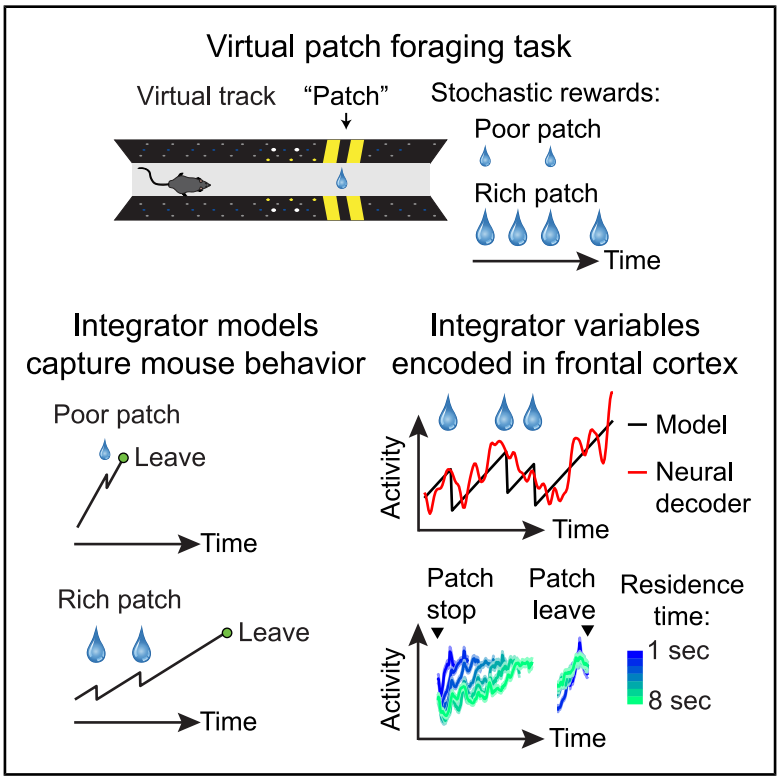
Neuron

Article

Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics

Graphical abstract



Highlights

- Mice in a virtual foraging task weigh time and reward to decide when to leave patches
- Value integration, scaled by slowly varying levels of patience, captures behavior
- Frontal cortex neurons show reward-suppressed ramps encoding decision variables
- Neural ramps suggest an integration mechanism guiding value-sensitive foraging

Authors

Michael Bukwich,
Malcolm G. Campbell,
David Zoltowski, ..., Jan Drugowitsch,
Scott W. Linderman, Naoshige Uchida

Correspondence

m.bukwich@ucl.ac.uk (M.B.),
mgcampb@fas.harvard.edu (M.G.C.),
uchida@mcb.harvard.edu (N.U.)

In brief

Bukwich and Campbell et al. show that mice integrate elapsed time and reward intake, scaled by a latent patience variable, to decide when to leave virtual "patches." Frontal cortex ramping activity matches the integration process, revealing a potential mechanism for how the brain converts dynamic evidence into value-based foraging decisions.

Bukwich et al., 2025, Neuron *113*, 1–18
October 15, 2025
https://doi.org/10.1016/j.neuron.2025.07.008

CellPress

Neuron

CellPress

Article

Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics

Michael Bukwich,[1,2,9,11,*] Malcolm G. Campbell,[1,2,11,*] David Zoltowski,[3,4] Lyle Kingsbury,[1,2] Momchil S. Tomov,[1,2,5,10] Joshua Stern,[1,2] HyungGoo R. Kim,[1,2,7,8] Jan Drugowitsch,[6] Scott W. Linderman,[3,4] and Naoshige Uchida[1,2,12,*]

[1]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA
[2]Center for Brain Science, Harvard University, Cambridge, MA 02138, USA
[3]Department of Statistics, Stanford University, Stanford, CA 94305, USA
[4]Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA 94305, USA
[5]Department of Psychology, Harvard University, Cambridge, MA 02138, USA
[6]Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA
[7]Center for Neuroscience Imaging Research, Institute for Basic Science, Suwon 16419, Republic of Korea
[8]Department of Biomedical Engineering, Sungkyunkwan University, Suwon 16419, Republic of Korea
[9]Sainsbury Wellcome Centre, University College London, London W1T 4JG, UK
[10]Present address: Motional AD LLC, Boston, MA 02210, USA
[11]These authors contributed equally
[12]Lead contact
*Correspondence: m.bukwich@ucl.ac.uk (M.B.), mgcampb@fas.harvard.edu (M.G.C.), uchida@mcb.harvard.edu (N.U.)
https://doi.org/10.1016/j.neuron.2025.07.008

SUMMARY

Patch foraging is a ubiquitous decision-making process in which animals decide when to abandon a resource patch of diminishing value to pursue an alternative. We developed a virtual foraging task in which mouse behavior varied systematically with patch value. Behavior could be explained by models integrating time and rewards antagonistically, scaled by a slowly varying latent patience state. Describing a mechanism rather than a normative prescription, these models quantitatively captured deviations from optimal foraging theory. Neuropixels recordings throughout frontal areas revealed distributed ramping signals, concentrated in the frontal cortex, from which multiple integrator models' decision variables could be decoded equally well. These signals reflected key aspects of decision models: they ramped gradually, responded oppositely to time and rewards, were sensitive to patch richness, and retained memory of reward history. Together, these results identify integration via frontal cortex ramping dynamics as a candidate mechanism for solving patch-foraging problems.

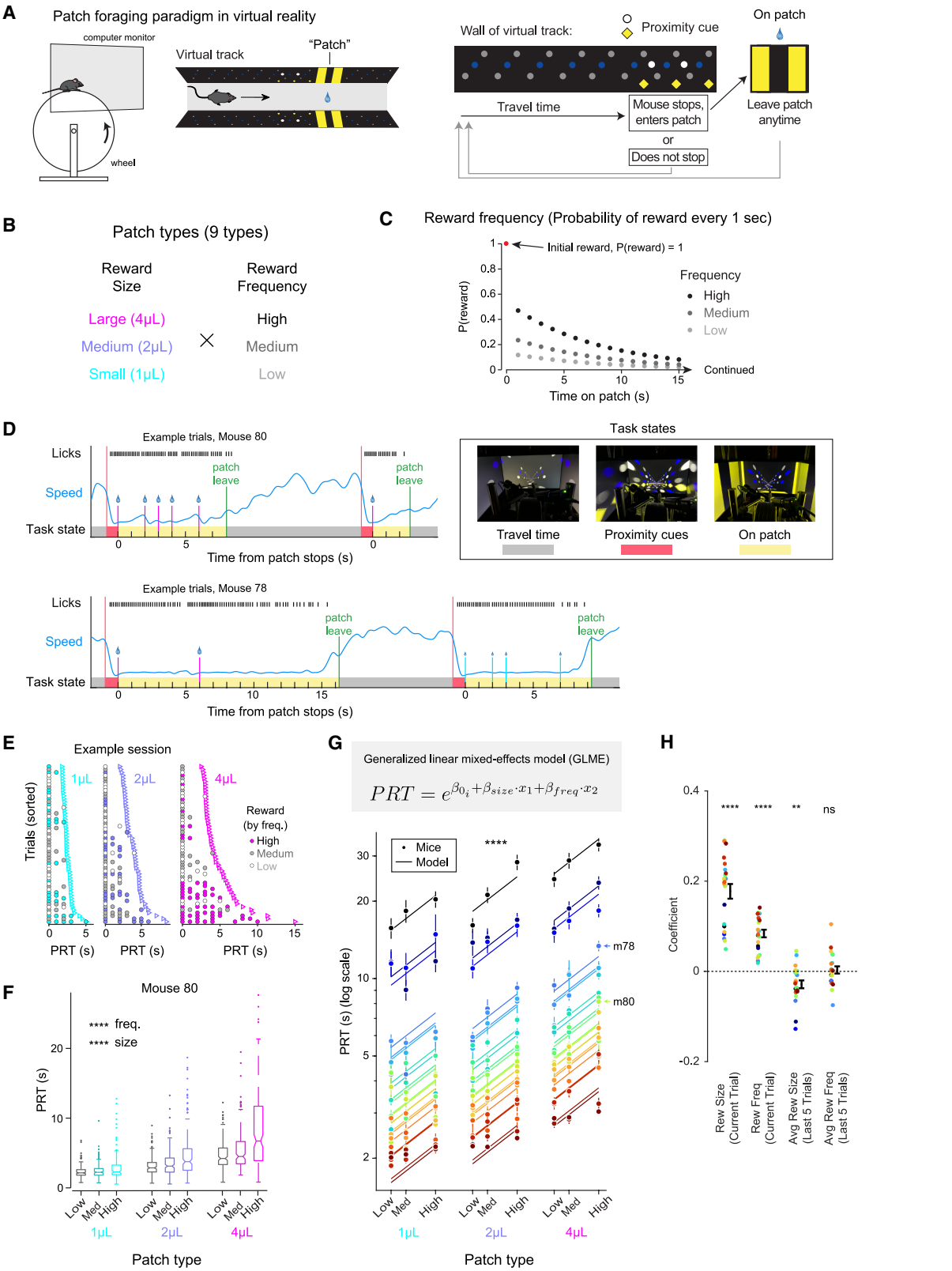
INTRODUCTION

Time is precious, so it should be spent judiciously. Animals regularly face this dilemma while patch foraging, during which they must decide when to leave locations with depleting resources to search elsewhere.[1,2] The marginal value theorem (MVT) dictates foragers should leave a patch when the instantaneous resource intake rate drops to the average rate of the environment.[3] MVT specifies optimality, but (1) whether animal behavior conforms precisely to MVT remains debated, and (2) biological processes underlying patch-leaving decisions remain unresolved.[1,4–6]

While some behavioral observations qualitatively support MVT,[7–12] animals often deviate from it.[5,13–18] One source of discrepancy is that MVT requires complete knowledge of instantaneous and overall reward rates, which is impracticable.[19,20] Relatedly, while MVT defines optimal behavior, it does not clarify the underlying decision processes. One candidate decision model is a "giving-up time rule" in which animals leave a patch after a predetermined amount of time elapses following resource intake.[14,21–24] Others employ integrate-to-threshold processes,[25] akin to drift-diffusion models accounting for perceptual decisions.[26–32] During foraging, these models track a decision variable (DV) integrating elapsed time and resource intake in opposite directions, triggering patch leaving when the DV reaches threshold. Competitive interactions between time and resource intake produce lengthier stays in richer patches and can support MVT-optimal decisions under idealized conditions.[33]

Previous studies implicate the frontal cortex in foraging decisions.[34] In a saccade-based task emulating patch foraging, neurons in the macaque dorsal anterior cingulate cortex (ACC) increased their pre-saccadic transient responses up to a certain level before patch leaves.[10] Human studies similarly identified

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008



(legend on next page)

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

regions in the frontal cortex tracking value during foraging-like decisions.[35,36] Rodent frontal cortex areas,[37,38] including ACC,[12,39] orbitofrontal cortex (OFC),[39,40] and secondary motor cortex (M2),[38,41] influence when-to-leave decisions. However, the mechanistic basis of scaling leave times with patch value remains unknown.

We therefore paired a virtual patch-foraging task with Neuropixels recordings throughout the frontal cortex. Our naturalistic decision-making paradigm balanced complexity with experimental control, facilitating ethologically relevant behavior while allowing us to probe its neural and algorithmic basis.

RESULTS

Virtual foraging task for mice

We developed a patch-foraging task in virtual reality (Figure 1), wherein head-fixed mice ran on a treadmill to traverse a linear corridor. After running a fixed distance, a visual "proximity cue" appeared, indicating the availability of a resource patch. If mice stopped running while the proximity cue was on, a patch trial was initiated with a visual cue, with water rewards delivered stochastically while mice remained on the patch. Patches delivered water droplets of one of three sizes (1, 2, or 4 μL), which remained consistent within a patch. Rewards were delivered probabilistically each second on the patch, with a starting probability (0.125, 0.25, or 0.5) exponentially decaying with time constant $\tau = 8\ s$ (Figures 1B and 1C). To encourage patch stopping, the first reward was delivered deterministically on all trials (at $t = 0$). While mice remained on patches, additional probabilistic rewards might occur at 1-s increments. Mice could leave patches anytime by traveling a fixed distance. By utilizing reward schedules with diminishing probability, rather than, for example, directly observable decays in reward size, mice were required to make decisions under uncertainty and were thus prevented from using heuristic strategies such as leaving once reward size dropped below some threshold.

Patch residence times are a scaled function of reward statistics

Mice consistently pursued water rewards (Figures 1D and S1B–S1D), showing higher patch residence time (PRT) in patches with larger and/or more frequent rewards (Figures 1E–1G). Alongside this value sensitivity, MVT predicts animals should leave patches sooner in richer environments. Concordantly, we found negative correlations between reward history and current PRT (Figure 1H) extending back at least 5 trials (Figure S1I).

There was large variability in willingness to wait across mice, yet their relative PRTs across patch types were approximately co-linear after log-scaling (Figure 1G, mice), indicating reward effects scale systematically with mean PRT (Figure S2E). We therefore fit a log-link generalized linear mixed-effects model (GLME) with fixed effects for reward size and frequency and a mouse-specific random intercept and found this model captured mean PRTs across conditions ($\beta_{size} = 0.168$, $\beta_{freq} = 0.103$, both $p < 0.0001$; Figure 1G).

We additionally observed substantial within-animal variability in PRTs over time, across (Figures 2A and S1F) and within sessions (Figure 2B). PRTs on successive patches were correlated (Figure 2B), suggesting a slowly fluctuating latent state, which we henceforth call "patience," describing varying mean levels of willingness to wait. Patience could not be explained by reward rate fluctuations, which negatively influenced future PRTs (Figures 1H and S2A–S2D).

We therefore used PRTs from successive patches to estimate subjects' latent degree of patience across trials (Figures 2C and 2D). We then transformed our GLME into a generalized linear model (GLM), replacing mouse-specific random intercepts with a regressor for latent patience (Figure 2E). With just four free parameters shared across mice, the GLM captured mean PRTs across patch types (Figure 2E) and robustly predicted single-trial PRTs (population, median $R^2 = 0.45$, Figure 2F; example mouse, Figure S2J). Importantly, GLM predictions reflect multiplicative interactions between latent state and reward statistics, demonstrating latent state modulates, rather than replaces, value-based performance.

Variability relative to normative MVT predictions

While mice exhibited qualitative agreement with MVT predictions (Figure 1H), we found systematic quantitative deviations from a baseline MVT model. MVT dictates that, if patches deplete monotonically, optimal foragers should leave once the instantaneous reward rate drops to the average

Figure 1. Mice foraging times qualitatively match MVT

(A) Schematic of virtual patch-foraging task.

(B) Combinations of reward sizes and frequencies yield nine patch types (Figure S1A).

(C) Probability of reward after 1-s intervals per frequency condition.

(D) Example trials from two mice. Mice stop in response to proximity cues to enter patches and receive stochastic water rewards. Brightness was increased for task state photos.

(E) Reward deliveries (circles) and patch-leave times (triangles) from example sessions grouped by patch reward size. Trials are sorted in ascending order of PRT from top to bottom. Circles' coloring indicates underlying frequency condition.

(F) PRT per patch type from an example mouse. Column groupings separate patches by reward size. Within columns, patches are split by reward frequency ($p < 0.0001$ for size and frequency, GLM). Box: median, 25%, and 75%; whiskers = 5%–95%; points = outliers.

(G) A generalized linear mixed-effects model (GLME) with log-link function predicts PRTs across conditions. Points indicate mouse PRTs with error bars showing SEM. Lines represent GLME fits. Colored by mouse ID "i."

(H) Coefficients from linear regression of normalized PRT on current trial reward size/frequency and average reward size/frequency of the preceding 5 trials. Points show individual mice, and black error bars show mean ± SEM over mice ($n = 22$). Coefficients on current reward size/frequency were positive, whereas coefficients on previous reward size were negative (current reward size coefficient = 0.18 ± 0.02, t test versus zero $p = 4.2 \times 10^{-10}$; current reward frequency coefficient = 0.084 ± 0.0084, $p = 2.2 \times 10^{-9}$; previous reward size coefficient = −0.029 ± 0.0087, $p = 0.0033$; $n = 22$ mice). Coefficient on previous reward frequency (a noisier indicator of patch quality) did not differ significantly from zero (previous reward frequency coefficient = 0.0032 ± 0.008, $p = 0.69$; $n = 22$ mice). ns, not significant, $^{**}p < 0.01$, $^{****}p < 0.0001$, t test versus zero.

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

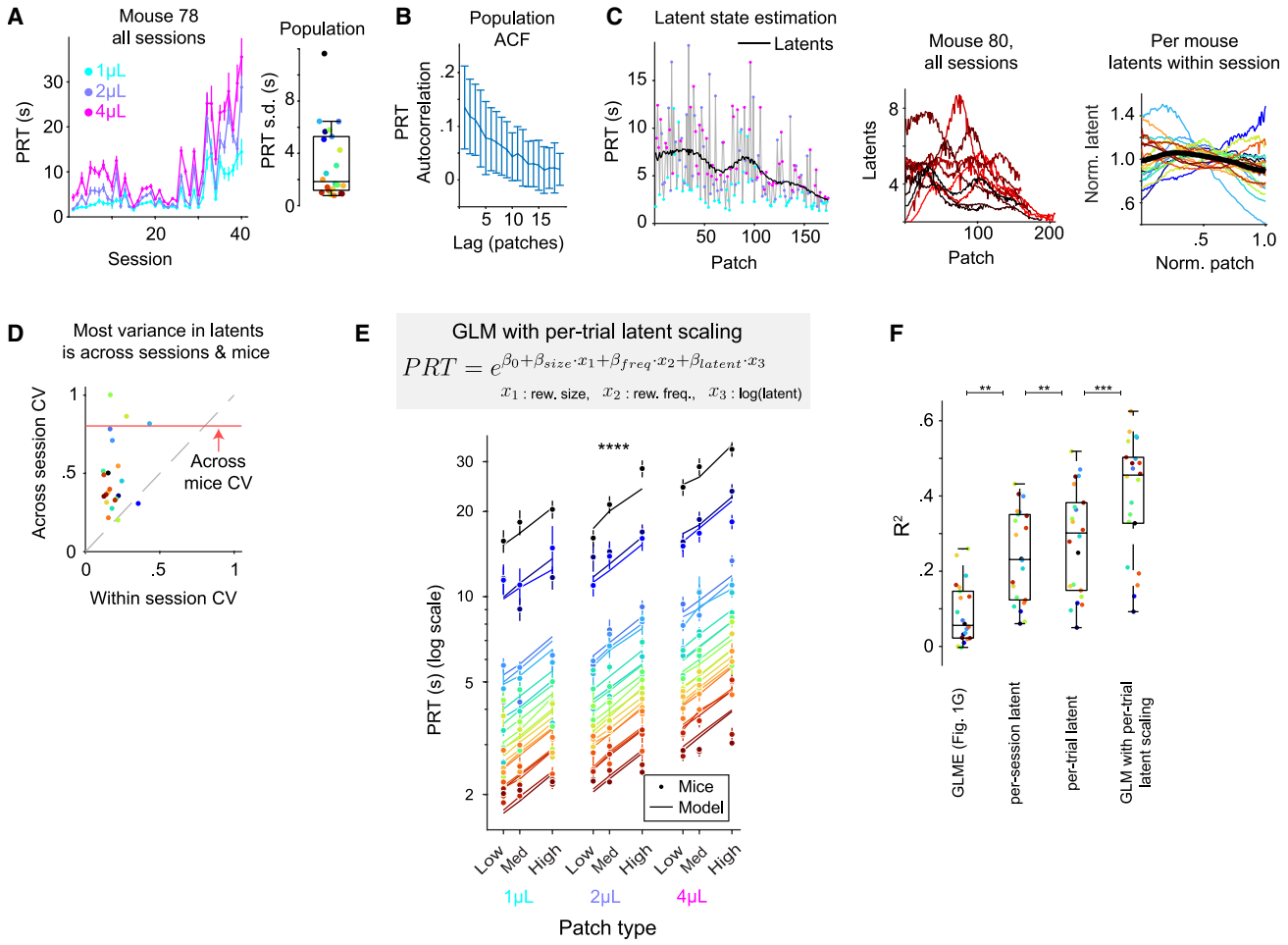


Figure 2. Scaling a common function explains PRT variance

(A) Left: mean PRT per μL across sessions, example mouse. Right: standard deviation of mean PRT across sessions per subject, colored by mouse ID.

(B) PRT autocorrelation over patches, mean across subjects. Error bars indicate standard deviation across subjects.

(C) Estimating latent "patience" state from surrounding PRTs. Left: example session. Gray trace tracks PRT with colored points indicating reward size for each patch. Black trace indicates estimated latent state (Gaussian filtered PRT, omitting the current trial). Middle: estimated latent state for all sessions from an example mouse. Right: mean normalized latent estimates across sessions per mouse. Black line shows mean across mice.

(D) Coefficient of variation (CV) of latent state within versus across sessions per mouse. For within-session measures, CV of latents across patches was calculated per session, and mean CV is shown. For across session measures, mean latent was calculated per session, and CV over these means is shown. For the across mice measure, mean latent was calculated per mouse, and CV over these means is shown (red line).

(E) Generalized linear model (GLM) scaled by latent patience estimates. Points and error bars show mean individual mouse PRTs ± SEM. Lines show GLM fit.

(F) R^2 statistics for predicting PRT using GLME (from Figure 1G), mean latent per session, per trial latent, and latent-scaled GLM. Ordered by ascending mean R^2, stars indicate significance of pairwise Wilcoxon signed-rank tests (p = 0.0044, 0.0027, 0.002, Bonferroni-adjusted).

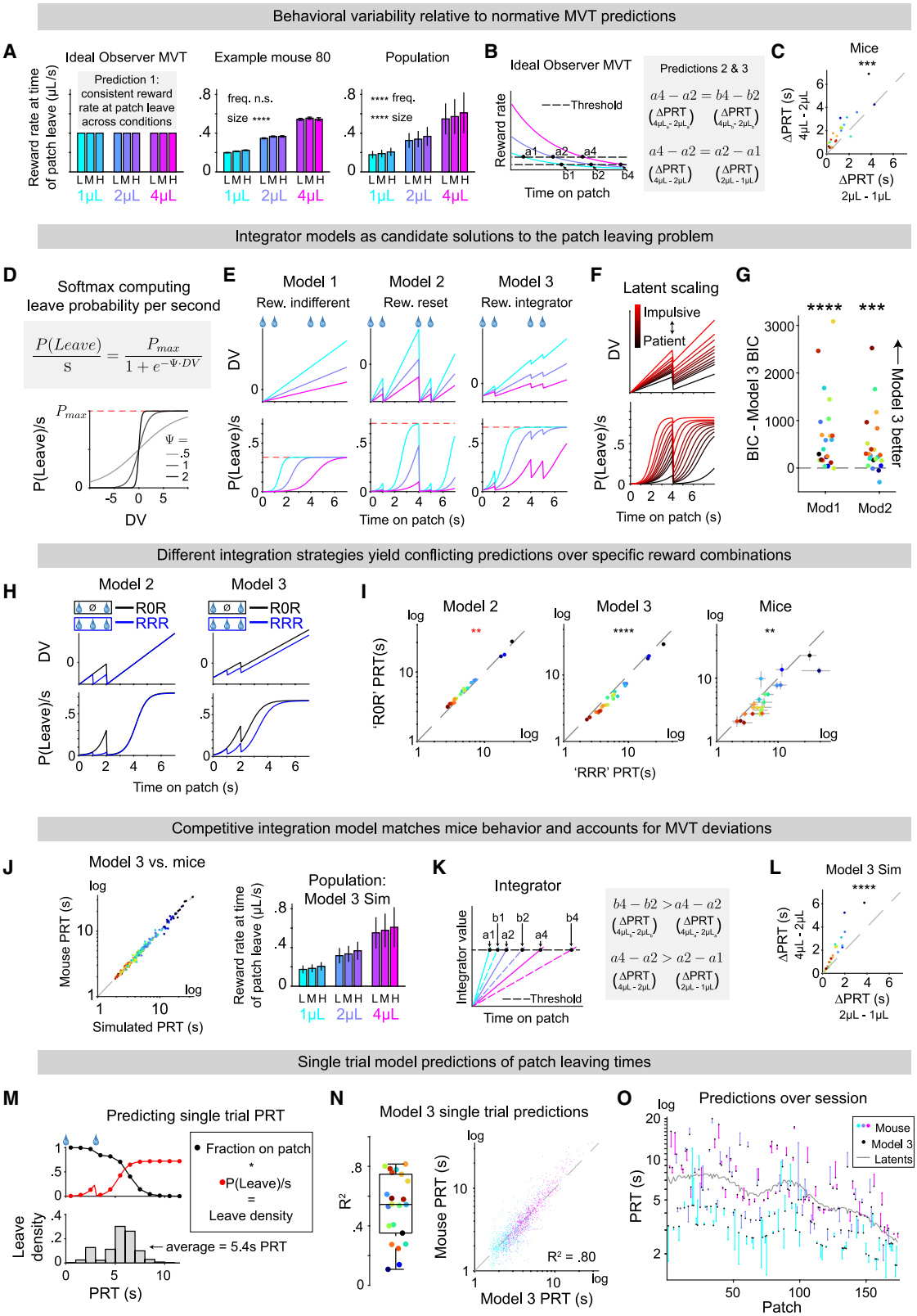
environmental reward rate (a common threshold, Figure 3A, left; prediction 1). By contrast, we found mice left high-value patches too early and low-value patches too late,[20] relative to an ideal observer, with the dominant effect driven by reward size (Figure 3A, middle: example mouse, $p < 0.0001$ for reward size and $p > 0.99$ for frequency, two-way ANOVA, Bonferroni-adjusted; right: population means, $p < 0.0001$ for size and frequency, LME).

We next considered more specific MVT predictions enabled by our task design. Because reward probability decayed exponentially with a fixed time constant (Figure 1C), PRT differences across patch types should remain constant, regardless of threshold (Figure 3B; prediction 2). Furthermore, because reward sizes were proportional, the difference in leave times between 4 and 2 μL patches should equal the difference between 2 and 1μL patches ($PRT_{4\mu L} - PRT_{2\mu L} = PRT_{2\mu L} - PRT_{1\mu L}$; Figure 3B; prediction 3). Neither prediction was borne out in the data. Instead, reward sensitivity grew with mean PRT (Figures 2E and S2), and $PRT_{4\mu L} - PRT_{2\mu L} > PRT_{2\mu L} - PRT_{1\mu L}$ (Figure 3C, $p = 0.0002$, Wilcoxon signed-rank test).

Importantly, we note that these analyses consider reward rate calculated objectively (μL/s), rather than subjective utility.[42–44] Converting into units of expected utility could potentially reconcile these results with MVT predictions (Figure S3B), though a

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

CellPress



(legend on next page)

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

property of the nonlinear transformation, such as utility function steepness, would need to covary with the leaving threshold (Figures S3C–S3E).

Competitive integration explains patch-leaving decisions

While extensions of MVT with nonlinear utility functions (Figures S3A–S3E) or other modifications[16,20,45,46] could potentially reconcile these findings, the model space is large and under-constrained by our data. We therefore explored a different category of explanation, namely integrator models,[33,47] which are not normative like MVT but instead describe a potential process for implementing the decision.

We devised three distinct integrator models to compare with mouse behavior (Figure 3E). In all three models, a DV ramps linearly over time in the absence of rewards, with slope scaled by reward size (passed through a power law function). The DV is transformed via a softmax to yield a leave probability. The models differ only in their reward responses. Model 1 is indifferent to rewards, tracking only time on patch. Model 2 fully resets, returning to baseline following each reward, memoryless of rewards before the most recent one.[22] Model 3 competitively integrates rewards versus time by decrementing the DV by a constant amount (R) at each reward delivery, thus preserving effects of multiple rewards over time. We scaled model parameters by each trial's latent estimate, which was critical for obtaining accurate fits (Figure S3F).

Model 1 can be ruled out a priori, as it contradicts the observed dependency of PRT on reward frequency, but was included for comparison. Models 2 and 3 could both potentially account for reward frequency dependence. We used the Bayesian information criterion (BIC) to compare model performance. BIC was lowest for model 3, demonstrating it best explained leaving decisions (Figure 3G; $p < 0.0001$, model 3 over model 1; $p = 0.0009$, model 3 over model 2; Wilcoxon signed-rank test, Bonferroni correction; cross-validated BIC gave similar results; Figure S3G).

To further differentiate models 2 and 3, we examined trial types wherein they make divergent predictions (Figure 3H): "RRR" trials, featuring rewards at $t = 0, 1,$ and 2 s, and "R0R" trials, featuring rewards at $t = 0$ and 2 s (no reward at $t = 1$). Because model 2 resets following reward, it predicts identical leaving behavior under both conditions (Figures 3H and 3I). Conversely, model 3 integrates reward history and preserves the effect of the $t = 1$ reward, predicting higher PRT on RRR over R0R patches, matching mice behavior (Figures 3H and 3I). Comparing "RR0" versus R0R PRTs again better matched model 3 (Figures S3H and S3I).

We devised a method to predict average PRT from each model fit given an arbitrary sequence of rewards (Figure 3M; STAR Methods). Models 2 and 3 performed similarly in this analysis, with a slight advantage to model 2 (mean R^2, model 2 = 0.54, model 3 = 0.53; Figures 3M–3O and S3N). We suspect

Figure 3. Competitive integration processes explain foraging behavior

(A) Instantaneous expected reward rate at time of patch leave across patch types, predicted by idealized MVT (left), example mouse (middle, Bonferroni-adjusted $p < 0.0001$ for size, $p > 0.99$ for frequency, two-way ANOVA), and population (right, $p < 0.0001$ for size and frequency, linear mixed-effects model). Error bars for the example mouse indicate SEM. Error bars for population average indicate standard deviation across mice.

(B) Schematic demonstrating MVT predictions for PRT per reward size for two different thresholds. Traces are colored by reward size. Dashed lines show two sample thresholds, a and b. Black dots indicate threshold crossing points. Gray box notes examples for prediction 2 (top) and prediction 3 (bottom).

(C) Mean PRT differences between $4-2\ \mu L$ and $2-1\ \mu L$ patches ($p = 0.0002$, Wilcoxon signed-rank test). Dashed line indicates unity.

(D) Sigmoid transformation of decision variable (DV) into leave probability per 1-s interval ($P(Leave)/s$), scaled by different inverse temperature values (Ψ, shades of gray). Red dashed line indicates P_{max}, the maximum $P(Leave)/s$ output.

(E) Example traces of DV (top) and corresponding $P(Leave)/s$ output (bottom) for integrator models over an example patch with rewards at $t = [0, 1, 4, 5]$ s. Red dashed line indicates P_{max}.

(F) Schematic demonstrating DV and $P(Leave)/s$ scaling by latent state, for an example patch with rewards at $t = [0, 4]$ s. Black/red traces indicate patches with higher/lower latent state estimates and ramp up less/more quickly.

(G) Relative BIC values for model fits across subjects ($p < 0.0001$, model 3 over model 1; $p = 0.0009$, model 3 over model 2, Wilcoxon signed-rank test, Bonferroni-adjusted).

(H) Schematics demonstrating differing models 2 (left) and 3 (right) predictions on patches with rewards at $t = [0, 2]$ s ("R0R" patches, black) versus $t = [0, 1, 2]$ s ("RRR" patches, blue).

(I) Per-subject mean simulated PRT for R0R versus RRR patches from model 2 (left), model 3 (middle), and empirical mice PRT (right; points and error bars indicate mean ± SEM). Model 3 and mice PRTs were higher for RRR versus R0R trials ($p < 0.0001$, model 3; $p = 0.0024$, mice; Wilcoxon signed-rank test). There was a small but significant effect of greater PRTs for R0R versus RRR trials for model 2 ($p = 0.0037$; red stars). This was due to selection bias over latent state (R0R trials tend to have higher latent state than RRR trials, lengthening PRT) and was in the opposite direction of the empirical mice data. Points are colored per mouse. Dashed line indicates unity.

(J) Left: mean PRT across patch types, per subject, from model 3 simulations versus empirical mouse PRT, log-scaled, colored per mouse ($R^2 = 0.985$, mean squared error = 0.413 s). Right: mean instantaneous expected reward rate at time of patch leave from model 3 simulations (as in "A").

(K) Schematic demonstrating integrator models can account for deviations from MVT predictions. Example traces are shown when patience is lower (solid lines) versus higher (dashed lines) for each reward size (color). Patch-leave times are determined by integrator value reaching threshold (black dashed line; points indicate threshold crossings). Gray box notes example violations of prediction 2 (top) and prediction 3 (bottom).

(L) Mean model 3 simulated PRT differences between $4-2\ \mu L$ and $2-1\ \mu L$ patches ($p < 0.0001$, Wilcoxon signed-rank test). Dashed line indicates unity.

(M) Schematic depicting how model-predicted PRT is calculated on a single trial (STAR Methods).

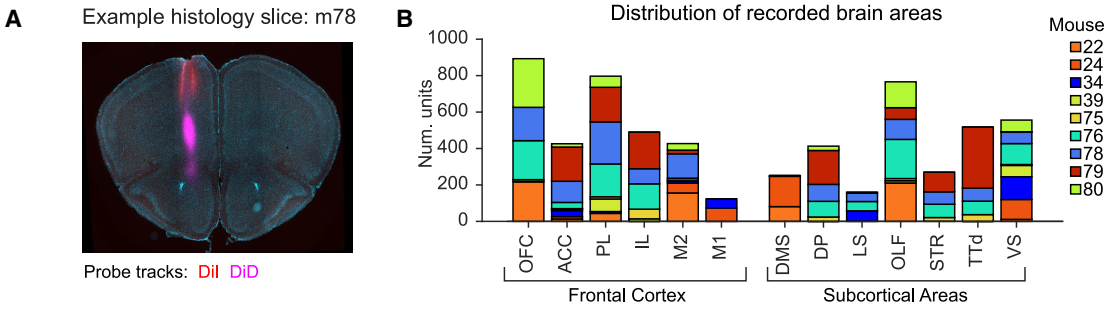
(N) Left: R^2 statistics for single-trial predictions from cross-validated model 3 fits across mice (left, median $R^2 = 0.544$). Box: median, 25%, and 75%; whiskers = 5%–95%. Right: single-trial model 3-predicted PRT (cross-validated) versus true PRT for an example mouse ($R^2 = 0.801$). Colors indicate reward size per patch.

(O) Model 3 predicted PRT and empirical mouse PRT across patches from example session in Figure 2C. Colored points indicate mouse PRT per reward size. Black points indicate model 3 cross-validated predicted PRT. Lines connecting points highlight the difference in predicted versus empirical PRT. Gray trace indicates latent state estimate for each patch.

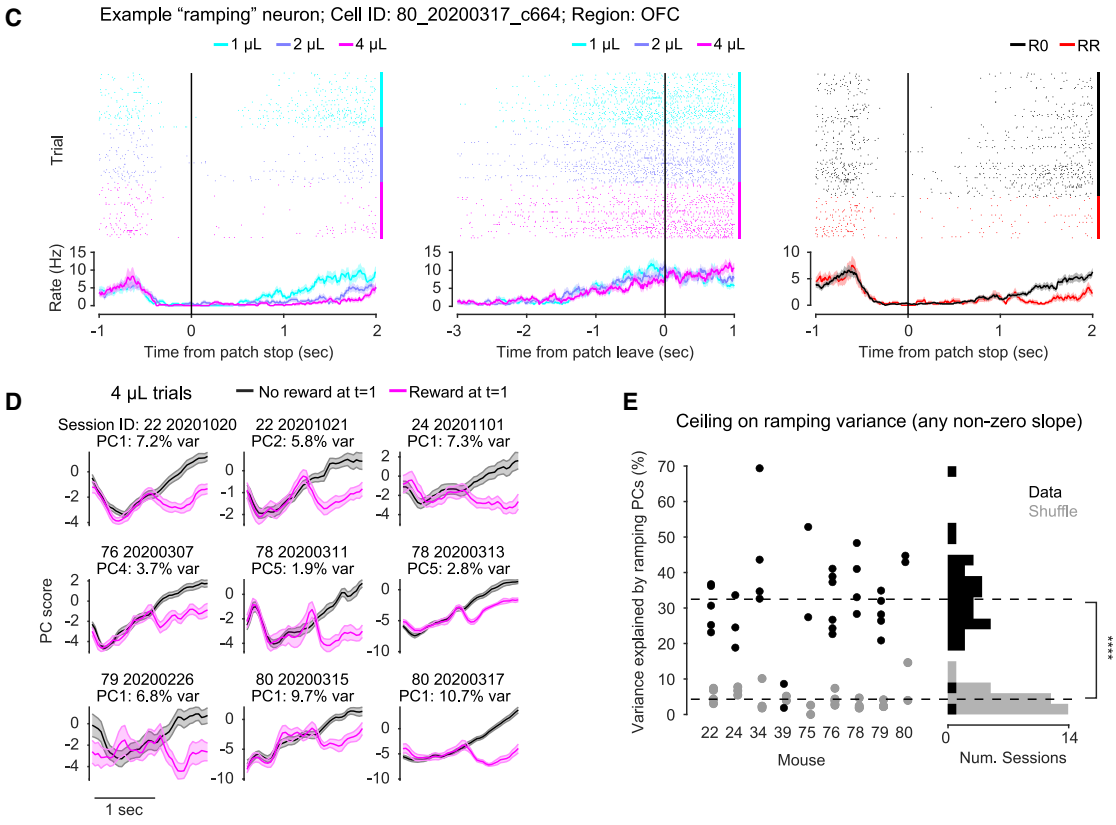
Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

CellPress

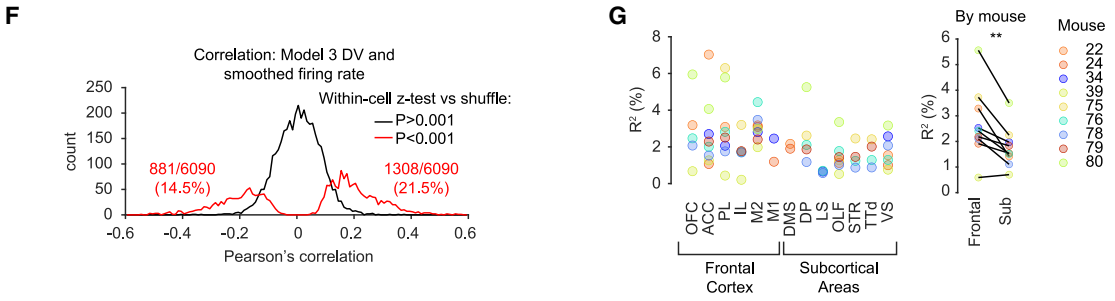
Neuropixels recordings across frontal cortex and underlying areas during patch foraging task



Neural activity contains ramping signals inhibited by rewards, at both the single neuron and population level



Activity of single neurons correlates with Model 3 decision variable



(legend on next page)

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

this analysis artificially favored model 2 by adding fictive time bins to the end of trials (STAR Methods).

Because of the way DVs change over time, integrator models naturally captured deviations from MVT (Figures 3J–3L; predictions 1–3). Figure 3K demonstrates this: given two integration processes with different slopes (e.g., 2 versus 4 μL traces shown, solid lines), proportionally reducing their slopes (corresponding to increased patience, dashed lines) results in a greater PRT difference between them, thus violating prediction 2. Additionally, unless slopes are precisely tuned across reward sizes, then $PRT_{4\mu L} - PRT_{2\mu L} \neq PRT_{2\mu L} - PRT_{1\mu L}$, thereby violating prediction 3. Consistent with Figure 3C, simulations replicated the observed $PRT_{4\mu L} - PRT_{2\mu L} > PRT_{2\mu L} - PRT_{1\mu L}$ (Figure 3L, $p < 0.0001$, Wilcoxon signed-rank test).

We additionally fit a version of model 3 with parameter fits shared across mice, with differences across mice driven solely by latent patience scaling. This model outperformed an alternative model with separate parameters per mouse but without latent scaling (Figures S3K and S3L). Latent patience scaling is thus sufficient to account for variability in PRT observed across subjects, suggesting individual differences might arise via scaling a common underlying algorithm.

Slow ramping is prominent in the frontal cortex

To investigate the neural underpinnings of the above foraging strategies, we recorded spiking activity of neurons during the task using Neuropixels probes throughout the frontal cortex and underlying subcortical areas (Figures 4A and 4B; n = 6,090 units from 33 recording sessions in 9 mice). We commonly observed neurons whose activity, like our behavioral models' DVs, slowly ramped as mice remained in patches, with reward modulation opposing the ramping direction (Figure 4C). Top principal components (PCs) of neural activity often similarly displayed reward-modulated ramping (Figure 4D). Average total variance explained by ramping PCs was 32.5% ± 2.2% (n = 33 sessions), significantly greater than a shuffle control (4.3% ± 0.5%, paired t test $p < 1.0 \times 10^{-13}$; Figure 4E).

The prevalence of reward-modulated ramping activity suggests a candidate for DV encoding. Indeed, 35.9% of individual neurons had activity significantly correlated with the model 3 DV (Figure 4F). Notably, model 3's DV explained more firing rate variance in frontal cortex compared with subcortical areas, suggesting a specialized role for frontal cortex (Figure 4G).

DV decoding reveals a neural ramping dimension with features of models 2 and 3

We investigated DV coding in high-dimensional neural population activity, fitting cross-validated linear models predicting DVs from joint activity of co-recorded neurons (Figures 5A and 5B).[41] DVs were reliably decoded in most sessions (Figure 5C). Although model 3 best explained behavior (Figure 3), DVs from models 1 and 2 were similarly decoded (mean coefficient of variation [CV] R^2 ± SEM for model 1 = 0.31 ± 0.03, model 2 = 0.38 ± 0.04, model 3 = 0.35 ± 0.03; n = 28 sessions after excluding 5 sessions with CV $R^2 < -0.1$ for all three models; STAR Methods; Figure 5C), more reliably in frontal cortex than subcortical areas (Figure 5G).

How can multiple DVs be decoded with similar accuracy? There are at least two possibilities, defining two ends of a spectrum: decoders could identify separate neural subspaces for each DV or a similar ramping dimension across models. Our analysis favors the latter, as single neuron regression coefficients were highly correlated across models (Pearson's correlation = 0.835 ± 0.018, mean ± SEM, n = 9 mice; Figure 5F). DVs themselves were highly correlated (models 2 versus 3 DV, Pearson's correlation = 0.843 ± 0.023; Figure 5D), and neural predictions even more so (Pearson's correlation = 0.922 ± 0.013; Figure 5F). This common ramping dimension captured the key feature shared across models: negative modulation of ramping slope by reward size (LME, reward size coefficient estimate = −0.061 ± 0.02, p = 0.0083 [model 1]; estimate = −0.073 ± 0.02, p = 0.0070 [model 2]; estimate = −0.064 ± 0.02, p = 0.0026 [model 3]; Figures 5H and 5I).

The models are distinguished by their reward responses (Figure 3E). We therefore examined each decoder's output at reward times and observed features of both models 2 and 3, regardless of which DV we asked the neural population to predict (Figures 5J and 5K). Linear models predicting reward responses from the pre-reward level had slopes between 0 (no dependence, as in model 3) and −1 (full reset, as in model 2), indicating

Figure 4. Reward-suppressed ramps are prevalent in the frontal cortex

(A) Example histology slice.

(B) Distribution of recorded brain areas.

(C) Example neuron with ramping activity suppressed by reward delivery. Left: PSTH aligned to patch stop, split by reward size (trials with reward at t = 1 omitted). Middle: PSTH aligned to patch leave, split by reward size. Right: PSTH aligned to patch stop, split by whether reward was delivered at 1 s (red) or not (black) (rewards of different sizes combined). Lines and shaded area = mean spike rate ± SEM over trials, colored by trial type.

(D) Hand-picked principal components (PCs) of neural activity showing integrator-like activity. For simplicity, only 4 μL trials are shown. Magenta/black traces are trials with/without reward at t = 1. Lines and shaded area = mean ± SEM over trials.

(E) Total variance explained by PCs with significant ramping slopes (black) versus a shuffle control (gray). ****$p < 0.0001$, data versus shuffle, sign rank test (n = 33 sessions).

(F) Histogram of correlation between single neuron firing rates and model 3 DV (Figure 3). Red/black indicate neurons with/without significant correlation versus shuffle control (z test $p < 0.001$).

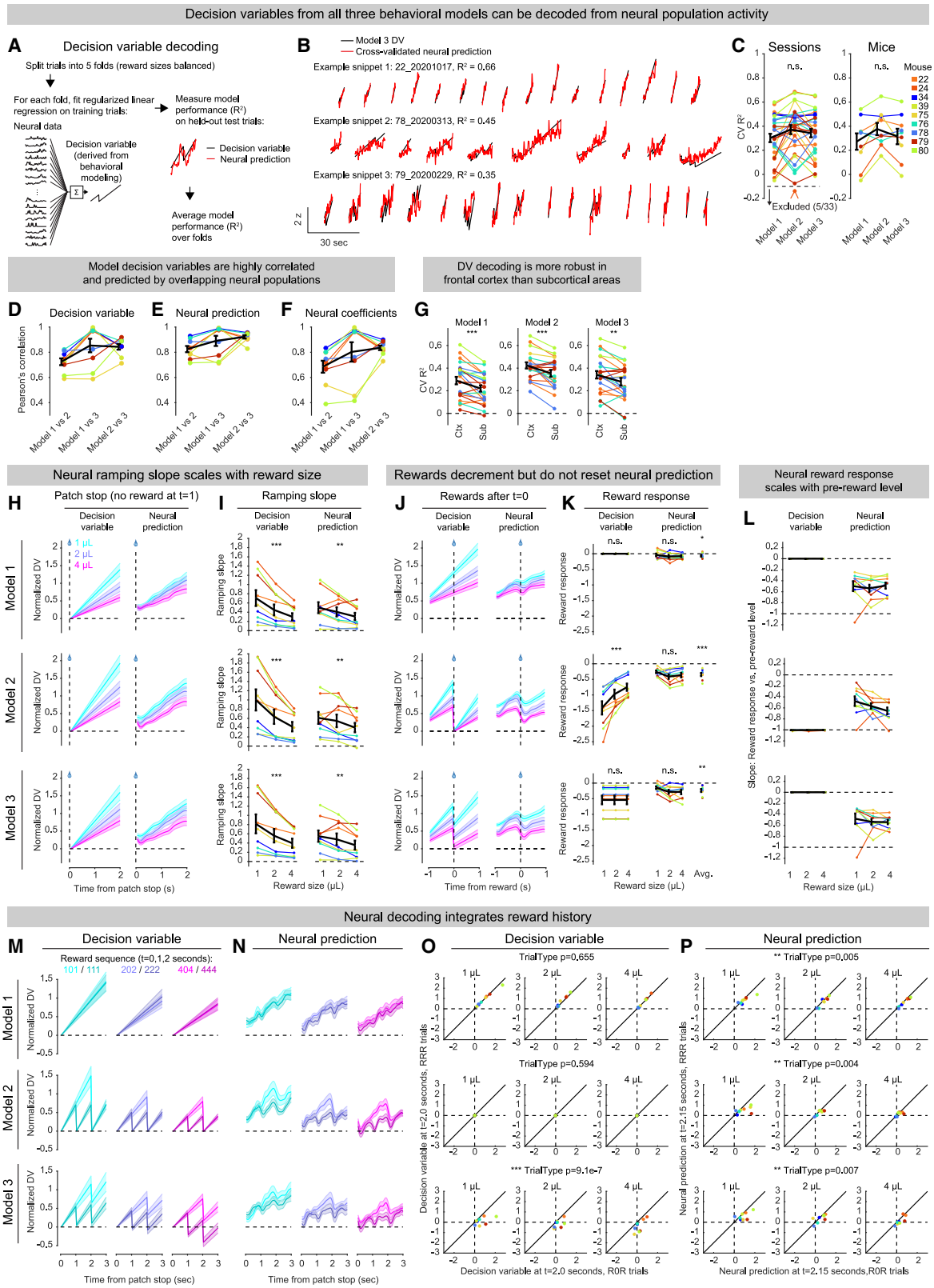
(G) R^2 between individual neurons' firing rates and the model 3 DV, by brain region. Points represent recording sessions (left) or mice (right). Frontal cortex areas had higher mean R^2 values than subcortical areas ($p < 0.01$, paired t test, n = 9 mice).

Frontal cortex areas: OFC, orbitofrontal cortex; ACC, anterior cingulate cortex; PL, prelimbic cortex; IL, infralimbic cortex; M2, secondary motor cortex; M1, primary motor cortex.

Subcortical areas: DMS, dorsomedial striatum; DP, dorsal peduncular area; LS, lateral septum; OLF, olfactory areas; STR, striatum; TTd, taenia tecta dorsal part; VS, ventral striatum.

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

CellPress



(legend on next page)

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

partial resetting (Figure 5L). Thus, reward responses are neither fully like model 2 nor model 3 but rather intermediate between them.

Model 3 tracks intra-patch reward history, whereas model 2 considers only time since the last reward. To assess reward history effects, we examined trials featuring rewards delivered at $t = 0$ and $t = 2$ (R0R trials) versus $t = 0, 1,$ and 2 (RRR trials). All three decoders' outputs following the last reward were more negative on RRR compared with R0R trials (Figures 5M–5P). Thus, like mouse behavior, neural activity tracks reward history—even when decoding analysis selects against it.

Latent state modulates decoder output at patch stop

Because latent state modulation was essential for behavioral model fits (Figures 2 and 3), we compared decoder output and latent state. We did not find a significant relationship between latent state and ramping slope (Figure S5F) or reward response magnitude (Figure S5G) across mice, though there were individual sessions with significant correlations (Figures S5F and S5G, right). By contrast, decoder output at time $t = 0$ negatively correlated with latent estimates (Figure S5H), hinting that latent state may modulate ramping dynamics via DC offset. However, because of limited latent state variance within sessions (Figure 2D), our data are likely underpowered to identify the effects of latent state on ramping dynamics. Indeed, population activity before patch stop could predict latent state traces from different behavioral sessions (Figures S5I–S5K), suggesting our recording sessions are not long enough to dissociate latent state from other slowly varying signals generically present in neural data.[48]

GLM modeling reveals functional clusters with oppositely signed ramping and reward responses

To gain a broader overview of neural activity, we fit a Poisson GLM to single neurons (Figures 6A and S6A–S6C). The GLM included the following regressors: session time and its square to capture slowly varying changes in firing rate; behavioral variables: position, speed, acceleration, lick rate, and change in lick rate; discrete events at patch stop, patch leave, and reward times, convolved with a raised cosine basis; and three ramping "accumulators" (time on patch, total reward, and time since reward).

We quantified performance via cross-validated percent deviance explained compared with a null model (Figure S6E, top; mean ± SEM = 5.2 ± 0.1, n = 6,090 neurons). Reward kernels and accumulators were termed "task variables" and used to identify task-related neurons. Percent deviance explained was compared between full models and models without task variables ("reduced models"), and neurons with deviance explained > 1% in this comparison were deemed "task-related" (Figure S6E, bottom; 1,458/6,090 neurons, mean % deviance explained versus reduced model ± SEM = 0.87 ± 0.03). Mirroring DV decoding performance, GLM performance was higher for frontal cortex areas than subcortical areas (Figure S6F, mean % deviance explained ± SEM, frontal cortex = 6.5 ± 1.0, subcortical areas = 4.1 ± 0.6, paired t test, $p = 0.016$, n = 9 mice). Accordingly, a greater fraction of frontal cortex neurons were labeled task-related (Figure S6G; frontal cortex: 916/3,156 [29.0%], subcortical areas: 482/2,934 [16.4%]). Aligning activity to patch stop and sorting by time of peak response revealed a "wave" of activity following each reward: a rapid reward response with variable latency, which gradually transitioned

Figure 5. DV decoder output shares features of models 2 and 3

(A) Schematic of decision variable (DV) decoding.

(B) Decoder output (red) versus true model 3 DV (black) for several contiguous patches from example recording sessions.

(C) CV R^2 for model DVs (models 1–3; Figure 3), by session (n = 28, left) or mouse (n = 9, right). 5/33 sessions with CV $R^2 < -0.1$ were excluded from further analysis. ns, not significant ($p > 0.05$, one-way ANOVA). In this panel and the remainder of the figure, black lines and error bars indicate means ± SEM.

(D) Within-session correlations between DVs for pairs of models, averaged across sessions per mouse.

(E) Same as (D), but for neural predictions (decoder outputs).

(F) Same as (D), but for neural regression coefficients.

(G) Comparison of R^2 between decoders using only units from the frontal cortex (Ctx) or subcortical areas (Sub). Each line represents a recording session, and colors represent mice. Black lines and error bars = mean ± SEM over sessions. $^{**}p < 0.01$, $^{***}p < 0.001$, paired t test (n = 23 sessions).

(H) PSTHs of DVs (left) and neural predictions (right; decoder output trained on those DVs) on trials with no reward at $t = 1$, which were used to estimate ramping slope. Lines and shaded areas = means ± SEMs over sessions. Colors indicate reward size. In (H)–(P), results are shown for behavioral models 1–3, arranged from top to bottom.

(I) Estimated ramping slope for DVs (left) and neural predictions (right). Colored lines show individual mice (slopes averaged over sessions). Black lines and error bars show means ± SEM over mice. $^{**}p < 0.01$, $^{***}p < 0.001$, reward size coefficient, linear mixed-effects model (LME).

(J) Same as (H), but aligned to reward deliveries, excluding rewards at $t = 0$.

(K) Reward responses of DVs (left) and neural predictions (right). Significance level reflects the reward size coefficient in an LME. For neural predictions, average response across reward sizes is also shown ("avg"), and stars indicate significance of a t test versus zero (n = 9 mice). ns, not significant ($p > 0.05$), $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

(L) Slope from linear regression of reward response on pre-reward level of DVs (left) and neural predictions from decoders (right). For all three decoders, slope was approximately −0.5 for all reward sizes, indicating partial resetting.

(M) PSTHs of DVs on trial types isolating the effect of reward history: trials with rewards at $t = 0$ and 2 s ("R0R") versus trials with rewards at $t = 0, 1,$ and 2 s ("RRR"). Lines and shaded areas indicate means ± SEMs over sessions. Trials are split by reward size and R0R versus RRR reward sequences, with lighter shades indicating R0R trials and darker shades indicating RRR trials.

(N) Same as (M), but showing neural predictions on R0R versus RRR trials.

(O) Comparison of the DV level just after the reward at $t = 2$ s between R0R (x axis) and RRR (y axis) trials for each reward size. p values for TrialType (R0R versus RRR) in an LME are shown for each behavioral model. By construction, only model 3 has a significant TrialType coefficient, indicating sensitivity to reward history.

(P) Same as (O), but for neural predictions. All decoders had significant TrialType coefficients, indicating a reward history effect on decoder output most consistent with model 3.

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

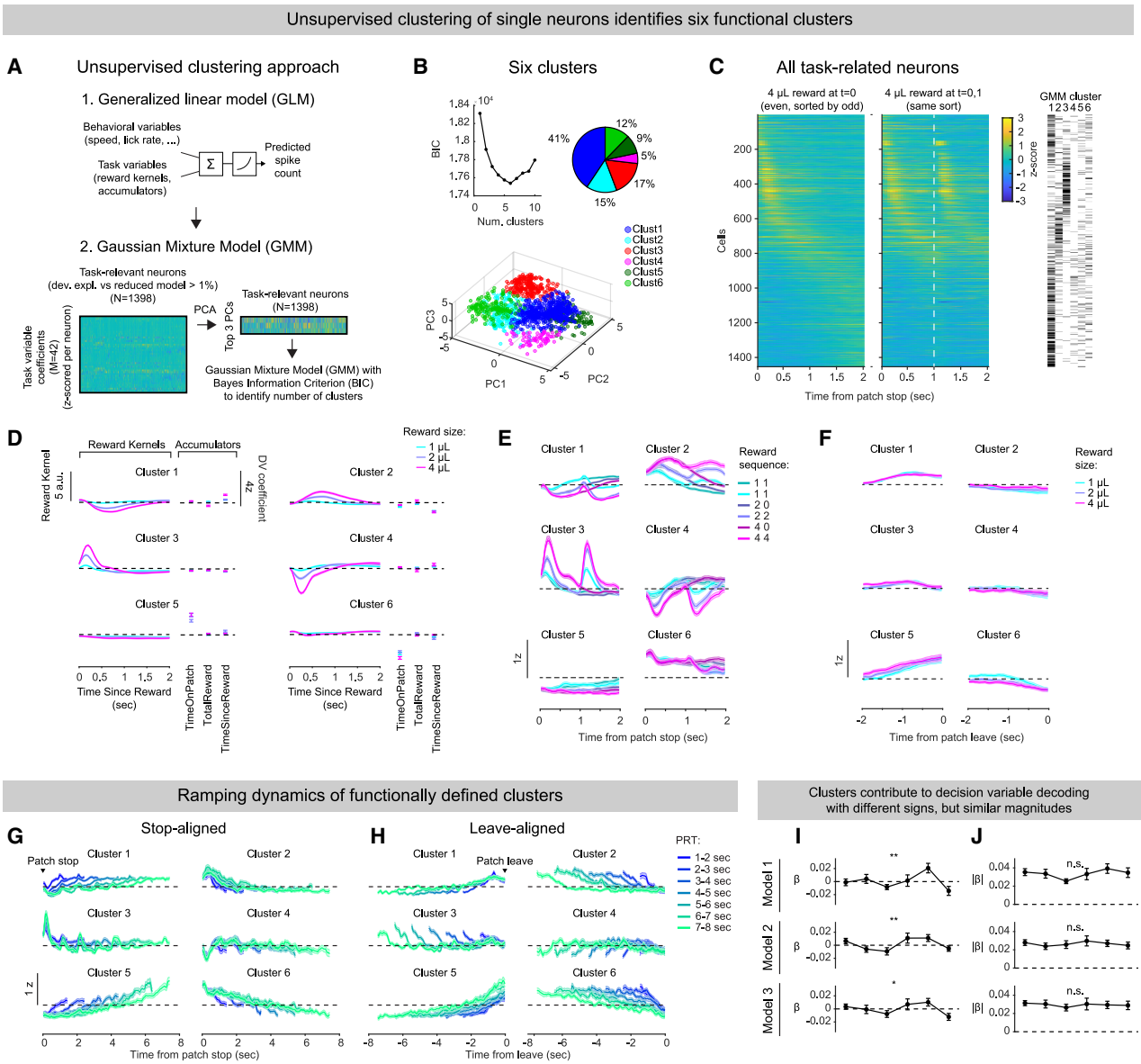


Figure 6. Functional clustering reveals module pairs with reciprocal integration dynamics

(A) Schematic of analysis approach: task-variable coefficients estimated via Poisson GLM are used to cluster neurons using a Gaussian mixture model (GMM).

(B) GMM clustering to identify clusters of neural activity patterns. Top left: BIC was used to select the number of clusters (minimum BIC: 6 clusters). Top right: percentage of neurons assigned to clusters. Clusters were ordered so patterns with similar shapes but opposite signs were adjacent (see E). Bottom: task-related neurons projected into the PC space used for clustering.

(C) Neural activity for task-related neurons on "40" trials (4 μL reward at 0 s, no reward at 1 s; left panel) or "44" trials (4 μL reward at 0 and 1 s; right panel; white dashed line indicates reward at 1 s). GMM cluster identity for each neuron is indicated on the right.

(D) Average GLM-predicted reward responses and Z scored accumulator coefficients. Z scored reward kernel coefficients were multiplied by corresponding basis functions and summed to generate the predicted reward response. In (D)–(H), lines indicated means, and shaded regions indicate SEM.

(E) Average PSTHs of Z scored neural activity following patch stop for each cluster, split by reward size and whether or not reward was delivered at 1 s.

(F) Average PSTHs of Z scored neural activity aligned to patch leave for each cluster.

(G) Average PSTHs of Z scored neural activity for each cluster, aligned to patch stop and split by PRT.

(H) Same as (G), but aligned to patch leave.

(I) Average coefficient (β) per neuron in linear DV decoding from models 1–3 (top to bottom; decoding as in Figure 5), by GMM cluster. Coefficients were averaged over neurons within session first, then averaged over sessions. Lines indicate means, and error bars indicate SEM over sessions (n = 28 sessions). Coefficients differed by GMM cluster for all three models (LME with fixed effect of cluster and random effect per session; model 1, p = 0.0044; model 2, p = 0.0033; model 3, p = 0.021; n = 28 sessions). $^{**}p < 0.01$, $^{*}p < 0.05$.

(J) Same as (I), but for absolute value of decoder coefficients ($|\beta|$), a measure of overall contribution to the decoder. $|\beta|$ did not differ between GMM clusters for any of the three models (model 1, p = 0.15; model 2, p = 0.94; model 3, p = 0.86; n = 28 sessions). ns, not significant.

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

into upward ramping activity that was inhibited by reward delivery (Figure 6C), reminiscent of model 2 and 3 DVs.

Next, we clustered task-related neurons by their GLM coefficients to characterize the space of activity (Figure 6A, bottom). Gaussian mixture model (GMM) clustering was performed on a dimensionality-reduced matrix of task-variable coefficients by neurons (top 3 PCs, 34.7% of variance; Figure S6D), with the number of clusters (K = 6) chosen to minimize BIC (Figure 6B). GMM clusters could be grouped into three pairs with similar shapes but opposite signs (Figure 6D): (1) a pair of clusters with slow negative/positive reward responses, scaled by reward size, and positive/negative coefficients on time since reward (clusters 1 and 2); (2) a pair with rapid reward responses, scaled by reward size, and minimal ramping (clusters 3 and 4); and (3) a pair with minimal reward coefficients but significant positive/negative coefficients on time on patch (clusters 5 and 6). Cluster coefficients were similar across mice (Figure S6I). PSTHs of cluster activity mirrored the oppositely signed ramping and reward responses of the behavioral model DVs (Figure 6E). Whereas clusters 3 and 4 consisted of rapid positive/negative reward responses with minimal ramping components, clusters 1, 2, 5, and 6 ramped up/down in the absence of rewards, either to similar thresholds prior to patch leave (clusters 1 and 2) or continuing to ramp up/down in trials with longer PRTs (a clock-like signal; clusters 5 and 6) (Figures 6F–6H).

Patterns of task-related activity captured by task-variable coefficients are unlikely to be explained by movement variables we measured, because these were included as regressors in the GLM. To further examine running speed contributions to ramping activity in our data, we fit GLMs to neural activity during patches and inter-trial intervals (ITIs) and found GLM coefficients for running speed were more correlated between odd and even patches than between patches and ITIs (Figure S4C), even when periods of high running speed were excluded from ITIs (Figure S4D). This suggests that ramping activity is not a low-level motor signal common to both patches and ITIs.

Neurons across GMM clusters contribute to DV coding

We investigated relationships between the six functional activity clusters and model DV decoding. One possibility is that decoding is driven by a small number of clusters. For example, average activity of cluster 1 qualitatively resembled model 3's DV, showing upward ramping decremented by reward delivery, and this cluster could be particularly useful for decoding ramping DVs. Alternatively, DV decoding could be accomplished through contributions across clusters, arguing for a distributed, population-level representation[49] that cannot be captured by activity patterns of single neurons. Examining individual neurons' coefficients in the linear decoders for model DVs favored this distributed representation: coefficients differed in sign across clusters but had similar magnitude (absolute value; Figures 6I and 6J).

State space modeling supports population ramping on single trials

We tested whether single-trial dynamics within clusters (e.g., Figure 7A) were better described as continuously varying ramps or discrete processes—an issue that has been controversial in perceptual decision-making.[50,51] While previous analyses focused on single neuron recordings, we reasoned that our simultaneously recorded neurons may help us resolve this issue for patch-foraging decisions.[52,53] We therefore fit a continuous ramping model and a discrete stepping model to neural population responses (population spike rate, summed across neurons within each cluster; Figures 7B and 7C). The ramping model had a pulse response to rewards and otherwise ramped in the absence of rewards. The stepping model had two discrete firing rate states, and the firing rate started in the low state and probabilistically jumped to the higher state following Markovian dynamics. Additionally, to capture reward responses, the firing rate returned to the low state after rewards, whereas otherwise the firing rate was fixed to remain in the high state.

We found all six GMM clusters were better fit by the ramping than the stepping model (mean difference in held-out log likelihood ± SEM [bits/spike], ramping – stepping model: cluster 1: 0.013 ± 0.003, $n = 16$ sessions with at least 10 cluster 1 neurons, t test versus zero $p = 0.00042$; cluster 2: 0.0079 ± 0.0018, $n = 10$, $p = 0.0025$; cluster 3: 0.018 ± 0.007, $n = 12$, $p = 0.024$; cluster 4: 0.0062, $n = 1$; cluster 5: 0.025 ± 0.012, $n = 5$, $p = 0.13$; and cluster 6: 0.0081 ± 0.0024, $n = 7$, $p = 0.020$; Figure 7D). This supports the idea that ramping neural signals are widespread in our dataset, distributed across multiple functionally defined clusters of neurons.

DISCUSSION

We sought to identify the algorithm underlying leaving decisions in a patch-foraging context long studied in the framework of optimal foraging theory. In our task, mouse behavior matched MVT qualitatively, with PRTs increasing with current patch richness and decreasing with previous patch richness (Figure 1H). However, stronger tests exploiting our task designs revealed systematic deviations from MVT predictions. Aiming to understand the mechanisms producing these behavioral patterns, we turned to integrator models, which are not normative like MVT but instead describe a decision process.[33] Integrator models scaled by a slowly varying "patience" variable captured behavior, including deviations from MVT (Figure 3), demonstrating that latent scaling of a common algorithm could produce the wide range of observed behaviors (Figure S3K).

Mirroring the behavioral models, neural data contained ramping signals with opposing influences of time and rewards at both the single neuron and population levels (Figures 4, 5, and 6). While model 3 best described the behavioral data, neural ramping activity reflected features of both model 2 (partial reset following rewards) and 3 (sensitivity to reward history), highlighting the difficulty of inferring the behavioral strategy directly from neural data (Figure 5). Ramping activity persisted for up to tens of seconds (Figure 7A), demonstrating the same mechanism could scale to produce both patient and impulsive choices. These results support continuous frontal cortex ramping dynamics as a candidate mechanism underlying patch-leaving decisions.

Integrator models account for MVT deviations

MVT provides a normative solution for when to leave a depleting patch, whereas integrator models offer heuristic

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

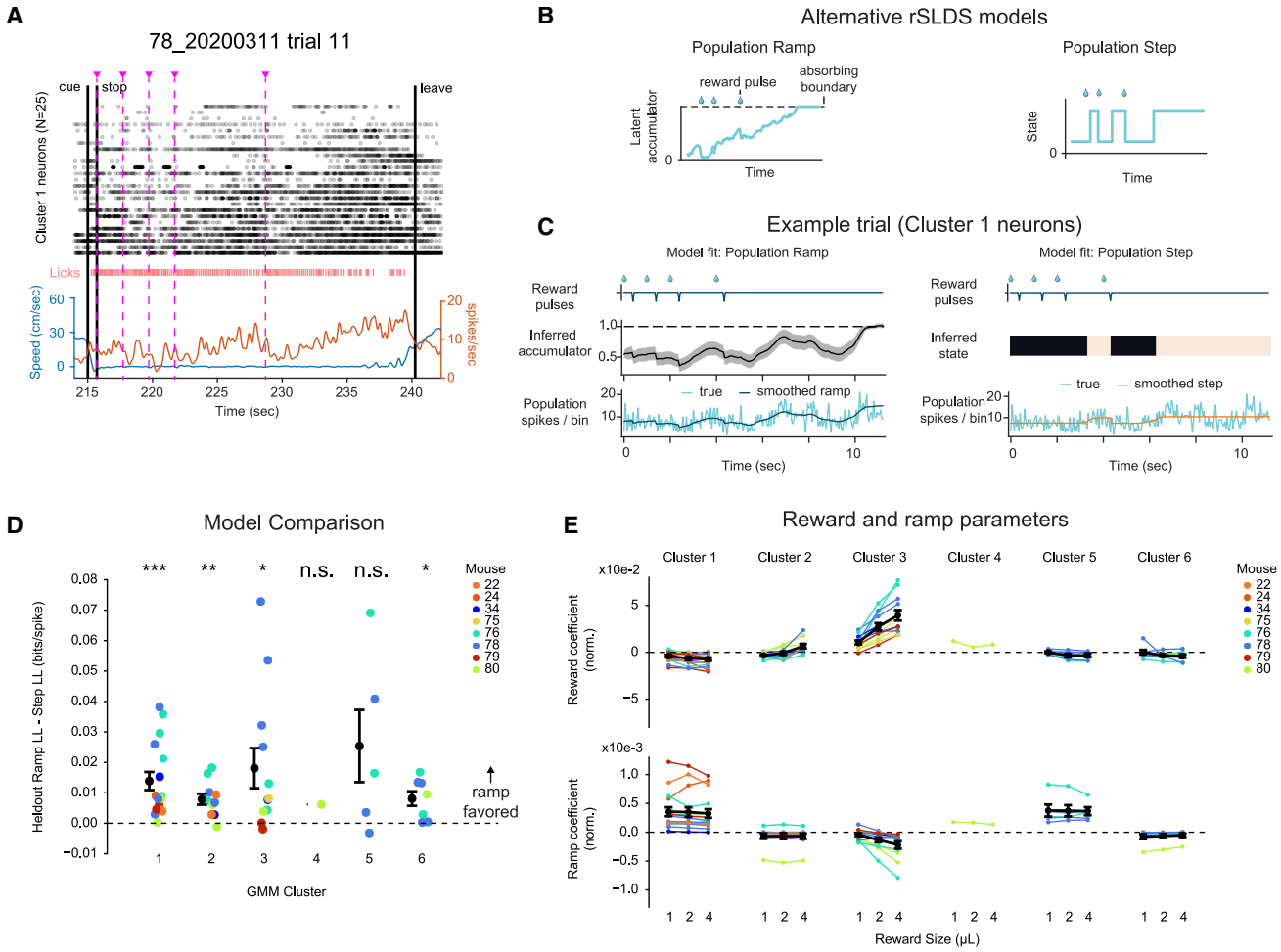


Figure 7. Functional clusters exhibit ramping activity

(A) Example trial showing simultaneously recorded cluster 1 neuron activity. Top: raster plot of cluster 1 neurons (n = 25). Middle: raster plot of mouse licks. Bottom: mouse speed (blue) and average firing rate of cluster 1 neurons (red). Magenta dashed lines: reward delivery (4 μL).

(B) Schematic of ramp and step models.

(C) Single-trial ramping and stepping model fits for cluster 1 neurons from an example session (80_20200317).

(D) Model comparison per GMM cluster. In (D) and (E), data points indicate sessions, colors indicate mice, and error bars indicate mean ± SEM across sessions. Stars indicate significance level of a t test versus zero per GMM cluster. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$; ns, not significant. LL: log likelihood.

(E) Reward coefficients and ramping slopes across reward size per GMM cluster.

approximations.[33] We focus on integrator models as candidate process-level explanations and show that they precisely capture the pattern of deviations from the normative solutions.

While mouse behavior qualitatively matched MVT (Figure 1H), our task design revealed quantitative discrepancies from theoretical expectations not readily explained by previous accounts—most notably systematic deviations from prediction 2. Because within-patch reward probability decayed exponentially with the same time constant across patch types, prediction 2 states that differences in PRT across patch types should not depend on the overall leaving threshold (Figure 3B). However, mice consistently violated this prediction, with PRT differences scaling with latent patience (Figures 2E, S2E, and S2H). While explanations such as nonlinear utility across reward sizes[42] could, in principle, account for observed deviations (Figure S3B), this would leave open the question of why the steepness of the utility function should scale in tandem with the overall threshold (Figures S3C–S3E). Alternative explanations, such as variability in risk sensitivity (cf.[43,54]) or biases in reward rate estimation,[20] elicit the same question of why that effect should covary with the leaving threshold. By contrast, we found that a single factor—latent scaling of the integration process—parsimoniously explains the covariance of relative waiting times across patch types with overall willingness to wait (Figures 3K and S3K).

Potential modulators of the latent patience state

While the current experiments cannot reveal what determines the latent patience state, we can rule out several simple explanations, including slow fluctuations in reward rate (Figures S2A–S2D). Satiation would be a sensible candidate,

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

but within-session latent state dynamics were heterogeneous across sessions, and even the direction of change varied (Figure 2C). Moreover, latent state varied across sessions (Figure 2D), whereas satiation primarily changed between the beginning and end of each session. Therefore, although satiation inevitably influences motivational state, it is also clearly not the sole factor.

The latent state we call patience is related to a broader literature on task engagement, arousal, and motivational state[55–61] and the idea that decision processes may be non-stationary.[62] Whereas other studies have focused on identifying discrete or abruptly changing latent behavioral states,[55,61,63,64] ours is a continuous, slowly changing variable,[60] likely influenced by neuromodulators, such as serotonin[65] or noradrenaline.[66]

In our behavioral modeling, we best captured the dependence of PRT on latent state by scaling ramping slopes (Figure S3F). However, the primary effect of latent state observed in our neural data was modulation of decoder output at patch stop, hinting at a DC offset (Figures S5F–S5H). Importantly, we observed the greatest variance in latent estimates across sessions and mice (Figure 2D), so future work tracking the same neurons across sessions is needed to better understand how patience modulates ramping activity.

A neural ramping dimension with partial reward resetting and memory of reward history

Decoding analysis identified a shared neural subspace containing ramping activity, from which models 1–3 DVs could be decoded similarly well (Figure 5). Opposite-signed ramping and reward responses were coupled, such that negative reward responses were observed even in the output of the model 1 decoder, which was trained not to respond to rewards (Figure 5K). Similarly, PCs with ramping activity on trials with no reward at $t = 1$ invariably showed opposite-signed reward responses on trials with reward at $t = 1$ (Figure 4D). While this does not disprove the existence of pure ramping signals without reward responses (e.g., cluster 5, Figure 6), it suggests that opposing ramping and reward responses—the core property of both models 2 and 3—are primary features difficult to separate in our neural data.

These results demonstrate how decoder failures can be as informative as successes. Similarly, the failure of the model 2 decoder output to reset following reward argues that fully resetting ramping dynamics are not easily extractable from the population activity. Instead, all three decoders retained a memory of reward history prior to the most recent reward (Figures 5M–5P). Since reward responses were not the only driver of decoder fits, this does not disprove the existence of a purely resetting ramping signal within our high-dimensional neural data, but it suggests the predominant ramping signals in our dataset do not have this property.

In the same spirit, we highlight an additional inconsistency between the decoder output and DVs at the time of patch leave (Figures S5D and S5E). Prior to patch leave, decoder output reaches a similar level across reward sizes, whereas the behavioral model DVs diverge. This is perhaps more indicative of a "decision bound" than a probabilistic leave decision as implemented in the behavioral models.

Reinforcement learning during foraging

Previous works analyzed foraging behavior in reinforcement learning (RL) frameworks.[42,67–69] We view integrator and RL models as different categories, not in tension with one another. Integrator and RL models share key commonalities. For one, variables in RL models often ramp over time. For example, the value of staying on a patch decays over time and increases with reward deliveries, mirroring our ramping DVs. Just as a mapping exists between integrators and MVT,[33] it may be possible to combine RL models and integrators, for example, by using RL to update integrator parameters.

Notably, over up to 30 days of training, we observed no change in relative PRT across patch types or mouse performance as measured by reward rate (Figures S1E and S1G). This suggests animals are not gradually learning the task's reward statistics, potentially ruling out RL models that feature such slow learning. Rather, the data are most consistent with mice implementing a default integration strategy, which may be adjusted by local reward history and other factors, such as patience.

The dependence of PRTs on previous reward size suggests mice track environment value by integrating recent reward history to modulate a threshold for patch leaving.[42,67] Experiments that explicitly vary environment value would be needed to rigorously test this hypothesis. Investigating how RL could optimize foraging decisions by tuning neural integration processes and/or synaptic weights based on reward history is a promising direction for future work.[70]

Implications of correlated behavioral and cognitive variables

Cognitive variables are notoriously difficult to separate from animal behavior. Indeed, several recent studies reported movement-related activity widely throughout the brain.[71,72] Task-related movements likely contribute to our ramping signals (Figures S4A and S4B) but cannot fully explain them, for the following reasons: (1) ramps were observed even in trials wherein mice remained stationary (Figure 7A), (2) there were significant non-zero GLM coefficients on accumulator variables even after movement-related activity had been regressed out (Figure 6), (3) ramping firing rates preceding leave time were similar across reward sizes despite different leaving speeds (Figures 6F and S1C), and (4) speed coding was qualitatively distinct in trial periods versus inter-trial-intervals (Figures S4C and S4D). An additional possibility is that correlations between DVs and behavior could, in part, reflect an embodied cognitive process.[73,74]

Although model 3 best explained mouse behavior in our task, DVs from each model could be similarly decoded from neural activity (Figure 5C), consistent with the notion of a "reservoir" of DVs in frontal cortex.[41] However, caution must be taken to consider potential contributions of task-related movements, such as licking, which is inherently more "model 2" (reset)-like. Excluding all movement-related activity (which is impracticable) risks discarding signals related to embodied cognitive processes. At the same time, it is circular to identify movements as "embodied cognition" because they match a particular model. Thus, we believe the neural data are of limited use in identifying the behavioral strategy. Instead, our argument for model 3

is behavioral (Figure 3), and we show that sufficient signals exist in neural activity to implement this model (Figure 5).

Ramping activity during decision-making

Ramping activity is prevalent during perceptual decisions across a range of organisms and brain areas.[30,53,75–77] It provides a mechanism for accumulating evidence,[30,53,78,79] making relative value decisions,[80] delineating episodic memory,[81] and determining action timing.[80,82–85] We observed ramping dynamics during naturalistic foraging, which can be understood as an evidence accumulation problem,[33] with time and rewards providing evidence for and against leaving, respectively. The flexibility of ramps to scale their slope and integrate categorically distinct variables in opposing directions, taken in the broader context, supports ramping activity as a ubiquitous feature of decision processes that unfold gradually over time.

RESOURCE AVAILABILITY

Lead contact

Requests for further information, resources, and reagents should be directed to and will be fulfilled by the lead contact, Naoshige Uchida (uchida@mcb.harvard.edu).

Materials availability

This study did not generate new, unique reagents.

Data and code availability

Data generated in this paper are available via figshare at https://doi.org/10.6084/m9.figshare.29802896. Custom analysis code is available at https://github.com/NaoUchida/Bukwich_Campbell_et_al_2025.

ACKNOWLEDGMENTS

We thank Ed Soucy of the Harvard Center for Brain Science Neurotechnology Core for engineering assistance, Emmanuel Garrison-Hooks for assistance with mouse training, and members of the Uchida lab for valuable discussions. We thank Dr. Matteo Carandini, Dr. Kenneth Harris, Dr. Andrew Peters, and other members of the Cortex lab for advice on Neuropixels recording. This work was supported by grants from the NIH Brain Initiative (U19NS113201 to S.W.L. and N.U.), the Simons Foundation (SCGB 697092 to S.W.L. and N.U.), the Sloan Foundation (S.W.L.), the McKnight Foundation (S.W.L.), NIH T32MH020017 (M.B.), NIH F32MH126505 (M.G.C.), and the Wu Tsai Neurosciences Institute (D.Z.).

AUTHOR CONTRIBUTIONS

M.B. and N.U. designed the foraging task. M.B. built behavioral and recording setups with assistance from H.R.K. M.B., M.G.C., L.K., and J.S. collected behavioral data. M.B., M.G.C., and J.S. analyzed behavior with feedback from N.U. M.B. and M.S.T. constructed integrator models with guidance from J.D. M.B., M.G.C., and L.K. collected Neuropixels data. M.G.C., M.B., and J.S. analyzed Neuropixels data with feedback from N.U. D.Z. performed recurrent linear dynamical systems modeling under the supervision of S.W.L. M.B., M.G.C., and N.U. wrote the manuscript. All authors reviewed the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice
- METHOD DETAILS
  - Surgery
  - Virtual reality setup
  - Patch-foraging paradigm in virtual reality
  - Electrophysiological recording and spike sorting
  - Histology
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Mouse inclusion criteria
  - Definition of latent state
  - Regression analyses modeling PRT
  - The Marginal Value Theorem (MVT)
  - Calculation of instantaneous expected reward rate
  - Behavioral model fitting
  - Diagnostic trial type groupings (e.g. RRR vs. R0R) and selection biases
  - Using behavioral models to patch-leaving times on single trials
  - Unit inclusion criteria
  - Spike waveforms
  - Generating z-scored firing rate traces for each unit
  - Principal components analysis (PCA) of neural activity
  - Definition of ramping principal components (PCs)
  - Correlation of single neuron firing rates with Model 3 decision variables
  - Linear models to predict behavioral decision variables from neural activity
  - Computing ramping slope of DVs and decoder output
  - Computing reward responses of DVs and decoder output
  - Computing dependence of reward response on pre-reward level
  - Measuring the effect of reward history on DVs and decoder output using diagnostic trial types
  - Predicting single-trial latent state from pre-stop neural activity
  - Generalized linear model (GLM) to predict neural activity from behavioral variables
  - Unsupervised clustering of GLM coefficients
  - Visualizing the population response of task-related neurons to diagnostic trial types
  - State space modeling of ramping and stepping dynamics

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.neuron.2025.07.008.

Received: October 3, 2024
Revised: April 2, 2025
Accepted: July 9, 2025

REFERENCES

1. Stephens, D.W., and Krebs, J.R. (1986). Foraging Theory (Princeton University Press). https://doi.org/10.2307/j.ctvs32s6b.
2. Kacelnik, A. (1984). Central Place Foraging in Starlings (Sturnus vulgaris). I. Patch Residence Time. J. Anim. Ecol. *53*, 283–299. https://doi.org/10.2307/4357.
3. Charnov, E.L. (1976). Optimal foraging, the marginal value theorem. Theor. Popul. Biol. *9*, 129–136. https://doi.org/10.1016/0040-5809(76)90040-x.

4. Pyke, G.H. (1984). Optimal Foraging Theory: A Critical Review. Annu. Rev. Ecol. Syst. *15*, 523–575. https://doi.org/10.1146/annurev.es.15.110184.002515.

5. Nonacs, P. (2001). State dependent behavior and the Marginal Value Theorem. Behav. Ecol. *12*, 71–83. https://doi.org/10.1093/oxfordjournals.beheco.a000381.

6. Mobbs, D., Trimmer, P.C., Blumstein, D.T., and Dayan, P. (2018). Foraging for foundations in decision neuroscience: insights from ethology. Nat. Rev. Neurosci. *19*, 419–427. https://doi.org/10.1038/s41583-018-0010-7.

7. Krebs, J.R., Kacelnik, A., and Taylor, P. (1978). Test of optimal sampling by foraging great tits. Nature *275*, 27–31. https://doi.org/10.1038/275027a0.

8. Pyke, G.H. (1978). Optimal Foraging in Hummingbirds: Testing the Marginal Value Theorem. Am. Zool. *18*, 739–752. https://doi.org/10.1093/icb/18.4.739.

9. Stephens, D.W., Brown, J.S., and Ydenberg, R.C. (2007). Foraging: Behavior and Ecology (University of Chicago Press). https://doi.org/10.7208/chicago/9780226772653.001.0001.

10. Hayden, B.Y., Pearson, J.M., and Platt, M.L. (2011). Neuronal basis of sequential foraging decisions in a patchy environment. Nat. Neurosci. *14*, 933–939. https://doi.org/10.1038/nn.2856.

11. Watanabe, Y.Y., Ito, M., and Takahashi, A. (2014). Testing optimal foraging theory in a penguin–krill system. Proc. Biol. Sci. *281*, 20132376. https://doi.org/10.1098/rspb.2013.2376.

12. Kane, G.A., James, M.H., Shenhav, A., Daw, N.D., Cohen, J.D., and Aston-Jones, G. (2022). Rat anterior cingulate cortex continuously signals decision variables in a patch foraging task. J. Neurosci. *42*, 5730–5744. https://doi.org/10.1523/JNEUROSCI.1940-21.2022.

13. Mellgren, R.L. (1982). FORAGING IN A SIMULATED NATURAL ENVIRONMENT: THERE'S A RAT LOOSE IN THE LAB. J. Exp. Anal. Behav. *38*, 93–100. https://doi.org/10.1901/jeab.1982.38-93.

14. Ydenberg, R.C. (1984). Great Tits and Giving-Up Times: Decision Rules for Leaving Patches. Behaviour *90*, 1–24. https://doi.org/10.1163/156853984X00533.

15. Marsh, B. (2004). Energetic state during learning affects foraging choices in starlings. Behav. Ecol. *15*, 396–399. https://doi.org/10.1093/beheco/arh034.

16. Wikenheiser, A.M., Stephens, D.W., and Redish, A.D. (2013). Subjective costs drive overly patient foraging strategies in rats on an intertemporal foraging task. Proc. Natl. Acad. Sci. USA *110*, 8308–8313. https://doi.org/10.1073/pnas.1220738110.

17. Carter, E.C., and Redish, A.D. (2016). Rats value time differently on equivalent foraging and delay-discounting tasks. J. Exp. Psychol. Gen. *145*, 1093–1101. https://doi.org/10.1037/xge0000196.

18. Kane, G.A., Bornstein, A.M., Shenhav, A., Wilson, R.C., Daw, N.D., and Cohen, J.D. (2019). Rats exhibit similar biases in foraging and intertemporal choice tasks. eLife *8*, e48429. https://doi.org/10.7554/eLife.48429.

19. McNamara, J. (1982). Optimal patch use in a stochastic environment. Theor. Popul. Biol. *21*, 269–288. https://doi.org/10.1016/0040-5809(82)90018-1.

20. Kilpatrick, Z.P., Davidson, J.D., and El Hady, A. (2021). Uncertainty drives deviations in normative foraging decision strategies. J. R. Soc. Interface *18*, 20210337. https://doi.org/10.1098/rsif.2021.0337.

21. McNair, J.N. (1982). Optimal Giving-Up Times and the Marginal Value Theorem. Am. Nat. *119*, 511–529. https://doi.org/10.1086/283929.

22. Krebs, J.R., Ryan, J.C., and Charnov, E.L. (1974). Hunting by expectation or optimal foraging? A study of patch use by chickadees. Anim. Behav. *22*, 953. https://doi.org/10.1016/0003-3472(74)90018-9.

23. Green, R.F. (1984). Stopping Rules for Optimal Foragers. Am. Nat. *123*, 30–43. https://doi.org/10.1086/284184.

24. Iwasa, Y., Higashi, M., and Yamamura, N. (1981). Prey Distribution as a Factor Determining the Choice of Optimal Foraging Strategy. Am. Nat. *117*, 710–723. https://doi.org/10.1086/283754.

25. Waage, J.K. (1979). Foraging for Patchily-Distributed Hosts by the Parasitoid, Nemeritis canescens. J. Anim. Ecol. *48*, 353–371. https://doi.org/10.2307/4166.

26. Usher, M., and McClelland, J.L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. Psychol. Rev. *108*, 550–592. https://doi.org/10.1037/0033-295x.108.3.550.

27. Ratcliff, R., and McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. Neural Comput. *20*, 873–922. https://doi.org/10.1162/neco.2008.12-06-420.

28. Bogacz, R., Brown, E., Moehlis, J., Holmes, P., and Cohen, J.D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. Psychol. Rev. *113*, 700–765. https://doi.org/10.1037/0033-295X.113.4.700.

29. Smith, P.L., and Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. Trends Neurosci. *27*, 161–168. https://doi.org/10.1016/j.tins.2004.01.006.

30. Gold, J.I., and Shadlen, M.N. (2007). The neural basis of decision making. Annu. Rev. Neurosci. *30*, 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038.

31. Brunton, B.W., Botvinick, M.M., and Brody, C.D. (2013). Rats and Humans Can Optimally Accumulate Evidence for Decision-Making. Science *340*, 95–98. https://doi.org/10.1126/science.1233912.

32. Drugowitsch, J., Moreno-Bote, R., Churchland, A.K., Shadlen, M.N., and Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. J. Neurosci. *32*, 3612–3628. https://doi.org/10.1523/JNEUROSCI.4010-11.2012.

33. Davidson, J.D., and El Hady, A. (2019). Foraging as an evidence accumulation process. PLOS Comput. Biol. *15*, e1007060. https://doi.org/10.1371/journal.pcbi.1007060.

34. Rudebeck, P.H., and Izquierdo, A. (2022). Foraging with the frontal cortex: A cross-species evaluation of reward-guided behavior. Neuropsychopharmacology *47*, 134–146. https://doi.org/10.1038/s41386-021-01140-0.

35. Kolling, N., Behrens, T.E., Mars, R.B., and Rushworth, M.F. (2012). Neural mechanisms of foraging. Science *336*, 95–98. https://doi.org/10.1126/science.1216930.

36. McGuire, J.T., and Kable, J.W. (2015). Medial prefrontal cortical activity reflects dynamic re-evaluation during voluntary persistence. Nat. Neurosci. *18*, 760–766. https://doi.org/10.1038/nn.3994.

37. Bari, B.A., Grossman, C.D., Lubin, E.E., Rajagopalan, A.E., Cressy, J.I., and Cohen, J.Y. (2019). Stable Representations of Decision Variables for Flexible Behavior. Neuron *103*, 922–933.e7. https://doi.org/10.1016/j.neuron.2019.06.001.

38. Murakami, M., Shteingart, H., Loewenstein, Y., and Mainen, Z.F. (2017). Distinct Sources of Deterministic and Stochastic Components of Action Timing Decisions in Rodent Frontal Cortex. Neuron *94*, 908–919.e7. https://doi.org/10.1016/j.neuron.2017.04.040.

39. Vertechi, P., Lottem, E., Sarra, D., Godinho, B., Treves, I., Quendera, T., Oude Lohuis, M.N., and Mainen, Z.F. (2020). Inference-Based Decisions in a Hidden State Foraging Task: Differential Contributions of Prefrontal Cortical Areas. Neuron *106*, 166–176.e6. https://doi.org/10.1016/j.neuron.2020.01.017.

40. Lak, A., Costa, G.M., Romberg, E., Koulakov, A.A., Mainen, Z.F., and Kepecs, A. (2014). Orbitofrontal cortex is required for optimal waiting based on decision confidence. Neuron *84*, 190–201. https://doi.org/10.1016/j.neuron.2014.08.039.

41. Cazettes, F., Mazzucato, L., Murakami, M., Morais, J.P., Augusto, E., Renart, A., and Mainen, Z.F. (2023). A reservoir of foraging decision variables in the mouse brain. Nat. Neurosci. *26*, 840–849. https://doi.org/10.1038/s41593-023-01305-8.

42. Constantino, S.M., and Daw, N.D. (2015). Learning the opportunity cost of time in a patch-foraging task. Cogn. Affect. Behav. Neurosci. *15*, 837–853. https://doi.org/10.3758/s13415-015-0350-y.

43. Constantinople, C.M., Piet, A.T., and Brody, C.D. (2019). An Analysis of Decision under Risk in Rats. Curr. Biol. *29*, 2066–2074.e5. https://doi.org/10.1016/j.cub.2019.05.013.

44. Glimcher, P.W., and Fehr, E. (2013). Neuroeconomics: Decision Making and the Brain, Second Edition (Academic Press).

45. Wajnberg, E., Bernhard, P., Hamelin, F., and Boivin, G. (2006). Optimal patch time allocation for time-limited foragers. Behav. Ecol. Sociobiol. *60*, 1–10. https://doi.org/10.1007/s00265-005-0131-7.

46. Carmel, Y., and Ben-Haim, Y. (2005). Info-Gap Robust-Satisficing Model of Foraging Behavior: Do Foragers Optimize or Satisfice? Am. Nat. *166*, 633–641. https://doi.org/10.1086/491691.

47. Shadlen, M.N., and Kiani, R. (2013). Decision Making as a Window on Cognition. Neuron *80*, 791–806. https://doi.org/10.1016/j.neuron.2013.10.047.

48. Elber-Dorozko, L., and Loewenstein, Y. (2018). Striatal action-value neurons reconsidered. eLife *7*, e34248. https://doi.org/10.7554/eLife.34248.

49. Shahidi, N., Franch, M., Parajuli, A., Schrater, P., Wright, A., Pitkow, X., and Dragoi, V. (2024). Population coding of strategic variables during foraging in freely moving macaques. Nat. Neurosci. *27*, 772–781. https://doi.org/10.1038/s41593-024-01575-w.

50. Latimer, K.W., Yates, J.L., Meister, M.L., Huk, A.C., and Pillow, J.W. (2015). Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. Science *349*, 184–187. https://doi.org/10.1126/science.aaa4056.

51. Zoltowski, D.M., Latimer, K.W., Yates, J.L., Huk, A.C., and Pillow, J.W. (2019). Discrete stepping and nonlinear ramping dynamics underlie spiking responses of LIP neurons during decision-making. Neuron *102*, 1249–1258.e10. https://doi.org/10.1016/j.neuron.2019.04.031.

52. Trautmann, E.M., Hesse, J.K., Stine, G.M., Xia, R., Zhu, S., O'Shea, D.J., Karsh, B., Colonell, J., Lanfranchi, F.F., Vyas, S., et al. (2023). Large-scale high-density brain-wide neural recording in nonhuman primates. Preprint at bioRxiv, 2023.02.01.526664. https://doi.org/10.1101/2023.02.01.526664.

53. Steinemann, N.A., Stine, G.M., Trautmann, E.M., Zylberberg, A., Wolpert, D.M., and Shadlen, M.N. (2024). Direct observation of the neural computations underlying a single decision. eLife *12*, RP90859. https://doi.org/10.7554/eLife.90859.

54. Eisenreich, B.R., Hayden, B.Y., and Zimmermann, J. (2019). Macaques are risk-averse in a freely moving foraging task. Sci. Rep. *9*, 15091. https://doi.org/10.1038/s41598-019-51442-z.

55. Hennig, J.A., Oby, E.R., Golub, M.D., Bahureksa, L.A., Sadtler, P.T., Quick, K.M., Ryu, S.I., Tyler-Kabara, E.C., Batista, A.P., Chase, S.M., et al. (2021). Learning is shaped by abrupt changes in neural engagement. Nat. Neurosci. *24*, 727–736. https://doi.org/10.1038/s41593-021-00822-8.

56. Pettit, N.L., Yuan, X.C., and Harvey, C.D. (2022). Hippocampal place codes are gated by behavioral engagement. Nat. Neurosci. *25*, 561–566. https://doi.org/10.1038/s41593-022-01050-4.

57. Allen, W.E., Chen, M.Z., Pichamoorthy, N., Tien, R.H., Pachitariu, M., Luo, L., and Deisseroth, K. (2019). Thirst regulates motivated behavior through modulation of brainwide neural population dynamics. Science *364*, 253. https://doi.org/10.1126/science.aav3932.

58. Vinck, M., Batista-Brito, R., Knoblich, U., and Cardin, J.A. (2015). Arousal and Locomotion Make Distinct Contributions to Cortical Activity Patterns and Visual Encoding. Neuron *86*, 740–754. https://doi.org/10.1016/j.neuron.2015.03.028.

59. Aston-Jones, G., and Cohen, J.D. (2005). Adaptive gain and the role of the locus coeruleus–norepinephrine system in optimal performance. J. Comp. Neurol. *493*, 99–110. https://doi.org/10.1002/cne.20723.

60. Cowley, B.R., Snyder, A.C., Acar, K., Williamson, R.C., Yu, B.M., and Smith, M.A. (2020). Slow Drift of Neural Activity as a Signature of Impulsivity in Macaque Visual and Prefrontal Cortex. Neuron *108*, 551–567.e8. https://doi.org/10.1016/j.neuron.2020.07.021.

61. Ashwood, Z.C., Roy, N.A., Stone, I.R., International Brain Laboratory, Urai, A.E., Churchland, A.K., Pouget, A., and Pillow, J.W. (2022). Mice alternate between discrete strategies during perceptual decision-making. In Nat. Neurosci., *25* (Z.C.), pp. 201–212. https://doi.org/10.1038/s41593-021-01007-z.

62. Roy, N.A., Bak, J.H., International; Brain Laboratory, Akrami, A., Brody, C.D., and Pillow, J.W. (2021). Extracting the dynamics of behavior in sensory decision-making experiments. Neuron *109*, 597–610.e596. https://doi.org/10.1016/j.neuron.2020.12.004.

63. Bolkan, S.S., Stone, I.R., Pinto, L., Ashwood, Z.C., Iravedra Garcia, J.M., Herman, A.L., Singh, P., Bandi, A., Cox, J., Zimmerman, C.A., et al. (2022). Opponent control of behavior by dorsomedial striatal pathways depends on task demands and internal state. Nat. Neurosci. *25*, 345–357. https://doi.org/10.1038/s41593-022-01021-9.

64. Calhoun, A.J., Pillow, J.W., and Murthy, M. (2019). Unsupervised identification of the internal states that shape natural behavior. Nat. Neurosci. *22*, 2040–2049. https://doi.org/10.1038/s41593-019-0533-x.

65. Lottem, E., Banerjee, D., Vertechi, P., Sarra, D., Lohuis, M.O., and Mainen, Z.F. (2018). Activation of serotonin neurons promotes active persistence in a probabilistic foraging task. Nat. Commun. *9*, 1000. https://doi.org/10.1038/s41467-018-03438-y.

66. Kane, G.A., Vazey, E.M., Wilson, R.C., Shenhav, A., Daw, N.D., Aston-Jones, G., and Cohen, J.D. (2017). Increased locus coeruleus tonic activity causes disengagement from a patch-foraging task. Cogn. Affect. Behav. Neurosci. *17*, 1073–1083. https://doi.org/10.3758/s13415-017-0531-y.

67. Shuvaev, S., Starosta, S., Kvitsiani, D., Kepecs, A., and Koulakov, A.A. (2020). R-learning in actor-critic model offers a biologically relevant mechanism for sequential decision-making. Adv. Neural Inf. Process. Syst. *33*, 18872–18882.

68. Wispinski, N.J., Butcher, A., Mathewson, K.W., Chapman, C.S., Botvinick, M.M., and Pilarski, P.M. (2023). Adaptive patch foraging in deep reinforcement learning agents. Preprint at arXiv, 2210.08085.

69. Kolling, N., and Akam, T. (2017). (Reinforcement?) Learning to forage optimally. Curr. Opin. Neurobiol. *46*, 162–169. https://doi.org/10.1016/j.conb.2017.08.008.

70. Pereira-Obilinovic, U., Hou, H., Svoboda, K., and Wang, X.-J. (2024). Brain mechanism of foraging: Reward-dependent synaptic plasticity versus neural integration of values. Proc. Natl. Acad. Sci. USA *121*, e2318521121. https://doi.org/10.1073/pnas.2318521121.

71. Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C.B., Carandini, M., and Harris, K.D. (2019). Spontaneous behaviors drive multidimensional, brainwide activity. Science *364*, 255. https://doi.org/10.1126/science.aav7893.

72. Musall, S., Kaufman, M.T., Juavinett, A.L., Gluf, S., and Churchland, A.K. (2019). Single-trial neural dynamics are dominated by richly varied movements. Nat. Neurosci. *22*, 1677–1686. https://doi.org/10.1038/s41593-019-0502-4.

73. Varela, F.J., Rosch, E., and Thompson, E. (1991). The Embodied Mind: Cognitive Science and Human Experience (The MIT Press). https://doi.org/10.7551/mitpress/6730.001.0001.

74. Hasnain, M.A., Birnbaum, J.E., Ugarte Nunez, J.L., Hartman, E.K., Chandrasekaran, C., and Economo, M.N. (2025). Separating cognitive and motor processes in the behaving mouse. Nat. Neurosci. *28*, 640–653. https://doi.org/10.1038/s41593-024-01859-1.

75. Roitman, J.D., and Shadlen, M.N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J. Neurosci. *22*, 9475–9489. https://doi.org/10.1523/JNEUROSCI.22-21-09475.2002.

76. Yartsev, M.M., Hanks, T.D., Yoon, A.M., and Brody, C.D. (2018). Causal contribution and dynamical encoding in the striatum during evidence accumulation. eLife *7*, e34929. https://doi.org/10.7554/eLife.34929.

77. Bahl, A., and Engert, F. (2020). Neural circuits for evidence accumulation and decision making in larval zebrafish. Nat. Neurosci. *23*, 94–102. https://doi.org/10.1038/s41593-019-0534-9.

78. Scott, B.B., Constantinople, C.M., Akrami, A., Hanks, T.D., Brody, C.D., and Tank, D.W. (2017). Fronto-parietal cortical circuits encode accumulated evidence with a diversity of timescales. Neuron *95*, 385–398.e5. https://doi.org/10.1016/j.neuron.2017.06.013.

79. Khilkevich, A., Lohse, M., Low, R., Orsolic, I., Bozic, T., Windmill, P., and Mrsic-Flogel, T.D. (2024). Brain-wide dynamics linking sensation to action during decision-making. Nature *634*, 890–900. https://doi.org/10.1038/s41586-024-07908-w.

80. Vijayan, V., Wang, F., Wang, K., Chakravorty, A., Adachi, A., Akhlaghpour, H., Dickson, B.J., and Maimon, G. (2023). A rise-to-threshold process for a relative-value decision. Nature *619*, 563–571. https://doi.org/10.1038/s41586-023-06271-6.

81. Tsao, A., Sugar, J., Lu, L., Wang, C., Knierim, J.J., Moser, M.-B., and Moser, E.I. (2018). Integrating time from experience in the lateral entorhinal cortex. Nature *561*, 57–62. https://doi.org/10.1038/s41586-018-0459-6.

82. Maimon, G., and Assad, J.A. (2006). A cognitive signal for the proactive timing of action in macaque LIP. Nat. Neurosci. *9*, 948–955. https://doi.org/10.1038/nn1716.

83. Murakami, M., Vicente, M.I., Costa, G.M., and Mainen, Z.F. (2014). Neural antecedents of self-initiated actions in secondary motor cortex. Nat. Neurosci. *17*, 1574–1582. https://doi.org/10.1038/nn.3826.

84. Hamilos, A.E., Spedicato, G., Hong, Y., Sun, F., Li, Y., and Assad, J.A. (2021). Slowly evolving dopaminergic activity modulates the moment-to-moment probability of reward-related self-timed movements. eLife *10*, e62583. https://doi.org/10.7554/eLife.62583.

85. Hanks, T., Kiani, R., and Shadlen, M.N. (2014). A neural mechanism of speed-accuracy tradeoff in macaque area LIP. eLife *3*, e02260. https://doi.org/10.7554/eLife.02260.

86. Aronov, D., and Tank, D.W. (2014). Engagement of neural circuits underlying 2D spatial navigation in a rodent virtual reality system. Neuron *84*, 442–456. https://doi.org/10.1016/j.neuron.2014.08.042.

87. Kim, H.R., Malik, A.N., Mikhael, J.G., Bech, P., Tsutsui-Kimura, I., Sun, F., Zhang, Y., Li, Y., Watabe-Uchida, M., Gershman, S.J., et al. (2020). A unified framework for dopamine signals across timescales. Cell *183*, 1600–1616.e25. https://doi.org/10.1016/j.cell.2020.11.013.

88. Jun, J.J., Steinmetz, N.A., Siegle, J.H., Denman, D.J., Bauza, M., Barbarits, B., Lee, A.K., Anastassiou, C.A., Andrei, A., Aydın, Ç., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. Nature *551*, 232–236. https://doi.org/10.1038/nature24636.

89. Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M., and Kenneth, D. (2016). Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. Preprint at bioRxiv, 061481. https://doi.org/10.1101/061481.

90. Pachitariu, M., Sridhar, S., and Stringer, C. (2023). Solving the spike sorting problem with Kilosort. Preprint at bioRxiv, 2023.2001.2007.523036. https://doi.org/10.1101/2023.01.07.523036.

91. Shamash, P., Carandini, M., Harris, K.D., and Steinmetz, N.A. (2018). A tool for analyzing electrode tracks from slice histology. Preprint at bioRxiv, 447995. https://doi.org/10.1101/447995.

92. Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. J. Stat. Softw. *33*, 1–22. https://doi.org/10.18637/jss.v033.i01.

93. Tay, J.K., Narasimhan, B., and Hastie, T. (2023). Elastic Net Regularization Paths for All Generalized Linear Models. J. Stat. Softw. *106*, 1. https://doi.org/10.18637/jss.v106.i01.

94. Driscoll, L.N., Pettit, N.L., Minderer, M., Chettih, S.N., and Harvey, C.D. (2017). Dynamic Reorganization of Neuronal Activity Patterns in Parietal Cortex. Cell *170*, 986–999.e16. https://doi.org/10.1016/j.cell.2017.07.021.

95. Linderman, S., Johnson, M., Miller, A., Adams, R., Blei, D., and Paninski, L. (2017). Bayesian learning and inference in recurrent switching linear dynamical systems. In Artificial Intelligence and Statistics. PMLR, pp. 914-922., pp. 914–922.

96. Zoltowski, D., Pillow, J., and Linderman, S. (2020). A general recurrent state space framework for modeling neural dynamics during decision-making. In International Conference on Machine Learning. PMLR, pp. 11680–11691.

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
Mouse: C57BL/6J	The Jackson Laboratory	RRID:IMSR_JAX:000664
Mouse: C57BL/6N	Charles River	RRID:IMSR_CRL:027
Software and algorithms		
Custom analysis code	This paper	https://github.com/NaoUchida/Bukwich_Campbell_et_al_2025
VirMEn (Virtual Reality MATLAB Engine)	Dmitriy Aronov[86]	NA
MATLAB (version 2019b or 2022a for analysis, 2011b for VirMEn)	MathWorks	https://www.mathworks.com
R (version 4.2.2)	The R Project for Statistical Computing	https://www.r-project.org
SpikeGLX (version 20190413-phase3B2)	Bill Karsh	https://github.com/billkarsh/SpikeGLX
Kilosort (versions 2 and 3)	Marius Pachitariu	https://github.com/MouseLand/Kilosort
glmnet (version 4-1.8)	Stanford University	https://glmnet.stanford.edu
Deposited data		
Behavioral and Neuropixels data	This paper	https://doi.org/10.6084/m9.figshare.29802896
Other		
Isosol (Isoflurane, USP)	Vedco	N/A
Ketoprofen (for analgesia)	Patterson Veterinary	Cat #07-803-7389
Buprenorphine	Patterson Veterinary	Cat #07-850-2280
Neuropixels 1.0 probes and acquisition system	Imec	https://www.neuropixels.org/

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

A total of 30 adult mice (C57/BL6j, 20 male, 10 female, age 2-5 months) were used in the experiments. Mice were housed on a 12 h dark/12 h light cycle (dark from 07:00 to 19:00) and performed the task in the dark period. All procedures were performed in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and approved by the Harvard Animal Care and Use Committee.

METHOD DETAILS

Surgery

Mice were anaesthetized using isoflurane (4% induction, 1-2% maintenance). For behavioral experiments, a custom titanium headplate was attached to the skull with Metabond (Parkell). For Neuropixels recordings, fiducial marks were made at the target sites for probe insertion using a fine-tipped pen, and a ground pin was inserted into the skull above contralateral cortex and attached with Metabond.

Virtual reality setup

Virtual reality setups were identical to those used in.[87] Three monitors (width 53 cm, height 30 cm) were placed in front and on either side of the animal. Virtual reality scenes were generated using VirMEn software in MATLAB[86] on a workstation computer (DELL Precision 5810). Mice were head-fixed on a cylindrical Styrofoam treadmill (diameter 20.3 cm, width 10.4 cm). The rotational velocity of the treadmill was recorded using a rotary encoder. The output pulses of the encoder were converted into continuous velocity signal using custom Arduino code running on a microprocessor (Teensy 3.2). Velocity was integrated within VirMEn to compute position. Water was delivered to the mouse from a spout placed in front of the mouse's mouth. Licks were monitored using an infrared sensor (OPB819Z, TT electronics). Voltage signals from the rotary encoder were digitized and recorded on the virtual reality computer using

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

a data acquisition system (PCIe-6323, National Instruments). Water timing and amount was controlled using a solenoid valve (LHDA 1221111H, The Lee Company) and a switch (2N7000, On Semiconductor), with TTL pulses generated by the virtual reality computer via the PCIe-6323 data acquisition card.

Patch-foraging paradigm in virtual reality

Mice were water-restricted to maintain body weight >85% of their baseline weight to ensure task motivation while maintaining health. Additional water supplementation was provided shortly following training sessions with a total of task and post-task water ranging from 1.0-2.4mL per day, determined individually per mouse. Mice that showed task motivation at higher weights were given greater amounts of total water supplementation accordingly to minimize any stress from water restriction. Body weights ranged from 85-105% of baseline pre-restriction weight across mice.

Early training

Mice were habituated to the treadmill prior to initial head-fixation, delivering occasional water droplets from the reward spout while mice were free to explore on top of the treadmill, which was held in place to prevent rotation. After mice showed consistent interest over 1-2 sessions in consuming water droplets from the spout (2-3 days of habituation), head-fixation was begun in the following session. Initial head-fixed sessions were run for durations of 5-10 minutes, with gradual increases in session duration of typically 5-10 minutes across consecutive sessions, until reaching a max session duration of 45-75 minutes, varied per mouse based on how long they were observed to show task engagement as evident by running and licking behavior. Mice were initially head-fixed for 2-3 consecutive daily sessions before being given the following day off, after which the next set of consecutive daily sessions was increased by +1-2 days, followed by another day off, and continued in that manner until reaching a typical training schedule of 6 days on, 1 day off.

Initial head-fixed sessions were run in the dark. During those sessions, water rewards were delivered within 60 sec intervals regardless of whether mice ran on the treadmill, to maintain positive reinforcement to rig fixation. To reinforce running behavior, mice were able to receive additional water rewards more readily via running, with water rewards delivered after traversing distances that were gradually increased across rewards.

After mice demonstrated consistent running behavior during the dark training (typically 2-3 sessions), mice began their initial training on the linear virtual track. After running a fixed distance, a visual "proximity cue" was illuminated on the track, followed by an automatic 1, 2, or 4 μL (randomly varied) water reward delivery after a delay of 0.25-1.0s following cue onset. Running distance required to trigger the proximity cue and subsequent reward was increased after each block of 20 trials was completed. When mice completed >40 trials in a training session, the following session was started with a greater required running distance. This training stage served to teach mice that they must run to obtain rewards and that the proximity cue indicated an upcoming reward delivery. We determined that mice had successfully learned the association between the proximity cue and reward after observing consistent behavioral signs of reward expectation in response to proximity cue onset (reduced running speed and licking). After mice progressed through increasing levels of running distance across sessions and demonstrated anticipatory slowdown and licking in response to the proximity cue onset prior to reward delivery, mice were transferred to the full version of the foraging task.

Full task

Mice traversed a virtual linear track wherein the periodic appearance of a visual proximity cue indicated that they could halt their running at that time to enter a virtual patch, in which water rewards could be obtained. Visuals for the linear track and proximity cues were the same as in earlier training. Each trial began with mice running along the track (travel time), and after running a fixed distance (∼2m; for two mice [34 and 39], the VR gain was set such that travel distance was ∼60 cm, with all other distances scaled accordingly), a proximity cue was illuminated. Contrary to earlier training, no automatic reward delivery occurred following the onset of the proximity cue. Instead, mice were required to pause their running on the treadmill while the cue was displayed, maintaining a near stationary positioning for ∼480ms, to obtain water reward(s). When mice achieved this stopping criteria, the surrounding visuals changed to indicate they had entered a patch, at which time an initial water reward was always delivered, droplet size = 1, 2, or 4 μL, depending on patch type. If mice failed to reduce their speed sufficiently to meet the stopping criteria and continued running through the proximity cue (distance =∼40cm), the cue was extinguished and a new trial started at the beginning of the track.

Nine patch types were pseudorandomly varied across trials (trial types balanced in blocks of 27 trials; Figure S1A). Visual cues were the same across all trials, irrespective of patch type. Patches differed in their reward statistics, with size and frequency of reward deliveries varied across patch types. Three values were used for each, yielding nine unique combinations for patch type. Reward sizes were 1, 2, or 4 μL, with size held constant across all rewards within a given patch. After the initial deterministic reward that was always delivered upon patch entry, additional reward events were drawn probabilistically following each one-second interval that elapsed on the patch. Values used for determining the probability of each reward delivery were dictated by an exponential decay function, which was scaled by reward frequency per patch type, with $N_0 = .125, .25, or .5$.

$$h(t) = N_0 e^{-\frac{t}{\tau}} \quad \text{(Equation 1)}$$

A time constant of 8 seconds was used across all patches ($\tau = 8s$). To obtain a discrete probability for reward delivery following each one-second interval, we integrate over the decay function:

$$P(reward)_t = \int_{t-1}^{t} N_0 e^{-\frac{s}{\tau}} ds \quad \text{(Equation 2)}$$

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

Mice were free to leave each patch at any time. Patches were delineated in virtual space, and leaving was triggered by crossing the virtual boundary (~20cm from patch entry location). Leaving the patch marked the start of a new trial at the beginning of the track.

Electrophysiological recording and spike sorting

Spiking data were collected using Neuropixels 1.0 single shank probes.[88] Neuropixels data were recorded with SpikeGLX (https://billkarsh.github.io/SpikeGLX/). Craniotomies were performed the day before recording and covered with Kwik-Cast (World Precision Instruments), and mice were allowed to recover overnight. On the day of recording, a 3d-printed piece was placed over the head of the mouse, blocking from the mouse's sight the experimenter manipulating the probe above the mouse's head. This was critical to maintain good behavioral performance on Neuropixels recording days.

Neuropixels probes were lowered using a Thorlabs micromanipulator (PT1-Z8) at 9 μm/sec. After reaching the target depth, the probe was allowed to settle for 30 minutes prior to starting the recording. Recordings lasted 34 minutes (minimum) to 74 minutes (maximum) and were terminated once the experimenter determined the mouse was no longer engaging with the task (e.g., running through patches without stopping).

Neuropixels data were spike sorted offline with Kilosort 2 or 3,[89,90] followed by manual curation in Phy (https://github.com/cortex-lab/phy).

Histology

After the last Neuropixels recording session, mice were perfused with phosphate buffered saline (PBS) followed by 4% paraformaldehyde in PBS. The brains were cut in 100 μm coronal sections using a vibratome (Leica). Brain sections were mounted on glass slides and stained with 4′, 6-diamidino-2-phenylindole (DAPI, Vectashield). Slides were imaged with a fluorescence microscope (Zeiss Axio Scan.Z1). Probe tracks were identified in histology images and aligned to a reference atlas using SHARP-Track[91] or software written by Dr. Andrew Peters (https://github.com/petersaj/AP_histology).

QUANTIFICATION AND STATISTICAL ANALYSIS

Unless otherwise noted, data were analyzed with custom code written in MATLAB (version 2019b or 2022a).

Mouse inclusion criteria

A total of 30 mice were trained for these experiments, 22 of which (73.3%) passed our inclusion criteria for analysis. For inclusion, mice were required to progress through the early training stage to the full task and complete at least 10 sessions with at least 20 trials in each. Sessions were included after mice achieved >50% of their maximal session reward rate. 2 mice failed to progress through the initial training, and 3 mice failed to complete the required number of eligible sessions after progressing to the full task. Of the remaining 25 mice that completed the required number of sessions, 3 were excluded from the analysis group due to their PRT varying randomly across patch types, indicating an indifference to reward statistics.

Definition of latent state

To obtain an approximate measure of mice's latent levels of patience across trials, we calculated a weighted average of PRT over successive trials. For each session, we ordered PRT by trial number. We then iterated over trials to compute the latent value per trial. First, we removed the PRT value associated with that trial, to prevent biasing the latent estimation. We then smoothed PRT across the session by applying a Gaussian filter (standard deviation = 10 trials, filter length = 30 trials). Artificial trials were added prior to the first trial of the session and following the last trial, using mean PRT from the first and last ten trials in the array, respectively, to avoid the filter being cutoff. After obtaining the smoothed PRTs across trials, we calculated the mean of the trials before and after the trial of interest and used that value as its latent patience estimate.

Regression analyses modeling PRT

We fit a generalized linear mixed-effects model (GLME) with a log link function to predict PRT per patch type across subjects (Figure 1G), given by the following equation:

$$PRT = e^{\beta_{0_i} + \beta_{size} * x_1 + \beta_{freq} * x_2} \quad \text{(Equation 3)}$$

β_{o_i} serves as a random effect, fit with a separate value per mouse (i). β_{size} and β_{freq} are fixed effects, each with a single parameter shared across the population. $x1$ and $x2$ indicate reward size and frequency, respectively, for each patch. The values for the three reward sizes and frequencies were standardized to $[1, 2, and\ 4]$ to allow for interpretation of size and frequency coefficients on the same scale. The fit was performed using the 'fitglme' function in MATLAB version R2022a with the Statistics and Machine Learning Toolbox. The model was fit using the Laplace approximation, and the optimization was performed with 'fminsearch'.

Additional GLME fits (same equation) were performed using 5-fold cross-validation and compared with alternative fits in which the trial labels (reward size, reward frequency, reward size and frequency, or mouse ID) were shuffled in the training set (Figure S1H). Trials were split into folds by ordering patches per subject and labeling them 1-5 in serial. The GLME was then

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

fit, as above, on the training folds. For the cross-validation and shuffles, GLME-predicted PRTs were calculated for patches in the held-out test folds.

We similarly fit a GLM using a log link function to predict PRT across patches by removing the per subject random effects term and adding a term to scale each patch by its latent 'patience' estimate (Figure 2E), given by the following equation:

$$PRT = e^{\beta_0+\beta_{size}*x_1+\beta_{freq}*x_2+\beta_{latent}*x_3} \quad \text{(Equation 4)}$$

In this formulation, $\beta_0$is the log-intercept. $x1$ and $x2$ are the standardized values for reward size and frequency for each patch, same as in the GLME. $x3$ is the log of the latent 'patience' estimate for each patch. A single set of values for the four free parameters was fit across the population. We next performed this same GLM fitting procedure, but fitting a separate set of parameters individually for each mouse (Figure S2I). We also performed several linear regressions, separately per mouse (Figures S2E, S2F, and S2H), in which the values for patch reward size and frequency were standardized the same as above.

To examine the influence of reward history on PRTs (Figures 1H, S1I, S2B, and S2C), we performed linear regression on normalized PRTs (PRT divided by mean PRT within each session), in an attempt to remove session-to-session variability and make coefficients more comparable across mice. Reward size was 1, 2, or 4 μL, reward frequency was coded as 1, 2, or 4 (low, medium, high), and intercept terms were included in the model. Normalized PRTs were linearly regressed on current reward size and frequency and either a 5-trial running average of previous reward sizes and frequencies (Figures 1H and S2B) or individual previous trials' reward sizes and frequencies (Figure S1I). Models were fit using the function 'fitlm' in MATLAB, separately for each session, then coefficients were averaged by mouse and reported at the mouse level in figures.

The Marginal Value Theorem (MVT)

In this study, we tested various predictions of MVT. Following the derivation of MVT by Charnov[3] and Stephens and Krebs,[1] here we describe the assumptions and predictions of MVT. We consider n patch types ($n = 1,2,3,\ldots,n$). A patch type i is defined by $r_i(t)$, the instantaneous reward rate as a function of the time on the patch, t. We assume that the animal encounters each patch type with the probability λ_i. We define $R_i(t_i)$ as the cumulative energy intake from staying in a patch type i for the duration t_i, where $R_i(t_i) = \int_0^{t_i} r_i(t)dt$. We also consider $E_{S,i}$, which is the energy cost per unit time while searching in a patch type i. The gain function $g_i(t)$ in a patch type i can be obtained by:

$$g_i(t_i) = R_i(t_i) - E_{S,i} \cdot t_i \quad \text{(Equation 5)}$$

We assume that the derivative of $g_i(t_i)$ is a monotonically decreasing function over time for all patch types.

We consider the time (T_T) and energy cost per unit time (E_T) of traveling between patches. The average time per trial to stay in a patch type (average travel time + average patch residence time) T_u is defined by:

$$T_u = T_T + \sum \lambda_i \cdot t_i \quad \text{(Equation 6)}$$

The average energy intake from a patch E_e is obtained by:

$$E_e = \sum \lambda_i \cdot g_i(t_i) \quad \text{(Equation 7)}$$

The overall rate of energy intake (E_n) (energy intake on patch + energy loss while traveling) is then described by:

$$E_n = \frac{E_e - T_T \cdot E_T}{T_u} = \frac{\sum \lambda_i \cdot g_i(t_i) - T_T \cdot E_T}{T_T + \sum \lambda_i \cdot t_i} \quad \text{(Equation 8)}$$

Assuming that the travel time is independent of the time spent in a patch, the above equation can be written from the standpoint of a patch type of interest j as:

$$E_n = \frac{\lambda_j \cdot g_j(t_j) + A}{\lambda_j \cdot t_j + B} \quad \text{(Equation 9)}$$

where A and B are not functions of t_j.

To obtain a set of t_j's that maximizes E_n, we differentiate E_n with respect to a given t_j (= patch residence time at patch j):

$$\frac{\partial E_n}{\partial t_j} = \frac{\lambda_j g_j'(t_j)\left[\lambda_j t_j + B\right] - \left[\lambda_j \cdot g_j(t_j) + A\right] \cdot \lambda_j}{\left(\lambda_j \cdot t_j + B\right)^2} \quad \text{(Equation 10)}$$

By setting $\frac{\partial E_n}{\partial t_j} = 0$, we obtain:

$$g_i'(t_j)\left[\lambda_j t_j + B\right] - \left[\lambda_j \cdot g_j(t_j) + A\right] = 0 \quad \text{(Equation 11)}$$

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

$$g_j'(t_j) = \frac{\lambda_j \cdot g_j(t_j) + A}{\lambda_j t_j + B} \quad \text{(Equation 12)}$$

The right hand-side of this equation is equal to the overall rate of energy intake E_n. Considering that $g_j'(t_j)$ is a monotonically decreasing function, this indicates that an optimal forager should leave a given patch i, when the instantaneous energy intake rate $g_j'(t_j)$ drops to the net rate of energy intake E_n. This is Charnov's marginal value theorem (MVT).

Importantly, while E_n, and therefore the patch residence time $(t_1, t_2, \cdots, t_j)$, depend on the values of travel time (T_T), travel cost (E_T), and search cost ($E_{S,i}$), the following equality must hold true regardless of the values of these parameters:

$$g_1'(t_1) = g_2'(t_2) = g_3'(t_3) = \cdots = g_n'(t_n) \quad \text{(Equation 13)}$$

Note that $g_j(t_j)$ contains the cost of staying in the patch j (i.e. $E_{S_j} \cdot t_j$) in addition to the cumulative rewards $R_j(t_j)$ as discussed above, with its derivative given by

$$g_j'^{(t_j)} = r_j(t_j) - E_{S_j} \quad \text{(Equation 14)}$$

If the energy cost per unit time in a patch E_{S_j} is the same across all patches, like in our task, the following equality should hold true (**Prediction 1**):

$$r_1(t_1) = r_2(t_2) = r_3(t_3) = \cdots = r_n(t_n) \quad \text{(Equation 15)}$$

Next, we consider the consequences of specific reward schedules used in our task such as exponentially-decreasing reward rates and the proportional scaling of overall reward rates between patches.

In our task, the instantaneous rate of reward for patch type i is:

$$r_i(t) = m_i \cdot f_i \cdot e^{-\frac{t}{\tau}} \quad \text{(Equation 16)}$$

Where $m_i = [1, 2, 4]$, $f_i = [0.125, 0.25, 0.5]$, and $\tau = 8$ sec.

For a pair of two patch types, i and j,

$$r_i(t) = m_i \cdot f_i \cdot e^{-\frac{t_j}{\tau}} \quad \text{(Equation 17)}$$

$$r_j(t) = m_j \cdot f_j \cdot e^{-\frac{t_j}{\tau}} \quad \text{(Equation 18)}$$

Now we define a such that $a = \frac{m_j \cdot f_j}{m_i \cdot f_i}$.

At MVT-optimal patch residence times, t_i and t_j:

$$r_i(t_i) = r_j(t_j) \quad \text{(Equation 19)}$$

Therefore,

$$m_i \cdot f_i \cdot e^{-\frac{t_i}{\tau}} = m_j \cdot f_j \cdot e^{-\frac{t_j}{\tau}} \quad \text{(Equation 20)}$$

$$-\frac{t_i}{\tau} = \log(a) - \frac{t_j}{\tau} \quad \text{(Equation 21)}$$

Thus,

$$t_j - t_i = \tau \cdot \log(a) \quad \text{(Equation 22)}$$

This means that although both t_i and t_j are sensitive to the net energy intake of the environment E_n, the difference in patch residence times is constant for any given pair of patches with exponentially decaying reward rates with the same time constant (τ), irrespective of E_n, as long as a is constant (**Prediction 2**).

Furthermore, with three patch types whose initial reward rates $m_i f_i$ increase proportionally such that:

$$m_1 f_1 : m_2 f_2 : m_3 f_3 = 1 : a : a^2 \quad \text{(Equation 23)}$$

Then the difference in MVT-optimal PRT will be constant across successive patch types, i.e.,

$$t_2 - t_1 = t_3 - t_2 = \tau \cdot \log(a) \quad \text{(Equation 24)}$$

That is, the difference in the patch residence times should be constant across the pairs with the same proportional increase in reward rates (**Prediction 3**). In this study, we used patches with proportional reward sizes (1, 2, and 4 μL). The difference in leave time between 4 and 2μL patches ($PRT_{4\mu L} - PRT_{2\mu L}$) should, therefore, be equal to the difference between 2 and 1 μL patches, for

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

patches with the same initial reward frequency. More specifically, this difference should be equal to a fixed number, $8 \cdot \log(2) = 5.5452$ sec.

Calculation of instantaneous expected reward rate

To determine whether mice are determining their patch-leaving times via reward rate estimation and setting a threshold as MVT predicts, we calculate the instantaneous expected reward rate at the time of patch departure from an ideal observer perspective (same as Lottem et al.[65]). We use this quantification rather than the true empirical reward rate since the underlying reward frequency condition is a hidden feature of each patch and cannot be observed directly. The model assumes full knowledge of the distribution of patch types in the task and their respective reward statistics. Reward size is therefore always known, as the first reward on each patch provides this information. For each frequency condition, f_i, likelihood is calculated based on the series of outcomes (rewards vs. omissions) observed across one second intervals within each patch $o_1, \ldots, o_n$. Assuming a uniform prior, we obtain the posterior distribution over frequency conditions, and the probability of reward delivery at time $t+1$ is (same as Lottem et al.[65]):

$$P(o_{t+1} = 1 | o_1, \ldots, o_t) = \sum_{i=1}^{3} P(o_{t+1} = 1 | f_i) * P(f_i | o_1, \ldots, o_t) \quad \text{(Equation 25)}$$

$$= \sum_{i=1}^{3} P(o_{t+1} = 1 | f_i) * \frac{P(o_1, \ldots, o_t | f_i)}{\sum_{j=1}^{3} P(o_1, \ldots, o_t | f_j)} \quad \text{(Equation 26)}$$

Probability of reward delivery at each timepoint per frequency condition is given by Equation 2 and aggregate probability across observations per patch is:

$$P(o_1, \ldots, o_t | f_i) = \prod_{k=1}^{t} P(o_k | f_i) \quad \text{(Equation 27)}$$

Lastly, to obtain instantaneous reward rate from expected reward probability, we scale it by reward size and consider that expected value over a one second interval as reward rate.

Behavioral model fitting

We created three models (Models 1-3) which differ in the way they integrate time and reward to calculate a decision variable (*DV*) (Figure 3). For each model, we compute the *DV* for each time bin and transform it through a softmax function to generate a predicted probability, $P(Leave)/s$, of the animal leaving the patch—i.e., the hazard rate of leaving (Figure 3D). As *DV* increases, so does $P(Leave)/s$. Because the model must account for decision processes unfolding over time durations spanning orders of magnitude, while fitting per one-second time bin, we constrained the maximum probability to P_{max}, which is a free parameter. As a result, the midpoint of the sigmoid, corresponding to instances when the $DV = 0$, is not an indifference point wherein $P(Leave)/s = 0.5$, but instead is the point at which $P(Leave)/s = \frac{P_{max}}{2}$.

Models 1 and 2 have five free parameters: X_0 sets the midpoint of the sigmoid. Ψ is inverse temperature. $P_{\max 0}$ constrains the models' $P(Leave)/s$ output. ω_0 is an exponent for power-law scaling of slope per reward size. λ_0 is an exponent for power-law scaling by latent state. Model 3 included those five, plus one additional free parameter for the value subtracted from the integrator per reward delivery, R.

Slope of the integrator model was scaled based on the patch's reward size. We fit this slope scaling by a power-law function to allow for varying sensitivity to reward size across mice. We used $\mu = [.5, 1, 2]$ for $[1, 2, 4]$ μL, respectively, to normalize slope to the medium reward size.

$$\omega_\mu = \mu^{\omega_0} \quad \text{(Equation 28)}$$

Slope was additionally scaled by the estimate of patience for each patch (patch = *i*). We again used a power-law function for this scaling, fitting the free parameter, λ_0, per mouse to account for differences in how dependent PRT was on estimated patience across mice. Values for patience across mice were mean normalized to ensure fit values across parameters could be straightforwardly compared across mice. Each patch's normalized latent, L_i, was scaled by λ_0 to determine its contribution to slope scaling.

$$\lambda_i = L_i^{\lambda_0} \quad \text{(Equation 29)}$$

Because variance in patch-leaving times increases when PRTs are longer, we also scaled $P_{\max}$ with patches' latent patience estimates. Because $0 < P_{\max} < 1$, we used a function that would reasonably allow scaling in either direction, while constraining the transformed value to remain between 0 and 1.

$$P_{max\ i} = \frac{P_{\max 0}}{\lambda_i \cdot (1 - P_{\max 0}) + P_{\max 0}} \quad \text{(Equation 30)}$$

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

The other free parameters, Ψ, and, for Model 3, R, were not scaled across patches. We chose to apply latent-state scaling to slope and P_{max}, and not the other parameters, because this consistently yielded the best fits (Figure S2F).

The models produce a decision variable (DV) per one second interval (t) by calculating its current integrator value, $X_{i,t}$ and subtracting X_0 from it:

$$DV = X_{i,t} - X_0 \quad \text{(Equation 31)}$$

The models differ in how they each track time, TOP (Time on Patch) or TSLR (Time Since Last Reward), and whether they integrate rewards delivered ($nRews$ = number of rewards delivered up to that point on the patch), to compute $X_{i,t}$:

Model 1, reward indifferent:

$$X_{i,t} = \frac{TOP}{\omega_\mu * \lambda_i} \quad \text{(Equation 32)}$$

Model 2, reward reset:

$$X_{i,t} = \frac{TSLR}{\omega_\mu * \lambda_i} \quad \text{(Equation 33)}$$

Model 3, reward integrator:

$$X_{i,t} = \frac{TOP}{\omega_\mu * \lambda_i} - nRews * R \quad \text{(Equation 34)}$$

To compute the model's output prediction, $P(Leave)/s$, the probability of leaving a patch per one second bin, the DV is scaled by inverse temperature and sigmoid transformed with the output constrained by $P_{\max}$:

$$P(Leave)_{i,t} / s = \frac{P_{\max\ i}}{1+e^{-\psi DV_{i,t}}} \quad \text{(Equation 35)}$$

We optimized free parameter fits per subject and compared the models using 'mfit' toolbox for MATLAB. Maximum a posteriori estimates were computed using MATLAB's 'fmincon' function using 20 random initializations, converging upon the set of parameter fits that maximized log likelihood of the models at predicting $P(Leave)/s$ across one second time bins. Ranges for the parameters used to fit were $X_0[-5:20]$, Ψ $[0:10]$, $P_{\max\ 0}[.01:.98]$, ω_0 $[0:2]$, $\lambda_0[0:4]$, and $R[0:20]$, with uniform priors for each. Resulting parameter fits are shown in Figure S3J. BIC values were computed for each model fit, and a protected exceedance probability was calculated for each model to estimate how prevalent the model is across subjects.

5-fold cross-validation was additionally performed across models. Trials were split into folds by ordering trials per subject and labeling them in serial 1-5. Fitting was then performed as above on the training folds, after which we computed the log predictive probability over trials from the held-out test folds.

Diagnostic trial type groupings (e.g. RRR vs. R0R) and selection biases

To compare effects of reward history, we subselected "diagnostic" patches with specific reward combinations. We grouped these trials based on reward/omission patterns encountered and labeled them with an alphanumeric code defined by that pattern such that characters indicate whether a reward ('R') or no reward ('0') was delivered at each time point eligible for reward delivery, following one-second intervals beginning with the first reward at time $t = 0$. For example, an 'RRR' label indicates rewards were delivered on that patch at $t = 0, 1, 2$ seconds, whereas 'R0R' trials featured reward deliveries at $t = 0$ and 2 seconds, with no reward at $t = 1$. Alternatively, when subselecting for patches of a particular reward size, we used reward size in microliters ('1', '2', or '4') to indicate reward delivery instead of an 'R'. Accordingly, a '44' trial indicates reward deliveries of $4\mu L$ at $t = 0$ and 1 seconds.

Behavioral analysis of PRTs following specific reward combinations (Figures 3H, 3I, S3H, and S3I) aimed to isolate effects of those rewards. Therefore, following the reward/omission pattern of interest, if any additional rewards were delivered, those patches were automatically disqualified for inclusion in the analysis (for example, an 'RRR0R' trial no longer isolates effects of the 'RRR' rewards since it includes an additional reward which would itself further influence mouse behavior). Because shorter waiting times provide less opportunity for disqualifying reward deliveries to occur, they would be overrepresented in the data if using this selection criteria, thereby resulting in biased PRTs that underestimate mice's true willingness to wait.

Furthermore, this selection effect is disproportionately larger on patch types with higher underlying reward frequency since they yield higher probabilities for disqualifying reward deliveries to occur. Because frequency conditions are not equally proportioned between RRR and R0R trials ('RRR' reward combinations are generated more often when underlying reward frequency is higher), leaving this effect uncorrected would produce systematic biases in PRT comparisons.

We therefore additionally subselected for patches wherein no more rewards would be delivered irrespective of how long mice might continue to wait ('depleted' patches). We were able to do so because we configured our experiments so that instead of waiting until the moments of possible reward delivery to assign probabilistic outcomes of reward/no-reward, we simulated those probabilistic draws in advance such that timing of the next potential reward delivery was always predetermined. We were thus able to use

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

these saved values to isolate depleted patches for inclusion in our analysis. This allowed us to fully control for this potential selection bias and more accurately estimate mice's true willingness to wait in our PRT calculations.

Neural analyses considering only activity during the time window of specific reward combinations were unaffected by whether additional rewards might be delivered later, so similar subselection was unnecessary (Figures 4D, 5M–5P, 6C, and 6E). We therefore included all trials with the reward combinations of interest for those neural analyses.

A second type of selection bias can occur due to nonuniformly sampling over the distribution of latent patience levels (Figure 3I, left; and Figure S3I, middle). This bias occurs when selecting patches based on potential reward events at later time points on patches. Because subjects are free to leave patches at any time, when selecting for reward occurrence, or lack thereof, at a particular time on patch, this inevitably selects for trials in which subjects were willing to remain on the patch through at least that event time, thereby disproportionately including more trials in which latent patience is higher. Caution was therefore taken to restrict diagnostic trial type comparison to consider only reward combinations over the first two seconds on patches to mitigate this bias.

Using behavioral models to patch-leaving times on single trials

To compute the average leaving time predicted from a given behavioral model on a single trial (Figures 3M–3O), we first calculated the "leave density" per each one-second time bin. For each time bin, the leave density is given by the product of the fraction of trials the model predicted to still be on the patch at the start of the time bin by that time bin's predicted leave probability: $Fraction\ on\ patch_{t,i} * P(Leave)_{t,i}/s = Leave\ density_{t,i}$. The predicted leaving time is then computed by averaging time bins, weighted by the leave density. Fictive time bins were added to the end of each trial so that bins tiled up to 300 seconds before averaging, with no rewards added to those bins.

Unit inclusion criteria

To be included for analysis, units from Neuropixels recordings had to have a minimum firing rate of 1 Hz within patches and additionally at least 1 Hz within each third of the session to ensure that they were tracked throughout the recording session.

Spike waveforms

Spike waveforms were extracted for each unit using C_Waves (https://billkarsh.github.io/SpikeGLX/). Spike width was defined as the difference between the time of the waveform minimum and the maximum following this minimum. A threshold of 0.433 ms was used to define Regular (>0.433 ms) versus Narrow (<0.433 ms) waveforms (Figure S6H).

Generating z-scored firing rate traces for each unit

For each unit, firing rate traces were generated by binning spikes into 50 ms bins. These traces were then smoothed by convolution with a causal filter, a half-Gaussian, where the standard deviation of the full Gaussian was 100 ms. For the linear decoding analysis only (Figures 5 and S5), a full-Gaussian filter was used, so that DVs which respond immediately to rewards could be captured by slightly delayed reward responses. Only in-patch times were selected, excluding the inter patch intervals, and this trial-concatenated, smoothed firing rate trace was z-scored. These z-scored single-neuron firing rate traces were used to create the heatmap in Figure 6C and averaged across neurons during different trial types to generate the plots in Figures 6E–6H.

Principal components analysis (PCA) of neural activity

Smoothed, z-scored firing rate traces were generated as in the section "Generating z-scored firing rate traces for each unit." PCA was performed on the resulting per-session matrix of neurons x in-patch time points (MATLAB function "pca").

Definition of ramping principal components (PCs)

To identify "ramping" PCs (Figure 4E), trials with no reward at $t = 1$ were selected. For each PC, slopes were computed on single trials by calculating the linear slope between $t = 0.5$ and $t = 2$ or patch-leave time, whichever came first. A sign rank test was performed to compare these single-trial slopes against 0, and PCs were deemed "ramping" if this test was significant at $p<0.05$. The total variance explained by these PCs was summed for each session and plotted (black points in Figure 4E). For a shuffle control (gray points in Figure 4E), neural activity traces were randomly rotated relative to task events, and the same analysis was performed, once per session. Because PCs with any non-zero ramping slope were included, not necessarily ramping patterns matching those in Figure 4D, this analysis gives a ceiling on ramping variance.

Correlation of single neuron firing rates with Model 3 decision variables

Smoothed, z-scored firing rate traces within patches were generated as in the section "Generating z-scored firing rate traces for each unit." Pearson's correlation coefficient was computed between each neuron's firing rate trace and the Model 3 DV computed as in Figure 3. With each neuron, firing rates were circularly permuted, 100 times per neuron, and neurons with a z-test p-value < 0.001 versus the distribution of correlation coefficients from shuffled data were deemed significant (red histograms, Figure 4F). R^2 was computed for each neuron and averaged over neurons within each brain region (Figure 4G).

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

Linear models to predict behavioral decision variables from neural activity

For each unit, spikes were binned into 50 ms bins and smoothed with a 100 ms Gaussian kernel to generate single-cell firing rate traces within each patch. Then, firing rate traces were z-scored across patches for each neuron, so that model coefficients could be compared across neurons. We then fit cross-validated linear models to predict each model's ramping decision variable (DV) from neural activity. The version of the behavioral model DVs used was that in which the ramping slope was scaled by latent state, which gave the best behavioral fits (Figure 3). DVs were z-scored before fitting decoder models, and zeroed to time t = 0 in PSTH plots (Figures 5H–5P).

Regularized linear models were fit using the function *lasso* in MATLAB with alpha = 0.1 (10% L1, 90% L2 regularization), predicting a given behavioral model variable from the activity of all recorded neurons at each time point. For cross-validation, patches were split into the same five folds as for the GLM (see below). Each of the five folds was predicted based on fits to the other four folds, and decoder output was concatenated across folds to produce a neural prediction of the decision variable for all trials in the session (examples shown in Figure 5B). Within each training fold, the regularization parameter lambda was selected by nested cross validation. Specifically, the training data was split into 10 folds and models were fit with a range of lambda values. Mean squared error was computed on nested test folds for each value of lambda, and lambda was chosen such that mean squared error was within 1 standard error of the minimum. The linear model was then fit on the entire training set to obtain β, the coefficients mapping neural activity to the behavioral variable, and R^2 was computed on the test set. R^2 and β were averaged across training folds to obtain final values for each recording session/neuron.

To compare decoding across brain regions, decoders were fit using neurons from either frontal cortex or subcortical areas (Figure 5G). For each within-session comparison, neurons were down-sampled so that each region had the same number of neurons. A minimum of 20 neurons per brain region per session were required. Only sessions with at least 20 neurons in each of the two brain regions were kept (23/33 sessions).

Computing ramping slope of DVs and decoder output

Single-trial ramping slopes of DVs and decoder output ("neural prediction") were computed in the same manner, as follows (Figures 5H and 5I). First, trials were selected with no reward at $t = 1$. Then, slopes were computed on each trial between 0.5 sec and min(2 sec, 0.5 sec before patch leave). The median single-trial slope for each reward size was taken with each session. In Figure 5I, slopes for each session were averaged within mouse, and data were plotted at the mouse level. Significance was assessed using the p-value for the reward size coefficient in a linear mixed effects model with random effect per mouse and fixed effect of reward size.

Computing reward responses of DVs and decoder output

For DVs, the reward response was calculated as the instantaneous change at reward delivery. For decoder output ("neural prediction," Figure 5J,K), the reward response was calculated as the difference in decoder output from 150 ms before to 150 ms after reward. A linear mixed effect model with random effect per mouse and fixed effect of reward size was used to estimate the effect of reward size on the reward response (Figure 5K; significance level of reward size coefficient indicated above each plot).

Computing dependence of reward response on pre-reward level

A linear model was fit relating the pre-reward level to the reward response, and the slope of these fits was examined to determine whether the DV or neural prediction shown no resetting (slope = 0, as in behavioral Models 1 and 3) or full resetting (slope = -1, as in behavioral Model 2) (Figure 5L). For DVs, the pre-reward level was defined as the level immediately preceding reward time. For neural predictions, the pre-reward level was defined as the level 150 ms before reward time, to account for Gaussian smoothing. For Models 1 and 3 decision variables, there is no dependence of reward response on pre-reward level (slope = 0). For Model 2, the decision variable is reset by rewards (slope = -1) (Figure 5L).

Measuring the effect of reward history on DVs and decoder output using diagnostic trial types

Trials were chosen with rewards at t = 0, 2 ('R0R' trials) or t = 0, 1, 2 ('RRR' trials) (Figures 5M and 5N). For DVs and neural prediction (decoder output), the level just after the reward at t = 2 was analyzed on these trials (t = 2 for DVs, t = 2.15 for neural prediction to account for neural processing delays and smoothing). A linear mixed effect model was fit to these data points, with random effects for each mouse, a fixed effect of reward size, and fixed effect of trial type (coded as 0 for R0R trials and 1 for RRR trials, so that a negative coefficient would indicate a suppression of the decision variable at $t = 2$ by reward at $t = 1$) (Figures 5O and 5P).

Predicting single-trial latent state from pre-stop neural activity

We used leave-one-out multivariate linear regression to predict the latent state on single trials from population neural activity preceding patch stop (Figures S5I, S5J, and S5K). Spikes were binned by the period of 0.5 seconds preceding patch stop, resulting in a matrix of N (neurons) x M (trials). Each row of this matrix, corresponding to an individual neuron, was z-scored. To denoise the data, the dimensionality of the z-scored matrix was reduced using PCA and PCs were kept accounting for 80% of the total variance. Then, for each trial, an elastic net multivariate linear model was trained on the remaining M-1 trials (function lasso in MATLAB, with alpha = 0.1 [10% L1 penalty, 90% L2 penalty] and lambda = 0.05 [regularization strength]) and used to predict the latent state on the

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

held-out trial ("within" session prediction, Figures S5J and S5K). As a control comparison, this was also done using neural PCs from different Neuropixels recording sessions ("across" session prediction, Figures S5J and S5K), keeping the number of trials corresponding to the shorter of the two sessions. For each prediction (within or across session), cross-validated R^2 was computed using $R^2 = 1 - \sum_i \frac{(y_i - \hat{y}_i)^2}{(y_i - \bar{y})^2}$, where i denotes trials, y_i is the latent state on trial i, $\hat{y}_i$ is the prediction of the latent state on trial i from the linear model, and $\bar{y}$ is the mean latent state in the session.

Generalized linear model (GLM) to predict neural activity from behavioral variables

Poisson GLMs were fit to spiking activity of single neurons using the glmnet toolbox (https://glmnet.stanford.edu).[92,93] To prepare data for GLM fitting, we binned spikes from each neuron into 50 ms bins aligned to patch stop and continuing until patch leave. The following variables were included as regressors in the GLM:

1. Session time: time in session, (time in session)^2
2. Behavioral variables: patch position, velocity, acceleration, lick rate, d(lick rate)
3. Patch-stop kernels: 6 kernels x 3 (1 set per reward size), peaks spanning 0-1 seconds from patch stop
4. Patch-leave kernels: 6 kernels x 3 (1 set per reward size), peaks spanning 0-1 second prior to patch leave
5. Reward kernels: 11 kernels x 3 (1 set per reward size), peaks spanning 0-2 seconds after each reward
6. Time on patch: 3 (1 per reward size). Time on patch was log scaled: $t_{scaled} = \ln(t+1)$
7. Time since last reward: 3 (1 per reward size). Time since reward was log scaled: $t_{scaled} = \ln(t+1)$
8. Total reward: 3 (1 per reward size). Total reward was log scaled: $R_{scaled} = \ln(R+1)$

Thus, a total of 85 variables were used in the GLM (86 including intercept). Kernels were a raised cosine basis with log scaling of the x-axis (time), created using MATLAB code from the Pillow lab (https://github.com/pillowlab/raisedCosineBasis). Parameters used were: logScaling = 'log', logOffset = 1.0, timeRange = [0 5].

Variables were z-scored prior to model fitting. Variables with one coefficient per reward size were z-scored across reward sizes, so that coefficients could be compared across reward sizes. Kernels were z-scored together.

Data were prepared for glmnet using custom MATLAB code, and then fit using glmnet (version 4.1-8) in R (version 4.2.2). Model fitting was performed on the Harvard Faculty of Arts and Sciences Researching Computing cluster (FASRC cluster, https://www.rc.fas.harvard.edu/).

Models were regularized with elastic net regularization with $\alpha = 0.9$ (90% L1, 10% L2) to encourage sparse coefficients.[94] For cross-validation, patches were split into 5 folds. Each patch was assigned in its entirety to a single fold. Folds were counterbalanced for patches of different reward size. Cross-validation with these folds was used to select the regularization parameter for each neuron (using the function *cv.glmnet*).

Cross-validated deviance explained was computed for each neuron as follows:

$$D_{expl} = 1 - \frac{D_{model}}{D_{null}} \quad \text{(Equation 36)}$$

Where D_{model} is the cross-validated model deviance and D_{null} is the cross-validated null deviance. These two terms were computed once per test fold and then averaged across folds. For each test fold j, D^j_{model} and D^j_{null} were computed using the formula for Poisson deviance:

$$D^j_{model} = \frac{2}{N_j}\sum_i \left[y_i \ln \frac{y_i}{\hat{y}_i} - (y_i - \hat{y}_i) \right] \quad \text{(Equation 37)}$$

$$D^j_{null} = \frac{2}{N_j}\sum_i \left[y_i \ln \frac{y_i}{\mu_j} - (y_i - \mu_j) \right] \quad \text{(Equation 38)}$$

Where N_j is the number of observations (time bins) in the jth test fold, y_i is the number of spikes in the ith time bin in the jth test fold, $\hat{y}_i$ is the predicted mean of the Poisson distribution in the ith time bin (computed using the model fit on the training data), and μ_j is the mean spike count per bin in the training data. For bins with no spikes, we set $0 \cdot \ln(0) = 0$.

To estimate the contribution of task variables (reward kernels, time on patch, total reward, and time since reward) to model fits, we fit a reduced model in which task variables were removed. Poisson deviance was computed for the reduced model and deviance explained by task variables was computed as $1 - \frac{D_{full\ model}}{D_{reduced\ model}}$. Neurons with deviance explained > 1% by this metric were deemed "task-related."

Unsupervised clustering of GLM coefficients

We used a Gaussian Mixture Model (GMM) to cluster the matrix of Task Variables Coefficients (M=42; z-scored per neuron) x Task-Related Neurons (n=1398; we excluded 60 neurons whose coefficients on task variables were all zero, leaving 1398 task-related

Please cite this article in press as: Bukwich et al., Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics, Neuron (2025), https://doi.org/10.1016/j.neuron.2025.07.008

neurons in this analysis) (Figure 6). The dimensionality of this matrix was reduced to 3 x 1398 using PCA (34.7% of variance; Figure S6D). The dimensionality-reduced matrix was clustered using a Gaussian Mixture Model (function fitgmdist in MATLAB, with parameters RegularizationValue = 0.4, MaxIter = 10000, Replicates = 10, TolFun = 10^{-6}) with K (the number of clusters) spanning 1-10. For each value of K, the Bayes Information Criterion (BIC) was calculated. K was selected to minimize BIC (K = 6; Figure 6B).

Visualizing the population response of task-related neurons to diagnostic trial types

Each neuron's spikes were binned, smoothed, sub-selected for in-patch periods, concatenated across patches, and z-scored (as in the section "Generating z-scored firing rate traces for each unit"). Patches were divided into three groups: odd 40 trials (trials with 4 μL reward at t = 0 but not t = 1), even 40 trials, and all 44 trials (trials with 4 μL reward at t = 0 and t = 1). Each neuron's z-scored activity was averaged within each of these three groups. Neurons were sorted based on the time of peak activity on odd 40 trials. The left panel of Figure 6C shows a heatmap of z-scored activity following same sort on even 40 trials, and the right panel shows the same sort on all 44 trials.

State space modeling of ramping and stepping dynamics

The ramping model was implemented as a constrained recurrent, switching linear dynamical system with Poisson emissions.[95,96] The model has a continuous latent state x and discrete latent state z that both evolve over time. The discrete latent state can be in one of two states corresponding to an accumulation state ($z_t = acc$)and a boundary state ($z_t = b$). The continuous latent state is initialized to $x_0 < 1$. In the accumulation state, it follows a drift-diffusion process with a separate slope per reward size and a shared diffusion variance σ^2. It also accumulates pulsatile reward inputs u_t with a separate accumulator weight per reward size. This corresponds to the following dynamics

$$x_t = x_{t-1} + \beta_{c,r} u_t + \beta_{c,s} + \epsilon_t, \epsilon_t \sim Normal\left(0, \sigma^2_{acc}\right) \quad \text{(Equation 39)}$$

where $\beta_{c,r}$ is the coefficient for rewards for reward size c, and $\beta_{c,s}$ is the slope for reward size c. When $\beta_{c,r} < 0$ the reward decrements the accumulator, although importantly $\beta_{c,r}$ is unconstrained during fitting and is initialized at 0. In the boundary state, the continuous dynamics have no constant slope and a small constant variance. However, because a salient reward event could alter an animal's decision to leave even after that intention was formed, rewards still influence the continuous state which can cause the model to leave the boundary

$$x_t = x_{t-1} + \beta_{c,r} u_t + \epsilon_t, \epsilon_t \sim Normal\left(0, \sigma^2_b\right) \quad \text{(Equation 40)}$$

where σ^2_b is fixed to 1e-4. Transitions between discrete states depend on the value of the continuous state relative to a fixed boundary threshold at $x = 1$. In particular, the discrete state transition distribution is

$$p(z_t | z_{t-1}, x_t) \propto \exp\left(\gamma\left(R_{z_{t-1}} + W x_{t-1}\right)\right) \quad \text{(Equation 41)}$$

$$R_{acc} = [0, -1]^T, R_b = [0, -0.99]^T, W = [0, 1]^T \quad \text{(Equation 42)}$$

and $\gamma = 50$.This setting of the parameters places a boundary at $x = 1$ with probabilistic transitions to and from the boundary state for values of the continuous state near the boundary. We note that the discrete state is slightly encouraged to stay in the boundary state, as in the discrete state the threshold for returning to the accumulation state decreases to $x = 0.99$. Overall, the generative model is

$$z_t | z_{t-1}, x_t \sim p(z_t | z_{t-1}, x_t) \quad \text{(Equation 43)}$$

$$x_t | z_t, u_t \sim \begin{cases} Normal\left(x_{t-1} + \beta_{c,r} u_t + \beta_{c,s}, \sigma^2_{acc}\right) \text{ if } z_t = z_{acc} \\ Normal\left(x_{t-1} + \beta_{c,r} u_t, \sigma^2_b\right) \text{ if } z_t = z_b \end{cases} \quad \text{(Equation 44)}$$

$$\lambda_t = \log(1 + \exp(C x_t)) \quad \text{(Equation 45)}$$

$$y_t \sim Poisson(\lambda_t) \quad \text{(Equation 46)}$$

The time-varying reward inputs are shifted in time such that u_t is equal to one 400ms after a reward and otherwise equal zero.

The parameters $\theta = (\beta_{(1:C,r)}, \beta_{(1:C,s)}, \sigma^2_{acc}, C)$ were fit using variational Laplace EM (vLEM),[96] while the other parameters were fixed to correspond to the ramping model structure. The vLEM algorithm returns an approximate posterior distribution $q(x_{1:t})$ that approximates the true posterior distribution over the continuous latent states $p(x_{1:T} | y_{1:T}, u_{1:T}, \theta)$. For model comparison, the held-out log-likelihood was estimated via importance sampling using $q(x_{1:t})$

$$\log p_\theta(y_{1:T}|u_{1:T}) \approx \log \frac{1}{S}\sum_{s=1}^{S} \frac{p_\theta\left(y_{1:T}, x_{1:T}^{(s)}\middle|u_{1:T}\right)}{q\left(x_{1:T}^{(s)}\right)}, x_{1:T}^{s} \sim q\left(x_{1:T}^{(s)}\right) \quad \text{(Equation 47)}$$

with S = 200 and where the discrete latent state is marginalized in the numerator when computing p_θ.

The stepping model was implemented as a constrained HMM with input-dependent transitions and trial-dependent rates. It has two discrete states corresponding to a "down" and "up" state. Here the time-varying inputs u_t are one-dimensional and equal to 1 when there is a reward at time t and are otherwise zero. As in the ramping model, the rewards are shifted 400ms in time. The model starts in the down state and in the absence of a reward transitions to the up state with probability p_s. When there is a reward, the model is forced to transition back to the down state. The generative model is

$$\pi(u_t) = \begin{cases} \begin{bmatrix} 1-p_s & p_s \\ 0 & 1 \end{bmatrix} \text{if } u_t = 0 \\ \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \text{if } u_t = 1 \end{cases} \quad \text{(Equation 48)}$$

$$z_t \middle| z_{t-1}, u_t \sim \pi(u_t)_{z_{t-1}} \quad \text{(Equation 49)}$$

$$y_t \middle| z_t \sim Poisson\left(\lambda_{z_t,i}\right) \quad \text{(Equation 50)}$$

where $\lambda_{z_t,i}$ is the firing rate for state z_t on trial i. Notably, we fit separate firing rate parameters for each state and trial to allow the model to account for variability in overall rates across trials. While the ramping model has condition-dependent parameters and the stepping model does not, the stepping model gains significant flexibility by having per-trial firing rate parameters.

The overall model parameters are $\theta = (\lambda_{:,1:I}, p_s)$ and are fit via maximum likelihood using the EM algorithm. The per-trial firing rates are initialized to the mean number of spike counts in the first and last 10 time bins of a trial for the down and up states, respectively. To evaluate the model on held-out trials, the step probability p_s was kept constant but we fit the per-trial firing rates $\lambda_{z_t,i}$ to the test trials. Given the per-trial firing rates on test trials, the held-out log-likelihood was computed as the log-likelihood of the test trials marginalizing out the discrete states. This is a strong baseline since it has access to the heldout data for fitting the per-trial firing rates, whereas the ramping model does not have per-trial parameters that are fit on the held-out data.

For both models, we fit the model to the summed spike count of neurons from the frontal cortex, split by GMM cluster (Figure 7D). For each GMM Cluster, only sessions with at least 10 neurons from that cluster were included (n = 16, 10, 12, 1, 5, 7 sessions, for clusters 1-6 respectively, out of 33 total sessions). Models were fit to the summed activity from all co-recorded neurons in the cluster (binned spike counts in 50ms bins). For each session, 80% of the trials were used for training the model parameters and the held-out loglikelihood is computed and reported on the remaining 20% of the trials. Both models were fit using code in the SSM (https://github.com/lindermanlab/ssm) and SSM-DM (https://github.com/davidzoltowski/ssmdm) code packages. To compare model coefficients across sessions with different numbers of neurons with different firing rates (Figure 7E), coefficients were normalized by the total firing rate of the cluster within each recording session (total spikes / num. time bins).

Neuron, Volume *113*

Supplemental information

Competitive integration of time and reward explains value-sensitive foraging decisions and frontal cortex ramping dynamics

Michael Bukwich, Malcolm G. Campbell, David Zoltowski, Lyle Kingsbury, Momchil S. Tomov, Joshua Stern, HyungGoo R. Kim, Jan Drugowitsch, Scott W. Linderman, and Naoshige Uchida

Figure S1

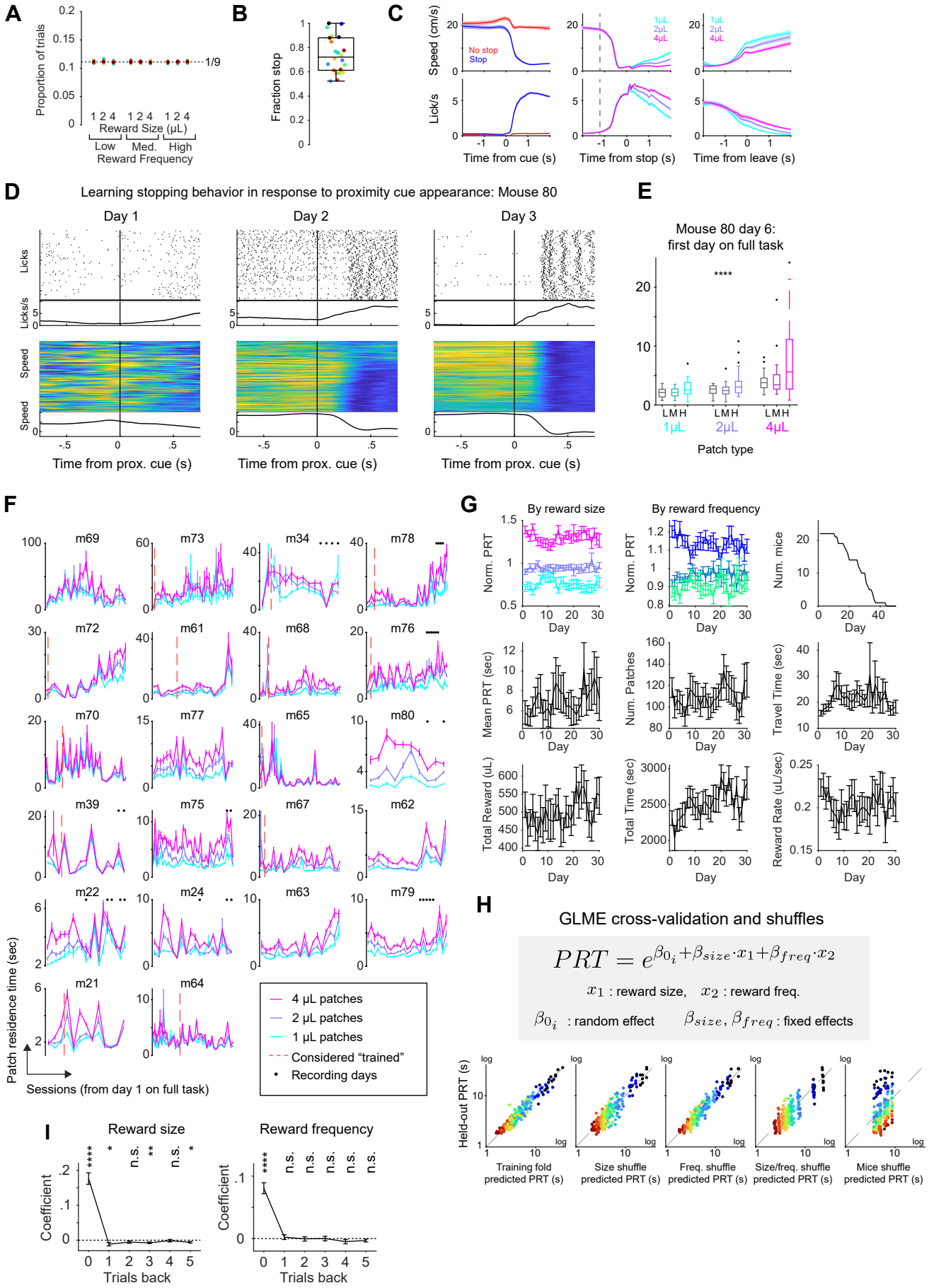
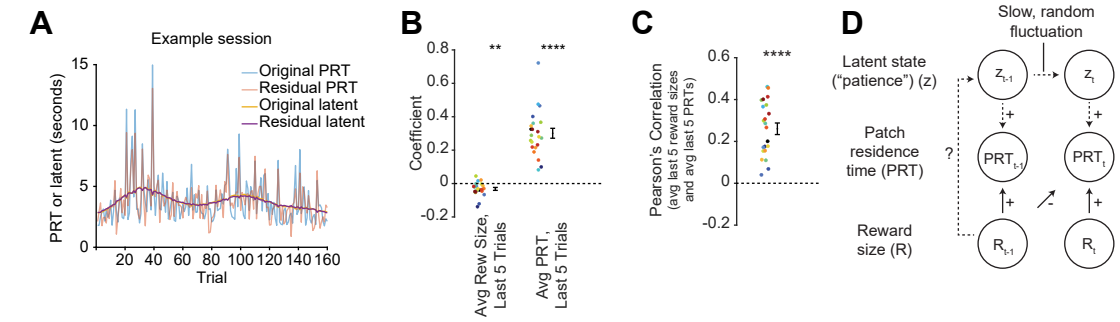

A
Proportion of trials
1/9
1 2 4 1 2 4 1 2 4
Reward Size (μL)
Low Med. High
Reward Frequency
B
Fraction stop
C
Speed (cm/s)
No stop
Stop
Lick/s
1μL
2μL
4μL
Time from cue (s)
Time from stop (s)
Time from leave (s)
D
Learning stopping behavior in response to proximity cue appearance: Mouse 80
Day 1
Day 2
Day 3
Licks
Lick/s
Speed
Time from prox. cue (s)
E
Mouse 80 day 6:
first day on full task
****
L M H
1μL 2μL 4μL
Patch type
F
m69 m73 m34 m78
m72 m61 m68 m76
m70 m77 m65 m80
m39 m75 m67 m62
m22 m24 m63 m79
m21 m64
Patch residence time (sec)
Sessions (from day 1 on full task)
4 μL patches
2 μL patches
1 μL patches
Considered "trained"
Recording days
G
By reward size
By reward frequency
Norm. PRT
Num. mice
Mean PRT (sec)
Num. Patches
Travel Time (sec)
Total Reward (uL)
Total Time (sec)
Reward Rate (uL/sec)
Day
H
GLME cross-validation and shuffles
$PRT = e^{\beta_{0_i} + \beta_{size} \cdot x_1 + \beta_{freq} \cdot x_2}$
x_1 : reward size, x_2 : reward freq.
β_{0_i} : random effect $\beta_{size}, \beta_{freq}$: fixed effects
Held-out PRT (s)
Training fold predicted PRT (s)
Size shuffle predicted PRT (s)
Freq. shuffle predicted PRT (s)
Size/freq. shuffle predicted PRT (s)
Mice shuffle predicted PRT (s)
I
Reward size
Reward frequency
Coefficient
n.s.
Trials back

Figure S1: Mouse behavior over training and task engagement (related to Figure 1).

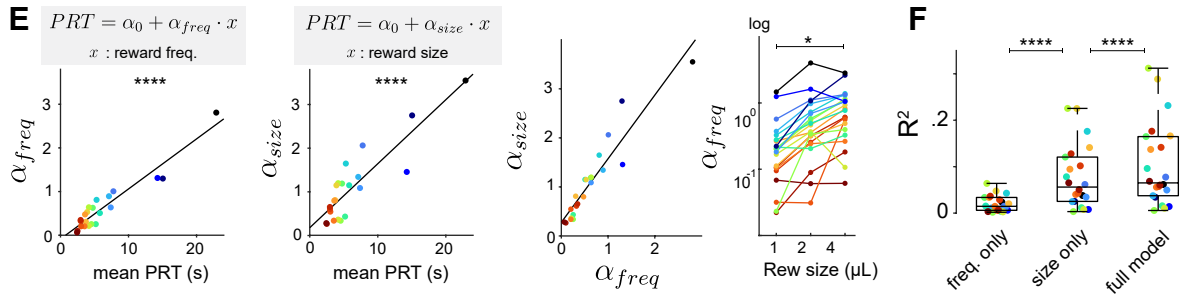
A) Proportion of trials of each type (3 reward sizes x 3 reward frequencies = 9 types) for each mouse. Each dot represents a mouse, and the dotted line shows $1/9^{th}$. Within each session, trial types were balanced in blocks of 27 trials.
B) Fraction of trials in which mice stopped in response to the proximity cue. Each dot represents a mouse.
C) Mean population running speed (Top) and lick rate (Bottom) in response to task events. (Left) Aligned to proximity cues appearance, split by trials in which mice successfully stopped (blue) or skipped (red) the patch. (Middle) Aligned to patch stop, split by reward size. Dashed line indicates mean time of proximity cue appearance. (Right) Aligned to time mice left the patch, split by reward size. Due to experimenter error, speed and lick signals could not be recovered for 2/22 mice. These mice were excluded from the plot.
D) Example mouse learning over first three days of pre-training in which proximity cue onset is followed by automatic reward delivery after a short delay. Patches are not introduced yet. Mouse learns that proximity cue is predictive of reward and begins showing anticipatory licking and slowdown in response to it.
E) First session full task with patches introduced, example mouse. Mice must stop in response to the proximity cue to enter a patch and earn rewards ($p < 0.0001$, linear regression of PRT on patch value). Column groupings separate patches by reward size. Within columns, patches are split by reward frequency.
F) Per subject PRTs across sessions, split by patch reward size. Error bars indicate standard error of the mean. Dashed red line indicates when mice task performance achieved inclusion criteria, after which sessions are included in the final data set. Mice with no dashed red line met inclusion criteria on the first day of the full task. Black dots mark recording sessions included for neural analysis.
G) Behavioral performance metrics over training. Norm. PRT = normalized PRT (PRT divided by mean PRT within session). Error bars indicate standard error of the mean over mice. Mice were trained for differing numbers of days; top rightmost plot shows the number of mice remaining on each training day.
H) Cross-validation of the GLME from Fig. 1G showing held-out PRT versus training fold predicted PRT. Training fold in the leftmost panel is unshuffled. Training fold trial labels for the remaining four were shuffled by reward size, frequency, size and frequency, and mouse ID, from left to right, respectively. Folds yielding the median mean squared error between predicted and held-out PRTs are shown for each shuffle type.
I) Linear regression of normalized PRT (PRT divided by mean PRT, within session) on reward size or frequency on the current and previous trials (similar to Fig. 1H but showing effects of previous trials individually; Methods). Lines and error bars show mean regression coefficients $\pm$ SEM over mice ($n = 22$). n.s. Not Significant, * $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$, one-sample t-tests of model coefficients versus zero ($n = 22$ mice for each test, each lag tested separately, no correction for multiple comparisons).

Figure S2

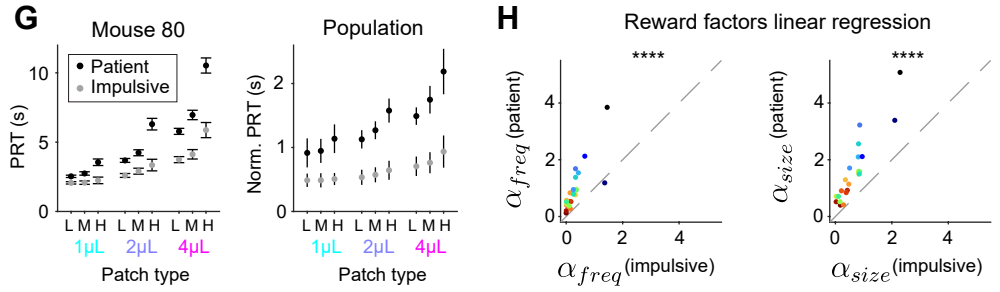
Relationship between reward history, current trial PRT, and latent state



Linear regression on reward factors: Reward sensitivity scales with overall 'patience' across mice



Splitting data per mouse by median latent: Reward sensitivity scales with latent state within mice



GLMs fit separately per mouse yields similar results to population fit

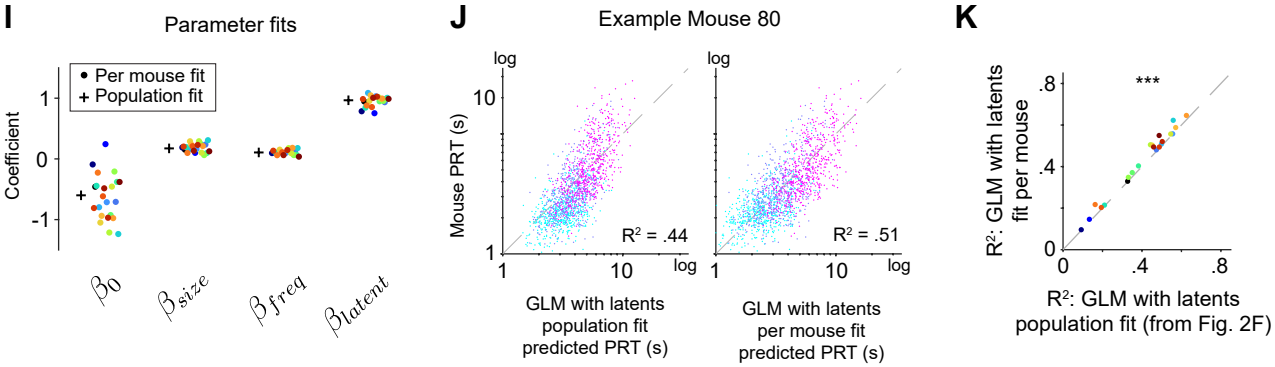


Figure S2: Reward sensitivity scales with patience across and within mice (related to Figure 2).

A) An example session showing how random fluctuations in reward size and frequency cannot explain the slowly fluctuating latent state, which we also call "patience." Original PRTs are shown in blue; residual PRTs after regressing out reward size and frequency are shown in red. The original latent, computed on the original PRTs, is shown in yellow; the residual latent, computed on the residual PRTs, is shown in purple. Note the close match between the original and residual latent. In this session, reward size and frequency explained 29.0% of PRT variance, but only 6.6% of latent variance. On average, reward size and frequency explained 19.8 ± 2.4% of PRT variance, and 3.9 ± 1.2% of latent variance (n = 22 mice).

B) Linear regression of normalized PRT (PRT divided by mean PRT, within session) on the average reward size and the average PRT of the preceding 5 trials (as in Fig. 1H but including previous PRTs for comparison). Regressions on previous reward sizes and PRTs were performed separately to avoid issues related to correlated regressors and plotted side by side for comparison. Even though past reward sizes were negatively correlated with current PRT, and past reward sizes were positively correlated with past PRTs (panel C), the coefficient on previous PRTs was positive (previous PRTs coefficient = 0.30 ± 0.03; n = 22 mice), suggesting the slowly fluctuating latent state is separable from reward rate. **** $p < 0.0001$, t-test versus zero.

C) Pearson's correlation between a moving average of reward size and a moving average of PRT (5-trial moving average) (mean Pearson's correlation ± SEM = 0.26 ± 0.03, n = 22 mice) **** $p < 0.0001$, t-test versus zero.

D) A diagram of the hypothesized relationships between rewards (current and previous), PRTs, and latent state. Rewards influence current PRTs positively, and future PRTs negatively. Both effects are qualitatively consistent with optimal foraging theory. At the same time, a latent state positively influences current PRTs and slowly fluctuates from trial to trial. We cannot completely rule out that rewards may *also* influence the latent state (question mark in the diagram) and thus have both a positive and negative influence on future PRTs, via different pathways. However, as shown in panel (A), the effect of rewards on latent state is small, if it exists.

E) Linear regressions of PRT on reward factors per mouse. Left: Regression coefficients for reward frequency and size scales with mean PRT. Middle: Effects of reward size and frequency scale with same ratio across mice. Right: Regression coefficients for PRT on reward frequency are larger for larger reward sizes.

F) R^2 values for linear regression across mice for models including reward frequency, reward size, and a full model including both frequency and size as well as their interaction. Stars indicate significance of pairwise Wilcoxon signed-rank tests (both $p < 0.0001$, Bonferroni-adjusted).

G) PRT per patch type with trials split into two latent state groupings: impulsive and patient, separated by median latent state, example mouse (Left) and population means (Right).

H) Linear regressions on reward factors, fit separately per latent state grouping across subjects ($p < 0.0001$, Wilcoxon signed-rank tests).

I) GLM with scaling by latent patience estimates, same as Fig. 2E except fitting with a separate set of parameters per mouse.

J) Single trial PRT predictions from GLM population fits (Left) and per mouse fits (Right), example mouse.

K) R^2 statistics for GLM fits per mouse yield modest improvements over population fit ($p = 0.0004$, Wilcoxon signed-rank test, Bonferroni-adjust).

Figure S3

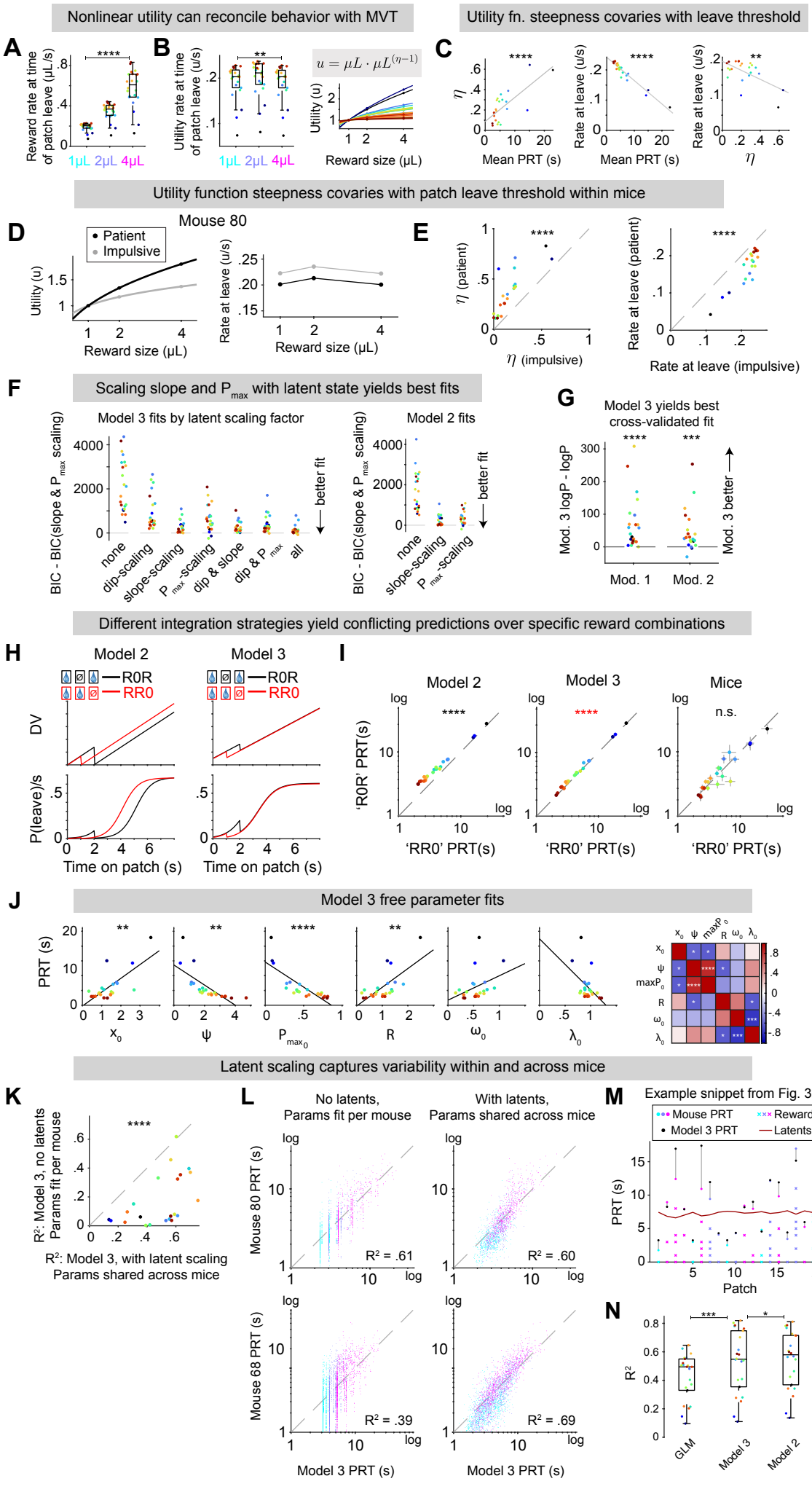

Nonlinear utility can reconcile behavior with MVT
Utility fn. steepness covaries with leave threshold
A
Reward rate at time of patch leave (μL/s)
1μL 2μL 4μL
B
Utility rate at time of patch leave (u/s)
$u = \mu L \cdot \mu L^{(\eta-1)}$
Utility (u)
Reward size (μL)
C
η
Mean PRT (s)
Rate at leave (u/s)
Utility function steepness covaries with patch leave threshold within mice
D
Mouse 80
Patient
Impulsive
Utility (u)
Reward size (μL)
Rate at leave (u/s)
E
η (patient)
η (impulsive)
Rate at leave (patient)
Rate at leave (impulsive)
F
Scaling slope and P_{max} with latent state yields best fits
Model 3 fits by latent scaling factor
Model 2 fits
BIC - BIC(slope & P_{max} scaling)
better fit
none
dip-scaling
slope-scaling
P_{max}-scaling
dip & slope
dip & P_{max}
all
G
Model 3 yields best cross-validated fit
Mod. 3 logP - logP
Mod. 3 better
Mod. 1
Mod. 2
Different integration strategies yield conflicting predictions over specific reward combinations
H
Model 2
Model 3
R0R
RR0
DV
P(leave)/s
Time on patch (s)
I
Model 2
Model 3
Mice
'R0R' PRT(s)
'RR0' PRT(s)
n.s.
J
Model 3 free parameter fits
PRT (s)
x_0
ψ
P_{max_0}
R
ω_0
λ_0
maxP$_0$
Latent scaling captures variability within and across mice
K
R^2: Model 3, no latents Params fit per mouse
R^2: Model 3, with latent scaling Params shared across mice
L
No latents, Params fit per mouse
With latents, Params shared across mice
Mouse 80 PRT (s)
Mouse 68 PRT (s)
R^2 = .61
R^2 = .60
R^2 = .39
R^2 = .69
Model 3 PRT (s)
M
Example snippet from Fig. 3O
Mouse PRT
Model 3 PRT
Reward
Latents
PRT (s)
Patch
N
R^2
GLM
Model 3
Model 2

Figure S3: Additional analyses of behavioral model fitting (related to Figure 3).

A) Instantaneous expected reward rate at time of patch leave per patch reward size, as computed by an ideal observer model, colored by mouse ID ($p < 0.0001$, linear mixed-effects model). Calculated same as in Fig. 3A.

B) Scaling utility as a nonlinear function of reward size can yield similar expected utility rates at time of patch leave across reward sizes. Rates were nearly equivalent across reward sizes (fits were devised to achieve this), though a small effect of reward size was still observed as significant ($p = 0.008$, linear mixed-effects model, left). Utility functions fit across rewards size per mouse, $u = \mu L \cdot \mu L^{(\eta-1)}$ (right). Values for utility-scaling exponent η were fit using fmincon in MATLAB by minimizing variance in mean resulting instantaneous expected utility rates across reward sizes. Utility per second is quantified in arbitrary units.

C) When converting from objective reward magnitude to units of subjective utility to reconcile observed behavior with MVT, utility-scaling exponent η scales with mean PRT across mice ($p < 0.0001$, linear regression, left). Fitted utility rates at patch leave (arbitrary units/s) scale inversely with mean PRT ($p < 0.0001$, linear regression, center). Utility rate at time of patch leave and utility-scaling exponent η covary inversely across mice ($p = 0.0035$, linear regression, right).

D) Example mouse 80 utility curves fit separately per latent state grouping, trials grouped into impulsive and patient groupings split by median latent patience estimate (left). Fitted utility rates at time of patch leave, split by latent state groupings (right).

E) Utility function fits across mice per latent state grouping. Fitted values for utility-scaling exponent η are higher in the patient latent groupings for all mice ($p < 0.0001$, Wilcoxon signed-rank test, left). Expected utility rates at time of patch leave are higher in the impulsive latent state groupings for all mice ($p < 0.0001$, Wilcoxon signed-rank test, right).

F) Model fits by scaling different factors by latent patient estimates. BIC values shown relative to the scaling method providing the best fits.

G) Cross-validation of model fits: relative log predictive probability of training fold model fits on held-out test folds, means across folds. For each fold, we used maximum likelihood estimation to obtain parameter fits across the four training folds. Those parameter fits were then used to generate predicted $P(Leave)/s$ across bins in the held-out test fold patches, over which we summed log predictive probability of patch leaving decisions. Model 3 outperformed Models 1 and 2 ($p < 0.0001$, Model 3 over Model 1; $p = 0.0008$, Model 3 over 2, Wilcoxon signed-rank test, Bonferroni correction).

H) Example schematics demonstrating how Models 2 (Left) and 3 (Right) make differing predictions for DV (Top) and $P(Leave)/s$ predictions (Bottom) on patches with rewards at $t = [0,2]$ sec ('R0R' patches, black) compared with patches with rewards at $t = [0,1]$ sec ('RR0' patches, blue).

I) Per subject mean simulated PRT for R0R versus RR0 patches from Model 2 fits (Left), Model 3 fits (Middle), and empirical mice PRT (Right). Points are colored per mouse. Axes are log-scaled. There was also a significant effect of greater PRTs for 'R0R' versus 'RR0' trials for Model 3 resulting due to a selection bias over latent patience values (Methods).

J) Model 3 free parameter fits versus mean PRT across mice (Left, X_0: $p = 0.009$, Ψ: $p = 0.0011$, P_{max}: $p < 0.0001$, R: $p = 0.0022$, ω: $p = 0.5828$, λ: $p = 0.2270$, linear regression with Bonferroni correction). Correlation matrix of Model 3 free parameter fit values across mice (Right: significant correlations indicated with asterisks, Bonferroni correction).

K) R^2 statistics for predicting single trial PRT per mouse with a latent-scaled model 3 using a single shared set of fit parameters across mice, versus model 3 fit separately per mouse but without latent scaling ($p < 0.0001$, Wilcoxon signed-rank test).

L) Single trial predictions from models fit in panel K. Example mouse 80 is shown (Top). Mouse with median fit difference across models, mouse 68, additionally shown (Bottom) as representative example since mouse 80 was an outlier for this fit comparison.

M) First 20 patches from Fig. 3O showing single trial mouse and Model 3 predicted PRTs with reward timings indicated per patch.
N) R^2 statistics for predicting single trial PRTs per mouse via the latent-scaled GLM fit per mouse from Fig. S2K, versus the single trial PRT predictions from Models 3 and 2, shown in ascending order of mean R^2 across the population. Asterisks indicate significance of pairwise Wilcoxon signed-rank tests (Model 3 over GLM, $p = 0.0002$; Model 2 over Model 3, $p = 0.026$, Bonferroni-adjusted)

Figure S4

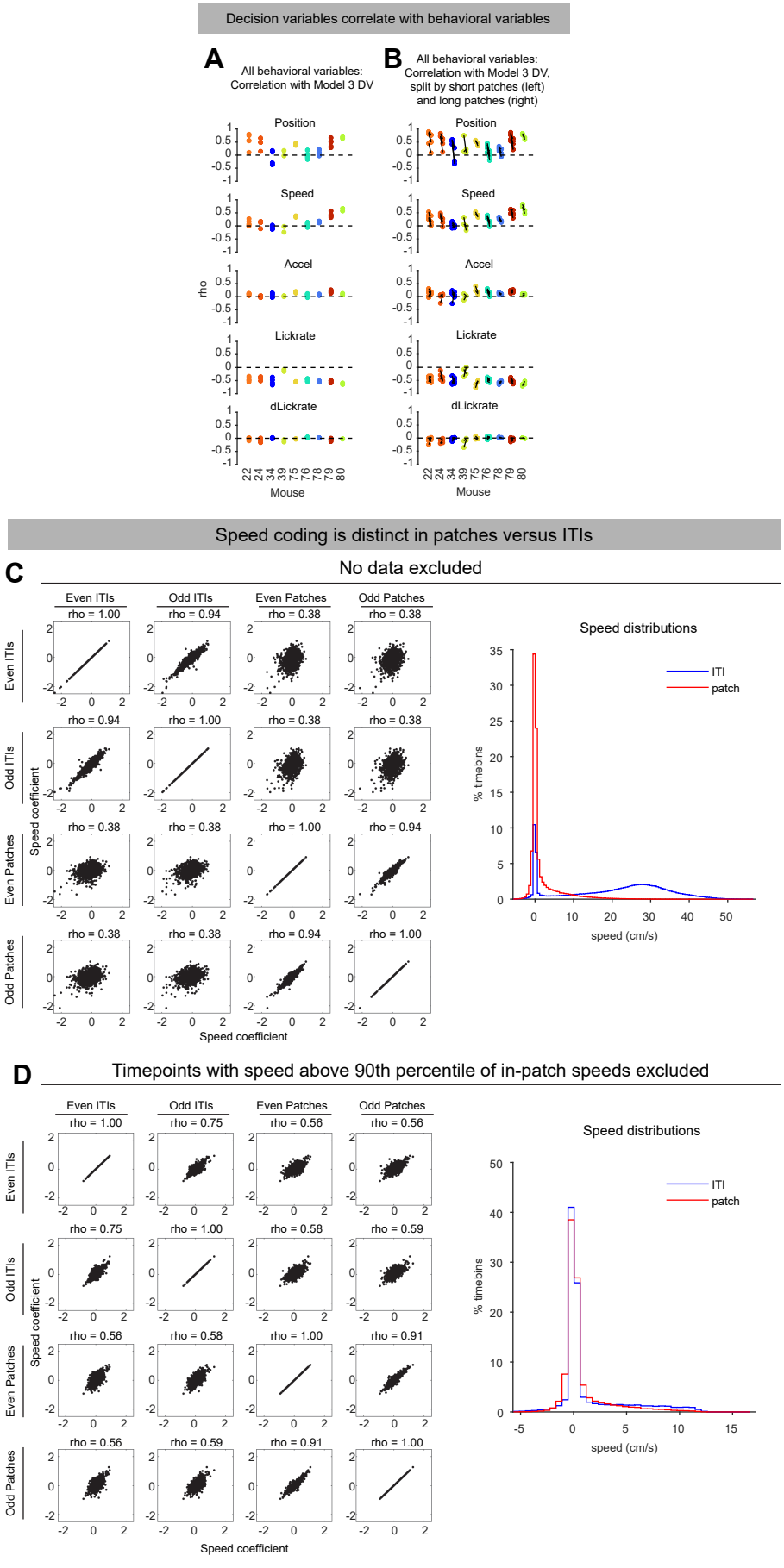

Decision variables correlate with behavioral variables
A
All behavioral variables: Correlation with Model 3 DV
B
All behavioral variables: Correlation with Model 3 DV, split by short patches (left) and long patches (right)
Position
Speed
Accel
Lickrate
dLickrate
rho
Mouse
22 24 34 39 75 76 78 79 80
Speed coding is distinct in patches versus ITIs
C
No data excluded
Even ITIs
Odd ITIs
Even Patches
Odd Patches
rho = 1.00
rho = 0.94
rho = 0.38
Speed coefficient
Speed distributions
ITI
patch
% timebins
speed (cm/s)
D
Timepoints with speed above 90th percentile of in-patch speeds excluded
rho = 0.75
rho = 0.56
rho = 0.58
rho = 0.59
rho = 0.91

Figure S4: Behavioral confounds (related to Figure 3, 4).

A) Pearson's correlation between Model 3 decision variable (Fig. 3) and multiple behavioral variables: position, speed, acceleration, lick rate, derivative of lick rate. Each data point represents a single Neuropixels recording session.
B) As in panel A, but split into short versus long patches (patch residence time above or below the median, per session). For each mouse, short patches are plotted on the left, and long patches on the right. Data points from the same session are connected by black lines.
C) GLMs fit to intertrial intervals (ITIs) revealed qualitatively distinct speed coding compared to patches, suggesting that ramping activity in patches is not a low-level motor response. GLMs contained only an intercept and speed term, fit to odd/even ITIs and odd/even patches separately. Speed traces were z-scored before fitting GLMs. Left: In scatter plots, each point represents a neuron, and "rho" above the plot indicates Pearson's correlation coefficient. Right: the distributions of running speed in ITIs (blue) and patches (red), pooled across all mice/sessions. Because running speeds were greater in ITIs compared to patches, we downsampled data to better match speed distributions (panel D).
D) Same as panel C, but with data downsampled to better match running speeds between patches and ITIs, by excluding all time points with running speed greater than 90^{th} percentile of in-patch running speeds (same cutoff applied to both ITIs and patches). While correlations of speed coefficients between ITI and patch periods increased, they remained substantially below correlations of coefficients fit to odd/even patches.

Figure S5

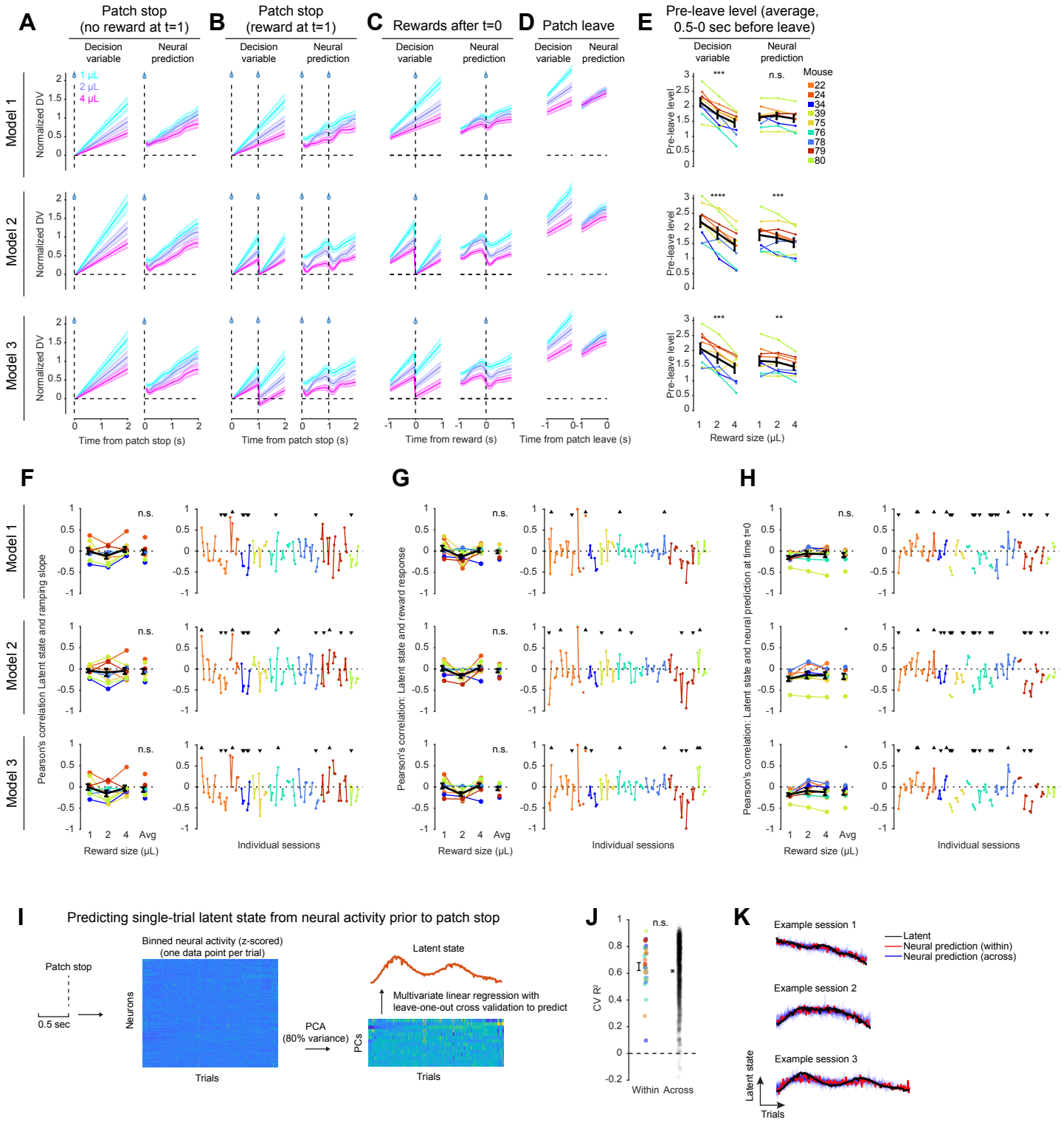

A
Patch stop (no reward at t=1)
B
Patch stop (reward at t=1)
C
Rewards after t=0
D
Patch leave
E
Pre-leave level (average, 0.5-0 sec before leave)
Decision variable
Neural prediction
Model 1
Model 2
Model 3
Normalized DV
1 µL
2 µL
4 µL
Time from patch stop (s)
Time from reward (s)
Time from patch leave (s)
Pre-leave level
Reward size (µL)
Mouse
22
24
34
39
75
76
78
79
80
n.s.
F
G
H
Pearson's correlation: Latent state and ramping slope
Pearson's correlation: Latent state and reward response
Pearson's correlation: Latent state and neural prediction at time t=0
Avg
Individual sessions
I
Predicting single-trial latent state from neural activity prior to patch stop
Binned neural activity (z-scored) (one data point per trial)
Patch stop
0.5 sec
Neurons
Trials
PCA (80% variance)
PCs
Latent state
Multivariate linear regression with leave-one-out cross validation to predict
J
CV R²
Within
Across
K
Example session 1
Example session 2
Example session 3
Latent
Neural prediction (within)
Neural prediction (across)
Latent state
Trials

Figure S5. Additional analyses of neural population decoding, including effects of latent state on decoder output (related to Figure 5).

A) PSTHs for model decision variables (left) and neural decoder predictions (right) on trials with no reward at $t = 1$. This panel is the same as Fig. 5H; it is reproduced here so that PSTHs aligned to multiple key timepoints in the trial can be seen side-by-side (A-D). Each panel is arranged in rows by behavioral model (Models 1-3, top to bottom). Colors indicate reward size. Lines and shaded areas indicate means $\pm$ SEM over sessions.

B) Same as (A), but for trials with reward at $t = 1$.

C) PSTHs aligned to reward deliveries after $t = 0$. Same as Fig. 5J, reproduced for comparison with panels A, B, and D.

D) PSTHs for model decision variables (left) and neural predictions (right) aligned to patch leave. Note that whereas decision variables diverge prior to patch leave, neural predictions are more similar, suggestive of a bound-reaching process in the neural data. A strict bound was not a feature of the behavioral models, which instead produced a leave probability as a sigmoid function of the decision variable.

E) Quantification of panel (D). Pre-leave level of model decision variables (left) and neural predictions (right) are shown for each reward size. Pre-leave level is defined as the average within the final 0.5 seconds before patch leave. Colored lines represent mice. In this and similar panels, values were averaged by session first and then by mouse, and statistical tests were performed at the mouse level. Black lines and error bars represent means $\pm$ SEM over mice. Stars indicate significance of the reward size coefficient in a linear mixed effects model with random effects per mouse and fixed effect of reward size. n.s. Not Significant, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

F) Within-session Pearson's correlation between ramping slope and latent state. For each recording session, the ramping slope and latent state were computed on each trial and correlated with each other, separately for trials of each reward size. By construction, decision variable ramping slopes were perfectly anti-correlated with latent state (not shown). This analysis measures the ability of neural decoders to capture that effect. **Left panels:** Correlations averaged by mouse, for behavioral models 1-3 (top to bottom). The average neural prediction correlation coefficient across reward sizes is shown at the right-most edge of the plot; none of these averages differed significantly from zero according to a t-test (all $p > 0.05$). **Right panels:** Correlations for individual sessions. For each session, the correlation between ramping slope and latent state is shown separately for each reward size (three data points, connected by a line). Sessions are colored by mouse. Triangles above individual session/reward size pairs indicate statistically significant correlations ($p < 0.05$), either positive (upward pointing triangles) or negative (downward pointing triangles). Statistically significant correlations were observed within some individual sessions.

G) Same as (F), but for decoder responses to rewards. At the mouse level, reward responses did not correlate with latent state on average for any of the three models.

H) Same as (F), but for decoder output at $t = 0$. The average correlation differed significantly from zero for Models 2 and 3 decoders but not Model 1 decoder (t-test, Model 1 $p = 0.11$, Model 2 $p = 0.032$, Model 3 $p = 0.038$; n = 9 mice). * $p < 0.05$ n.s. Not Significant, t-tests.

I) Schematic of approach used to predict latent state from neural population activity on single trials (see Methods for further details).

J) Cross-validated (CV) R^2 did not differ between models trained on neural data from the same session ("Within") or different sessions ("Across"), suggesting that neural population activity can decode arbitrary traces that vary slowly with similar statistics to the latent state (mean CV $R^2 \pm$ SEM, Within = 0.65 ± 0.03 [n = 33 sessions], Across = 0.62 ± 0.01 [n = 1056 across-session pairs], unpaired t-test $p = 0.39$). This does not rule out that latent state may be represented in

neural spiking activity, but further tests, for example with longer sessions, would be needed to reject the null hypothesis that decoding is driven by generic slowly varying neural population activity. n.s. Not Significant.

K) Example latent state traces from individual sessions (black) along with the neural prediction from co-recorded neural activity (red, “Within”) as well as neural predictions from other Neuropixels recording sessions (blue, “Across”), showing that predictions were similar Within versus Across session.

Figure S6

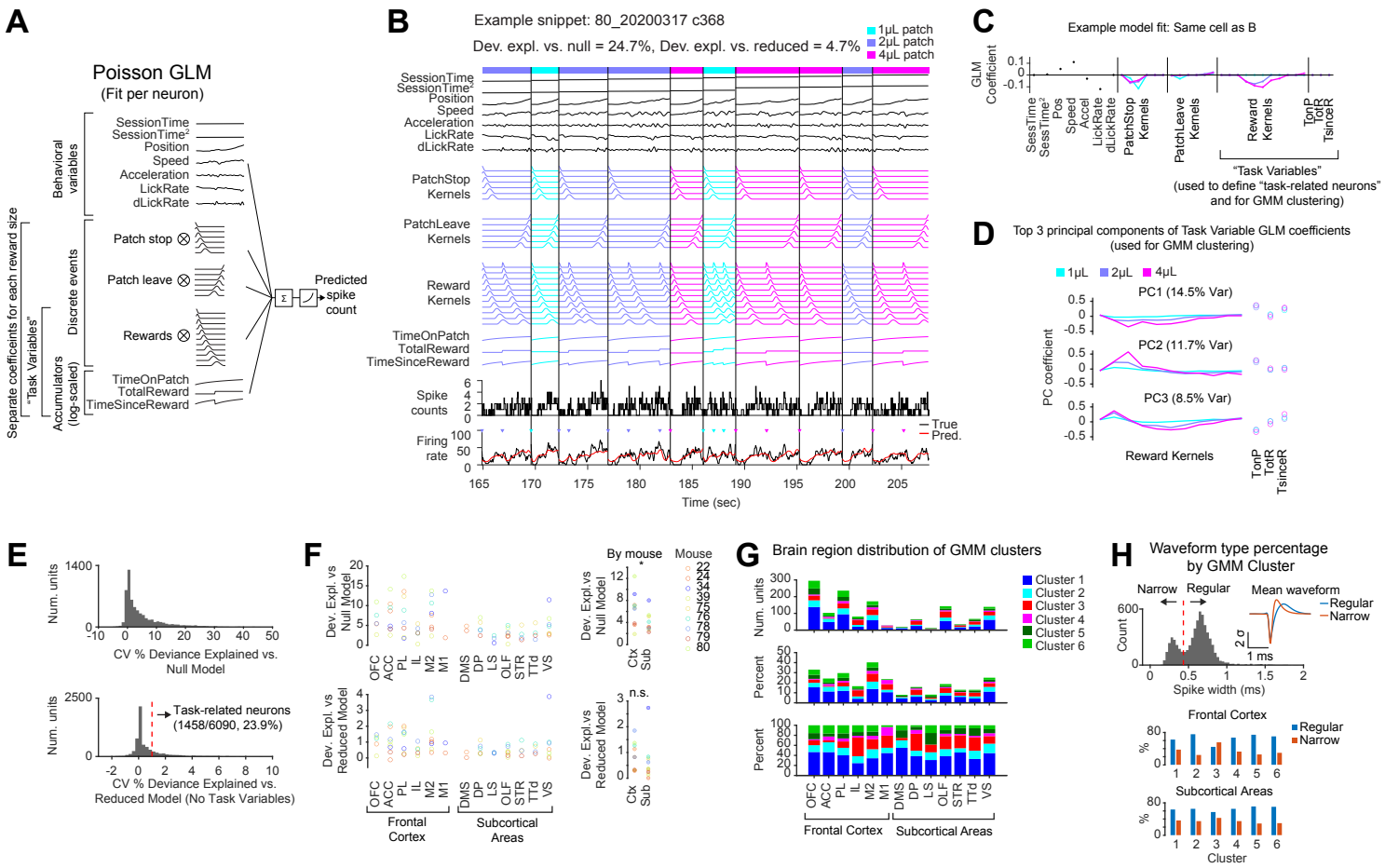

A
Poisson GLM
(Fit per neuron)
Behavioral variables
SessionTime
SessionTime²
Position
Speed
Acceleration
LickRate
dLickRate
Discrete events
Patch stop
Patch leave
Rewards
Accumulators (log-scaled)
TimeOnPatch
TotalReward
TimeSinceReward
Separate coefficients for each reward size
"Task Variables"
Predicted spike count
B
Example snippet: 80_20200317 c368
Dev. expl. vs. null = 24.7%, Dev. expl. vs. reduced = 4.7%
1μL patch
2μL patch
4μL patch
PatchStop Kernels
PatchLeave Kernels
Reward Kernels
Spike counts
Firing rate
True
Pred.
Time (sec)
C
Example model fit: Same cell as B
GLM Coefficient
"Task Variables"
(used to define "task-related neurons" and for GMM clustering)
D
Top 3 principal components of Task Variable GLM coefficients
(used for GMM clustering)
PC coefficient
PC1 (14.5% Var)
PC2 (11.7% Var)
PC3 (8.5% Var)
Reward Kernels
E
Num. units
CV % Deviance Explained vs. Null Model
Task-related neurons (1458/6090, 23.9%)
CV % Deviance Explained vs. Reduced Model (No Task Variables)
F
Dev. Expl. vs Null Model
Dev. Expl. vs Reduced Model
By mouse
Mouse
n.s.
Frontal Cortex
Subcortical Areas
G
Brain region distribution of GMM clusters
Num. units
Percent
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
H
Waveform type percentage by GMM Cluster
Narrow
Regular
Mean waveform
Count
Spike width (ms)
Frontal Cortex
Subcortical Areas
Cluster
I
Average GMM cluster coefficients by mouse
Mouse 22 (N=166 task-relevant neurons)
Mouse 24 (N=30 task-relevant neurons)
Mouse 34 (N=109 task-relevant neurons)
Mouse 39 (N=17 task-relevant neurons)
Mouse 75 (N=41 task-relevant neurons)
Mouse 76 (N=330 task-relevant neurons)
Mouse 78 (N=372 task-relevant neurons)
Mouse 79 (N=135 task-relevant neurons)
Mouse 80 (N=198 task-relevant neurons)
5 a.u.
1 sec.
Reward kernels
Accumulators (TimeOnPatch, TotalReward, TimeSinceReward)
Reward size:
1 μL
2 μL
4 μL

Figure S6: Additional information related to GLM fitting and GMM clustering (related to Figure 6).

A) Schematic of the Poisson GLM.
B) A snippet of an example cell showing the structure of the Poisson GLM: Regressors (top) and neural activity (bottom). Inverted triangles in the bottom panel indicate reward delivery. Vertical lines indicate the boundaries between patches; only on-patch times were included. Colors indicate reward size. Reward Kernels, Time on Patch, Total Reward, and Time Since Reward were considered "Task Variables."
C) GLM coefficients for the example cell shown in B. Task Variables were used for clustering (Reward Kernels, Time on Patch, Total Reward, and Time Since Reward; the latter three are also termed "accumulators"). Only "task-related" neurons were included in clustering (panel D), and further were required to have at least on non-zero Task Variable GLM coefficient.
D) The top 3 PCs of the matrix of Task Variable coefficients (accounting for 34.7% of variance). Task Variable coefficients were z-scored prior to PCA.
E) **Top:** Cross-validated percent deviance explained versus null model (intercept only) for all neurons in the dataset. **Bottom:** Cross-validated percent deviance explained versus reduced model (no task variables) for all neurons in the dataset. Neurons with >1% deviance explained versus reduced model were designated as "task-related."
F) **Top:** Average percent deviance explained versus null model by brain region. Within each brain region, neurons were averaged by mouse. Colors indicate mice. **Bottom:** Average percent deviance explained versus reduced model by brain region.
G) **Top panel:** Number of neurons in each GMM cluster, split by brain region. **Middle panel:** Percent of all recorded neurons in each brain region assigned to each cluster. **Bottom panel:** Percent of all task-related neurons in each brain region assigned to each cluster. All task-related neurons were assigned to a GMM cluster, which is why the bars in this panel all reach 100%.
H) **Left:** Histogram of spike widths for all units in the dataset. A threshold of 0.433 ms was used to define Narrow versus Regular waveforms. Inset: The mean waveform for Regular and Narrow units. **Right:** Unit waveforms by GMM cluster. Especially in Frontal Cortex, Cluster 3 was enriched for units with Narrow waveforms. 56% of Frontal Cortex Cluster 3 neurons were narrow spiking, versus 32% of Frontal Cortex neurons from the other 5 clusters. In Subcortical Areas, these numbers were 43% versus 34%.
I) Same as Fig. 6D, but for individual mice: Average z-scored GLM task variable coefficients for each GMM cluster. Reward kernel coefficients were multiplied by corresponding basis functions and summed to generate the predicted reward response.