Taken together, these data demonstrate that the temporal dynamics of rate-based encoding varied with respect to brain area, direction and variable. Notably, positively encoding neurons in basal ganglia structures had particularly strong phasic encoding of prospective reward during the offer window, but the peak encoding of the majority of populations was spread across the entire task.

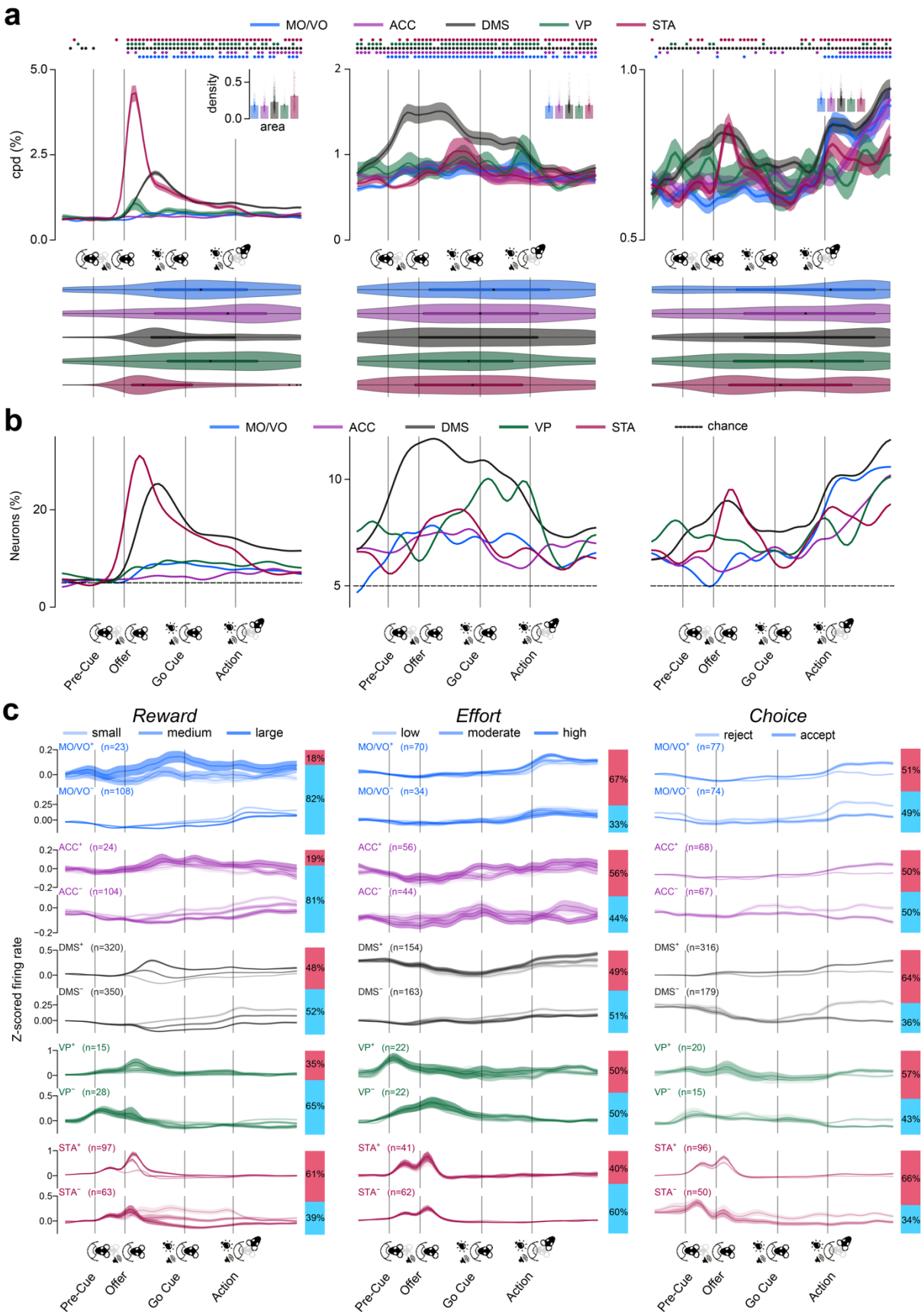
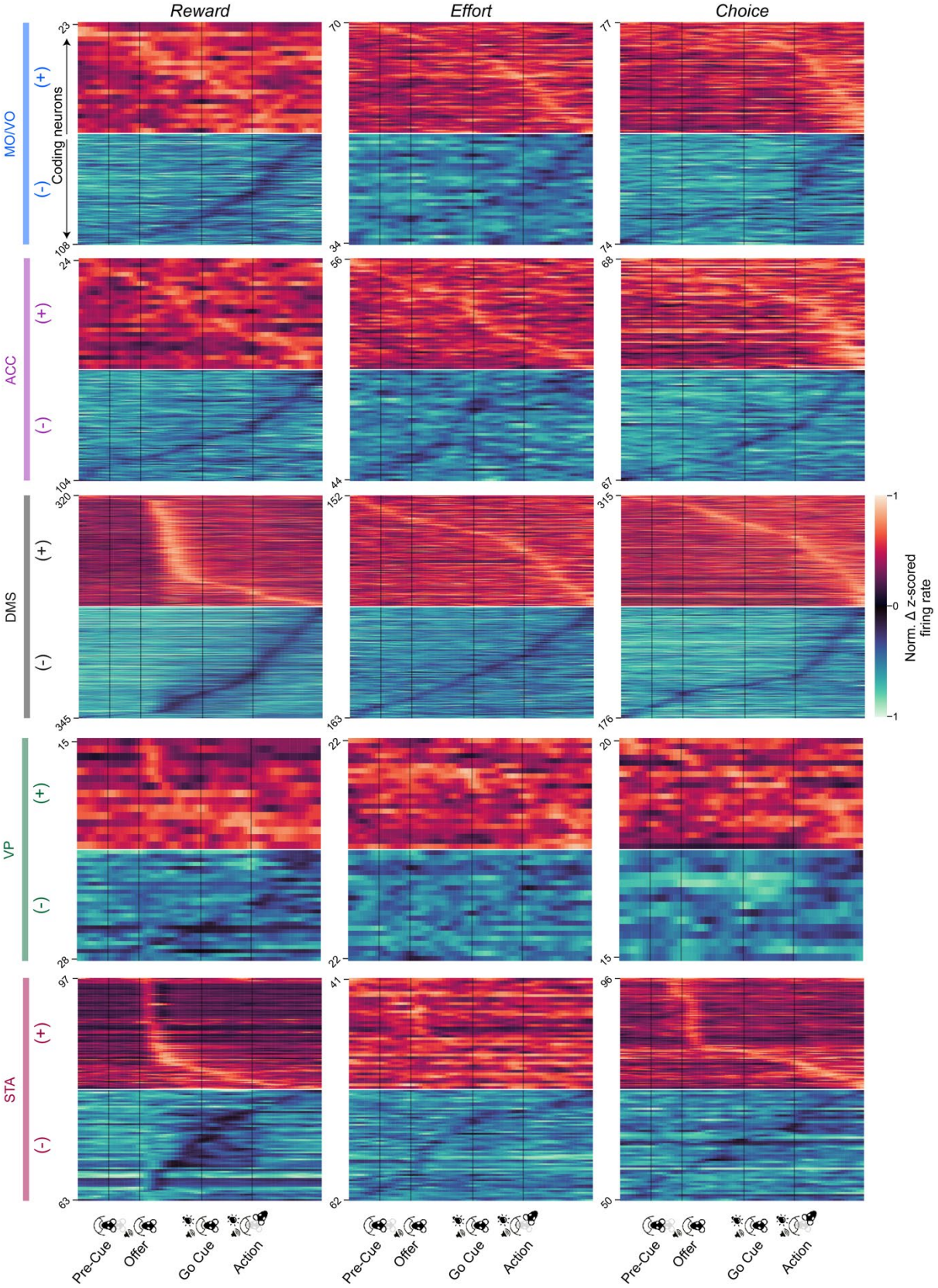


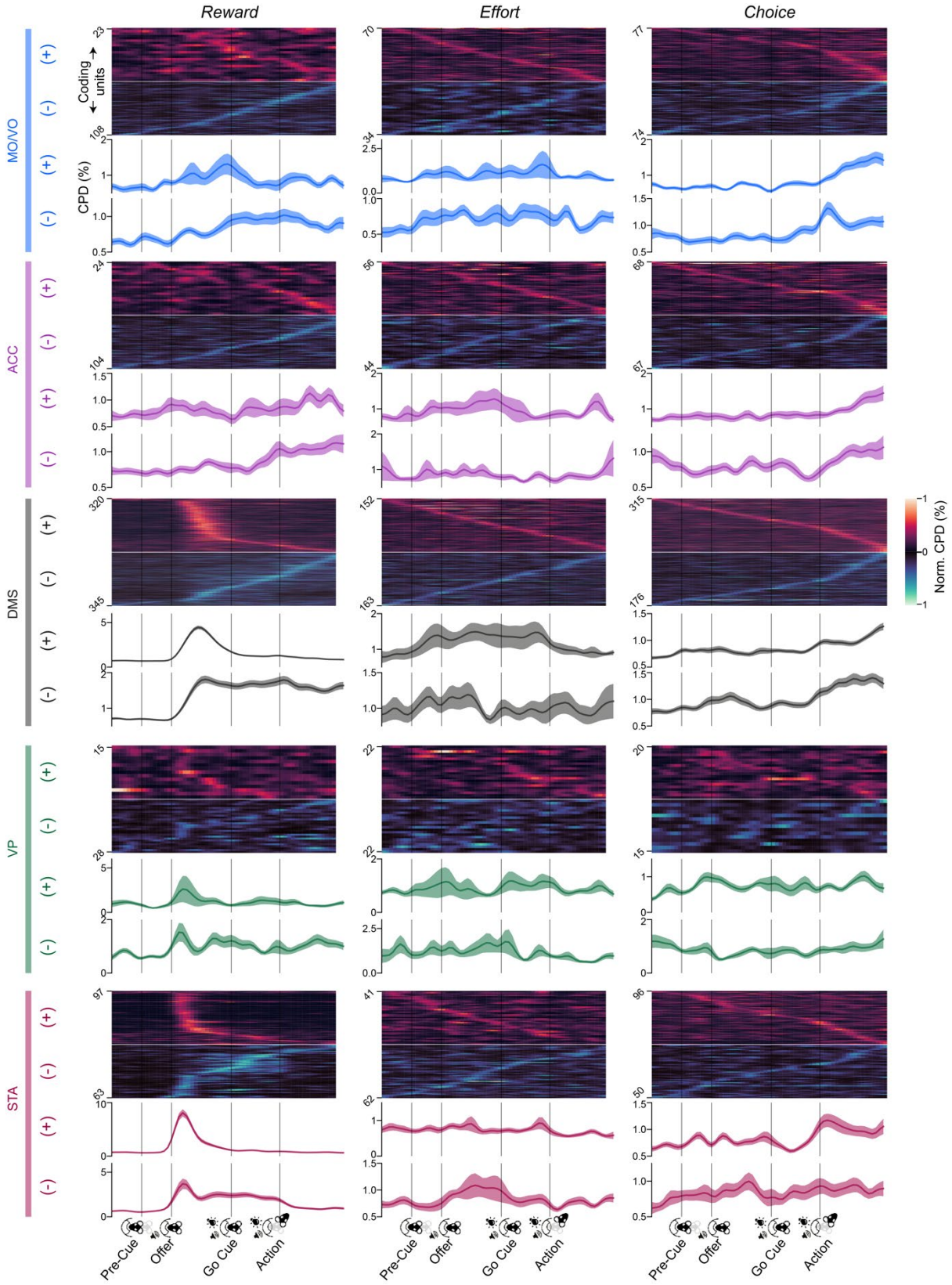
Figure 3. Dynamic and distributed single neuronal representations of reward, effort and choice across the frontal-basal ganglia network. a, upper, population average CPD time series, depicted as means (solid lines) ±

SEM (shaded areas) across all units, for each decision variable (columns). Colour-coded segments above indicate time points where the population average CPD was significantly higher than expected by chance in each region using a significance threshold of $p < .05$ (permutation testing, fdr corrected across time points). Insets depict individual unit (coloured dots) and population average (bars) ± SEM phasicness scores for each area. Phasicness comparisons for reward: STA vs all other regions: all $|z| > 5.71$; $p < .001$, two-sided Mann-Whitney U test. DMS vs all other regions: all $|z| > 3.16$ $p < .002$. All other $|z| < 0.58$, $p > .050$. Choice phasicness tests: MO/VO: $z = 4.12$, $p < .001$; ACC: $z = 3.48$, $p < .001$; DMS: $z = 10.18$, $p < .001$; STA and VP: both $z < 0.93$, $p > .05$. Lower, latencies to peak CPD, depicted as kernel density estimates (violins) and superimposed boxplots. Instances of peak reward CPDs occurring before offer cue onset were detected at below-chance rate (< 3.72%, $p > .05$, one-sided binomial test). Reward CPD phasicness tests against bootstrapped null distribution (online methods): MO/VO: $z = 1.28$, $p > .050$; ACC: $z = 2.27$; $p = .023$; DMS: $z = 2.12$, $p = .027$; VP: $z = 0.49$, $p > .050$; STA: $z = 6.80$, $p < .001$. **b,** Instantaneous recruitment of neurons dedicated to signalling any decision variable with a significance cut-off of $p < .05$. **c,** Population-averaged z-scored firing rates, grouped by positive (above) or negative (below) tuning valence and plotted on each level of a decision variable (dark lines) ± SEM (shaded areas). Data in each group are averaged across other conditions. Curves in a-c are depicted over a warped trial timeline beginning 0.4s before nose-poke entry up to the median corridor entry time (online methods) and are smoothed with a gaussian filter with a standard deviation of 1.0.



Supplemental Figure 3. Dynamic single neuronal representations of reward, effort and decision across the frontal-basal ganglia network. Data are depicted as normalized z-scored firing rate differences for each neuron (row) between the highest and lowest category of each decision variable and sorted by the time bin of peak difference. Heatmap and PSTH data are grouped by positive (above) or negative (below) tuning valence. Neurons

were selected on the basis of significant regression coefficients compared to trial-shuffled null distributions (online methods). Time series for each neuron are smoothed with a Gaussian kernel of 1.5 standard deviation.



Supplemental Figure 4. Dynamic and variable single neuronal representations of reward, effort and decision across the frontal-basal ganglia network. Heatmap data are depicted as CPD values for each

significantly coding neuron (row) and sorted by the time bin of peak CPD, while PSTH data represent mean ± SEM across all significantly coding neurons. Heatmap and PSTH data are grouped by positive (above) or negative (below) tuning valence. Neurons were selected on the basis of significant regression coefficients compared to trial-shuffled null distributions (online methods). Time series for each neuron on the heatmaps and for population-average traces are smoothed with a Gaussian kernel of 1.5 standard deviation.

Neurons across the cortico-basal ganglia network form cell assemblies during cost-benefit decision-making

Such diversity in the single neuronal firing rate read-outs poses a challenge at the output interpretation level. How can levels of reward, effort, and intention to accept/reject these offers be inferred from spatiotemporally varying inputs? One candidate mechanism of co-ordination across varying single neuronal signals is the formation of co-active assemblies, defined as groups of neurons that synchronise their activity in short time windows. The activity patterns of such assemblies have been proposed to reflect the underlying synaptic connectivity of the constituent neurons and provide a higher-order level of computation [32]. We thus examined whether cell assemblies could be detected in the cortico-basal ganglia network during cost-benefit decision-making and, if so, whether they could encode decision-relevant variables through the co-ordination of diverse populations.

We first addressed whether such short-latency cofiring was a common feature of these brain areas during our task. Many pairs of units across all anatomical structures were highly cross-correlated within a short interval (+/- 50 ms) around 0s lag (Fig. 4a-b), consistent with a physiologically meaningful readout for downstream neurons [32]. These short-latency correlations were dependent on the precise temporal relationship between spike trains, as randomly jittering the spike trains between 0 to 250ms reduced the baseline to peak difference in cross correlation by 71% (Fig. 4c). We used principal component analysis (PCA) to determine the extent of spike timing coordination across groups of simultaneously recorded cells. The added variance explained by the first principal component was much larger for the unjittered z-scored binned spike train (Fig. 4d), as was the mean of the variance added by the first 5 principal components (Fig. 4d). This indicates that a significant proportion of neural activity in the frontal cortical-basal ganglia network is coordinated at the fine temporal scale necessary for the formation and expression of cell assemblies [32, 35, 37, 41]. Coactivating assemblies of neurons were identified by PCA followed by independent component analysis (PCA-ICA) on the binned spike trains of neurons from all recorded regions during a time window spanning the start of the offer cue and the end of the action window (online methods; [42, 43]. This approach yielded a total of 727 assemblies. Most had 'member neurons' (i.e., neurons with a high weighting in the assembly pattern) restricted to a single brain region (~87% Fig. 4e, g). The remaining assemblies had member neurons that spanned multiple structures (Fig. 4f, g), including a small proportion (12/727) which were made up of member

neurons spanning 3 or more brain regions (e.g., Fig. 4f). The proportion of assemblies that had member units from within a single structure was significantly greater than chance (Fig. 4g).

We then examined whether such assemblies provide a unified read-out of canonical decision variables. To this end, the mean expression of the coactivity pattern was modelled as a weighted linear sum of reward, effort and choice (online methods). Strikingly, nearly half (49%) of all identified assemblies had a significant degree of selectivity for the prospective reward, current effort cost and choice, either individually or conjointly (Fig. 4h). Jittering the spike trains preserved the mean firing rates over the trial and almost all the rate-coding within the window of analysis (Fig. S5), but substantially disrupted coding at the level of the assembly (Fig. 4i). Overall, assemblies thus provided a substrate of encoding that co-ordinates activity across populations of neurons but is independent of the underlying correlated changes of firing rate at task-relevant timescales (see Fig.1)

Such assemblies can be formed of neurons from individual brain areas [35, 37] or can span multiple regions of the frontal-basal ganglia network [34]. To address how each region contributed to assemblies, we examined the likelihood of neurons from any structure forming assemblies regardless of their underlying tuning properties. We found that DMS neurons were more - but ACC neurons less - likely to be members of co-active assemblies (Fig. 4j), compared to a null hypothesis of assembly member neurons being equally distributed across structures. We then examined the relationship of neurons from each area to assemblies coding for one of the canonical decision variables. Though the proportions of neurons contributing to task-tuned assemblies were relatively even across regions, DMS and ACC neurons were more or less likely, respectively, to form reward- or choice-coding assemblies compared to assemblies that did not encode any decision variable (Fig. 4k). Cross structural assemblies were not significantly more likely to encode any of the decision variables (Fig. 4l). Together, assemblies were formed by neurons from across the entire network, with DMS neurons in particular having an important role in the expression of reward- and choice-related co-activity.

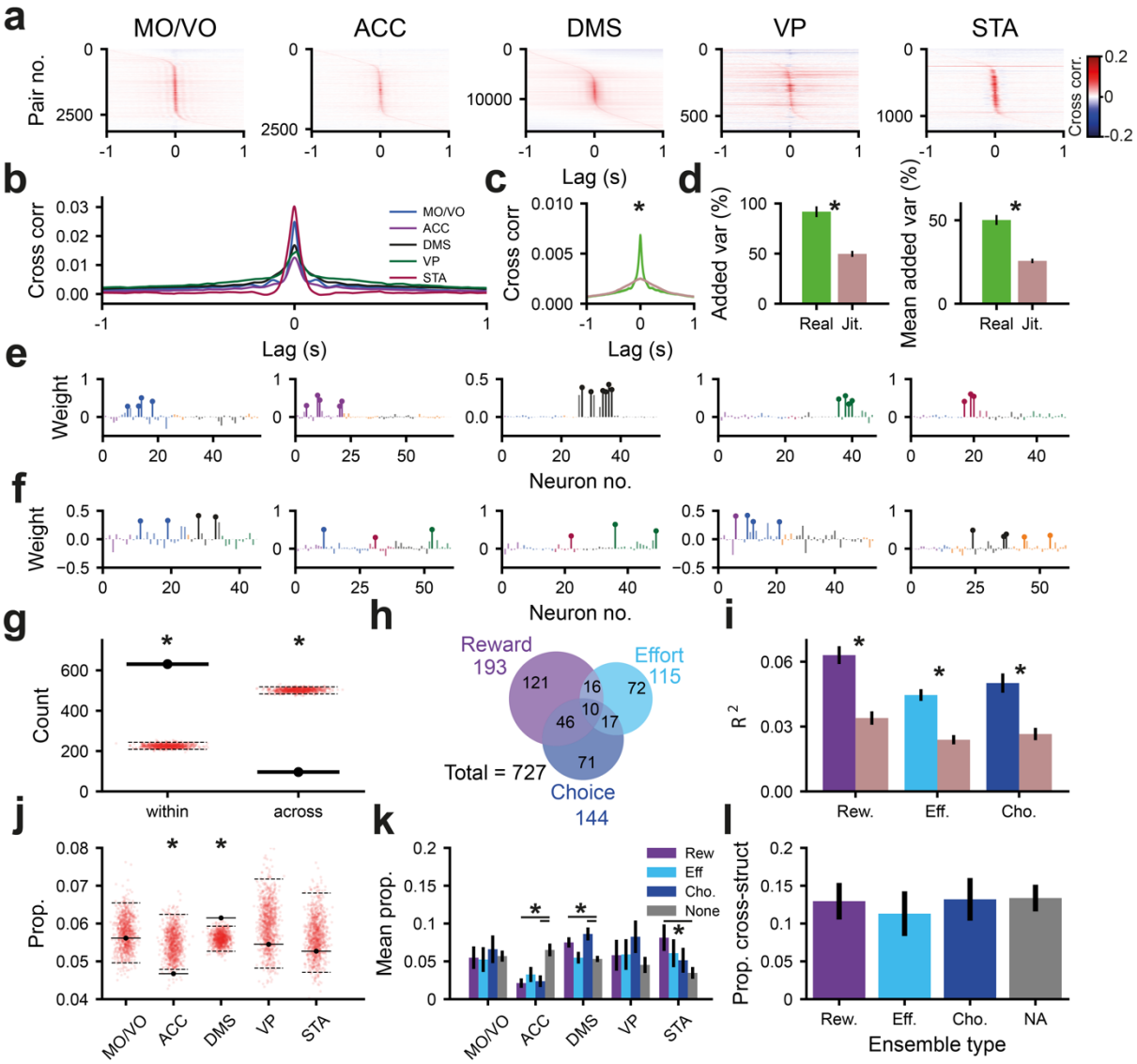
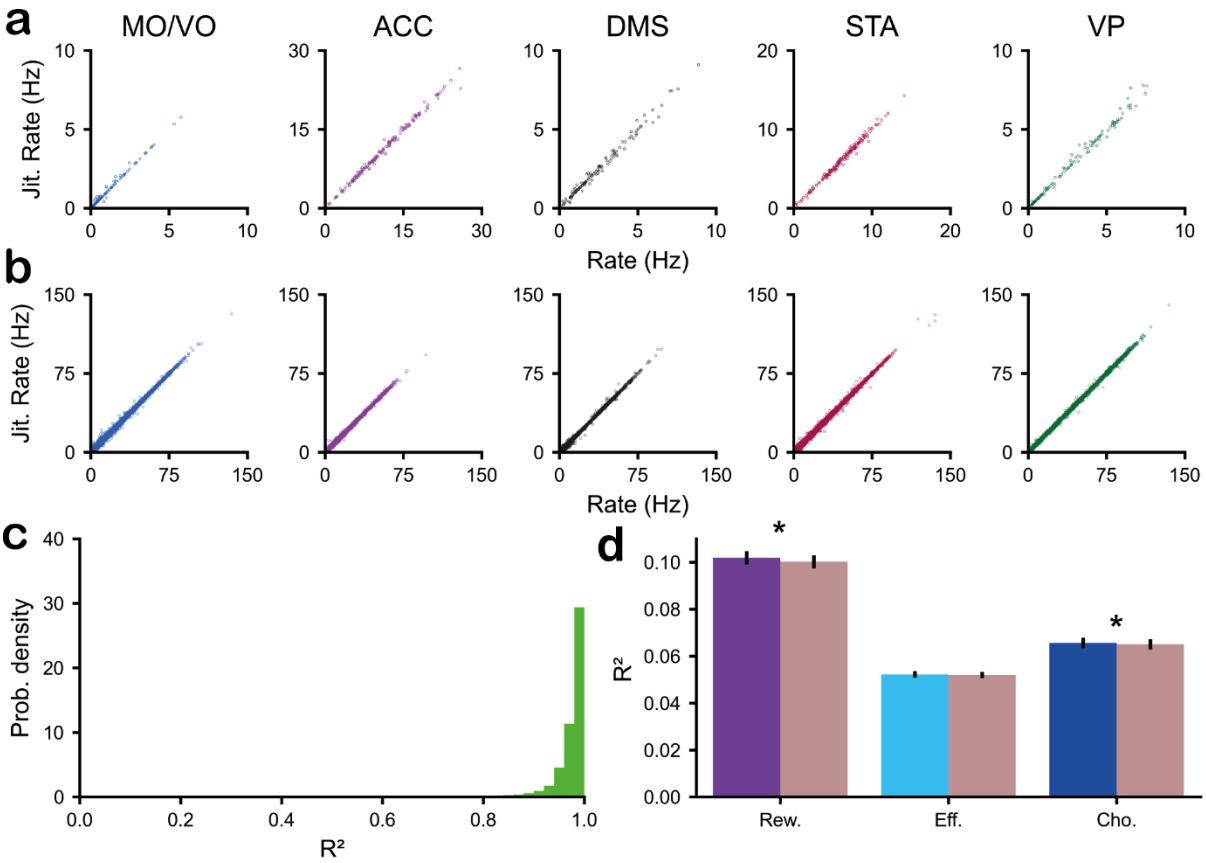


Figure 4. Cofiring assemblies across the frontal-basal ganglia network represent combinations of reward, effort and decision. a, heat map showing the cross correlation between pairs of units in the MO, ACC, DMS, VP and STA. **b,** the mean cross correlation between pairs of units in the MO, ACC, DMS, VP and STA. **c,** The mean cross correlation between pairs of units across all structures with (orange) or without (blue) random jittering of spike trains in the range -250ms to 250ms (0ms peak height, $p<10^{-13}$, Wilcoxon signed-rank test). **d,** The mean variance added over that of a single unit for the first eigenvector (left, $p<10^{-12}$, Wilcoxon signed-rank test) and mean variance added of the first 5 eigenvectors (right, $p<10^{-12}$, Wilcoxon signed-rank test) with or without random jittering of spike trains in the range -250ms to 250ms **e,** lollipop plot showing example assembly patterns with members solely within the MO/VO, ACC, DMS, VP and STA (left to right). Member neurons are marked with a circle on top of the bar indicating their weight. Color scheme as in **b**. **f,** example assembly patterns that span multiple anatomical regions. Member neurons are marked with a circle on top of the bar indicating their weight. Color scheme as in **b**. **g,** the number of assembly patterns that span multiple anatomical structures (black marker) as compared to the proportion expected due to chance when shuffling the anatomical labels of neurons (red dots). The black dashed lines mark the 2.5th and 97.5th percentile of this shuffled distribution. **h,** Venn diagram showing the number of assemblies which have their coactivity significantly modulated by reward, effort and decision outcome. **i,** the variance in reward ($p<10^{-19}$, Wilcoxon signed-rank test), effort ($p<10^{-13}$, Wilcoxon signed-rank test) or decision outcome (pseudo-R^2 , $p<10^{-15}$, Wilcoxon signed-rank test), explained by assembly pattern expression strength with (brown) or without (blue) random jittering of spike trains in the range of -250ms to

250ms for significantly encoding assemblies **j,** the mean proportion of SUs from different brain regions which were members of each assembly as compared to the mean proportion when anatomical labels were shuffled amongst neurons (red dots). The 2.5thand 97.5th percentiles for shuffled distributions were marked with dashed black lines. **k,** the mean proportion of SUs from different brain regions in assemblies separated by their encoding properties. The proportion of neurons from each anatomical structure in encoding assemblies was compared to assemblies which did not encode canonical variables (significance for p < .050 determined by permutation test, Benjamini-Hochberg procedure used to control for multiple comparisons). **l,** the proportion of cross-structural assemblies by assembly encoding type (p>0.05 for reward, effort and decision vs non-coding, z-test). * p < .050.



Supplemental Figure 5. Cofiring assemblies across the frontal-basal ganglia network represent combinations of reward, effort and decision. a, The trial-by-trial jittered firing rate plotted against the real firing rate for example neurons from the MO, ACC, DMS, STA and VP. **b**, The trial-by-trial jittered firing rate plotted against the real firing rate pooled across all neurons in the MO, ACC, DMS, STA and VP (brain structures ordered as in **a**). **c**, The distribution of the R^2 for the correlation between the trial-by-trial jittered firing rate and the real firing rate for each of the recorded neurons. **d**, The R^2 between the firing rate of unjittered (brown) as compared to jitttered spike trains (blue) in the decision window for reward (p=4.3 x 10-5, Wilcoxon signed-rank test), effort (p = .068, Wilcoxon signed-rank test) and decision (p = .038, Wilcoxon signed-rank test) for significantly coding neurons.

Assemblies that encode decision variables are comprised of individual neurons with diverse tuning properties.

Both individual neurons and assemblies could encode and integrate the canonical decision variables. This calls to question the degree of independence between the identified temporal and rate codes. To

address this, we asked whether ensembles that encoded a given decision variable were comprised mostly of neurons that *rate-coded* the same variable. We found examples of assemblies where the rate-coding of individual neurons clearly diverged from that of the overall ensemble (Fig. 5a-f). To quantify the intersection of the encoding properties of single neurons (Fig. 2f) and assemblies (Fig. 4h), we defined the proportion of each type of rate-coding neuron that contributed to each type of encoding assembly (Fig. 5g). In all cases, single units that encoded a given variable were more likely to contribute to assemblies that encoded the same variable than any other assembly type, including non-encoding assemblies (Fig 6g). However, only a small proportion of the variance in the assembly pattern expression strength dynamics (<18%, for reward, effort and choice) could be explained by changes in the firing rate of member neurons (Fig. 5h), demonstrating that this bias was not a consequence of correlated, task-related firing rate across assembly members. Interestingly, 20% of single units of all coding types contributed to assemblies encoding each of the other two decision variables. Additionally, units encoding reward or choice were more likely to be members of assemblies encoding choice or reward, respectively, than assemblies that did not encode the decision variables (Fig. 5g). Taken together, these findings suggest that though a member neuron's individual tuning often aligns with that of their assembly, assembly pattern expression strength changes were poorly explained by firing rate dynamics. Furthermore, neurons often contributed to assemblies encoding a different parameter to their own. Co-activity is thus separable from rate code and can integrate units with diverse tuning properties.

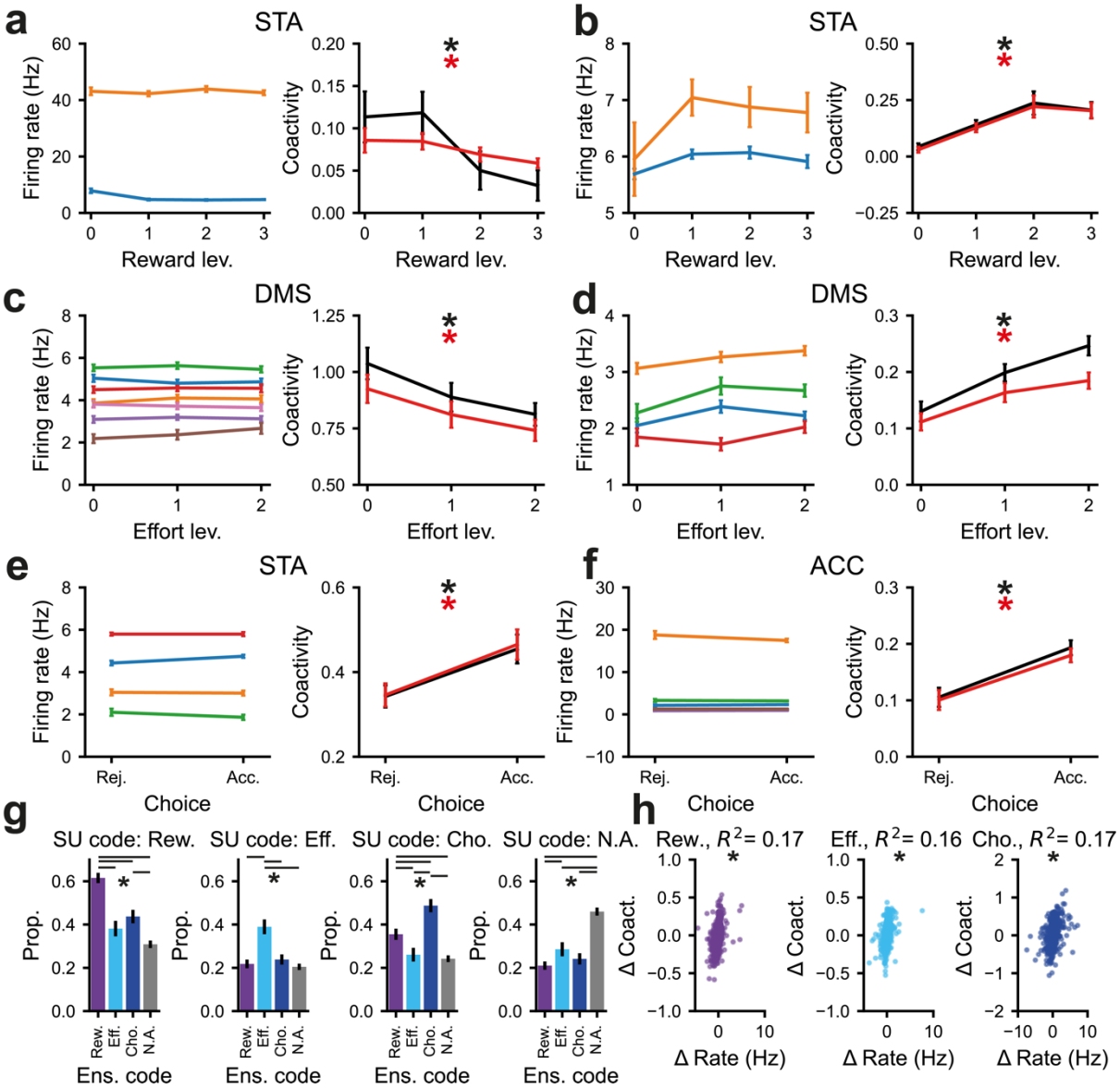


Figure 5. Decision-related encoding by co-firing assemblies can be orthogonal to the rate coding of the individual neurons that comprise them. a and **b,** The firing rate of the member neurons that make up assemblies (left). They show no significant change across reward levels. Assembly coactivity (right) computed using the full assembly pattern weights (black) or restricting weights to only member neurons (red). Both the coactivity for the full assembly pattern and for members only is significantly shaped by the level of offered reward. **c** and **d,** The firing rate of the member neurons that make up assemblies (left). They show no significant change across effort levels. Assembly coactivity (right) computed using the full assembly pattern weights (black) or restricting weights to only member neurons (red). Both the coactivity for the full assembly pattern and for members is significantly determined by the level of effort. **e** and **f,** The firing rate of the member neurons that make up assemblies (left). They show no significant change with choice outcome. Assembly coactivity (right) computed using the full assembly pattern weights (black) or restricting weights to only member neurons (red). The choice outcome has a significant effect on the coactivity pattern of both the full and members-only assembly pattern. For a-f, * p<.050, significance was determined using linear regressions with firing rate or coactivity as the dependent variable. The CPD of reward, effort and choice outcome was compared to the null distribution from trial shuffled data. **g,** the proportion of member neurons of reward, effort, decision and non-encoding assemblies which rate code reward (left), effort (middle-left), decision outcome (middle-right) or do not rate code for the canonical variables (right). * p <.050, z-test, Benjamini-Hochberg procedure used to control for multiple comparisons. **h,** the change in the

firing rate of member neurons across the levels of reward (**i**) and effort (**ii**) and decision (**iii**) outcome only explains a small proportion (<20%) of the change in assembly expression strength. * $p < .050$ in linear regression.

Encoding of reward, effort and choice by cortico-basal ganglia assemblies aligns with relevant task events.

Our analyses of assemblies to this point defined encoding using expression strength across the entire task window. If assemblies provide population-level encoding that has ongoing influence on cost-benefit decisions akin to that proposed for single neurons, their temporal dynamics should align with key task epochs. Whereas single neuron rate dynamics are defined by the neuron firing at a specific task point at a specific magnitude, an equivalent change in assembly strength dynamics indicates that millisecond-level *synchronisation* is consistent with regard to decision variable magnitude and task-point.

To examine whether this was the case, we analysed the time course of encoding (as measured by CPD) for significant assemblies with respect to the mean occurrence of the reward offer, nose-poke exit and action initiation. Significance at a given timepoint was defined in relation to assemblies constructed from jittered spike trains (Fig. S5; online methods). Reward and effort were encoded at above-chance levels during virtually all time-bins from the reward offer until the end of the decision point, whereas choice-encoding assemblies were significant from shortly after the nose-poke exit to the trial end (Fig. 6a). As found for single unit rate-coding, reward and choice encoding were more phasic than effort (Fig. 6a), although peak encoding for all variables was spread across the relevant task points (Fig. 6b). As seen for rate-coding of individual neurons, assemblies could encode positively (e.g. maximum expression strength for maximum reward) or negatively (e.g. maximum expression strength for minimum reward). Interestingly, the times course of assemblies that negatively encoded reward was more sustained/less phasic than for positively encoding assemblies (Fig. 6c, d), resembling that of individual DMS units. Overall, both single unit firing rates (Fig. 3; Fig. S3-4) and assembly strength (Fig. 6) could dynamically encode the decision variables on task-relevant timescales. Importantly, however, the assembly-level encoding spanned multiple neurons, many of which did not individually encode the same parameter as their assembly (Fig. 5). Cost-benefit decision variables in these cortico-basal ganglia structures were thus dynamically encoded and integrated both locally at the level of single unit and emergent assemblies across the network.

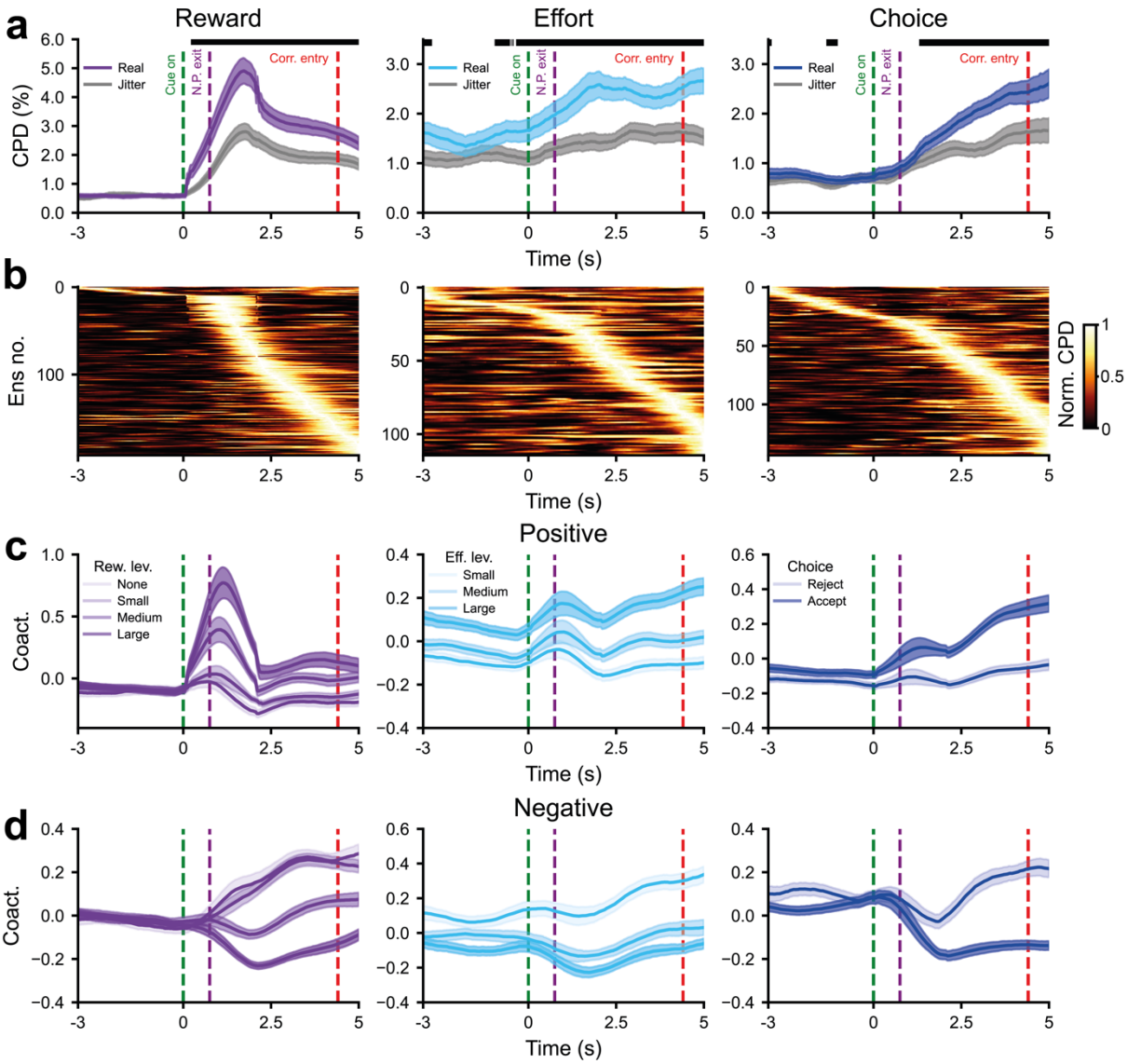


Figure 6: Co-firing assemblies can dynamically encode decision variables on task-relevant timescales. a, the CPD for reward, effort and choice regressed onto the time-series of the expression strength of cell assemblies either with (grey) or without (blue) the random jittering of the spike train in the range of -250ms to 250ms for reward, effort and choice coding assemblies (left to right). Time of 0s is the onset of the reward cue and periods of time where the CPD is significantly higher for the unjittered than the jittered assembly pattern expression strength were marked with a blue bar (Wilcoxon signed-rank test, $p < .050$). The Benjamini-Hochberg procedure was used to control for multiple comparisons. **b**, heat plot where rows represent the normalised coefficient of partial determination (CPD) for reward, effort and choice regressed onto the time-series of the expression strength of cell assemblies for each reward, effort and choice coding assemblies respectively (left to right). CPD was normalised by dividing it by its peak value for each assembly. Time of 0s is the onset of the reward cue. **c** and **d**, the z-scored, smoothed (using a gaussian of std = $2 / \sqrt{12}\ s$) coactivity across the different levels of decision variables for positively (**c**) and negatively (**e**) reward, effort and choice (left to right) encoding assemblies. Increasing levels of reward and effort and accept trials were shown in darker colour. Time of 0s is the onset of the reward cue.

Discussion

Cortico-basal ganglia circuits play a key role in choosing actions based on information about prospective costs and benefits. The partially segregated loops that comprise these circuits are still often assumed to primarily be a substrate for serial rate coding, with neurons in each structure across the circuit performing a specific computation. Recently, many authors have questioned whether this level of serial organisation can fully account for value-based computations in these circuits [29, 30, 44]. To investigate, alternative, parallel encoding schemes however, it is necessary to record many neurons across the network simultaneously. Utilizing such recordings, we found that some populations, particularly striatal neurons, were mostly locked to specific task point and that single neurons could integrate information about reward and effort. Most neurons, however, were engaged in a wide range of different responses across the task space. We demonstrate that these same neurons could also form cell assemblies that could encode the same decision parameters by utilising the synchronisation of spiking activity. Together, these findings suggest a hierarchical organisation of cost-benefit computation, where individual neurons with diverse rate-coding properties are integrated into higher-order, emergent assemblies that can encode decision-related parameters across the circuit at millisecond timescales.

A behavioural paradigm to delineate the influence of reward and effort of subsequent choice

Deciding whether to expend energy based on the trade-off between effort and reward is fundamental to adaptive behaviour. To investigate the neural correlates of such decision-making we developed an "accept-reject" decision task inspired by diet selection foraging [45] and previous effort-based decision-making paradigms [46, 47]. Importantly, our task design separated out the initial presentation of the reward offer from subsequent initiation and accept/reject decision, allowing us to dissociate neural activity driven by these factors. Decisions to pursue reward were robustly influenced both by the reward offer and the amount of effort needed to obtain it. Though effort was not explicitly signalled at the beginning of trials, it nevertheless had a consistently negative effect on choice throughout trial blocks. Choices were further modulated by increasing reward count which, together with the results of the outcome devaluation experiment, indicated that rats were flexibly titrating their behaviour to changes in the subjective value of the food rewards. Interestingly, while the magnitude of

prospective reward increased premature nose-poke exit rates, it also *increased* action initiation latencies after successful nose-pokes, indicating that rats employed restraint in proportion to the value of the offers. Overall, the task provided a robust platform for interrogating the neural correlates of cost-benefit decision-making.

Single neuron encoding of cost-benefit parameters is specialised in some populations, but also widely distributed.

A central question regarding the neural correlates of cost-benefit decision-making is whether decision variables are represented locally or distributed across multiple brain areas [7, 29, 30, 48]. Taken to the extreme, if each population of neurons were responsible for specific computations, then it would be expected that neurons in each brain regions would have uniform relationships to individual/combinations of decision-variables at specific task points. The activity of many single neurons in the striatum and subthalamic area positively encoded reward in this way, with responses tightly locked to the offer of prospective reward. This finding is in line with the established role of striatal neurons, under the modulation of midbrain dopaminergic inputs from the midbrain, in reward prediction [49]. However, we found that at least 20-30% of neurons in all five structures encoded *each* decision parameter over the entire task window. Interestingly, negatively encoding neurons in the striatum and subthalamic area also had this temporally distributed activity. Thus, our results show at any given task point following the reward offer, there were neurons in each structure encoding reward and effort, and later encoding choice. Irrespective of the relationship of single neurons to cell assemblies (see below), this finding supports parallel and distributed processing of cost-benefit parameters at the single unit level.

Our behavioural results clearly demonstrate that reward and effort were used to determine the animal's choice. Whether single neurons in the recorded areas can process and/or integrate both parameters is thus a key question as to how they contributed to such decisions. Several previous studies have suggested that orbital and cingulate cortical neurons perform this role[21, 50]. In contrast, we found that dorsomedial striatal neurons were most likely to multiplex combinations of reward and effort encoding. Striatal neurons have previously been shown to integrate reward with other parameters [51, 52], but not reward and effort. The ability of striatal neurons to integrate reward and effort is unlikely to be a downstream effect of dopaminergic modulation, as several studies have shown a weak relationship between dopamine release and effort (reviewed in [53]). Given that effort here was altered over blocks that lasted for minutes,

the ability of striatal neurons to integrate information present on slower timescales may have been underestimated, with frontal cortical regions more typically assigned this role [54].

Overall, our results suggest that neurons with uniform, phasic responses can be found in specific brain areas, but that they are superimposed on a background of more temporally dispersed network-level encoding. While the former has been relatively well characterised, dispersed coding, particularly across the wider basal ganglia network may have been significantly underestimated.

Cortico-basal ganglia assemblies could represent network-level processing of cost-benefit parameters.

Here we demonstrate that neurons across all the recorded cortico-basal ganglia regions can form cell assemblies, and that these assemblies can encode decision parameters. Cell assemblies have been mostly described in hippocampal neurons during spatial navigation [32, 41] and during the representation contextual information [37], but to our knowledge, they have rarely been described in cortico-basal ganglia networks. In those rare examples, the expression strength of assemblies has been linked with spatial encoding and/or discrimination[34, 55]. Here we provide evidence that variability in the strength of cell assembly expression in the cortex, striatum and other basal ganglia structures can individually and conjointly encode multiple levels of prospective reward and effort over timescales that are relevant to the utilisation of that information to behaviour.

Cell assemblies have specific properties that could be as important in value-based learning and decision-making as they are in spatial learning and navigation. Specifically, assemblies promote firing within the membrane time constant (10-30ms) that is optimal to discharge a downstream neurons/assemblies and for strengthening synaptic weights between those neurons[32]. Instead of acting as individual processing units, neurons in cell assemblies, both within and across structures, represent information cooperatively via their synchrony. This synchronous firing can produce a composite downstream affect that cannot be achieved by single neurons alone [32, 33, 41]. Therefore, the importance of these findings is that they demonstrate that value-based variables can be represented on a spatiotemporal scale than cannot be achieved by single neurons, allowing integration of network-level connectivity akin to that in hippocampal networks, where the key property of the assembly is to simultaneously

mobilize enough pyramidal neurons to discharge the target neurons and promote synaptic plasticity between those neurons.

We provide further evidence to support the recent demonstration that assemblies can span multiple brain structures [34]. While we detected considerably more cross-structural assemblies in absolute numbers than that study, they made up a smaller proportion of assembly patterns overall. The technical challenge of recording large numbers of neurons across structures means that such cross-structural assemblies are probably underrepresented, with actual cell assembly size likely comprised of hundreds of neurons [56]. Thus, because we sample only a small proportion of assembly members, we likely compute a noisy estimate of the coactivity that arises in the complete cell assembly.

Assemblies that encoded a given decision variable were often comprised of single neurons coding for either a different parameter or none of the canonical variables at all (see schematic representation in Fig. 7). This finding supports the idea that such assemblies are emergent: the activity of the population is more than a simple sum of the individual members, echoing previous findings in the hippocampus [37]. This has important implications for future studies of decision-making in these circuits, as it demonstrates that neurons that do not have clear, temporally consistent rate-coding responses (Fig. 2-3) can still play an important role at the population level (Fig. 7). This view is not mutually exclusive to neurons of a specific type having a highly having a specialized role within the network, as striatal neurons appear to do here. Rather, we suggest that higher-level emergent activity constitutes a different level of organisation that facilitates parallel representations and/or computations across neurons and structures in ways that rate-coding alone cannot.

A hierarchical relationship between single unit and assembly encoding could reflect the underlying scale synaptic connectivity formed during learning.

One key outstanding question is how encoding at the level of single neurons and cell assemblies is related. In this regard, a key aim of this study was to provide a detailed analysis of both levels of computation. Few would doubt that the firing rate of a single neuron represents an important level of computation in the brain. Single neuron tuning must ultimately result from integration of synaptic inputs and weights of those inputs to the neuron, which will evolve and consolidate during learning under the influence of neuromodulators such as dopamine. As demonstrated here, such single neuron responses can achieve complex computations, such as the encoding

of multiple task variables that occur across different timescales. Nevertheless, while the origin of inputs to a single neuron can arise from disparate and distant brain areas, such computations can only result from the combination of synapses to that neuron.

Cell assemblies, on the other hand, must reflect interactions on a larger spatial scale. Importantly, connections between assembly members need not be direct. Striatal neurons had the largest propensity to form assemblies here, but connections between striatal projection neurons are weak [57]. Striatal assemblies are therefore more likely to be formed from common cortical and thalamic inputs that are the dominant drivers of striatal activity [27, 58] but require convergence of many neurons to drive reliable output [59]. Thus, even when an assembly was not defined as cross-structural in our data, it was highly likely to have reflected a much wider pool of connected neurons across structures. This framework shares common ground with the Hopfeldian approach, where information is represented by coordinated neural activity, conceptualised as an attractor basin in a low dimensional manifold [60]. Within this approach, the expression strength of each of the identified assemblies can be thought of as dimensions of a latent space describing the neural activity [61]. This latent space is particularly sensitive to the coactivity of neurons and, importantly, cannot be wholly described by the second-by-second changes in firing rate [62]. In fact, in the absence of fine temporal coactivity, the position along each dimension of this latent space carries substantially less information about decision relevant variables.

We suggest that the relationship between single neuron rate and assembly coding represents different levels of a hierarchical organisation. Assemblies would be higher in the hierarchy in terms of incorporating a larger number of individual neurons and having lower dimensionality. However, within a Hebbian framework, development the tuning curves of individual neurons and their relationships to those of their efferent and afferent partners will likely play a key role in the formation of those assemblies. Phasically active neurons, such as those in the striatum and subthalamus, may be particularly powerful in this regard. Similarly, the expression of assemblies should reinforce the rate-related properties of the individual members by consolidating the synaptic connections on which they rely. Applying such a framework could further reveal how large areas of the brain are coordinated to implement cost-benefit decisions.

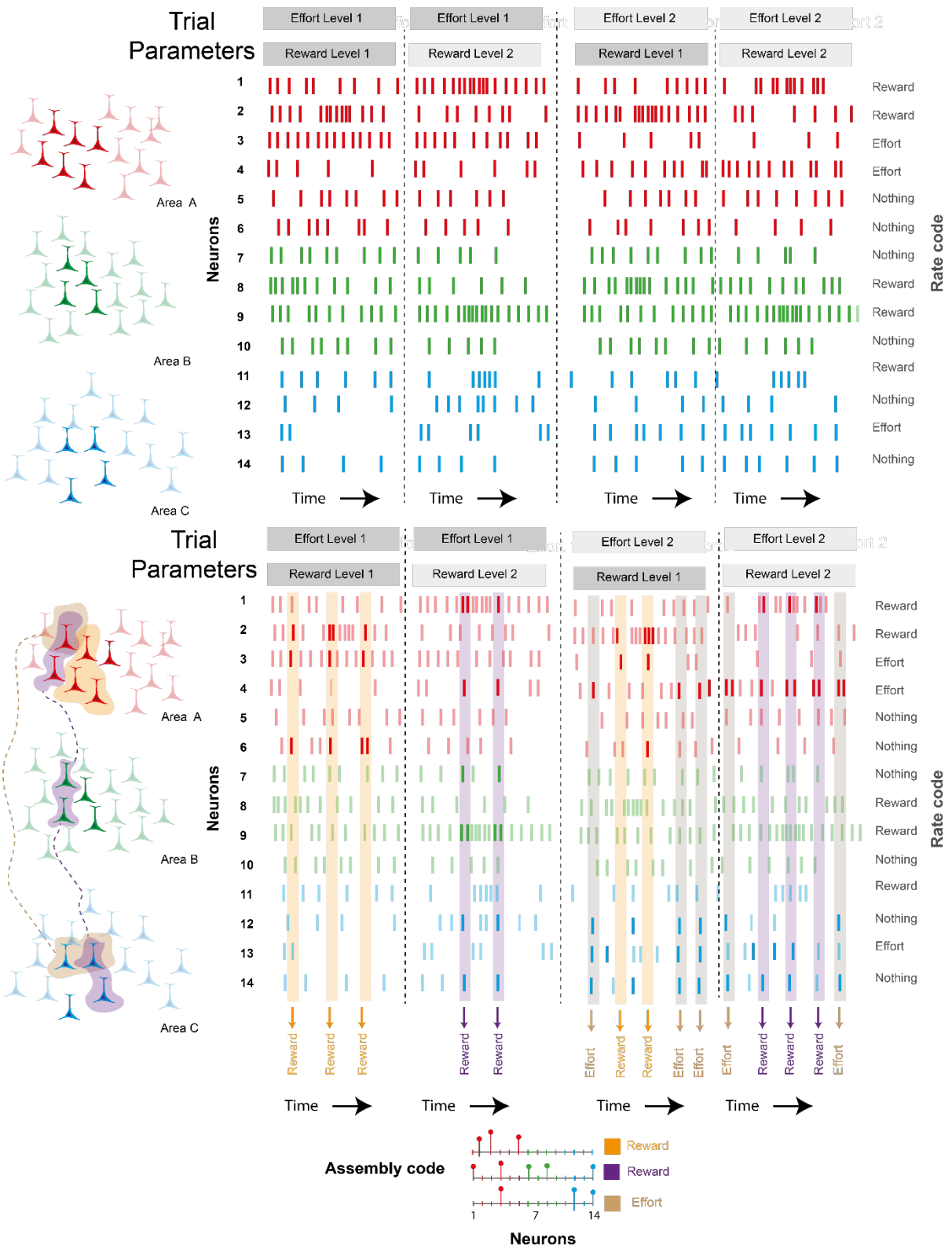


Figure 7. Schematic representation of relationships between detected rate and assembly codes and their possible relationships to the wider network. a, Schematic representation of rate responses in 3 populations of neurons (denoted by colour) from different brain areas during trials with different combinations of prospective reward and effort. Neurons can encode the decision parameters and/or transient rate changes that occur consistently in part of the trial (e.g. neurons 1 and 2) and/or global changes in firing rate during specific conditions (e.g. neurons 3 and 4). The recorded neurons (in bold) are assumed to reflect

the rate-responses of the wider population (faded) within that structure. **b,** Assembly-coding can occur simultaneously during the rate coding shown in A. Neurons with all types of rate coding (reward, effort, nothing) can contribute to assemblies that encode specific combinations of effort and reward. Spikes that contribute/or a consequence of assemblies can occur inside or outside or rate-coding periods. Assemblies can be formed within and/or across different areas and a given neuron can participate in assemblies encoding different parameters (e.g. neuron 14). The recorded neurons (bold) represent a small part of the "real" assembly which likely comprises many neurons in the wider of population of neurons within and outside the recorded structures.

References

1. Adam, R., et al., *Dopamine reverses reward insensitivity in apathy following globus pallidus lesions.* Cortex, 2013. **49**(5): p. 1292-303.
2. Hauber, W. and S. Sommer, *Prefrontostriatal circuitry regulates effort-related decision making.* Cereb Cortex, 2009. **19**(10): p. 2240-7.
3. Levy, R. and B. Dubois, *Apathy and the functional anatomy of the prefrontal cortex-basal ganglia circuits.* Cereb Cortex, 2006. **16**(7): p. 916-28.
4. Manohar, S.G. and M. Husain, *Human ventromedial prefrontal lesions alter incentivisation by reward.* Cortex, 2016. **76**: p. 104-20.
5. Munster, A. and W. Hauber, *Medial Orbitofrontal Cortex Mediates Effort-related Responding in Rats.* Cereb Cortex, 2018. **28**(12): p. 4379-4389.
6. Pennartz, C.M., et al., *Corticostriatal Interactions during Learning, Memory Processing, and Decision Making.* J Neurosci, 2009. **29**(41): p. 12831-8.
7. Rushworth, M.F., et al., *Frontal cortex and reward-guided learning and decision-making.* Neuron, 2011. **70**(6): p. 1054-69.
8. Uslaner, J.M., P. Yang, and T.E. Robinson, *Subthalamic nucleus lesions enhance the psychomotor-activating, incentive motivational, and neurobiological effects of cocaine.* J Neurosci, 2005. **25**(37): p. 8407-15.
9. Walton, M.E., et al., *Functional specialization within medial frontal cortex of the anterior cingulate for evaluating effort-related decisions.* J Neurosci, 2003. **23**(16): p. 6475-9.
10. Abitbol, R., et al., *Neural mechanisms underlying contextual dependency of subjective values: converging evidence from monkeys and humans.* J Neurosci, 2015. **35**(5): p. 2308-20.
11. Balewski, Z.Z., et al., *Value dynamics affect choice preparation during decision-making.* Nat Neurosci, 2023. **26**(9): p. 1575-1583.
12. Cowen, S.L., G.A. Davis, and D.A. Nitz, *Anterior cingulate neurons in the rat map anticipated effort and reward to their associated action sequences.* J Neurophysiol, 2012. **107**(9): p. 2393-407.
13. Diehl, G.W. and A.D. Redish, *Differential processing of decision information in subregions of rodent medial prefrontal cortex.* Elife, 2023. **12**.
14. Hillman, K.L. and D.K. Bilkey, *Neurons in the rat anterior cingulate cortex dynamically encode cost-benefit in a spatial decision-making task.* J Neurosci, 2010. **30**(22): p. 7705-13.
15. Lopatina, N., et al., *Medial Orbitofrontal Neurons Preferentially Signal Cues Predicting Changes in Reward during Unblocking.* J Neurosci, 2016. **36**(32): p. 8416-24.

16. Nakamura, K., et al., *Differential reward coding in the subdivisions of the primate caudate during an oculomotor task.* J Neurosci, 2012. **32**(45): p. 15963-82.
17. Nougaret, S., C. Baunez, and S. Ravel, *Neurons in the Monkey's Subthalamic Nucleus Differentially Encode Motivation and Effort.* J Neurosci, 2022. **42**(12): p. 2539-2551.
18. Porter, B.S., K.L. Hillman, and D.K. Bilkey, *Anterior cingulate cortex encoding of effortful behavior.* J Neurophysiol, 2019. **121**(2): p. 701-714.
19. San-Galli, A., et al., *Primate Ventromedial Prefrontal Cortex Neurons Continuously Encode the Willingness to Engage in Reward-Directed Behavior.* Cereb Cortex, 2018. **28**(1): p. 73-89.
20. Tachibana, Y. and O. Hikosaka, *The primate ventral pallidum encodes expected reward value and regulates motor action.* Neuron, 2012. **76**(4): p. 826-37.
21. Kennerley, S.W., et al., *Neurons in the frontal lobe encode the value of multiple decision variables.* J Cogn Neurosci, 2009. **21**(6): p. 1162-78.
22. Arber, S. and R.M. Costa, *Networking brainstem and basal ganglia circuits for movement.* Nat Rev Neurosci, 2022. **23**(6): p. 342-360.
23. Roseberry, T.K., et al., *Cell-Type-Specific Control of Brainstem Locomotor Circuits by Basal Ganglia.* Cell, 2016. **164**(3): p. 526-37.
24. Kable, J.W. and P.W. Glimcher, *The neurobiology of decision: consensus and controversy.* Neuron, 2009. **63**(6): p. 733-45.
25. Padoa-Schioppa, C., *Neurobiology of economic choice: a good-based model.* Annu Rev Neurosci, 2011. **34**: p. 333-59.
26. Rangel, A. and T. Hare, *Neural computations associated with goal-directed choice.* Curr Opin Neurobiol, 2010. **20**(2): p. 262-70.
27. Peters, A.J., et al., *Striatal activity topographically reflects cortical activity.* Nature, 2021. **591**(7850): p. 420-425.
28. Guo, Z.V., et al., *Maintenance of persistent activity in a frontal thalamocortical loop.* Nature, 2017. **545**(7653): p. 181-186.
29. Cisek, P., *Making decisions through a distributed consensus.* Curr Opin Neurobiol, 2012. **22**(6): p. 927-36.
30. Hunt, L.T. and B.Y. Hayden, *A distributed, hierarchical and recurrent framework for reward-based choice.* Nat Rev Neurosci, 2017. **18**(3): p. 172-182.
31. Rushworth, M.F., et al., *Valuation and decision-making in frontal cortex: one or many serial or parallel systems?* Curr Opin Neurobiol, 2012. **22**(6): p. 946-55.
32. Buzsaki, G., *Neural syntax: cell assemblies, synapsembles, and readers.* Neuron, 2010. **68**(3): p. 362-85.
33. Hebb, D.O., *The organization of behavior; a neuropsychological theory*. The organization of behavior; a neuropsychological theory. 1949, Oxford, England: Wiley. xix, 335-xix, 335.
34. Oberto, V.J., et al., *Distributed cell assemblies spanning prefrontal cortex and striatum.* Curr Biol, 2022. **32**(1): p. 1-13 e6.
35. van de Ven, G.M., et al., *Hippocampal Offline Reactivation Consolidates Recently Formed Cell Assembly Patterns during Sharp Wave-Ripples.* Neuron, 2016. **92**(5): p. 968-974.
36. Wilson, M.A. and B.L. McNaughton, *Reactivation of hippocampal ensemble memories during sleep.* Science, 1994. **265**(5172): p. 676-9.
37. El-Gaby, M., et al., *An emergent neural coactivity code for dynamic memory.* Nat Neurosci, 2021. **24**(5): p. 694-704.
38. Pessiglione, M., et al., *Why not try harder? Computational approach to motivation deficits in neuro-psychiatric diseases.* Brain, 2018. **141**(3): p. 629-650.
39. Paxinos, G. and C. Watson, *The rat brain in stereotaxic coordinates*. 2006: Elsevier.

40. Shamash, P., et al., 2018.
41. Harris, K.D., et al., *Organization of cell assemblies in the hippocampus.* Nature, 2003. **424**(6948): p. 552-6.
42. Lopes-dos-Santos, V., et al., *Neuronal assembly detection and cell membership specification by principal component analysis.* PLoS One, 2011. **6**(6): p. e20996.
43. Lopes-dos-Santos, V., S. Ribeiro, and A.B. Tort, *Detecting cell assemblies in large neuronal populations.* J Neurosci Methods, 2013. **220**(2): p. 149-66.
44. Christensen, A.J., T. Ott, and A. Kepecs, *Cognition and the single neuron: How cell types construct the dynamic computations of frontal cortex.* Curr Opin Neurobiol, 2022. **77**: p. 102630.
45. Stephens, D.W.a.K., J.R. , *Foraging Theory.* 1986, Princeton: Princeton University Press.
46. Le Heron, C., M.A.J. Apps, and M. Husain, *The anatomy of apathy: A neurocognitive framework for amotivated behaviour.* Neuropsychologia, 2018. **118**(Pt B): p. 54-67.
47. Muller, T., et al., *Neural and computational mechanisms of momentary fatigue and persistence in effort-based choice.* Nat Commun, 2021. **12**(1): p. 4593.
48. Yoo, S.B.M. and B.Y. Hayden, *Economic Choice as an Untangling of Options into Actions.* Neuron, 2018. **99**(3): p. 434-447.
49. Oyama, K., et al., *Reward prediction error coding in dorsal striatal neurons.* J Neurosci, 2010. **30**(34): p. 11447-57.
50. Kennerley, S.W. and J.D. Wallis, *Evaluating choices by single neurons in the frontal lobe: outcome value encoded across multiple decision variables.* Eur J Neurosci, 2009. **29**(10): p. 2061-73.
51. Bloem, B., et al., *Multiplexed action-outcome representation by striatal striosome-matrix compartments detected with a mouse cost-benefit foraging task.* Nat Commun, 2022. **13**(1): p. 1541.
52. Rueda-Orozco, P.E. and D. Robbe, *The striatum multiplexes contextual and kinematic information to constrain motor habits execution.* Nat Neurosci, 2015. **18**(3): p. 453-60.
53. Walton, M.E. and S. Bouret, *What Is the Relationship between Dopamine and Effort?* Trends Neurosci, 2019. **42**(2): p. 79-91.
54. Klein-Flugge, M.C., A. Bongioanni, and M.F.S. Rushworth, *Medial and orbital frontal cortex in decision-making and flexible behavior.* Neuron, 2022. **110**(17): p. 2743-2770.
55. Trouche, S., et al., *A Hippocampus-Accumbens Tripartite Neuronal Motif Guides Appetitive Memory in Space.* Cell, 2019. **176**(6): p. 1393-1406 e16.
56. Ainsworth, M., et al., *Rates and rhythms: a synergistic view of frequency and temporal coding in neuronal networks.* Neuron, 2012. **75**(4): p. 572-83.
57. Koos, T., J.M. Tepper, and C.J. Wilson, *Comparison of IPSCs evoked by spiny and fast-spiking neurons in the neostriatum.* J Neurosci, 2004. **24**(36): p. 7916-22.
58. Hunnicutt, B.J., et al., *A comprehensive excitatory input map of the striatum reveals novel functional organization.* Elife, 2016. **5**.
59. Ramanathan, S., et al., *Synaptic convergence of motor and somatosensory cortical afferents onto GABAergic interneurons in the rat striatum.* J Neurosci, 2002. **22**(18): p. 8158-69.
60. Barack, D.L. and J.W. Krakauer, *Two views on the cognitive brain.* Nat Rev Neurosci, 2021. **22**(6): p. 359-371.
61. Langdon, C., M. Genkin, and T.A. Engel, *A unifying perspective on neural manifolds and circuits for cognition.* Nat Rev Neurosci, 2023. **24**(6): p. 363-377.

62. Stringer, C., et al., *Spontaneous behaviors drive multidimensional, brainwide activity.* Science, 2019. **364**(6437): p. 255.

Acknowledgements
We would like to thank T. Akam, M. Rothwell and B. Micklem for methodological and technical assistance and D. Dupret for feedback during the project. **Funding:** This work was supported by The Clarendon Fund and Mary Somerville Clarendon Graduate Scholarship in conjunction with the Department of Experimental Psychology (award SFF1819_CB2_MSD_1196514 to OH), the Medical Research Council UK (award MC_UU_00003/6 to AS), the Oxford National Institute for Healthcare Research (NIHR) Biomedical Research Centre to S.G. and the Wellcome Trust (Sir Henry Wellcome Fellowship 209120/Z/17/Z to CGM; 202831/Z/16/Z and 214314/Z/18/Z to MEW). **Author contributions:** Conceptualisation: OH, MEW, AS and SGM; Technical development: OH, RT and CGM; Investigation: OH, BP and IG; Analysis: OH, BP and IG; Resources and funding acquisition: AS and MEW; Writing - original draft: OH, IG, and AS; Writing - Reviewing & Editing: OH, IG, BP, RB, CGM, HC, SGM, MEW, AS; Supervision: AS, MEW, SGM and HC. **Competing interests**: The authors declare no competing interests.

Materials and Methods

Subjects

All procedures were carried out in accordance with the UK Animals (Scientific Procedures) Act (1986) and its associated guidelines. We used 12 group-housed Lister Hooded naive rats (Charles River, U.K.) weighing 300-390g at the beginning of training. Rats were maintained on a twelve-hour light/dark cycle (lights on at 7 AM). During testing periods, rats were food restricted to a target weight range of 85-90% of their free-feeding weight. Water was available *ad libitum* in their home cages. Following implantation with multielectrode drives, animals were singly housed.

Behavioural apparatus

The operant box was designed in Fusion360 (Autodesk) and consisted of two 17.6cm diameter custom-designed 3D-printed plastic semi-circular arenas, connected by a 130 cm-long Perspex corridor with walls on either side. The corridor had 6 equidistantly placed infra-red (IR) light-gates to register the rats' position and 3 equidistant slots for barriers to increase the travel cost between the ends of the corridor. Each arena housed a custom designed 3D-printed plastic nose-poke and reward port, positioned symmetrically on either side. Nose-poke and reward ports contained a infrared light-gate and a white LED house light. Reward ports were connected with 20-mg food pellet dispensers (Med-associates) fitted outside the walls of the arenas using medical-grade silicone tubing. Each reward port housed a moveable floor that was able to obstruct the port. The electronics were controlled by a Teensy 3.6 microcontroller that ran on custom-written firmware.

Behavioural paradigm

Rats initiated trials by poking and holding in the nose-poke port of the starting arena ("pre-cue"). The starting arena was randomly allocated on the 1st trial. After a pre-cue period (0.3-0.4s), an auditory cue ("offer") was again presented from the distal end of the apparatus to signal the available reward at end of the track from where they initiated. The available reward was pseudo-randomly varied trial-by-trial (4 sizes: nil, low, mid, high, corresponding to zero, one, three or six 20mg sucrose pellets), signaled by a distinct auditory cue. The cue-reward associations were counter-balanced across subjects. Rats were required to maintain nose-poking for another 0.6-0.8s until the onset of a visual Go cue, signaled by a white light LED. Premature exiting of the nose-poke either during the pre-cue or offer period would elicit a 5-s error time-out, indicated by a flashing house-light in the nose-poke, followed by a 5s ITI. If rats exited prematurely during the auditory cue, the same cue would be repeated on the next nose-poke attempt. A grace period of up to 100ms during nose-poking - where nose-poke exits did not count towards an error - allowed for minor adjustments in body position. To accept the offer, rats needed to enter the corridor (cross a light-gate 24cm away from the arena) within a time window; otherwise, that offer was rejected. The cost of pursuing reward was pseudo-randomly varied across blocks of 17-23 trials (3 levels: low, moderate, high), implemented using varying numbers of barriers (0-2 or 1-3 in implanted or non-implanted animals,

respectively) in the corridor. The barriers were 0.5 cm thick and were 12 cm and 18 cm tall at their lowest and highest points respectively, with a 45-degree slant mid-way between to generate a more effortful zig-zag of running between the ends of the corridor.

Behavioural training

Rats were handled daily for a week prior to training. Training consisted of 5 steps in both versions of the task, summarized in table 3.

Step 1 ("magazine training") consisted of rats being introduced into the apparatus with bedding from their home cages and receiving 20 rewards in each food dispenser, which they were free to consume in 20 minutes.

In Step 2 ("cue-reward association training") - entry of the starting arena as detected by entry/presence in the arena of latest reward delivery would trigger one of 4 auditory cues, predicting the magnitude of the promised reward in the distal reward port. The offer cue was played until the animal made an entry into the distal reward port. Next, the reward port LED was activated for a 1s pre-reward delay, followed by delivery of the reward corresponding to offer cue. 2.5s after reward delivery, the moveable reward port floor was elevated in that same arena, resulting in the blocking of that port to indicate that reward would not be available there on the next trial. After the rats had reached a criterion of initiating 120 trials in £ 60 minutes, rats were taken to step 3.

In step 3 ("nose-poke training"), in v1, rats had to learn to sustain a nose-poke with a jittered 1-1.2s length in order to elicit the offer cue. If rats sustained a 0.3-0.4s-long nose-poke ("pre-cue"), one of 4 auditory cues would be played. Importantly, rats were not free to leave until another 0.6-0.8s later ("offer") once a visual Go cue, signalled by a white light LED in the nose-poke port, had turned on. This was implemented to ensure a movement-free period following offer presentation for electrophysiological analyses. The pre-cue nose-poke requirement was incremented from 0.1s to 0.4s in increments of 25 ms after every successful trial. Once animals had learned to sustain successful pre-cue nose-pokes, the cue nose-poke requirement was incremented from 0.0s to 0.8s in steps of 20 ms after every successful trial. Once animals had learned to complete 120 trials with the full 1.2s nose-poke requirement, jitters of 0.1s ("pre-cue") and 0.2s ("offer") were introduced. In order to allow for small changes in body posture that may have caused the nose-poke sensor to stop registering an entry, a 0.1s grace period was introduced which did not result in a nose-poke error. Importantly, if a nose-poke was interrupted for more than 0.1s (the grace period) during the "offer" window, the same offer cue would be repeated during the next trial. This was not an intentional feature, but rather the side-effect of using a pre-determined sequence of offers. Once rats had reached a criterion of initiating 120 trials within 60 minutes, they were taken on to step 4.

In step 4 ("barrier training") rats needed to nose-poke to elicit offer cues as outlined previously, but now needed to additionally learn to scale barriers when traversing the corridor between the arenas. Barriers were added in one by one in blocks of 40 trials, starting from 1 barrier and

going up to a maximum of 3 barriers. Once rats had successfully managed to complete 120 trials within 60 mins, they were taken on to step 5.

In step 5 ("final training for non-implanted animals") *r*ats were required to either accept or reject a distal reward offer available at the other side of the corridor (fig. 2). After the onset of the Go cue, to accept the offer, rats needed to enter the corridor within a 3s time window as determined by a light-gate placed near the border of the arena; otherwise, that offer is rejected. The effort cost which was fixed across blocks of ~20 trials (3 levels: low – 1 barrier; moderate – 2 barriers; high – 3 barriers; table 3). Upon corridor entry, they were allowed a further 5s to traverse the corridor and reach the distal reward port. To ensure average reward rate was minimally affected by rats' accept/reject choices, ITIs were adjusted daily on 'accept' trials by subtracting the previous day's cohort median run time.

In step 5 ("post-surgery re-training"), the corridor entry and traverse times were increased to 15s and 10s respectively post-surgery to account for the additional constraints on implanted rats with a recording tether. Additionally, to avoid collisions between the implant and the barriers, the number of barriers per each level of effort was reduced by 1. Rats were considered ready for the recording experiment once they managed to complete 120 trials within 60 minutes.

Outcome devaluation experiment

Following standard testing, outcome devaluation was performed on a subset of rats (n=6, first cohort) in a within-subjects regular Latin square design. In devaluation sessions, subjects were placed in a solitary "feeding" cage with *ad libitum* lab chow for 1h prior to the start of the testing session. Rats were placed in a solitary "control" cage for 1h on control sessions. The outcome devaluation lasted 3 days altogether, out of which days 1 and 3 were testing days and day 2 was a baseline training day.

Multielectrode drive

For single unit recordings we used a bespoke driveable stereotrode device weighing ~20g. The implant was comprised of a lid which housed a 128-pin connector designed for Intan 64-channel head-stages (Intan Technology), a plastic corpus that housed the electrode movement mechanisms, glass tubing and a base. All components were assembled in house. For assembly, the elements of the corpus were first 3D printed and post-processed with UV radiation and propanol. Second, holes were drilled in the corpus to allow threading of electrode wires and the driving mechanisms. Third, driving support mechanisms were created using metal rods, plastic shuttles and screws. Fourth, glass tubing was introduced into the driving mechanisms as electrode supports. Fifth, strands of 50-micron diameter tungsten wire (California Fine Wire) were twisted and heated to create 4- and 6-strand electrodes (tetrodes and hextrodes, respectively). Sixth, these were threaded through the glass tubing and the rest of the implant corpus. Seventh, ground wire stubs (connected to ground wires on the animal during surgery, see below) were introduced into the implant using a dedicated set of stationary holes. Finally, the implant was sealed and the electrodes cut to length by hand.

Three of the subjects (gh001, gh009, gh010) reported here received an implant that targeted all 5 regions, whereas one subject (oh011) received a modified drive without the VP electrodes (table 1). Electrodes in the MO/VO, ACC, and VP were arranged as a row along one sagittal plane. DMS electrodes were as rows along two sagittal planes and STA electrodes were arranged in triangular fashion around a centre point. The ML and AP positions of each electrode in the drive were planned around a centre-point co-ordinate obtained as an average between readings in two different versions of a commonly used rat brain atlas [1, 2] that visually offered the most exposure in each structure, while also avoiding the ventricles (Fig. S2a). Relative electrode positions in the AP and ML dimensions were planned around centroid co-ordinates.

Table 1 | Overview of the number of electrodes and channels in each subject and targeted recording area.

	Number of electrodes (channels) in targeted area				
	MOFC	ACC	DMS	VP	STN
oh011	4 (32)	4 (16)	5 (30)	0 (0)	3 (18)
gh001	4 (32)	4 (16)	4 (24)	2 (12)	3 (18)
gh009	4 (32)	4 (16)	5 (30)	2 (12)	3 (18)
gh010	4 (32)	4 (16)	5 (30)	2 (12)	3 (18)

Implantation of multielectrode drives for recordings in behaving rats

Electrode implantations were performed in a subset of expert animals (n=4; table 1). Surgeries were performed under isoflurane anaesthesia (5% for induction, 1.5-2.5% maintenance) and oxygen (2L/min). Subcutaneous local anaesthetic (Marcaine, 2mg/kg, 2.5 mg/mL) and a non-steroidal anti-inflammatory drug (Metacam, 1mg/kg, 5mg/mL) were administered at the beginning of each surgery. Post-operative opioid analgesia (Vetergersic, 0.3 mg/mL, 0.03 mg/kg) was administered subcutaneously for three consecutive post-operative days.

Craniotomies were created above each of the targeted areas and *dura mater* thoroughly removed from the visible brain surface. The drive was aligned on the animals' heads by using the centre-point of the STA electrode bundle as a reference. Next, the drive was lowered to target depth on a stereotaxic frame (David Kopf Instruments) based on the DV reading of the STA electrode bundle from brain surface. The implant was then secured to skull-attached screws using adhesive dental cement and dental acrylic. Recordings were referenced to two supra-cerebellar screws which were connected to the implant during surgery. Neuronal recordings began at least 2 weeks after surgery. Electrode placement was confirmed using histology.

Histological data analysis

To verify electrode placements, after the experiments were completed, animals were deeply anaesthetised with isoflurane (5%) and intraperitoneal pentobarbital (3 mL, Pentoject, 200 mg/mL and perfused trans-cardially with phosphate buffered saline (PBS) followed by fixative solution (paraformaldehyde dissolved in PBS, 4%, wt/vol). For post-mortem fixation, the sectioned heads were further submerged in fixative for another 24h, following which brains

were extracted and washed 3 times in PBS and stored in a PBS solution containing 0.02% sodium azide until sectioning. Brains were sectioned at 0.1mm thickness using a vibratome (Leica), immunohistologically stained and mounted in Vectashield. Unfiltered "bright-field" epifluorescence microscope images were using 5x magnification and Zen Blue Pro (Zeiss).

Trode localisation and histological verification

To estimate all the possible recorded areas and exclude cells yields from neighbouring regions, we (1) estimated each identifiable electrode tract's DV end-depth from our epifluorescence imaging data, (2) averaged this across all identifiable electrode tracts to get an average end-depth, (3) estimated each day's targeted structure based on a log of putative lowering distances averaged across electrodes within a region. We did not restrict the cell yield for the STA based on histological data. For other structures we restricted our cell yields to only the MO or VO for MO/VO, Cg1 and Cg2 for the ACC, CPu for the DMS; and VP only for the VP.

Electrophysiology data acquisition and processing

Wideband and accelerometer data were recorded, amplified and digitised on 2 tethered Intan 64-channel headstages at 20,000 Hz using an Intan acquisition board (Intan Technologies). No hardware filtering was applied during recording. Files were saved in binary format. Each electrode was lowered by at least 125 mm up to 2h ahead of each recording session to ensure identification of unique units every day.

Behavioural data analysis

All data were analysed in Python. A binomial mixed effects model using the pymer4 library [3] was used to regress the following fixed effects onto trial-by-trial choice:

- *reward*, coded as 0, 1, 3 or 6
- *effort*, coded as 1, 2, or 3 (0, 1, or 2 in implanted animals)
- *reward x effort interaction*, coded as the multiplication of those vectors
- *reward count*, coded as the number of pellets earned since the start of the session.

To safeguard against type I errors, we aimed to fit a maximal random effects structure whenever possible by modelling each predictor as a random, in addition to a fixed, slope [4]. To account for day-by-day and subject-by-subject covariation in the data, random slopes were paired with a random intercept for either the testing day or the subject ID. Whenever model fitting failed to converge, random effects that fit close to zero variance were removed to simplify the random effects' structure. The resulting decision model for the non-implanted cohort had the following formula:

accept = reward + effort + reward x effort + reward count + (reward| date) + (reward | subject),

while the implanted rat decision model was formulated as follows:

accept = reward + effort + reward x effort + reward count + (reward + effort + reward count| date) + (reward + effort + reward count | subject)

For the outcome devaluation experiment, we added pre-feeding and its interaction with reward and effort as an additional predictor, resulting in the following formula:

accept = reward + effort + pre-feeding + reward x effort + reward x pre-feeding + effort x pre-feeding + reward x effort x pre-feeding (1| date) + (1| subject)

All predictors were standardised to facilitate side-by-side comparison. P-values were obtained using the likelihood ratio test.

Firing rate modulation during task execution

For visualising task event-locked firing rates (Fig. 2b), spiking counts were quantified in 10ms non-overlapping bins sampled in 10 ms increments -400ms to +400ms around the initiation of the nose-poke (termed "Pre-Cue"), -400ms to +400ms around the onset of the offer cue (termed "Offer"), -500ms to +500ms around the onset of the Go cue (termed "Go Cue") and -500ms pre- to +1000ms following nose-poke exit (termed "Action"). Data were subsequently averaged across trials of the same magnitude of offered reward, the required effort and prospective decision, each time averaging across the other two variables. Peri-stimulus time histogram (PSTH) data were smoothed with a gaussian filter of standard deviation 5.0 for display purposes.

To facilitate the alignment of data on the task (Fig. 3; Fig S3-4), time on trial was transformed into a uniform bin space by dividing the Pre-Cue epoch (-400 ms before Pre-Cue up to Offer) into 16 bins, the Offer epoch (from Offer up to Go Cue) into 16 bins, the Go Cue epoch (from Go Cue up to Action) into 13 bins and the Action epoch (from Action to corridor entry on accepted trials or each animal's daily median accepted trial corridor entry latency on rejected trials) into 18 bins. Firing rates were quantified in 50 ms bins centered around each such bin and z-scored relative to a concatenated series of ITI data from the ongoing effort block.

Individual neuron regression analyses

General linear models (GLMs) were employed to quantify the influence of reward, effort and prospective decision on z-scored individual unit firing rates. For this, the loading (b) of each variable on z-scored firing rates were obtained using the closed-form least-squares solution:

$$b = (X^T X)^{-1} X^T Y$$

where X is a design matrix of shape [trials x variables] and Y is a vector of z-scored firing rates of shape [trials x 1] from one recording day. The 4 regressors used in the design matrix X were constructed as follows:

- Intercept: a vector of ones
- Reward: a vector containing the values 1, 3 or 6 corresponding to small, medium or large reward trials.

- Effort: a vector containing the values 1, 2 or 3 corresponding to low, moderate or high effort trials.
- Decision: a vector comprised of the values -0.5 and +0.5 corresponding to rejected or accepted trials.

All regressors except the intercept were z-scored to themselves before fitting the models and obtaining coefficient loadings. From these GLMs, coefficients of partial determination (CPD), reflecting the unique variance explained by variable, were obtained as follows:

$$\mathrm{CPD_i} = 100 \times \left(\frac{\mathrm{SSE_{\sim i}} - \mathrm{SSE_i}}{\mathrm{SSE_{\sim i}}}\right)$$

where $\mathrm{SSE_{\sim i}}$ and $\mathrm{SSE_i}$ represent the sum of squared errors of models that, respectively, do not or do include the regressor of interest. Using CPD ensured that the encoding strength for a given variable was not secondary to another (e.g. higher firing rate/encoding on accept trials due to more of these trials having higher prospective reward). To identify significant coding, permutation testing was used. For each neuron and decision variable, 1,000 shuffled datasets were generated by randomly shuffling the trial indices within each recording day. Such shuffling preserves potential autocorrelations between task variables and firing rates and is thus an appropriate estimate of the chance-level outcome. Individual neuron regression coefficient p-values were calculated as the fraction of (1,000) permuted CPD values which exceeded a neuron's non-permuted CPD.

Significant neurons (Fig. 2b; Fig 3c: Fig. S3-4) were identified on the basis of CPD p-values less than 0.05 across in one time bin abutting the onset of the offer cue and entry into the corridor (on accepted trials) or the session median corridor entry latency (on rejected trials). Individual neuron examples in Fig. 2b were selected if they met this criterion and displayed some consistency in the discrimination response over time based on visual assessment. Instantaneous population CPD p-values (Fig. 3a) were calculated as the fraction of permutations for which the (1,000) shuffled population average CPDs were larger than the non-permuted population mean CPD. The Benjamini-Hochberg (BH) procedure was applied to correct for false discovery rate across tested time points.

The proportions of significantly coding neurons between regions were compared pair-wise between areas using the z-test of proportions with BH correction for the number of unique comparisons.

Quantification of CPD phasicness (fig. 3a insets)

To characterise the degree to which individual unit CPDs magnitudes were temporally localised vs distributed across the trial timeline, we calculated *phasicness.* This is related to *sparsity* as defined by previous studies [5, 6] and indexed the extent to which the CPD time

series for a population of neurons was dominated by a small number of time points. For each neuron in a given area, it was defined as follows:

$$\frac{\sqrt{n} - \sum |v_i|}{\sqrt{n} - 1}$$

where n is the length of the CPD vector (i.e., the number of time points) and v_i is the CPD vector's i^{th} element. For this analysis, the input CPD time series was a unit vector where the sum of the squares of all elements added up to one. Phasicness distributions were compared between brain areas using the two-sided Mann-Whitney U test with BH correction for the number of pairwise comparisons. All Mann-Whitney U test results reported in this work as expressed as a Z-score obtained as:

$$z = \frac{U - \frac{n_A n_B}{2}}{\sqrt{\frac{n_A n_B (n_A + n_B + 1)}{12}}}$$

where U is the Mann-Whitney U test statistic and n_A and n_B represent the sample sizes in the two comparison groups.

Quantification of latency to peak coding response and phasic responding properties (fig. 3a, lower panel)

To quantify the latency to peak coding response, we calculated the time bin of maximal CPD magnitude for each neuron and plotted the distribution across all neurons as violin plots using the python Seaborn library. To evaluate whether these peak CPD latencies tended to co-occur close-by in time (phasic responding) more than would expected by chance, we compared these values with a null-distribution formed by bootstrapped sampling of all observed unique latency values 1,000 times with replacement. The null and real distributions were compared using the Mann-Whitney U test and the false-discovery rate corrected for by using the BH correction.

Co-activity analysis

Assembly patterns were determined using the combination of principal component analysis and independent component analysis (PCA-ICA) as detailed earlier [7]. This analysis was conducted using neurons from all structures simultaneously. This allowed for the unsupervised extraction of assembly patterns which could be contained entirely within an anatomical structure or could span multiple brain regions. The spike trains for all units during time periods where the animal was actively engaged with the task were binned at 25ms, z-scored and concatenated into a *n* x *B* matrix, where *n* was the number of neurons and *B* was the number of 25ms bins. The element of this matrix at position (i, b) was the z-scored spike count of neuron i in bin b.

PCA was then applied to the z-scored spike count matrix. The number of significant patterns embedded in the z-scored spike counts was determined using the Marcenko-Pastur law. This is an analytical result from random matrix theory, which bounds the magnitude of the eigenvalues from PCA that can occur due to chance. Principal components with eigenvalues

above the upper limit of the bound set by the Marcenko-Pastur law were classified as significant. Principal components with an eigenvalue greater than this threshold captured more of the correlation which existed between the binned firing of neurons than should occur if all neurons spiked independently of one and other [5].

The z-scored spike count matrix was then projected onto the significant principal components and ICA was applied to the resulting matrix. The output of ICA was an unmixing matrix. This unmixing matrix was then projected back onto the original basis, where each dimension represents the binned z-scored spike count of an individual neuron. The columns of the resulting matrix represent assembly patterns. These assembly patterns contain a weight for each neuron. The greater this weight is, the greater the contribution the neuron makes to the identified coactivity pattern [5].

There are several advantages to using the combination of PCA and ICA as opposed to using either technique alone [5, 7]. For instance, PCA has been found to group several assembly patterns into the first principal component. PCA also artificially constrains identified assembly patterns to be orthogonal to each other, meaning that a high weighting of a neuron in one assembly pattern limits the weighting of that neuron is subsequently identified patterns. Finally, PCA is only based on pairwise correlations and cannot identify higher order correlations. However, if ICA was applied directly to the z-scored binned spike count matrix, one could identify the same number of assembly patterns as there are neurons. This could result in the extraction of spurious patterns. To address these caveats, ICA is applied to the z-scored binned spike count matrix following dimensionality reduction using PCA.

In order to determine the expression of assembly patterns across time, the outer product of each assembly pattern was determined [5].The expression strength (R) of assembly pattern k can be determined as follows:

$$R_k = z(t)^T P_k z(t)$$

where P_k is the outer product of assembly pattern k with elements in the main diagonal set to 0 , and z(t) is a column vector containing the smoothed z-scored firing rate of neurons at time t. The main diagonal of P_k must be set to zero to ensure that assembly pattern expression strength cannot be driven solely by isolated increases in the firing rate of individual units. The instantaneous firing rate of neurons in $z(t)$ were computed by convolving the spike train of neuron with a gaussian kernel and then z-scoring the resulting timeseries for each neuron. The standard deviation of the gaussian kernel was set to 25 / $\sqrt{12}$ ms to match the 25ms bin used for PCA-ICA [8]. Convolution with a gaussian kernel allows for greater temporal resolution when compared to the 25ms bins which were used to define the assembly patterns via PCA-ICA.

Linear and logistic regression was conducted using either the python module statsmodel or in custom written pytorch functions. The general strategy utilized in the linear regressions was to take an epoch in each trial, determine the value of variables such as the spike rate or the mean coactivity in this epoch and use this to predict reward, effort or choice outcome across trials, or vice versa.

For identifying coding ensembles, the mean expression strength of an ensemble was calculated from an interval starting with offer cue onset and ending with either corridor entry or, in instances where the animal did not enter the corridor, the median duration taken for corridor entry on that recording session. The mean expression strength was then fitted as the dependent variable in a linear regression with reward, effort and choice outcome as independent regressors. Significance was determined by comparing the CPD of reward, effort and choice to that of the $95th$ percentile of the null distribution where reward and effort levels or choice outcomes were shuffled across trials. After identifying ensembles with significant modulation by the level of reward, effort or choice, we probed the time course of this modulation. This was achieved by fitting regression models with reward, effort and choice outcome as independent variables to the smoothed assembly pattern expression strength at different time points. Smoothing was achieved by convolution with a Gaussian with a standard deviation of $2s\ /\ \sqrt{12}$. The CPD for reward, effort and choice was compared to that of linear models fitted to the assembly pattern expression strength in jittered spike trains.

Data Accessibility Statement: Data is not available at this Preprint stage due to ongoing work.

References

1. Paxinos, G. and C. Watson, *The Rat Brain Atlas in Stereotaxic Coordinates*. 4th ed. 1998: Elsevier.
2. Paxinos, G. and C. Watson, *The Rat Brain Atlas*. 6th ed. 2006: Elsevier.
3. Jolly, E., *Pymer4: Connecting R and Python for Linear Mixed Modeling.* Journal of Open Source Software, 2018. **3**(31).
4. Barr, D.J., et al., *Random effects structure for confirmatory hypothesis testing: Keep it maximal.* J Mem Lang, 2013. **68**(3).
5. van de Ven, G.M., et al., *Hippocampal Offline Reactivation Consolidates Recently Formed Cell Assembly Patterns during Sharp Wave-Ripples.* Neuron, 2016. **92**(5): p. 968-974.
6. McHugh, S.B., et al., *Adult-born dentate granule cells promote hippocampal population sparsity.* Nat Neurosci, 2022. **25**(11): p. 1481-1491.
7. Lopes-dos-Santos, V., et al., *Neuronal assembly detection and cell membership specification by principal component analysis.* PLoS One, 2011. **6**(6): p. e20996.
8. Kruskal, P.B., et al., *A binless correlation measure reduces the variability of memory reactivation estimates.* Stat Med, 2007. **26**(21): p. 3997-4008.