Cortico-Striatal Beta-Oscillations as a Marker of Learned Reward Value

Abbreviated title: (Beta-Oscillations Represent Reward Value)

Koloski Francoeur M.[1,2], Hulyalkar S. [1,2], Tang T. [1,2], Wu X. [1,2], Fakhraei, L.[1,2], Mishra J. [2] & Ramanathan., D.S[1,2].

1. Mental Health Service, VA San Diego Healthcare Syst., La Jolla, CA, 92161

2. Dept. of Psychiatry, UC San Diego, La Jolla, CA, 92093;

Corresponding author: Dhakshin Ramanathan (dramanathan@ucsd.edu)

Number of pages: 69

Number of figures/ tables: 8

Number of words abstract: 207

Number of words introduction: 648

Number of words discussion: 1482

Conflict of Interest: The authors declare no competing financial interests.

Acknowledgments: This work was supported by the following awards: Burroughs Wellcome fund 1015644 (to DR), 1R01MH123650 (to DR), start-up funds from the UCSD Department of Psychiatry (to DR), and a training award T32MH018399 (to MFK). Histopathology was performed in part by Tissue Technology Shared Resource supported by CCSG Grant P30CO23100.

Abstract

Single neuron correlates of reward value have been observed in brain regions along the cortico-striatal pathway including ventral striatum, orbital, and medial prefrontal cortex. Brain imaging studies in humans further validate these findings and suggest that value is represented in a reward network opposed to a particular brain region. To explore how a learned reward value signal may be communicated across a widespread, dynamic network, we measured local field potentials of male Long-Evans rats performing three distinct behavioral tasks, each probing a different aspect of reward-processing. Our goal was to identify a common electrophysiology property linked with reward value. We found that oscillations at beta frequencies that occurred during the reward outcome period served as a robust marker of learned reward value. Beta-oscillations signaled positive reward valence and scaled in predictable ways with subjective reward value modulated by temporal discounting and reversal learning. Oscillatory signatures of reward-processing were observed across canonical reward-regions including orbitofrontal cortex, anterior insula, medial prefrontal cortex, ventral striatum, and amygdala. Activity of single-units in orbitofrontal cortex were also selectively modulated at beta frequencies during positive valence reward outcome. This data suggests that beta-oscillations reflect learned reward value in a distributed network that may serve as a stable and robust bio-marker for future studies.

Significance Statement: Using large-scale measurements of local field potentials, we found that activity within a cortico-striatal circuit, operating at beta/gamma frequencies, was associated with reward valence, and scaled with temporal discounting and reversal learning.

Introduction

Reward-processing comprises the set of neural systems associated with appetitive, motivational, or pleasurable stimuli (Dalley et al., 2004; Berridge and Kringelbach, 2008). Deficits in reward-processes are linked with learning and decision-making impairments and likely contribute to anhedonia, amotivation, and substance abuse problems observed in various psychiatric conditions (Dalley et al., 2004; Pujara & Koenigs, 2014). Identifying preclinical bio-makers of reward-processing will help assess behavioral deficits and expand treatment options that are currently limited (Berridge and Kringelbach, 2008; Pujara and Koenigs, 2014; Whitton et al., 2015; Bilderbeck et al., 2020).

The ventral striatum, in particular the nucleus accumbens, is connected to the medial prefrontal cortex, orbitofrontal cortex, and basolateral amygdala through cortico-striatal-limbic reward-network projections (Groenewegen et al., 1997; Schultz et al., 2000; Dalley et al., 2004; Bayer and Glimcher, 2005; Abler et al., 2006; Berridge and Kringelbach, 2008; Schoenbaum et al., 2009; Haber and Knutson, 2010; Chau et al., 2015; Haber, 2016). The ventral striatum and its extended network are innervated by midbrain dopamine neurons from the ventral tegmental area, contributing to reward-processing behaviors through reward-prediction error signals (Berridge and Robinson, 1998; Bayer and Glimcher, 2005; Humphries and Prescott, 2009; Haber and Knutson, 2010; Francois et al., 2014; Snyder et al., 2020). Prior research, including single-unit

recordings in animals and neuroimaging in humans, has described common and distinct aspects of reward-processing amongst these disparate brain regions.

Single-unit activity is high-dimensional. Neurons from any brain region encode a diverse array of task-related processes (Cui et al., 2016; Francoeur and Mair, 2018; Williams et al., 2018; MacDowell and Buschman, 2020). Neurons in ventral striatum, prefrontal and orbitofrontal cortex fluctuate during reward anticipation and delivery (van der Meer and Redish, 2009; van Duuren et al., 2009; Goldstein et al., 2012; Atallah et al., 2014; Levcik et al., 2017; Francoeur and Mair, 2018, 2019; Constantinople et al., 2019) and signal different types (Carelli et al., 2000; Schultz et al., 2000), magnitudes (Schultz et al., 2000; Roesch et al., 2011; Goldstein et al., 2012; Simon et al., 2015), and locations of reward (van der Meer and Redish, 2009). Low-dimensional representations of population activity provide a more robust, stable, and simpler framework to identify neuro-behavioral relationships and can be compared with human neuroimaging data (Williams et al., 2018; MacDowell and Buschman, 2020; Abbaspourazad et al., 2021). Local field potentials (LFP) offer an opportunity to bridge micro- and macroscopic levels of brain activity and may represent low-dimensional population level features of single-units (Ray et al., 2008; Hall et al., 2014; Cui et al., 2016; Huang et al., 2019; Abbaspourazad et al., 2021).

We have previously used multi-site LFP recordings to characterize networks operating at distinct oscillatory frequencies to support behavioral inhibition and default-mode-like processing (Fakhraei et al., 2021b, 2021a). Here we utilize our multi-site LFP approach to identify electrophysiology markers linked with reward-processing. To increase the

behavioral relevance of any electrophysiological markers identified, we examined data from three distinct behavioral tasks: 1) A go/wait behavioral inhibition task to identify signals related to the valence of feedback (reward vs. no reward) (**Fig. 1A**); 2) A temporal discounting task to identify signals scaling with subjective value driven by reward magnitude (high vs. low reward) and temporal delay (0.5 to 20s) (**Fig. 1B**); 3) Finally, a probabilistic reversal learning task to identify signals modulated by the learned probability of reward delivery (high vs. low-probability) (**Fig. 1C**). On each task, we first examined activity in the lateral orbitofrontal cortex (lOFC), a brain region consistently identified for its role in evaluating reward outcomes and expectancies to drive adaptive behavior (Dalley et al., 2004; Winstanley et al., 2004a; Schoenbaum et al., 2009; Dalton et al., 2016; Constantinople et al., 2019; Wassum 2022). Next, we analyzed pertinent oscillatory markers in 12 brain regions comprising the cortico-striatal-limbic pathway (**Table 1**). We provide evidence of oscillatory activity at beta and high-gamma frequencies found consistently across our three tasks that modulates with expected reward value.

Methods

Ethics Statement

This research was conducted in strict accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocol was approved by the San Diego VA Medical Center Institutional Animal Care and Use Committee (IACUC, Protocol Number A17-014).

116

117 Experimental Design

118 **Subjects:** 36 male Long-Evans rats obtained from Charles River Laboratories were
119 used for these experiments. When received, rats were ~ one month old weighing 150g.
120 Behavioral training was initiated two weeks after arrival. Rats were housed in pairs
121 during training prior to electrode implantation, and individually housed thereafter, in a
122 standard rat cage (10 x 10.75 x 19.5 in, Allentown, NJ, USA) with free access to food
123 and on a standard light cycle (lights on at 6 am / off at 6 pm). During behavioral training,
124 animals underwent water scheduling (free access to water for two hours/day) to
125 maintain motivation for water reward in the task. Water was unrestricted on non-training
126 days and rats were weighed weekly to ensure that water scheduling did not lead to
127 reduced food intake. Different cohorts of rats were trained to perform one of three tasks
128 designed to measure distinct aspects of reward-processing: 11 rats with multi-site LFP
129 probes were trained on a go/wait response inhibition task, 10 on a temporal discounting
130 task, and 7 on a probabilistic reversal learning (PRL) task. Additionally, 8 rats trained on
131 the go/wait task were used for single-unit recordings to supplement LFP findings.
132 Subjects with chronic implants were monitored daily for signs of infections, injuries, and
133 bleeding.

134

135 **Operant Chamber and Training:** The same custom-designed operant chamber was
136 used for all three tasks. The chamber had five nose-ports (NP), each with an LED, IR
137 sensor and metal cannula for water delivery. The chamber also contained two auditory
138 tone generators, a house-light, a screen to display visual stimuli, and five peristaltic

139 stepper motors/water pumps that delivered the water rewards into NPs. The chamber
140 was 6.2 x4.7x 6.23 inches with a ceiling opening that allowed electrophysiology tethers
141 to move freely. Simulink (Mathworks) installed directly onto a Raspberry Pi system
142 controlled the behavioral programs. Behavioral outputs from the operant chambers were
143 synchronized with electrophysiological signals using lab-streaming-layer, a protocol
144 designed to integrate multiple behavioral and physiological streams into a common
145 timing stream (Ojeda et al., 2014; Buscher et al., 2020). The design, operation and
146 software control of this chamber has been described previously (Buscher et al., 2020).
147 Animals first went through a pre-training period (5-10 sessions), to learn that a NP with
148 an LED "on" signaled an available response port; that responding in an available NP
149 would trigger a water reward; and finally that there was a sequential nature to the task
150 (animals start a "trial" by first entering the middle NP (3), after which they could use
151 either of the neighboring ports (2 or 4) to respond and collect an immediate reward).
152 This standard pre-training paradigm was used for all three final behavioral paradigms.
153 Animals advanced to the next stage of training when they consistently performed ≥100
154 trials in a 60 min session.
155
156 **Behavioral Tasks:**
157 *Go/Wait Response Inhibition Task.* The visual-cue go/wait task was used to observe
158 brain activity associated with positive valence on successful go-cue trials (rewarded)
159 compared to unsuccessful wait-cue trials (not rewarded) (**Fig.1A**). Animals began a trial
160 by entering the middle NP (3), ensuring animals were in an identical position on every
161 trial when the visual stimulus appeared. After a fixed delay of 30ms, a visual stimulus

appeared on the screen denoting the trial as either a “go-cue” trial (animal required to respond within 2s to attain a reward) or a “wait-cue” trial (animal required to withhold response for 2s to attain a reward)**.** The stimulus remained on the screen until the animal responded. If animals responded correctly, a water reward was delivered into the middle NP (3) after a delay of 400ms. If animals responded incorrectly, the house light flashed for a 5 second “time-out” period and no reward was given. Rewards consisted of 20 μL of water delivered over a two second period using a stepper-motor (the motor sound provides an instantaneous cue regarding reward delivery). After water delivery, there was a 5s inter-trial-interval before the next trial began. The trials were distributed randomly as 25% “go-cue” and 75% “wait-cue” trials. Animals were trained for ~14 weeks until behavior typically stabilized (>80% accuracy on go-cue trials), after which they were implanted with electrodes (described below). We waited two weeks for animals to recover from surgery prior to resuming water-scheduling. LFP analyses are based on data from 60 recording sessions from 11 rats. Single-unit analyses are based on data from 62 recording sessions from 8 rats.

Temporal Discounting Task. A different cohort of animals were trained on a temporal discounting task to contrast electrophysiology activity for different reward magnitudes (high vs. low) delivered at increasing temporal delays (0.5 to 20s) (**Fig. 1B**). Generally, temporal discounting tasks center around choosing between a low-value reward delivered immediately, or a high-value reward delivered after a delay. In our version of the task, subjects chose between a low-value (1x) reward delivered immediately (500ms after response) or a high-value (3x) reward delivered at a variable delay. In separate

behavioral sessions the high-value delay ranged from 500ms to 20s after the response. Each session began with 6 forced-choice trials, orienting the rat to both the low-value (NP 2) and high-value (NP 4) options. The houselights were on, and LED lights signaled the available response port, alternating between low-value (NP 2) response and high-value (NP 4) response. Reward following either response was delivered immediately (500ms) after response during forced-choice trials. After the 6 trials were complete, the houselights dimmed and rats began the full, self-paced, trial sequence. Response port (2 and 4) LEDs were on, signaling the rat to choose. Selecting the low-value response port (2) turned on the houselights, the middle NP LED (3), and a tone (500ms duration) to indicate a choice was made. A small reward (10μL delivered over a 1s duration) was delivered immediately (500ms after response) from NP 3. Selecting the high-value response port turned on the houselights, the middle NP LED (3), and a tone (500ms duration) signaled the choice. Depending on the session, there was a variable delay (0.5s, 1s, 2s, 5s, 10s, 20s) until the high-value reward (30μL over a 3s duration) was delivered out of NP 3. The motor delivering water made an audible sound, to cue reinforcement delivery onset and amount of reward. The high-value delay alternated between behavioral sessions but remained the same throughout the entire (60 min) session. The houselights turned off when water was delivered out of NP3 and a 5s inter-trial interval began after water delivery. To learn to discriminate magnitude differences, rats were trained on the immediate delay condition (500ms) for both high-value and low-value choices. Once they showed a clear preference for the high-value choice (≥70% high-value responses/session) and consistently performed ≥100 trials, they were advanced to the other delay conditions. Training (including pre-training sessions) on

208 average lasted18 sessions across 5 weeks. After implantation we waited two weeks to
209 allow animals to recover from surgery before electrophysiology recording began.
210 Recording sessions lasted 60 minutes and occurred 3-4 days a week. LFP analyses are
211 based on data from 124 recording session from 10 rats. There was an average of 20
212 recording sessions per delay condition (two sessions at each delay condition per rat;
213 N=19 sessions at 0.5s; N=23 sessions at 1s; N=15 sessions at 2s; N=28 sessions at 5s;
214 N=18 sessions at 10s; N=21 20 s).
215
216 *Probabilistic Reversal Learning Task.* The probabilistic reversal learning task (PRL) was
217 used to examine brain activity associated with learned reward likelihood (high vs. low-
218 probability of reward) and tests the subjects' ability to update information after reward
219 contingencies are reversed (**Fig. 1C**). In our version of the PRL task, rats must choose
220 between two nose ports: the high-probability choice ("target") delivers water 80% of the
221 time, and the low-probability choice ("non-target") only 20% of the time. The PRL task is
222 self-paced. Each trial began with houselights off and the middle NP LED (3) on. Once a
223 rat responded, LEDs in NP 2 and 4 turned on, indicating an available choice between
224 the response ports. Each NP is randomly assigned as the target or non-target NP in
225 each session. Selecting the target choice led to 2s (20μL) of water on 80% of trials and
226 no water only 20% of the time. Selecting the non-target choice led to 2s (20μL) of water
227 only 20% and no water 80% of the time. A response in NP 2 or 4 caused the other LED
228 to turn off. On rewarded trials, the houselights remained off and water was delivered out
229 of the selected NP (2 or 4) 500ms after the response. There was a 5s inter-trial interval
230 that started with water delivery. On unrewarded trials, a tone (500ms in duration)

231 signaled no water delivery, the houselights turned on, and a 5s inter-trial interval began.
232 Throughout a session, the NP contingencies reversed based on the rats' behavior.
233 Reversals occurred when a rat made 80% target responses (rewarded or non-
234 rewarded) over a 10-trial moving window (8 of the last 10 responses are "targets"). To
235 perform PRL effectively, rats must respond appropriately to correct feedback ("target"
236 rewards; "non-target" no reward) while also ignoring misleading feedback due to the
237 probabilistic nature of our task ("target" no reward; "non-target" reward). Rats received
238 at least two weeks of pre-training (described above) prior to surgical implantation of LFP
239 probes but were naïve to the PRL task. Two weeks after surgery rats began training on
240 the PRL task. Performance was measured by counting the number of reversals/
241 sessions, the target choice percentage, and win-stay behavior (propensity to choose the
242 same NP after receiving a reward on the previous trial). On average, rats took 15
243 sessions to train on the PRL task (~3.5 weeks). Once rats were consistently performing
244 at least one reversal and performing ≥100 trials we started to record LFP. Rats ran 60
245 min sessions 3-4 days a week. LFP analyses are based on data from 36 recording
246 session from 7 rats (average of 5 sessions per rat).
247
248 Surgery
249 Surgeries were performed with the 'sterile tip' method, all instruments were autoclaved
250 prior to start, and were conducted under isoflurane anesthesia (SomnoSuite, Kent
251 Scientific, CT, USA). Animals received a single dose of Atropine (0.05 mg/kg) to
252 diminish respiratory secretions during surgery, a single dose of Dexamethasone (0.5
253 mg/kg) to decrease inflammation, and 1mL of 0.9% sodium chloride solution prior to

254 surgery. The area of incision was cleaned with 70% ethanol and iodine solution. A local
255 anesthetic, Lidocaine (max .2cc), was injected under the skin at the incision site while
256 the animal was anesthetized but before surgery initiation. The fabrication and
257 implantation procedures of our custom fixed field potential and single-unit probes are
258 described in detail (Francoeur et al., 2021).
259
260 **LFP Probe Implantation:** Briefly, our LFP probe targets 32 different brain areas
261 simultaneously. 50µm tungsten wire (California Fine Wire, CA, USA) used for our
262 electrodes was housed in 30-gauge stainless steel metal cannula (Mcmaster-Carr,
263 Elmhurst, IL, USA) cut 8-9mm long. Each cannula (8) contained four electrode wires cut
264 to their unique D/V length. During surgery, 8 holes were drilled in the skull (one for each
265 cannula) at predetermined stereotactic locations (see **Supp Fig. 1**). Additional holes
266 were drilled for a ground wire and anchor screws (3-8). The ground wire was soldered
267 to an anchor screw and inserted above cerebellum. Electrodes were slowly lowered to
268 desired depth, pinned to the EIB board, and secured with superglue followed by
269 Metabond (Parkell, NY, USA). The entire head stage apparatus was held to the skull
270 and encased with dental cement (Stoelting, IL, USA).
271
272 **Single-unit Probe Implantation:** To record single-units we used a 32-channel
273 stationary array with microwires arranged in a brush-like formation (see Francoeur et
274 al., 2021). Initial preparation of the animal and location of ground screws was identical
275 to the LFP probe surgical procedures described above. A cranial window with diameter
276 of 2mm was drilled with a 0.7mm micro drill (Stoelting, IL, USA) centered at the OFC

277 target location. Three probes targeted ventral OFC (AP:+3.5mm, ML: +/-1.5mm, DV:
278 5.0mm), four targeted lateral OFC (AP: +3.5mm, ML:+/-2.5mm, DV: 5.0mm), and one
279 implant had 16 electrodes in each region. The implant was lowered to desired depth
280 slowly under stereotactic control. A thin layer of superglue was applied to the skull
281 followed by a layer of Metabond (Parkell Inc., NY, USA) to seal the craniotomy. The
282 implant was secured to anchor screws and attached to the dry skull with dental cement
283 (Stoelting, IL, USA). The ground wire was pinned to the channel on the EIB board, and
284 the remaining exposed wires covered in dental cement.
285
286 At the conclusion of surgery, the skin was sutured closed, and rats were given a single
287 dose (1mg/kg) of buprenorphine SR for pain management. Rats recovered from surgery
288 on a heating pad to control body temperature and received sulfamethoxazole and
289 trimethoprim in their drinking water (60mg/kg per day for 8 days) to prevent infections.
290
291 Electrophysiology
292 LFP data was recorded using a 32-channel RHD headstage (Intan Technologies, CA,
293 USA; Part C3324) coupled to a RHD USB interface board (Intantech, Part C3100) and
294 SPI interface cable. We used plug-in GUI (Open Ephys) software for acquisition. Data
295 was recorded at 1Khz, with a band-pass filter set at 0.3 to 999 Hz during acquisition.
296 Physiology data was integrated with behavioral data using a lab-streaming-layer (LSL)
297 protocol (Ojeda et al., 2014), as described previously (Buscher et al., 2020).
298

Single-unit data was recorded using a 32-channel RHD headstage with signal amplified using a PZ5 Neurodigitizer and RZ2 bioamp processor (TDT, FL, USA). Recorded signals were processed using Synapse software (TDT) at a sampling rate of 25KHz, high-pass filter of 300Hz and low-pass filter of 3000Hz. Behavioral markers were also integrated using LSL protocol.

Statistical Analysis

LFP Time Frequency Analysis: We carried out standard pre-processing and time frequency (TF) analyses using custom MATLAB scripts and functions from EEGLAB (Fakhraei et al., 2021b, 2021a; Francoeur et al., 2021). 1) Data epoching: We first extracted time-points for events of interest during each task. This paper focuses on neural activity time-locked to response/reward (opposed to trial onset) to examine neural activity during the reward-feedback period. Time-series data was extracted for each electrode (32), for each trial and organized into a 3D matrix (electrodes, times, trials). 2) Artifact removal: Activity was averaged across the time/electrodes to get a single value for each trial. Trials with activity greater than 4X standard deviation were treated as artifact and discarded. 3) Median reference: At each time-point, the “median” activity was calculated across all electrodes (32) and subtracted from each electrode as a reference. 4) Time-Frequency Decomposition: A trial by trial time-frequency decomposition (TF decomposition) was calculated using a complex wavelet function implemented within EEGLAB (newtimef function, using Morlet wavelets, with cycles parameter set to capture frequency windows of between 2 to 150 Hz (2 to 70 Hz in the go/wait task) and otherwise default settings used). We calculated the analytic amplitude

322 of the signal (using the abs function). 5) Baseline normalization: To measure evoked
323 activity (i.e. change from baseline) we subtracted, for each electrode at each frequency,
324 the mean activity within a baseline window (1000-750ms prior to the start of the trial). 6)
325 Trial averaging: We next calculated the average activity across trials for specific trial
326 types at each time-point and frequency for each electrode, thus creating a 3D matrix
327 (time, frequency, and electrode) for each behavioral session. Trials of interest were
328 different for each task: In the go/wait task we analyzed go-cue rewarded trials vs. wait-
329 cue unrewarded trials; temporal discounting task we separated high and low-reward
330 choice at each delay condition; and PRL task we separately analyzed target/ non-target
331 choices and whether those choices were rewarded or non-rewarded (Fig. 1). 8)
332 Comparison across animals: Prior to averaging across sessions/animals, we "z-scored"
333 the data recorded from each behavioral session by subtracting the mean and dividing
334 by the standard deviation of activity in each electrode (at each frequency) over time. Z-
335 scoring was helpful for normalizing activity measured from different animals prior to
336 statistical analysis. These pre-processing steps resulted, for each session, in a 3D time-
337 frequency-electrode matrix of dimensions 200x139x32 which was used for further
338 statistical analyses as described below.
339
340 We performed two main types of statistical analyses on the time-frequency-electrode
341 (TFE) data. 1) Trial-type mean: We analyzed the mean activity at each frequency and
342 evaluated the statistical significance of this mean for trial types of interest (see above).
343 Mean activity was calculated at each time and frequency point for each electrode across
344 behavioral sessions. Statistical significance was estimated with a one-sample, two-

sided t-test (ttest function in MATLAB), compared with the null hypothesis of Z=0. Because we had already performed a “baseline” subtraction (as described above), this analysis captured whether there was a significant increase or decrease in activity compared to baseline. FDR-correction was applied to the entire time-frequency-electrode matrix (FDR-corrected p-value threshold set to 0.05). 2) Trial-type contrast: For some analyses we were interested in the “contrast” of activity between trial-types (i.e., [go-cue correct – wait-cue incorrect] in the go/wait task). Contrasts were performed by subtracting the mean TFE data matrix for different trial types estimated from each session. We then calculated the average increase between trial types across sessions for the contrast. Statistical significance was performed with a one-sample, two-sided t-test applied to this contrast, using a null-hypothesis of 0 (i.e., no difference between trial types). FDR-correction was performed across all times-frequencies-electrodes. We performed similar analyses also for average evoked power within a particular frequency band of interest. We used the following frequency bands: delta power (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (15 -30 Hz), gamma (40-70Hz) and high gamma (70-150 Hz).

Data from the go/wait and PRL tasks was time-locked to response. We analyzed the full two second window of reward-feedback (500-2500ms after response). To account for the variable delay-to-reward conditions in the temporal discounting task, data was time-locked to reward delivery. We analyzed the first second of activity post-reward (0ms to 1000ms after reward onset) to control for the difference in water delivery between the high (3s) and low (1s) reward magnitudes. The mean / SEM of that data was calculated

368 across sessions, and we used a one-sample, two-sided t-test (again compared to the
369 null hypothesis that Z=0) to evaluate significance of the means. We applied a Bonferroni
370 correction to the p-values from this data. To identify oscillatory frequencies active during
371 reward-feedback, we used ANOVAs to compare mean normalized power averaged
372 across all electrodes, sessions, and time. To further understand neurobehavioral
373 correlations in the temporal discounting task, we used regression models within a
374 general linear model framework, performed in MATLAB. The mean oscillatory frequency
375 power (from 0 to 1000ms after reward) on a particular session was the dependent
376 variable and the overall likelihood of choosing the high-reward choice on that session
377 was the independent variable. We did not control for delay condition as we already
378 determined in subsequent analyses that it was indeed a significant factor in modulating
379 reward-related activity. We calculated correlations across all sessions/ animals for each
380 of the 12 reward-related brain regions, followed by FDR correction. General linear
381 models (GLM) used in this paper were simple regression models in which the slope (β
382 value) and the significance (p-value) are reported for each brain region of interest.
383
384 **Single-unit Analyses:** Single-unit activity was recorded in 8 animals performing the
385 go/wait task. We extracted behavioral markers of interest, LFP streams, and spiking
386 data. Neural data was cleaned and referenced off-line using Wave_Clus v.2.5., a
387 Matlab-based spike-sorting program (Quiroga et al., 2004). Signals were processed as
388 two polytrode (16- electrode) groups with median referencing applied to each channel
389 (see Francoeur et al., 2021). A threshold of spike detection was set at 5 times standard
390 deviation of voltage potential for each channel. Broken channels, with large impedances

beyond 10 MOhm were excluded from referencing/ clustering. Spikes in each polytrode group, as identified by Wave_Clus, were examined manually for characteristics of single-units (average spiking rate within the whole session was more than 0.5Hz, fewer than 1.5% inter-spike interval violations (<3ms), waveform resembling action potentials as opposed to sinusoidal noise artifacts, and clusters distinct from others in the principal component space) (Swindale and Spacek, 2014; Rey et al., 2015; Harris et al., 2016). Spikes meeting criteria were time-locked with behavioral events in MATLAB.

MATLAB functions were used to create peri-stimulus time histograms and raster plots at trial types of interest (go-cue rewarded / wait-cue unrewarded trials). Activity was time-locked to response (time point 0). PSTH (spikes/second) were made in 25ms bins from -2s to +2s after response and gaussian filtered for smoothing. Activity was baseline normalized by subtracting the average firing rate during the pre-response baseline (first 750ms from trial onset) on go-cue correct trials from firing rate in subsequent time bins. Units from recording days with at least 30 trials and a minimum baseline firing rate of 2 spikes/s were further analyzed for their task-related activity. A unit was considered “task- modulated” if the average firing rate was 2 standard deviations above or below the baseline activity for > 75 consecutive ms in either go-cued correct or wait-cued incorrect trials. Task-modulated units were categorized further based on the timing of their activation/ suppression relative to response. Anything with significant (+/- 2 standard deviations) activation or suppression from -500 to 0ms from response was considered pre-response activity (action). Anything with later activation or suppression (0 to 2000ms) was considered post-response (outcome) activity.

414

415 Finally, we calculated spike-field coherence (SFC) to relate spiking activity to field-
416 potential oscillations. SFC ranges from 0 (spikes operate independently from LFP) to 1
417 (spikes are completely aligned to LFP) (Buzsáki et al., 2012; Ray, 2015; Lashgari et al.,
418 2018). To calculate SFC, we matched spike times to the continuous LFP signal from an
419 electrode with minimal artifacts by multiplying the sampling rate of the LFP (1.0173).
420 Next, we down sampled to normalize non-standard sampling frequencies and epoched
421 the LFP data to align with task events (described above). SFC was analyzed separately
422 at each oscillatory frequency (delta, theta, alpha, beta, gamma) for each trial type (go-
423 cue correct and wait-cue incorrect), time-locked to response. Units were categorized as
424 "high" or "low" SFC by using a median split based on beta frequency SFC values from
425 500-2500ms after response (reward feedback) on go-cued correct trials. Units one
426 standard deviation above the median SFC value were categorized as "high" and units
427 one standard deviation below were categorized as "low". A two-way ANOVA was used
428 to compare firing rate on go-cue correct vs. wait-cue incorrect trials x SFC value at beta
429 frequencies (classified as "high" or "low").

430

431 **Histology:** Histological analyses were completed for 22/ 28 rats with LFP implants
432 (go/wait task N=6/11; temporal discounting N=9/10; probabilistic reversal learning
433 N=7/7) and for 5/8 rats with single-unit implants. At completion of recording sessions
434 wire tips were marked by passing 12μA current for 10s through each electrode (Nano-Z,
435 Neuralynx, MO, USA). Rats were sacrificed under deep anesthesia (100 mg/kg
436 ketamine, 10 mg/kg xylazine IP) by transcardiac perfusion of physiological saline

437 followed by 4% formalin. Brains were extracted and immersed in 4% formalin for 24
438 hours and then stored in 30% sucrose until ready to be sectioned. Tissue was blocked
439 in the flat skull position using a brain matrix (RWD Life Science Inc., CA, USA). Brains
440 with field potential probes were sectioned frozen in the coronal plane at 50μm thick.
441 Brains with single-unit electrodes were paraffin embedded and sectioned 20 μm thick
442 due to diameter difference in wires (processed by Tissue Technology Shared Resource;
443 CCSG Grant P30CA23100). Brain slices were Nissl stained using thionin to identify the
444 course of the electrode tracks. Sections were processed with a slide scanner at 40x
445 magnification (Zeiss, Oberkochenn, Germany; Leica Biosystems, IL, USA). Positions of
446 electrodes were inferred by matching landmarks in sections to the rat atlas (Paxinos &
447 Watson, 2013) when electrode tips could not be identified.
448
449 **Results**
450
451 Beta Frequency Oscillations Linked with Positive Valenced Feedback
452 A full description of behavior on the go/wait task in animals with LFP probes can be
453 seen in our prior publication describing inhibition and stimulus-response oscillatory
454 signatures (Fakhraei et al., 2021a). Animals were shown two visual stimuli, one which
455 required an immediate response (go-cued trials) and the other which required the
456 animal to withhold from responding for 2s (wait-cued trials) (**Fig. 2A**). Across behavioral
457 sessions, animals performed better on go-cued trials compared to wait-cued trials (**Fig.
458 2B**). Animals could generally distinguish between go and wait-cues indicated by a
459 significant difference in reaction times ($t_{(59)}$ =17, $p<.001$) and poorer accuracy on wait-

460 cued trials ($t_{(59)}$ =18, $p<.001$). On go-cued trials animals had a mean reaction time of
461 610 +/- 160ms and correctly responded within 2s of the visual cue on 94.0 +/- 9.2% of
462 trials (data averaged across 60 sessions from 11 animals). On wait-cued trials animals
463 took longer to respond (1700 +/- 460ms) and were able to correctly wait on 41.0 +/-
464 23.0% of trials **(Fig. 2B)**.
465
466 The primary goal of our first analysis was to identify electrophysiological markers that
467 differentiated between positive and negative feedback on the go/wait task. Feedback on
468 correct trials consisted of water delivery 400ms after the response at the rate of
469 10µl/sec, for a duration of 2s. Feedback on incorrect trials consisted of a 5s flashing
470 house-light and an auditory 1000Hz tone with no water delivery. We began by analyzing
471 electrophysiological activity, time-locked to the response, from lateral OFC (lOFC)- a
472 cardinal reward-processing brain region. We analyzed mean TF power across sessions
473 (N=60) from correct go-cued trials (animals received a go-cue and responded within two
474 seconds) and incorrect wait-cued trials (animals received a wait cue but responded
475 within two seconds). The other two trial types (go-cued incorrect and wait-cued correct)
476 had fewer trials and therefore were not included in the subsequent analyses. During the
477 reward feedback period (starting 400ms after the response denoted by the vertical red
478 line in **Fig. 2C,D**) we observed an increase in beta (15-30 Hz) and gamma (50-70 Hz)-
479 frequency activity that occurred on rewarded trials (go-cue correct). There was far less
480 feedback-related activity on unrewarded trials (wait-cue incorrect). The activity that was
481 present on unrewarded trials decayed to baseline quickly when no water was delivered
482 (**Fig. 2C**). We also observed post-response/pre-feedback alpha-frequency activity (8-12

483 Hz) on rewarded and non-rewarded trials (**Fig. 2C**). The normalized power more clearly
484 distinguishes between peak timing of different frequencies (**Fig. 2D**). On rewarded trials
485 the peak beta activity in lOFC occurred at 805ms after the response (~400ms after
486 reward onset) and lasted for around 2 seconds- the approximate time of reward delivery
487 (**Fig. 2D**). The peak timing of alpha activity was 302ms after response on rewarded
488 trials and 272ms after response on non-rewarded trials (**Fig. 2D**). As this peak occurred
489 prior to the reward delivery (400ms after response) and was observed in both correct
490 and incorrect trial types, we surmised alpha activity may have been linked with some
491 aspect of post-response processing or reward anticipation, and not reward-feedback
492
493 To examine frequency differences more closely, we took the average power across
494 delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (15-30 Hz), and gamma (50-70 Hz)
495 frequencies during a two second reward-feedback window from 500-2500ms after
496 response (corresponding to the timepoint of reward delivery). We used a two-way
497 ANOVA to compare reward-feedback power across frequencies and reward feedback
498 (go-cued correct trials/ wait-cued incorrect trials). We found a main effect of frequency
499 ($F_{(4,650)}$=14.6, *p<.001*), main effect of trial type ($F_{(4,650)}$=1.74, *p<.001*), and a significant
500 interaction between frequency and trial type ($F_{(4,650)}$=5.77, *p<.001*). Post-hoc analyses
501 (Bonferonni corrected) revealed lOFC power at theta, alpha and beta frequencies were
502 different between rewarded and non-rewarded trials (**Fig. 2E**). The mean difference (go-
503 cue correct – wait-cue incorrect) in theta-frequency power was -2.86, SE of difference=
504 0.517, *corrected p<0.001.* Similarly, alpha activity was also significantly greater on
505 incorrect trials (mean difference= -2.68, SE of difference= 0.517, *corrected p<0.001*).

506 The mean difference in beta frequency activity was 1.17, SE of difference= 0.517,
507 *corrected p= <0.05*, on go-cue correct – wait-cue incorrect trials and thus was the only
508 frequency band that positively modulated power with reward valence. Beta power was
509 significantly greater on go-cue correct trials (mean power= 1.34, SEM=0.380) and
510 showed almost no modulation from baseline on wait-cue incorrect trials (mean power=
511 0.06, SEM= 0.10) during reward-feedback (**Fig. 2E**).
512
513 To better understand the spatial distribution of the positive valence beta activity beyond
514 lOFC, we next analyzed beta power from 12 putative reward-related brain regions (M2,
515 A32D, A32V, vOFC, ALM, LFC, Ains, lOFC, VMS, NAcS, NAcC, BLA) (**Table 1**). As
516 seen in lOFC, there was a main effect of trial type (go-cue correct vs. wait-cue incorrect)
517 ($F_{(11,1560)}$=92.5, *p<0.001,* two-way ANOVA) on beta frequency power (**Fig. 2F**). The
518 average activity on rewarded trials across the 12 brain regions was 1.20 (SEM=0.05)
519 compared to 0.18 (SEM=0.03) on non-rewarded trials. There was also a main effect of
520 area ($F_{(11,1560)}$=10.6, *p<0.001,* two-way ANOVA) but no significant trial x area interaction
521 ($F_{(11,1560)}$=1.52, *p=0.117*). Planned post-hoc (Bonferroni corrected) tests revealed a
522 significant difference in the contrast between trials (rewarded – non-rewarded) in the
523 A32D, BLA (*p<0.001*), M2, LFC, Ains, LOFC, and VMS (*p<0.05*) electrodes. BLA had
524 the greatest beta power on rewarded trials (M=3.12) and the largest difference between
525 rewarded and unrewarded trials (mean difference =1.80, SE of difference= 0.368,
526 corrected *p=<0.001*) (**Fig. 2F**).
527
528 <u>Beta-Oscillations Related to Single-unit Activity in OFC During Reward Feedback</u>

Despite the name, “local” field potentials are challenging to properly localize (Pesaran et al., 2008; Ray et al., 2008; Buzsáki et al., 2012). To better understand whether the beta frequency activity observed during positive reward-feedback was related to local spiking activity within a particular brain region, we recorded single-units from the OFC of 8 male Long-Evans rats performing the go/wait task. The version of the task used for single-unit recordings was slightly modified (due to an error in coding the task), from a 400ms delay between response and reward as noted above, to only a 30ms delay. As we were focused on the period post-feedback, this error did not substantially affect our analyses or interpretation of results. We recorded 376 neurons across 62 session (5.81 units +/- 0.03 per session). While performance on the task differed in these animals compared to those with LFP implants, rats were still able to discriminate between go-cue and wait-cue trials. Reaction time was different between trial types ($t_{(61)}$=7.8, $p<.001$): 800 +/- 180ms on go-cue trials and 1100 +/- 330ms on wait-cue trials. Accuracy on go-cue trials was 73.0 +/-24.0%, compared to 30.0 +/- 20.0% on wait-cue trials ($t_{(61)}$=8.7, $p<.001$) (**Fig. 3A**).

After excluding sessions with a limited trial number (< 30 trials) and units with low firing rates (< 2 spikes/s), 228 units were included for subsequent analyses. 125 neurons (33%) were defined as task-modulated based on our criteria of an increase/decrease of two standard deviations above baseline for >75 consecutive ms. This included peak firing rate increases or decreases both prior to the response (action-related) or after the response (outcome-related) (**Fig. 3B**). The average peak firing rate activity of action-related neurons was 375ms before the response (time 0). The average peak firing rate

552 of outcome-related neurons was 225ms after response (~195ms after reward onset).
553 103 neurons were not task-modulated based on our criteria.
554
555 Our main goal for this analysis was to study whether single-units were modulated by
556 reward-related beta-oscillations. We used spike-field-coherence (SFC) to assess the
557 relationship between spiking and oscillatory activity during the reward-feedback period
558 (0 to 2000ms after response) (Howe et al., 2011; Buzsáki et al., 2012; Ray, 2015;
559 Lashgari et al., 2018). Units with significant task-related suppression or missing LFP
560 data stream were not included in SFC analysis (173 units remaining). We observed that
561 outcome-related neurons with greater firing rate on go-cue correct trials compared to
562 wait-cue incorrect trials ("correct preferring") showed increased beta frequency SFC
563 modulation on correct trials during reward-feedback (**Fig. 3C**). Outcome-related neurons
564 with greater firing rate on wait-cue incorrect trials ("incorrect preferring") did not show as
565 great of SFC modulation at beta frequencies (**Fig. 3C**). We used a two-way ANOVA to
566 statistically test this observation. We grouped neurons into two categories based on
567 their beta SFC value during the reward-feedback period. Neurons one standard
568 deviation higher-than-average SFC were compared to neurons one standard deviation
569 lower-than-average SFC. We found a main effect of SFC category (high vs. low)
570 ($F_{(1,314)}$= 5.11, p=*0.024*) on reward-feedback firing rate, and a significant interaction (F
571 $_{(1,314)}$= 4.45, p=*0.036*) between SFC category and trial type (go-cue correct vs. wait-cue
572 incorrect) (**Fig. 3D**). Neurons in the "high" SFC group had an average firing rate of 0.66
573 +/-0.41 spikes/s on go-cue correct trials compared to the "low" SFC group neurons
574 which had an average firing rate of -0.78+/-0.30 spikes/ s. On wait-cue incorrect trials,

575 firing rate was not modulated based on SFC value. Neurons in the “high” SFC group
576 had an average firing rate of 0.20 +/- 0.27 spikes/s on wait-cue incorrect trials and “low”
577 SFC neurons had an average of 0.15 +/- 0.17 spikes/s. Importantly, the firing rate of
578 “high” and “low” SFC neurons was similar on non-rewarded (wait-cue incorrect) trials
579 but was significantly different on rewarded (go-cue correct) trials (mean difference [high-
580 low]= 1.44, *corrected p=0.004*) (**Fig. 3D**).
581
582 Beta Power Reflects Dimensions of Reward Value
583 Our data from the go/wait task suggested that beta activity within lOFC and other
584 reward-related regions is a marker of positive valence (i.e. rewards). Single-units from
585 OFC with higher reward-locked beta SFC are also more likely to be positively
586 modulated by reward while those with low reward-locked SFC are more likely to be
587 suppressed by rewards. Given prior work, we hypothesized that beta-oscillations on the
588 go/wait task might reflect: 1) expected reward valuation; 2) reward-prediction error
589 which occurs during the presentation of a reward but are typically in the opposite
590 direction of expected reward value (i.e., higher when the expected value is less); 3)
591 some non-specific aspects of reward consumption unrelated to encoded value or
592 prediction. To further explore these hypotheses, we recorded LFP activity on a new set
593 of animals (N=10) trained to perform a temporal discounting task (**Fig. 4A**). On this task,
594 animals were given the choice of a low-value reward delivered immediately (with a fixed
595 standard delay of 500ms after response), or a higher-value reward delivered at variable
596 delays of 500ms up to 20 seconds. To allow for greater numbers of trials for
597 electrophysiological analysis, delays on the high-reward condition were kept constant

598 throughout each session but varied across sessions. Reward-value was encoded as
599 either a 1s (10μL) or 3s (30μL) period of reward delivery. Previous work suggests
600 reward value is negatively influenced by temporal costs associated with earning a
601 reward (Winstanley et al., 2004b; Kable and Glimcher, 2007; Roesch et al., 2011;
602 Lefner et al., 2021). In the context of this task, the relative value of the low reward
603 choice increased, and the relative value of the high reward choice decreased as the
604 length of the delays and thus the temporal cost of high reward is increased. Results
605 from the temporal discounting task are based on 128 total sessions (average for each
606 rat was 2 sessions/ variable delay). Behaviorally, animals' preference shifted from high-
607 value choice to the low-value choice as the delay to reward increased ($F_{(5,45)}$ =30.9, p
608 =<*0.001,* ANOVA) (**Fig. 4B**). When delays of each choice were the same (500ms),
609 animals strongly prefer the 3x magnitude reward (90.4 +/- 1.4 % high-value choices per
610 session). When the high-value reward follows a 20s delay, rats only select the high
611 reward choice 24.2 +/- 6.8% of trials, showing a shifted preference for the immediate,
612 low-value reward. We do see individual differences emerge in the average rate of
613 discounting across delays. As seen in the group average, there is a main effect of delay
614 (p <*0.001*), but also a main effect of individual rat on the amount of high-value choices
615 made per session ($F_{(9,45)}$=7.02, p<*0.001,* two-way ANOVA).There was no interaction of
616 individual rat x delay on discounting behavior (p=*0.15*) (**Fig. 4B**).
617
618 The first question we asked was whether OFC beta power is modulated by expected
619 reward value. Specifically, we hypothesized that if beta reflects an aspect of value, then
620 power should be greater for the high (3x) compared to the low (1x) reward magnitude

621 when delays were the same (500ms for both). We analyzed only the first second of
622 activity post-reward to ensure that, for both trial types, animals were receiving the same
623 quantity of reward during the period of analysis. Thus, we surmised that any differences
624 in activity would have to reflect differences in expected reward value and not reward
625 consumption per se. We first investigated data across all frequencies (delta: 1-4 Hz;
626 theta: 4-8 Hz; alpha: 8-12 Hz; beta: 15-30 Hz; gamma: 50-70 Hz; and high gamma: 70-
627 150 Hz) at the lOFC electrode. We measured high-gamma on this task as a proxy for
628 single-unit activity (Ray et al., 2008). Using a two-way ANOVA, we found a main effect
629 of frequency ($F_{(5,216)}$ =26.5, $p<0.001$), but no main effect of trial type (high vs. low-value
630 choice) ($p=0.107$) (**Fig. 4C**). There was however a significant trial type by frequency
631 interaction ($F_{(5, 216)}$= 5.59, $p<0.001$). Post-hoc analyses (Bonferonni corrected) detected
632 that only beta and high-gamma frequencies had greater lOFC power on high-value
633 choices. The mean difference in beta frequency power between high and low reward
634 choices was 3.72, SE of difference= 1.01, *corrected p=0.002*, and the mean difference
635 at high-gamma frequencies was 2.86, SE of difference= 1.01, *corrected p=0.03* (**Fig.**
636 **4C**).
637
638 The next question we asked is whether reward-locked beta power was sensitive to the
639 temporal delays of reward. If beta power reflected reward value, we hypothesized that
640 the difference between high and low-value rewards should be modulated with
641 increasing delays, reflecting the discounted value of the delayed high-value choice. We
642 indeed observed this relationship in lOFC using a two-way ANOVA to statistically
643 measure whether there was an effect of delay length (0.5,1 ,2,5,10,20s) and trial type

644 (high vs. low reward) on beta power during reward feedback. We observed a main
645 effect of delay ($F_{(5,2724)}$=20.6, *p<0.001*) , a main effect of trial type ($F_{(1,2724)}$=8.05,
646 *p=0.005*), and a significant interaction between the two factors ($F_{(5,2724)}$=26.8, *p<0.001*)
647 (**Fig 3D**). Post-hoc (Bonferroni corrected) tests at each delay condition showed that at
648 low delays (0.5s, 1s) there was significantly greater beta activity for high-value choices
649 (0.5s delay, mean difference [high-low] in power = 2.13, SE of difference= 0.302,
650 *corrected p =<0.001*; 1s delay, mean difference = 0.833, SE of difference= 0.23,
651 corrected *p = 0.014*). At a moderate delay (2s) there was no significant mean difference
652 [high-low] =- 0.777, SE of difference= 0.362, corrected *p =.192*); and with longer delays
653 (5s, 10s, 20s) there was greater beta power on low-value choices (5s delay mean
654 difference [high-low] =-0.657, SE of difference= 0.246, corrected *p=0.046*; 10s delay
655 mean difference =-1.83, SE of difference= 0.320, corrected *p =<0.001* ; 20s delay
656 mean difference = -1.80, SE of difference= 0.300, corrected *p =<0.001*) (**Fig. 4D**). We
657 performed a second follow-up analysis to look at the effects of delay separately for the
658 low-value and high-value trials at the lOFC electrode. Beta activity was significantly
659 different across delay conditions for high-value choices ($F_{(5,122)}$=4.14, *p=0.002,* ANOVA)
660 (**Fig. 4E**). Post-hoc Bonferroni comparisons confirmed significant differences in beta
661 power at the smallest 500ms delay compared to 5s, 10s, and 20s delays (*corrected p=*
662 *<0.05*) . Importantly, beta power was not different across delay conditions on low-reward
663 choices (*p=0.720)* even though the low reward is preferred compared to high reward at
664 increasing high-value choice delays (**Fig. 4E**). Thus, beta power in lOFC significantly
665 decreases as value of high reward is decreased but is not modulating the increased
666 number of low value responses as high value reward increases.

667

668 We next calculated the difference in power on high-value and low-value choices (high-
669 low reward) across 12 putative reward regions (M2, A32D, A32V, vOFC, ALM, LFC,
670 Ains, lOFC, VMS, NAcS, NAcC, BLA) to understand whether this effect was specific to,
671 or even strongest within, the lOFC (**Table 1**). In the following analysis, the total number
672 reported may be less than 12 brain areas x 128 sessions because in some sessions a
673 particular electrode may not have provided usable data (noise/ broken channels, etc.).
674 Using a two-way ANOVA, we observed a significant effect of delay ($F_{(5,1176)}$= 32.3 , p
675 <0.001) but no significant main effect of region ($p=.981$) or region x delay interaction
676 ($p>.999$) (**Fig. 4F**). High-gamma reward feedback oscillations also show a similar
677 pattern in power-related changes (**Supp Fig. 2A**). We found a main effect of delay
678 ($F_{(5,1176)}$=11.1, p <0.001) and no significant main effect of brain region ($p=0.989$) or
679 interaction ($p>.999$). Thus, temporal discounting of beta and high-gamma power seems
680 to reflect a value signal that is dispersed broadly across areas of the cortico-striatal
681 reward network.

682

683 Finally, to understand whether beta-oscillatory activity within this reward network was
684 related to choice (i.e., preference for selecting either the high or low-value choice), we
685 performed a logistic regression analysis with mean beta frequency power on a particular
686 session as the dependent variable and the overall likelihood of choosing the high-value
687 choice that session as the independent variable. A positive beta value indicated that as
688 beta power increased, the percent of high-value choices increased on that session as
689 well. We ran this analysis for each of the 12 brain regions, followed by FDR correction,

using power from the difference (high-low reward) between trials. Strikingly, all 12 brain areas showed higher and significant (FDR-corrected) positive relationships with high-reward choice when the contrasting beta power on high and low-value responses (**Fig. 4G**). This suggests that the relative difference in beta power between high and low reward reflects value-related choice behaviors across reward-related regions- that is, the larger the difference in value between those signals, the more likely animals were to select the high-value reward.

Beta Power Reflects Reward Certainty and Updates after Reversal

On a probabilistic reversal learning (PRL) task, subjects first learned that one response leads to a high-probability of reward ("target") and an alternate response would lead to a low-probability of reward ("non-target"), then subjects flexibly updated this representation after contingencies were reversed. In our version of this task, each day one response port would be randomly assigned to start as the target NP (rewards delivered 80% of the time) while the other, non-target, port would deliver rewards 20% of the time and reverse when 8 out of the last 10 trials were target choices (regardless of reward outcome) (**Fig. 5A**). Thus, on any session, subjects needed to dynamically modulate their behavior to track reward contingencies. The following analyses are from 122 behavioral sessions, from 7 male rats. The minimum number of behavioral sessions/ rat was 7 and on average rats performed 17 PRL sessions. 36 PRL sessions included LFP data (average= 5 LFP sessions/ rat). On the first session rats performed an average of 1.33 reversals (SEM=0.211) and each animal showed significant improvement in number of reversals across time ($t_{(6)} = 4.39$, $p=.007$, paired t-test) (**Fig.**

713 **5B**). On the last session rats performed an average of 10.7 reversals (SEM=2.03). Rats
714 tended to make the same choice after receiving a reward (termed a "win-stay"
715 response). On 64.2 +/- 10.9% of rewarded trials, rats returned to the same response
716 port on the subsequent trial (**Fig. 5B**).
717
718 Based on data gathered from the temporal discounting task, we hypothesized that beta
719 power represents subjective value and is dynamically adjusted according to the value
720 representation within a specific context. Thus, we expected to see greater beta power
721 on target rewards (high reward probability), compared to non-target rewards (low
722 reward probability). We predicted that beta-oscillations would dynamically track the
723 choice leading to the higher-expected value and shift after a reversal. To study this, we
724 compared oscillatory activity when rewards were delivered for the high-probability
725 choice (target reward) to when the same reward was delivered for the low-probability
726 choice (non-target reward). On the lOFC electrode, there was a main effect of
727 frequency ($F_{(5,420)}$=9.94, $p<0.001$, two-way ANOVA) and a main effect of trial type
728 ($F_{(5,420)}$=12.9, $p<0.001$,two-way ANOVA), but no interaction ($p=0.763$) (**Fig. 5C**). The
729 mean difference between trials (target - non-target rewards) was greatest at beta and
730 high-gamma frequencies. Beta power was greater for high-probability of reward (mean
731 difference of [target reward – non-target reward] = 0.934, SE of difference= 0.447).
732 High-gamma power was also greater for target trials compared to non-target trials,
733 (mean difference= 1.17, SE of difference= 0.447) (**Fig. 5C**). The shaded error plot
734 illustrates the pattern of greater beta power for high-probability compared to low-
735 probability reward at the lOFC electrode (**Fig. 5D**). To confirm that there was a

736 significant reward effect as seen in the previous tasks, we compared beta frequency
737 power for rewarded and non-rewarded outcomes at both trial types (target and non-
738 target) (**Fig. 5E**). Consistent with our findings that beta activity signals positive reward
739 valence, on the PRL task beta power was greater on rewarded (M=1.135, SEM=2.45)
740 compared to non-rewarded trials (M=-0.435, SEM= 0.775) ($t_{(70)}$=3.67,$p<0.001$,unpaired
741 t-test) (**Fig. 5E**).
742
743 We next examined changes in beta power across the other 12 electrodes. Comparing
744 beta power on target and non-target rewards in the PRL task, we found a significant
745 main effect of trial type ($F_{(11,722)}$= 15.5, $p<0.001$, two-way ANOVA), main effect of area
746 ($F_{(11,722)}$= 5.49, $p<0.001$, two-way ANOVA), but no significant interaction ($p=0.716$) (**Fig.**
747 **5F**). The total number of sessions in the following analyses may be less than expected,
748 because on certain sessions a particular electrode may have been missing LFP data.
749 Post-hoc (Bonferroni corrected) tests revealed the main effect of area was influenced by
750 increased power on target choices within anterior insula and lOFC brain regions
751 compared to others. Pooling data across all 12 reward areas, beta power was greater
752 on targets (mean beta power= 1.3 SEM=0.26) than non-targets (mean beta power=0.59
753 SEM=0.13) ($t_{(11)}$=5.3, $p<0.001$, paired t-test). We performed a similar analysis to see if
754 this activity was associated with subsequent behavior by the animal. We compared
755 beta power for “win-stay” trials (rewarded trials in which animals chose the same
756 response on the subsequent trial) and “win-go” trials (rewarded trials in which animals
757 chose the different response on the subsequent trial) in all 12 electrodes. We observed
758 a main effect of trial type ($F_{(11,709)}$= 4.92, $p=.027$) and main effect of area ($F_{(11,709)}$=4.76,

759 *p=<0.001*), and no significant interaction between trial type and area (*p=0.991*) (**Fig.**
760 **5G**). Post-hoc (Bonferroni corrected) tests revealed the main effect of area was
761 influenced by increased power on win-stay choices within anterior insula, lOFC, and
762 A32V brain regions compared to others. Pooling data across all 12 brain regions, beta
763 power was greater on win-stay trials (mean beta power=1.2 SEM=0.25) than win-go
764 trials (mean beta power=0.49 SEM=0.16) ($t_{(11)}$=6.9, *p<0.001*). High-gamma frequency
765 power also showed a main effect of trials (target vs. non-target) ($F_{(11,722)=}$ 33.2, *p=.001)*
766 and main effect of area ($F_{(11,722)=}$ 2.93, *p=.001)* but no trial x area interaction (*p=0.979*)
767 (**Supp Fig. 2B**). Similarly, high-gamma power was significantly different for win-stay
768 trials than win-go trials ($F_{(11,709)=}$ 10.7, *p=.001*) and had a main effect of area ($F_{(11,709)=}$
769 2.21, *p=.012*) (**Supp Fig. 2C**).
770
771 Finally, pooling data across all 12 brain regions, we examined how beta power reflected
772 a change in reward contingencies. We used a two-way ANOVA to compare beta power
773 on rewarded and non-rewarded target choices before and after a reversal. Specifically,
774 we analyzed the last four rewarded trials pre-reversal and the first four rewarded trials
775 post-reversal from the "new" target; and the same for non-rewarded trials. Average
776 activity pre and post-reversal beta power was calculated separately at each electrode
777 for each reversal pooled across animals/sessions. Analyzing the data this way we found
778 a main effect of reward outcome ($F_{(1,332)}$=12.0*, p<0.001*) on beta power in reward
779 regions, no main effect of reversal (pre vs. post) (p=0.133), but a significant interaction
780 between reward outcome and reversal ($F_{(1,332)}$=13.4*, p<0.001)* (**Fig. 5H**). Post-hoc
781 (Bonferroni corrected) comparisons show a selective decrease in beta power after a

782 reversal that only occurs on rewarded target trials. The mean difference of beta power
783 pre vs. post reversal was 0.521, SE of difference= 0.062, *corrected p <0.001* on
784 rewarded trials. The mean difference of beta power pre-post reversal was -0.108, SE of
785 difference= 0.062, *corrected p=0.166*, on non-rewarded trials (**Fig. 5H**). This is largely
786 consistent with the idea that beta power reflects the expected outcome value which,
787 immediately after a reversal, is still low for the "new" target. Beta-oscillations are
788 selectively accumulating evidence about rewarded outcomes and modulating
789 expectancy by tracking repeated positive outcomes and does not meaningfully reflect a
790 signal related to the lack of reward (expected or unexpected).
791
792 Verification of LFP Probe Locations at Target Brain Areas
793 Coronal sections stained with thionine to capture cell bodies were used to verify the
794 electrode placement in target brain regions. For each cannula (1-8), a graphical
795 representation of a rat brain atlas (Paxinos & Watson, 2013) shows the identified center
796 of recording sites at each DV location (four per cannula) (**Supp Fig. 1**). Colored dots
797 represent the task the animals belong to (green: go/wait N= 6/11; pink: temporal
798 discounting N= 9/10; blue: PRL N=7/7). An example coronal slice at the corresponding
799 AP location is also shown for each cannula placement with magnification of each track
800 in the brain. The table includes the AP, ML, and DV coordinates for all 32 electrodes
801 and their corresponding nomenclature. The location of single-unit OFC recording
802 electrodes is also shown (B) from a range of +4.2 AP through +3.25 AP relative to
803 bregma. The LO/VO subdivisions are outlined on the example coronal sections taken

804 from the rat brain atlas. Electrodes span both divisions. The graphical representation
805 includes all electrode tracks (N=5/8 rats) (**Supp Fig. 1**).
806
807 **Discussion**
808
809 We observed changes in beta (15-30 Hz) and high-gamma (>70 Hz) frequencies that
810 scaled with multiple aspects of reward value. Each task contributed something unique to
811 our findings and using different cohorts of animals offered replication for greater
812 certainty of our findings. Measuring LFP on a behavioral inhibition task (go/wait), we
813 identified beta power-related changes that signaled positive valence (rewarded) trials
814 during reward-feedback. Firing rates of single-units in OFC were also modulated at beta
815 frequencies during positive reward outcome. On a temporal discounting task, beta
816 power corresponded to subjective reward value and on a PRL task beta was heightened
817 on trials where reward delivery was more certain and was responsive to changes in
818 reward contingencies. Generally, medial prefrontal cortex, lateral OFC, and anterior
819 insula had the greatest beta power during positive reward feedback. M2 and ALM
820 consistently show lowest levels of beta power during reward outcome and do not
821 differentiate between trial contrasts. Subtle variations in activity between tasks may
822 represent an example of functional segregation between cortical subdivisions seen
823 previously on other reward-guided tasks (Dalton et al., 2016; Hardung et al., 2017;
824 Francoeur and Mair, 2019; Turner and Parkes, 2020; Salehinejad et al., 2021). Lateral
825 divisions of cortex are thought to be important for tasks relying heavily on inhibition,
826 error detection and behavioral flexibility whereas medial/ventral cortex may be more

central in temporal discounting and decision-making (Turner and Parkes, 2020; Salehinejad et al., 2021). Specific task features, like visual predictive cues in the go/wait task, may also recruit different brain regions. The BLA, critical for forming stimulus-outcome memories (Wassum, 2022), was most robust during the go/wait task, potentially a reflection of the richer stimulus associations used to form predictive reward memories. We generally see evidence of a beta reward-processing signal broadly throughout the cortico-striatal network, but there are instances where distinct brain regions are more/less engaged based on task-dependent features.

Data from each of these tasks, when considered alone, could have multiple explanations – however, the similar relationships observed across animals/tasks help define the role of beta-oscillations in reward-processing. The most trivial explanation is that beta activity reflects an artifact time-locked with reward delivery such as movement (i.e. muscle/EMG-related contamination during reward consumption) or electrical noise associated with water delivery. This does not easily explain why on matched delays (500ms) on the temporal discounting task there was greater beta power for high-value compared to low-value trials during the first second following reward delivery when movements and electrical noise would be matched. Similarly, data from the PRL task indicates beta power was greater for high-probability responses which also has identical reward delivery to low-probability responses.

A different possibility is that the beta activity is physiological in nature but reflects a motor, not reward, process. Beta-oscillations have been well-characterized within motor

850 cortex (Witham et al., 2007; Elisabeth Kilavik et al., 2012; Feingold et al., 2015; Khanna
851 and Carmena, 2017; Luhmann et al., 2021) and dorsolateral striatum (Jenkinson and
852 Brown, 2011; Feingold et al., 2015; Schwerdt et al., 2020; Luhmann et al., 2021) and
853 tend to be largest pre/post-movement, but are classically reduced during movement
854 (Hammond et al., 2007; Engel and Fries, 2010; Elisabeth Kilavik et al., 2012; Khanna
855 and Carmena, 2017). This functional description fits with observations of beta activity in
856 Parkinson's disease patients who have trouble initiating movements and show
857 increased beta-oscillations related to symptom severity (Hammond et al., 2007;
858 Jenkinson and Brown, 2011; Khanna and Carmena, 2017; McGregor and Nelson, 2019;
859 Schwerdt et al., 2020). Thus, one explanation is that increased beta power reflects
860 motor inhibition during reward consumption. However, our data is not compatible with
861 this hypothesis in a few ways. First, sensorimotor beta-oscillations, as previously
862 described, are more localized within motor and dorsal striatum, whereas we observe
863 oscillations (and single-units related to beta-oscillations) more strongly within
864 orbitofrontal and ventral striatal projection areas. Second, as before, we believe our
865 data comparing high vs. low value reward (temporal discounting) and high-probability
866 vs. low-probability reward (PRL) argues against this interpretation, as it is unclear why
867 animals would be more stationary when consuming rewards on these trials where motor
868 requirements should be matched.
869
870 If beta-oscillations reflect reward-processing, then what might they represent at the time
871 of reward-feedback? We hypothesize that beta-oscillations reflect activity within a
872 cortico-striatal network that drives optimal decision-making based on expected reward-

873 value. We provide evidence that beta-oscillations during reward feedback modulate
874 activity based on task variables such as reward magnitude, temporal delay, and
875 probability of reward. Growing research has identified beta-oscillations outside of
876 sensorimotor networks related to attention (Shin et al., 2017; Schmidt et al., 2019), top-
877 down processing (Buschman and Miller, 2007; Engel and Fries, 2010), working memory
878 (Siegel et al., 2009; Marco-Pallarés et al., 2015; Spitzer and Haegens, 2017; Schmidt et
879 al., 2019) and outcome evaluation (Pesaran et al., 2008; Torrecillos et al., 2015).
880 Consistent with our findings, beta-oscillations during reward-feedback have been
881 observed in humans and animals (Cohen et al., 2007; Pesaran et al., 2008; van der
882 Meer and Redish, 2009; Howe et al., 2011; Rothé et al., 2011; HajiHosseini and
883 Holroyd, 2015; Marco-Pallarés et al., 2015; Zavala et al., 2018; Patai et al., 2022). EEG
884 and MEG measures in humans find beta-oscillations during positive-valence reward
885 within frontal-striatal circuits (Cohen et al., 2007; Marco-Pallares et al., 2008;
886 HajiHosseini and Holroyd, 2015; Marco-Pallarés et al., 2015; Zavala et al., 2018; Patai
887 et al., 2022). Similarly, increased beta power in cortico-striatal regions has been
888 observed in rodents approaching reward locations (Howe et al., 2011; Samson et al.,
889 2017) that was modulated by reward magnitude (Samson et al., 2017) and stabilized
890 with task experience (van der Meer and Redish, 2009; Howe et al., 2011).
891
892 It is possible that beta-oscillations are mediated by dopaminergic tone. Dopamine
893 activity is linked with both reward-prediction and reward-prediction errors (RPE) (Bayer
894 and Glimcher, 2005; HajiHosseini and Holroyd, 2015; Snyder et al., 2020).
895 Reinforcement learning models suggest that the expected value of a particular trial is

896 inversely related to RPE: as the expected value goes up, the RPEs are less (Bayer and
897 Glimcher, 2005). Previous research explored the possibility of frontal beta-oscillations
898 as an RPE signal but found that stimuli signaling expected rewards elicited more beta
899 power than unexpected rewards; the inverse of an RPE (HajiHosseini and Holroyd,
900 2015). An inverse correlation between beta activity and RPE could explain the
901 decreased beta modulation on high-value choices as delay to reward increases (more
902 unpredictable) and why beta power does not change on low-value choices even when it
903 is preferred (predictable delay). Similarly, on the PRL task, beta power was elevated on
904 high-probability choices (target) but rapidly decayed following a reversal when reward
905 contingencies were unknown and expectancy signals were not defined. Thus, the
906 reward-related beta signal is not equivalent to an RPE but may still be related to
907 expectation. Strikingly, an inverse relationship between dopamine and beta-oscillations
908 has been observed in motor cortex and dorsolateral striatum (Bayer and Glimcher,
909 2005; Schwerdt et al., 2020); a relationship contingent upon task parameters, reward
910 history, and brain area (Schwerdt et al., 2020). Therefore, a similar relationship between
911 beta-oscillations and dopamine may exist within ventral striatum and cortical projecting
912 areas. In this way, beta signals may represent a common modality of communication
913 across distributed cortico-striatal networks: cortico-thalamic-basal ganglia pathways for
914 motor controls (McGregor and Nelson, 2019; Luhmann et al., 2021) and cortico-striatal-
915 limbic pathways for reward-processing (Schultz et al., 2000; Dalley et al., 2004; Abler et
916 al., 2006; Berridge and Kringelbach, 2008; Haber and Knutson, 2010; Chau et al.,
917 2015). A common striatal-beta generating mechanism could explain how increases in
918 attention, motor inhibition, and reward-processing information are linked to beta-

oscillations in distributed brain networks (Khanna and Carmena, 2017; Schmidt et al., 2019).

We acknowledge there are many subsequent analyses to be completed for each task. Here, we present a comprehensive overview of reward-processing activity in all tasks opposed to the fine intricacies of each. Future analyses will investigate network-level connectivity to determine whether beta-oscillations originate in brain areas, like the striatum, or if they are an emergent property of cortico-striatal networks. Moreover, investigations will need to extend beyond power measures to include phase dynamics which can determine temporal relationships between brain areas. Across our tasks, we also see evidence of increased high-gamma power during reward-feedback. Much like theta-gamma coupling is linked to learning and memory in rodents (Tort et al., 2009; Tamura et al., 2017; Zhang et al., 2019), beta-high-gamma coupling may be linked to reward-processing or reflect spike coherence. Additionally, our results are limited to only male rats. We are now repeating this set of studies in a balanced cohort of male/females to understand whether these findings generalize across sexes. Finally, further analysis of movements/video-tracking would lend greater certainty to our findings and rule out movement-related artifacts. Based on the preponderance of evidence across animals and tasks, we propose that beta-oscillations during reward-feedback may present a phenotype that can be used to identify disturbed reward-related processing deficits in psychiatric disorders or brain injury.

940 Abbaspourazad H, Choudhury M, Wong YT, Pesaran B, Shanechi MM (2021)
941 Multiscale low-dimensional motor cortical state dynamics predict naturalistic reach-
942 and-grasp behavior. Nature Communications 12.
943 Abler B, Walter H, Erk S, Kammerer H, Spitzer M (2006) Prediction error as a linear
944 function of reward probability is coded in human nucleus accumbens. Neuroimage
945 31:790–795.
946 Atallah HE, McCool AD, Howe MW, Graybiel AM (2014) Neurons in the ventral striatum
947 exhibit cell-type-specific representations of outcome during learning. Neuron
948 82:1145–1156.
949 Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative
950 reward prediction error signal. Neuron 47:129–141.
951 Berridge KC, Kringelbach ML (2008) Affective neuroscience of pleasure: Reward in
952 humans and animals. Psychopharmacology 199:457–480.
953 Berridge KC, Robinson TE (1998) What is the role of dopamine in reward: hedonic
954 impact, reward learning, or incentive salience?
955 Bilderbeck AC, Raslescu A, Hernaus D, Hayen A, Umbricht D, Pemberton D, Tiller J,
956 Søgaard B, Sambeth A, van Amelsvoort T, Reif A, Papazisis G, Pérez V, Elices M,
957 Maurice D, Bertaina-Anglade V, Dawson GR, Pollentier S (2020) Optimizing
958 Behavioral Paradigms to Facilitate Development of New Treatments for Anhedonia
959 and Reward-processing Deficits in Schizophrenia and Major Depressive Disorder:
960 Study Protocol. Frontiers in Psychiatry 11.
961 Buscher N, Ojeda A, Francoeur M, Hulyalkar S, Claros C, Tang T, Terry A, Gupta A,
962 Fakhraei L, Ramanathan DS (2020) Open-source raspberry Pi-based operant box

963 for translational behavioral testing in rodents. Journal of Neuroscience Methods
964 342:108761.
965 Buschman TJ, Miller EK (2007) Top-down versus bottom-up control of attention in the
966 prefrontal and posterior parietal cortices. Science (1979) 315:1860–1864.
967 Buzsáki G, Anastassiou CA, Koch C (2012) The origin of extracellular fields and
968 currents — EEG, ECoG, LFP and spikes. Available at:
969 www.nature.com/reviews/neuro.
970 Carelli RM, Ijames SG, Crumling AJ (2000) Evidence That Separate Neural Circuits in
971 the Nucleus Accumbens Encode Cocaine Versus " Natural " (Water and Food)
972 Reward. The Journal of Neuroscience 20.
973 Chau BKH, Sallet J, Papageorgiou GK, Noonan MAP, Bell AH, Walton ME, Rushworth
974 MFS (2015) Contrasting Roles for Orbitofrontal Cortex and Amygdala in Credit
975 Assignment and Learning in Macaques. Neuron 87:1106–1118.
976 Cohen MX, Elger CE, Ranganath C (2007) Reward expectation modulates feedback-
977 related negativity and EEG spectra. Neuroimage 35:968–978.
978 Constantinople CM, Piet AT, Bibawi P, Akrami A, Kopec C, Brody CD (2019) Lateral
979 orbitofrontal cortex promotes trial-by-trial learning of risky, but not spatial, biases.
980 Available at: https://doi.org/10.7554/eLife.49744.001 [Accessed June 12, 2022].
981 Cui Y, Liu LD, McFarland JM, Pack CC, Butts DA (2016) Inferring cortical variability
982 from local field potentials. Journal of Neuroscience 36:4121–4135.
983 Dalley JW, Cardinal RN, Robbins TW (2004) Prefrontal executive and cognitive
984 functions in rodents : neural and neurochemical substrates. 28:771–784.

985 Dalton GL, Wang NY, Phillips AG, Floresco SB (2016) Multifaceted contributions by
986 different regions of the orbitofrontal and medial prefrontal cortex to probabilistic
987 reversal learning. Journal of Neuroscience 36:1996–2006.
988 Elisabeth Kilavik B, Trachel R, Confais J, Takerkart S, Riehle A (2012) Context-Related
989 Frequency Modulations of Macaque Motor Cortical LFP Beta Oscillations. Cerebral
990 Cortex 22:2148–2159 Available at:
991 https://academic.oup.com/cercor/article/22/9/2148/420115.
992 Engel AK, Fries P (2010) Beta-band oscillations-signalling the status quo? Current
993 Opinion in Neurobiology 20:156–165.
994 Fakhraei L, Francoeur M, Balasubramani P, Tang T, Hulyalkar S, Buscher N, Claros C,
995 Terry A, Gupta A, Xiong H, Xu Z, Mishra J, Ramanathan DS (2021a) Cognition and
996 Behavior Mapping Large-Scale Networks Associated with Action, Behavioral
997 Inhibition and Impulsivity. Available at: https://doi.org/10.1523/ENEURO.0406-
998 20.2021.
999 Fakhraei L, Francoeur M, Balasubramani PP, Tang T, Hulyalkar S, Buscher N, Mishra
1000 J, Ramanathan DS (2021b) Electrophysiological Correlates of Rodent Default-
1001 Mode Network Suppression Revealed by Large-Scale Local Field Potential
1002 Recordings. Cerebral Cortex Communications 2:1–16 Available at:
1003 https://academic.oup.com/cercorcomms/article/2/2/tgab034/6265044.
1004 Feingold J, Gibson DJ, Depasquale B, Graybiel AM (2015) Bursts of beta oscillation
1005 differentiate postperformance activity in the striatum and motor cortex of monkeys
1006 performing movement tasks. Proc Natl Acad Sci U S A 112:13687–13692.

1007 Francoeur MJ, Mair RG (2018) Representation of actions and outcomes in medial
1008 prefrontal cortex during delayed conditional decision-making: Population analyses
1009 of single neuron activity. Brain and Neuroscience Advances 2:239821281877386.
1010 Francoeur MJ, Mair RG (2019) Effects of choice on neuronal activity in anterior
1011 cingulate, prelimbic, and infralimbic cortices in the rat: Comparison of serial lever
1012 pressing with delayed nonmatching to position. European Journal of Neuroscience:
1013 14643 Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/ejn.14643.
1014 Francoeur MJ, Tang T, Fakhraei L, Wu X, Hulyalkar S, Cramer J, Buscher N,
1015 Ramanathan DR (2021) Chronic, Multi-Site Recordings Supported by Two Low-
1016 Cost, Stationary Probe Designs Optimized to Capture Either Single-unit or Local
1017 Field Potential Activity in Behaving Rats. Frontiers in Psychiatry 12.
1018 Francois J, Huxter J, Conway MW, Lowry JP, Tricklebank MD, Gilmour G (2014)
1019 Differential contributions of infralimbic prefrontal cortex and nucleus accumbens
1020 during reward-based learning and extinction. Journal of Neuroscience 34:596–607.
1021 Goldstein BL, Barnett BR, Vasquez G, Tobia SC, Kashtelyan V, Burton AC, Bryden DW,
1022 Roesch MR (2012) Behavioral/Systems/Cognitive Ventral Striatum Encodes Past
1023 and Predicted Value Independent of Motor Contingencies.
1024 Groenewegen HJ, Wright CI, Uylings HBM (1997) The anatomical relationships of the
1025 prefrontal cortex with limbic structures and the basal ganglia.
1026 Haber SN (2016) Corticostriatal circuitry. Available at: www.dialogues-cns.org.
1027 Haber SN, Knutson B (2010) The reward circuit: Linking primate anatomy and human
1028 imaging. Neuropsychopharmacology 35:4–26.

1029 HajiHosseini A, Holroyd CB (2015) Sensitivity of frontal beta oscillations to reward
1030 valence but not probability. Neuroscience Letters 602:99–103.
1031 Hall TM, Nazarpour K, Jackson A (2014) Real-time estimation and biofeedback of
1032 single-neuron firing rates using local field potentials. Nature Communications 5.
1033 Hammond C, Bergman H, Brown P (2007) Pathological synchronization in Parkinson's
1034 disease: networks, models and treatments. Trends in Neurosciences 30:357–364.
1035 Hardung S, Epple R, Jäckel Z, Eriksson D, Uran C, Senn V, Gibor L, Yizhar O, Diester I
1036 (2017) A Functional Gradient in the Rodent Prefrontal Cortex Supports Behavioral
1037 Inhibition. Current Biology 27:549–555.
1038 Harris KD, Quiroga RQ, Freeman J, Smith SL (2016) Improving data quality in neuronal
1039 population recordings. Nature Neuroscience 19:1165–1174.
1040 Howe MW, Atallah HE, Mccool A, Gibson DJ, Graybiel AM (2011) Habit learning is
1041 associated with major shifts in frequencies of oscillatory activity and synchronized
1042 spike firing in striatum. PNAS 108 Available at:
1043 www.pnas.org/cgi/doi/10.1073/pnas.1113158108.
1044 Huang C, Ruff DA, Pyle R, Rosenbaum R, Cohen MR, Doiron B (2019) Circuit Models
1045 of Low-Dimensional Shared Variability in Cortical Networks. Neuron 101:337-
1046 348.e4.
1047 Humphries, Prescott TJ (2009) The ventral basal ganglia, a selection mechanism at the
1048 crossroads of space, strategy, and reward. Progress in Neurobiology 90:385–417
1049 Available at: http://eprints.whiterose.ac.uk/155180
1050 Jenkinson N, Brown P (2011) New insights into the relationship between dopamine,
1051 beta oscillations and motor function. Trends in Neurosciences 34:611–618.

Kable JW, Glimcher PW (2007) The neural correlates of subjective value during intertemporal choice. NATURE NEUROSCIENCE 10 Available at: http://www.nature.com/natureneuroscience.

Kable JW, Glimcher PW (2009) The Neurobiology of Decision: Consensus and Controversy. Neuron 63:733–745.

Khanna P, Carmena JM (2017) Beta band oscillations in motor cortex reflect neural population signals that delay movement onset.

Lashgari R, Li X, Marateb HR, Jahed M, Zarei M, Daliri MR (2018) Introducing a Comprehensive Framework to Measure Spike-LFP Coupling. Frontiers in Computational Neuroscience | www.frontiersin.org 12:78 Available at: www.frontiersin.org.

Lefner MJ, Magnon AP, Gutierrez JM, Lopez MR, Wanat MJ (2021) Delays to reward delivery enhance the preference for an initially less desirable option: role for the basolateral amygdala and retrosplenial cortex. Journal of Neuroscience 41:7461–7478.

Levcik D, Sugi AH, Pochapski JA, Baltazar G, Pulido LN, Villas-Boas C, Aguilar-Rivera M, Fuentes-Flores R, Nicola SM, da Cunha C (2017) Title: Nucleus accumbens neurons encode initiation and vigor of reward approach behavior Abbreviated title: NAc neurons encode spontaneous reward approach. Available at: https://doi.org/10.1101/2021.01.12.425739.

Luhmann HJ, Arce-Mcshane FI, Litvak V, Barone J, Rossiter HE (2021) Understanding the Role of Sensorimotor Beta Oscillations. Available at: www.frontiersin.org.

1074 MacDowell CJ, Buschman TJ (2020) Low-Dimensional Spatiotemporal Dynamics
1075 Underlie Cortex-wide Neural Activity. Current Biology 30:2665-2680.e8.
1076 Marco-Pallares J, Cucurell D, Cunillera T, García R, Andrés-Pueyo A, Münte TF,
1077 Rodríguez-Fornells A (2008) Human oscillatory activity associated to reward-
1078 processing in a gambling task. Neuropsychologia 46:241–248.
1079 Marco-Pallarés J, Münte TF, Rodríguez-Fornells A (2015) The role of high-frequency
1080 oscillatory activity in reward-processing and learning. Neuroscience and
1081 Biobehavioral Reviews 49:1–7.
1082 McGregor MM, Nelson AB (2019) Circuit Mechanisms of Parkinson's Disease. Neuron
1083 101:1042–1056.
1084 Ojeda A, Bigdely-Shamlo N, Makeig S (2014) MoBILAB: An open source toolbox for
1085 analysis and visualization of mobile brain/body imaging data. Frontiers in Human
1086 Neuroscience 8.
1087 Patai EZ, Foltynie T, Limousin P, Akram H, Zrinzo L, Bogacz R, Litvak V (2022) Conflict
1088 Detection in a Sequential Decision Task Is Associated with Increased Cortico-
1089 Subthalamic Coherence and Prolonged Subthalamic Oscillatory Response in the β
1090 Band. The Journal of Neuroscience 42:4681–4692.
1091 Pesaran B, Nelson MJ, Andersen RA (2008) Free choice activates a decision circuit
1092 between frontal and parietal cortex. Nature 453:406–410.
1093 Pujara M, Koenigs M (2014) Mechanisms of reward circuit dysfunction in psychiatric
1094 illness: Prefrontal-striatal interactions. Neuroscientist 20:82–95.

Quiroga R, Nadasdy Z, Ben-Shaul Y (2004) Unsupervised Spike Detection and Sorting with Wavelets and Superparamagnetic Clustering. Available at: http://direct.mit.edu/neco/article

Ray S (2015) Challenges in the quantification and interpretation of spike-LFP relationships. Current Opinion in Neurobiology 31:111–118.

Ray S, Crone NE, Niebur E, Franaszczuk PJ, Hsiao SS (2008) Behavioral/Systems/Cognitive Neural Correlates of High-Gamma Oscillations (60-200 Hz) in Macaque Local Field Potentials and Their Potential Implications in Electrocorticography. Available at: http://erl.neuro.jhmi.edu/mpsoft.

Rey HG, Pedreira C, Quian Quiroga R (2015) Past, present and future of spike sorting techniques. Brain Research Bulletin 119:106–117.

Roesch MR, Bryden DW, Kalenscher T, Weber B (2011) Impact of size and delay on neural activity in the rat limbic corticostriatal system. Available at: www.frontiersin.org.

Rothé M, Quilodran R, Sallet J, Procyk E (2011) Coordination of high gamma activity in anterior cingulate and lateral prefrontal cortical areas during adaptation. Journal of Neuroscience 31:11110–11117.

Salehinejad MA, Ghanavati E, Rashid MHA, Nitsche MA (2021) Hot and cold executive functions in the brain: A prefrontal-cingular network. Brain and Neuroscience Advances 5:239821282110077.

Samson RD, Lester AW, Duarte L, Venkatesh A, Barnes CA (2017) Cognition and Behavior Emergence of-Band Oscillations in the Aged Rat Amygdala during

Discrimination Learning and Decision Making Tasks. Available at: http://dx.doi.org/10.1523/ENEURO.0245-17.2017.

Schmidt R, Ruiz MH, Kilavik BE, Lundqvist M, Starr PA, Aron AR (2019) Beta oscillations in working memory, executive control of movement and thought, and sensorimotor function. Journal of Neuroscience 39:8231–8238.

Schoenbaum G, Roesch MR, Stalnaker TA, Takahashi YK (2009) A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. Nature Reviews Neuroscience 10:885–892.

Schultz W, Tremblay L, Hollerman Jeffrey R. (2000) Reward-processing in Primate Orbitofrontal Cortex and Basal Ganglia.

Schwerdt HN, Amemori K, Gibson DJ, Stanwicks LL, Yoshida T, Bichot NP, Amemori S, Desimone R, Langer R, Cima MJ, Graybiel AM (2020) Dopamine and beta-band oscillations differentially link to striatal value and motor control. Available at: https://www.science.org.

Shin H, Law R, Tsutsui S, Moore CI, Jones SR (2017) The rate of transient beta frequency events predicts behavior across tasks and species. Elife 6 Available at: https://doi.org/10.7554/eLife.29086.001.

Siegel M, Warden MR, Miller EK (2009) Phase-dependent neuronal coding of objects in short-term memory.

Simon NW, Wood J, Moghaddam B (2015) Action-outcome relationships are represented differently by medial prefrontal and orbitofrontal cortex neurons during action execution. Journal of Neurophysiology 114:3374–3385 Available at: www.jn.org

1140 Snyder SH, Ottenheimer DJ, Bari BA, Sutlief E, Fraser KM, Kim TH, Richard JM, Cohen
1141 JY, Janak PH (2020) A quantitative reward prediction error signal in the ventral
1142 pallidum. Nature Neuroscience 23:1267–1276 Available at:
1143 https://doi.org/10.1038/s41593-020-0688-5.
1144 Spitzer B, Haegens S (2017) Beyond the status quo: A role for beta oscillations in
1145 endogenous content (RE)activation. eNeuro 4.
1146 Swindale N v., Spacek MA (2014) Spike sorting for polytrodes: A divide and conquer
1147 approach. Frontiers in Systems Neuroscience 8.
1148 Tamura M, Spellman TJ, Rosen AM, Gogos JA, Gordon JA (2017) Hippocampal-
1149 prefrontal theta-gamma coupling during performance of a spatial working memory
1150 task. Nature Communications 8 Available at:
1151 www.nature.com/naturecommunications.
1152 Torrecillos F, Alayrangues J, Kilavik E, Malfait N (2015) Behavioral/Cognitive Distinct
1153 Modulations in Sensorimotor Postmovement and Foreperiod-Band Activities
1154 Related to Error Salience Processing and Sensorimotor Adaptation.
1155 Tort ABL, Komorowski RW, Manns JR, Kopell NJ, Eichenbaum H (2009) Theta-gamma
1156 coupling increases during the learning of item-context associations. Available at:
1157 www.pnas.orgcgidoi10.1073pnas.0911331106.
1158 Turner KM, Parkes SL (2020) Prefrontal regulation of behavioural control: Evidence
1159 from learning theory and translational approaches in rodents. Neuroscience and
1160 Biobehavioral Reviews 118:27–41.

1161 van den Bos W, Rodriguez CA, Schweitzer JB, Mcclure SM (2014) Connectivity
1162 Strength of Dissociable Striatal Tracts Predict Individual Differences in Temporal
1163 Discounting. Available at: http://www.fil.ion.ucl.ac.uk/spm
1164 van der Meer MAA, Redish AD (2009) Covert expectation-of-reward in rat ventral
1165 striatum at decision points. Frontiers in Integrative Neuroscience 3.
1166 van Duuren E, van der Plasse G, Lankelma J, Joosten RNJMA, Feenstra MGP,
1167 Pennartz CMA (2009) Single-cell and population coding of expected reward
1168 probability in the orbitofrontal cortex of the rat. Journal of Neuroscience 29:8965–
1169 8976.
1170 Wassum KM (2022) Amygdala-cortical collaboration in reward learning and decision
1171 making. eLife 11: e80926.
1172 Whitton AE, Treadway MT, Pizzagalli DA (2015) Reward-processing dysfunction in
1173 major depression, bipolar disorder and schizophrenia. Curr Opin Psychiatry 28:7–
1174 12.
1175 Williams AH, Kim TH, Wang F, Vyas S, Ryu SI, Shenoy K v., Schnitzer M, Kolda TG,
1176 Ganguli S (2018) Unsupervised Discovery of Demixed, Low-Dimensional Neural
1177 Dynamics across Multiple Timescales through Tensor Component Analysis.
1178 Neuron 98:1099-1115.e8.
1179 Winstanley CA, Theobald DEH, Cardinal RN, Robbins TW (2004a) Contrasting Roles of
1180 Basolateral Amygdala and Orbitofrontal Cortex in Impulsive Choice.
1181 Winstanley CA, Theobald DEH, Cardinal RN, Robbins TW (2004b) Contrasting Roles of
1182 Basolateral Amygdala and Orbitofrontal Cortex in Impulsive Choice.

Witham CL, Wang M, Baker SN (2007) Cells in somatosensory areas show synchrony with beta oscillations in monkey motor cortex. European Journal of Neuroscience 26:2677–2686 Available at: www.blackwell-synergy.com.

Zavala B, Jang A, Trotta M, Lungu CI, Brown P, Zaghloul KA (2018) Cognitive control involves theta power within trials and beta power across trials in the prefrontal-subthalamic network. BRAIN 141:3361–3376 Available at: https://academic.oup.com/brain/article-abstract/141/12/3361/5144595.

Zhang L, Lee J, Rozell C, Singer AC (2019) Sub-second dynamics of theta-gamma coupling in hippocampal CA1. Available at: https://doi.org/10.7554/eLife.44320.001.

Figures

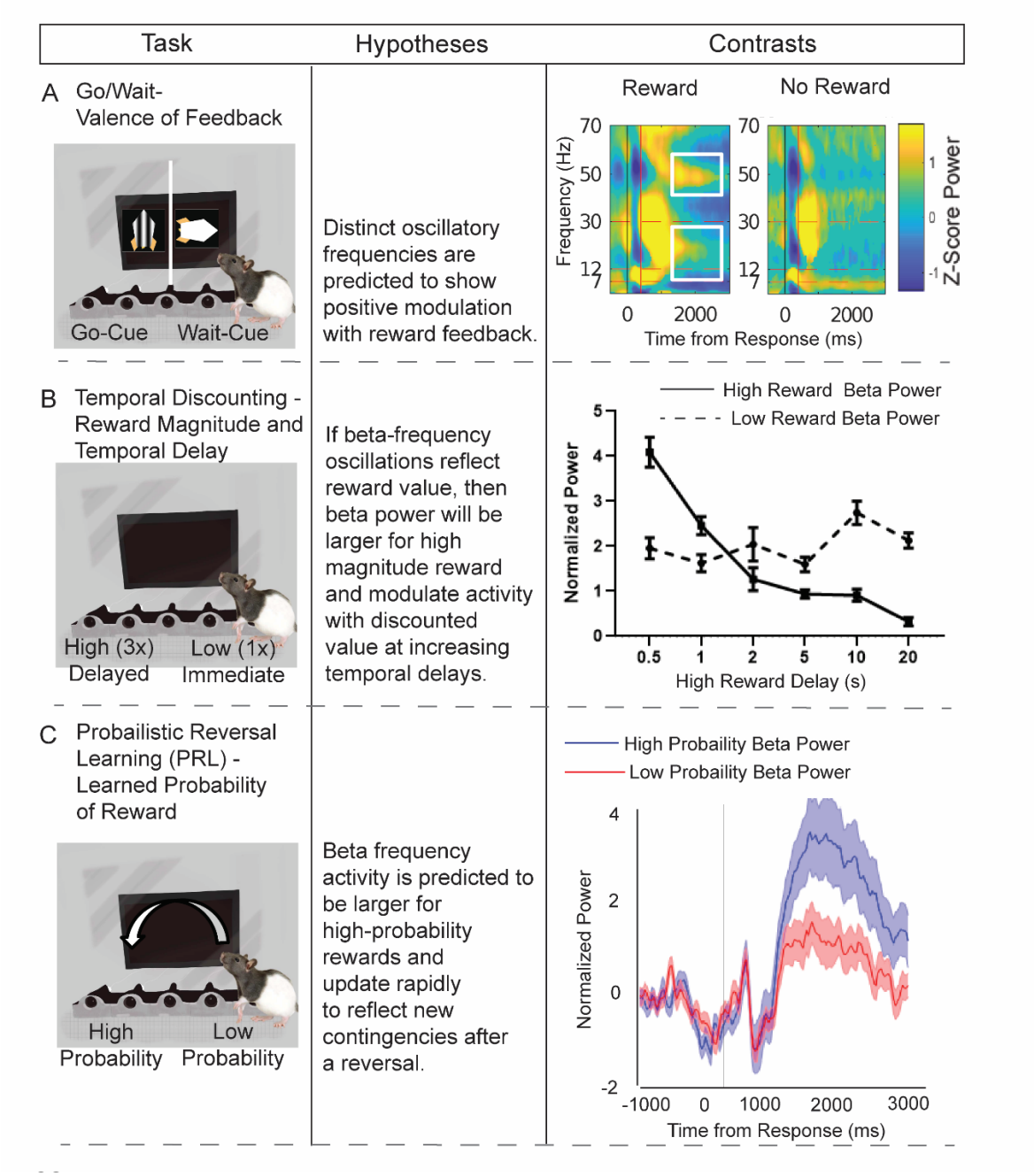


Figure 1: Overview of Tasks Used to Find Markers of Reward Value.

A graphical representation of the hypotheses and contrasts for each task. (A) The go/wait response inhibition task was used to measure oscillatory activity based on valence of reward feedback (positive/ reward or negative/ no reward). Beta and high-

gamma frequencies have greater power (z-scored) during reward outcome (red line= time of reward delivery). (B) The temporal discounting task measures brain activity related to reward magnitude (high vs. low) and temporal delay of reward (0.5, 1, 2, 5, 10, 20s). Beta power for high (3x magnitude) reward decreases at increasing temporal delays (solid black line). (C) The probabilistic reversal learning task was used to examine reward-related activity based on likelihood of reward (high vs. low) and how brain activity updated to reflect reversing reward contingencies. Beta power is larger during reward outcome (black line= time of reward delivery) of a high-probability (blue average and SEM) choice compared reward outcome of a low-probability choice (red average and SEM).

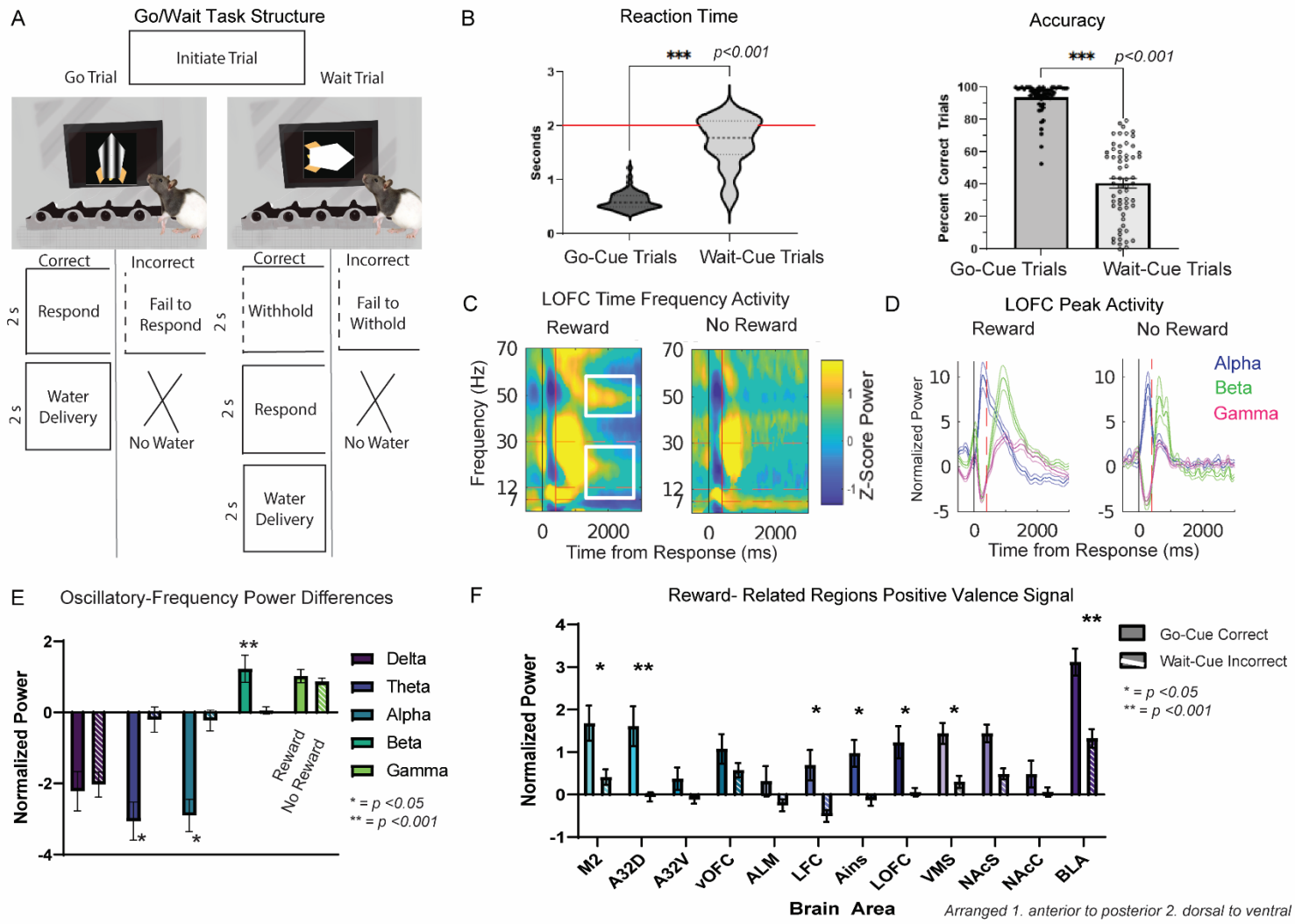


Figure 2: Positive Valence Representation on the Response Inhibition Task. (A) Trial structure of the go/wait task. Animals were shown two visual stimuli: a striped vertical rocket indicated a go-cue trial (respond within 2s) and a solid horizontal rocket denoted a wait-cue (withhold from responding for 2s). On correct trials rats were given 2s (10μL/s) of water. (B) Behavior on the go/wait task of animals with LFP implants (N= 60 sessions). Violin plot shows the median (thick dotted line), interquartile range (thin dotted line), and distribution shape of reaction times (s) on go-cue and wait-cue trials.

Red horizontal line represents the 2s time to response on go-cue trials and time needed to withhold on wait-cue trials. Bar plots shows the mean and SEM of proportion correct trials on go-cue (respond within 2s) and wait-cue (withhold for 2s) trials. Dots show individual values per session. (C) Time-frequency plots of z-scored lOFC power time-locked to response for frequencies ranging from 0-70Hz on correct go-cue trials (reward) and incorrect wait-cue trials (no reward). Red vertical line marks the time of reward delivery. White squares represent the elevated beta and high-gamma oscillatory power during reward outcome. (D) Line plots show mean and SEM activity for lOFC on correct go-cue and incorrect wait-cue trials. Separate lines are shown for alpha (8-12 Hz; blue), beta (15-30Hz; green) and gamma frequencies (50-70Hz; pink) with peak activity time-locked to the response. Red vertical line marks time of reward delivery. (E) Average lOFC power across delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (15-30 Hz), and gamma (50-70 Hz) frequencies during the reward-feedback period (500-2500ms after response). We used a two-way ANOVA to compare power across frequencies and reward outcome (go-cue correct and wait-cue incorrect). Rewarded trials are solid bars and no reward trials are striped bars. (F) Beta power from 12 putative reward-related brain regions on go-cue correct (solid) and wait-cue incorrect trials (striped) during the reward-feedback period (1550ms-2550ms). The mean and SEM for each trial type and brain region of interest are shown. Brain areas are shown in order from 1. anterior to posterior and 2.dorsal to ventral.

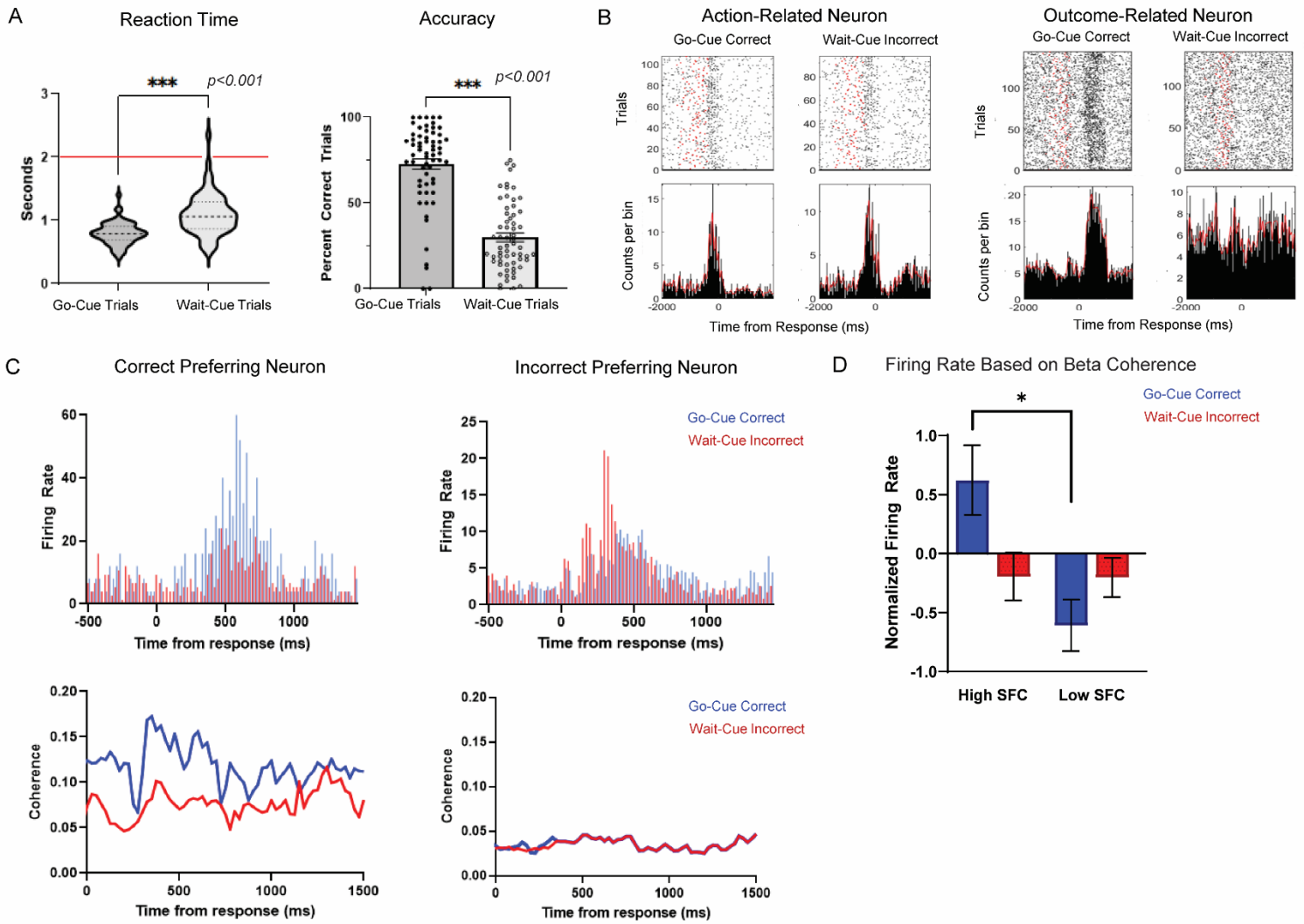


Figure 3: Single-Units Related to Positive Valence Reward Outcome.

(A) Behavior on the go/wait task of animals with single unit implants (N= 62 sessions). Violin plot shows the median (dark dotted line), interquartile range (light dotted line), and distribution shape of reaction times (s) on go-cue trials and wait-cue trials. Red line drawn at 2s represents the time to respond on go-cue trials and the time needed to withhold on wait-cue trials. Bar plots show proportion of correct trials on go-cue and wait-cue trials. Dots show individual values per session. (B) Individual examples of an

action-related neuron (peak firing rate increases or decreases prior to the response) and an outcome-related neuron (peak firing rate increases or decreases after the response). Activity of each neuron is plotted for go-cue correct (reward) and wait-cue incorrect (no reward) trials time-locked to the response (time 0). Raster plots (top-panel) show spiking activity across trials (each horizontal line). Red dots indicate trial onset. Peri-event stimulus histograms (bottom panel) show firing rate (counts per bin) time-locked to response. Red lines indicate the mean activity of the unit across trials. (C) Individual examples of a correct preferring (more activity on rewarded trials) and an incorrect preferring (more activity on non-rewarded trials) unit. Firing rate (counts/bin) from go-cue correct (blue) and wait-cue incorrect (red) trials are plotted on top of each other time-locked to response for comparison (top-panel). Examples of beta-frequency SFC are shown for the same correct and incorrect preferring neurons on go-cue correct (blue) and wait-cue incorrect (red) trials (bottom-panel). Coherence values are time-locked to response. (D) Normalized firing rate (spikes/s) of neurons split into “high” or “low” groups based on their beta-SFC values during reward-feedback. Firing rates are plotted separately for go-cue correct (blue) and wait-cue incorrect (red) trials. Error bars indicate SEM.

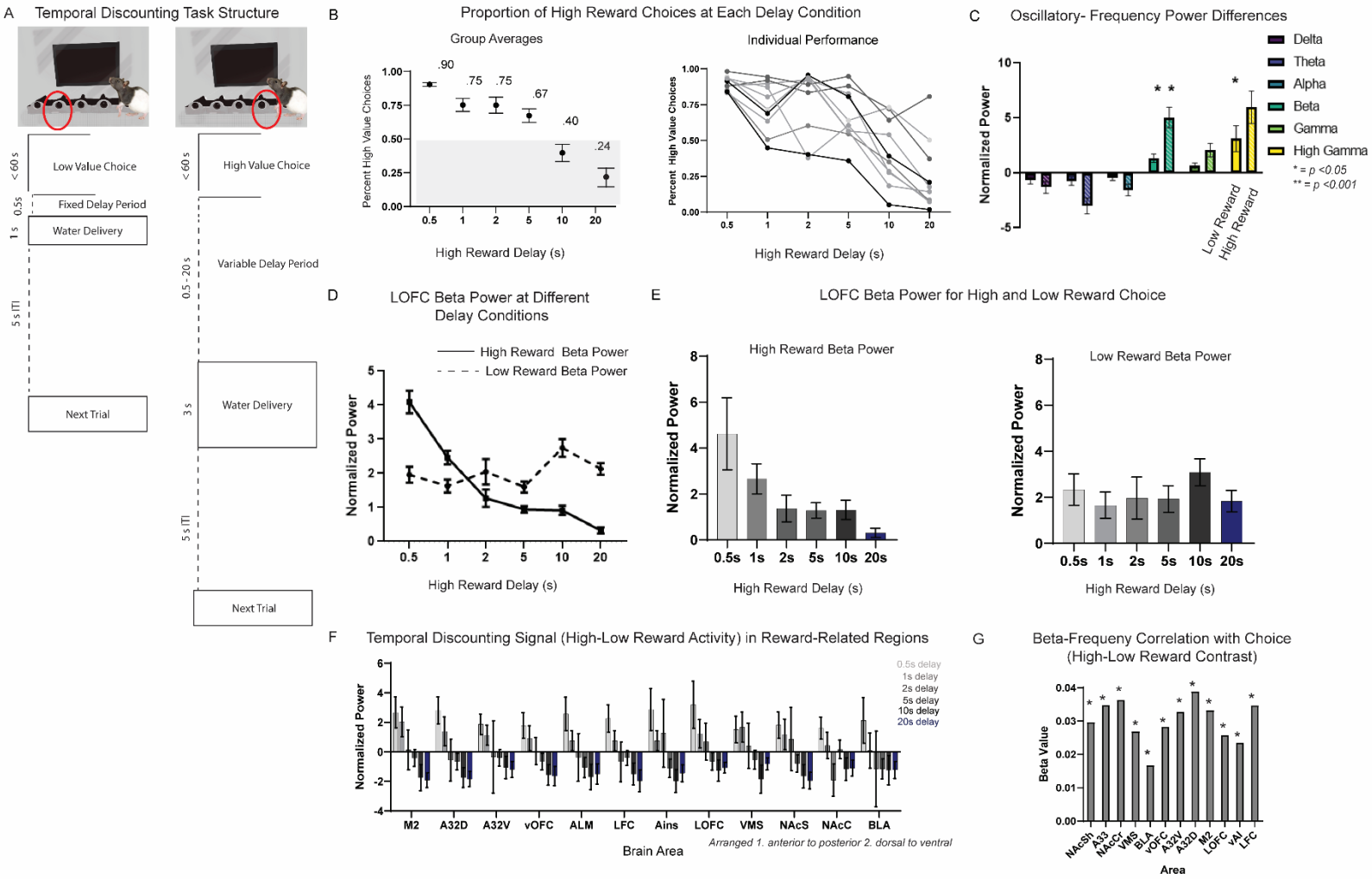


Figure 4: Subjective Value Representation on the Temporal Discounting Task. (A) Trial structure of the temporal discounting task. Animals were given the choice of a low magnitude reward (10μL) delivered immediately (delay of 500ms after response), or a high magnitude reward (30μL) delivered at variable delays of 0.5s up to 20s. Delays on the high-value choice were kept constant throughout each session but varied across sessions. (B) Behavior on the temporal discounting task shown as proportion of high-value choices per session (N=124 sessions). The group mean and SEM are shown at

each delay condition (0.5, 1, 2, 5, 10, and 20s). A horizontal line is drawn at 0.5 to indicate when proportion of high-value choices fall below the 50/50 mark. The average proportion of high-value choices at each delay condition is also plotted separately for each rat to show individual differences in discounting rates. (C) Average power across delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (15-30 Hz), low gamma (50-70 Hz) and high-gamma (70-150Hz) frequencies on the lOFC electrode one second after reward. We used a two-way ANOVA to compare power across frequencies and trial type (high-value or low-value choice) at equal delays (500ms). Error bars represent SEM. (D) The normalized beta power on the lOFC electrode at each high-value delay condition (0.5, 1, 2, 5, 10, 20s). Power is averaged across the first second following reward delivery and the mean/SEM are plotted separately for high-reward choices (solid line) and low-reward choices (dashed line). (E) LOFC beta-power is separately plotted for high-reward and low-reward choice for each high-value delay condition (0.5, 1, 2, 5, 10, 20s). Bar plots show mean and SEM. ANOVAs show a significant decrease in beta power for high-value choice that occurs at longer delay conditions whereas power for low-value choice is not different across delay conditions. (F) The difference in power on high reward and low reward choices (high-value choice – low-value choice) across 12 brain regions. Beta power during reward-feedback (averaged activity one second after response) is plotted at each delay condition (0.5, 1, 2, 5, 10, 20s) for each brain region. Brain regions are organized from 1. anterior to posterior and 2. dorsal to ventral. (G) Beta values from the logistic regression analysis for 12 brain regions using power for the difference of trial types (high-value – low-value). Asterisks indicate brain regions with significant p-values after FDR-correction

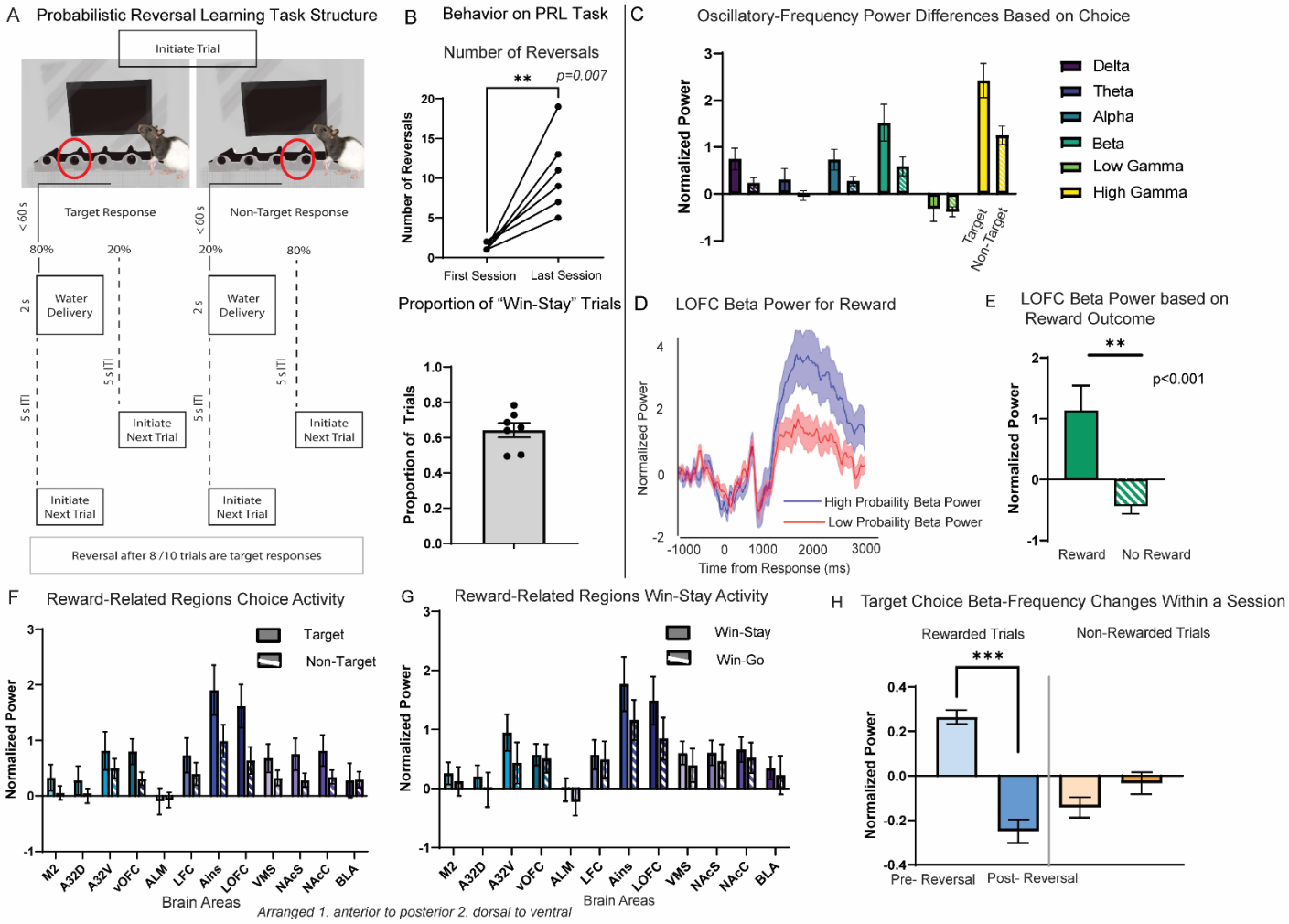
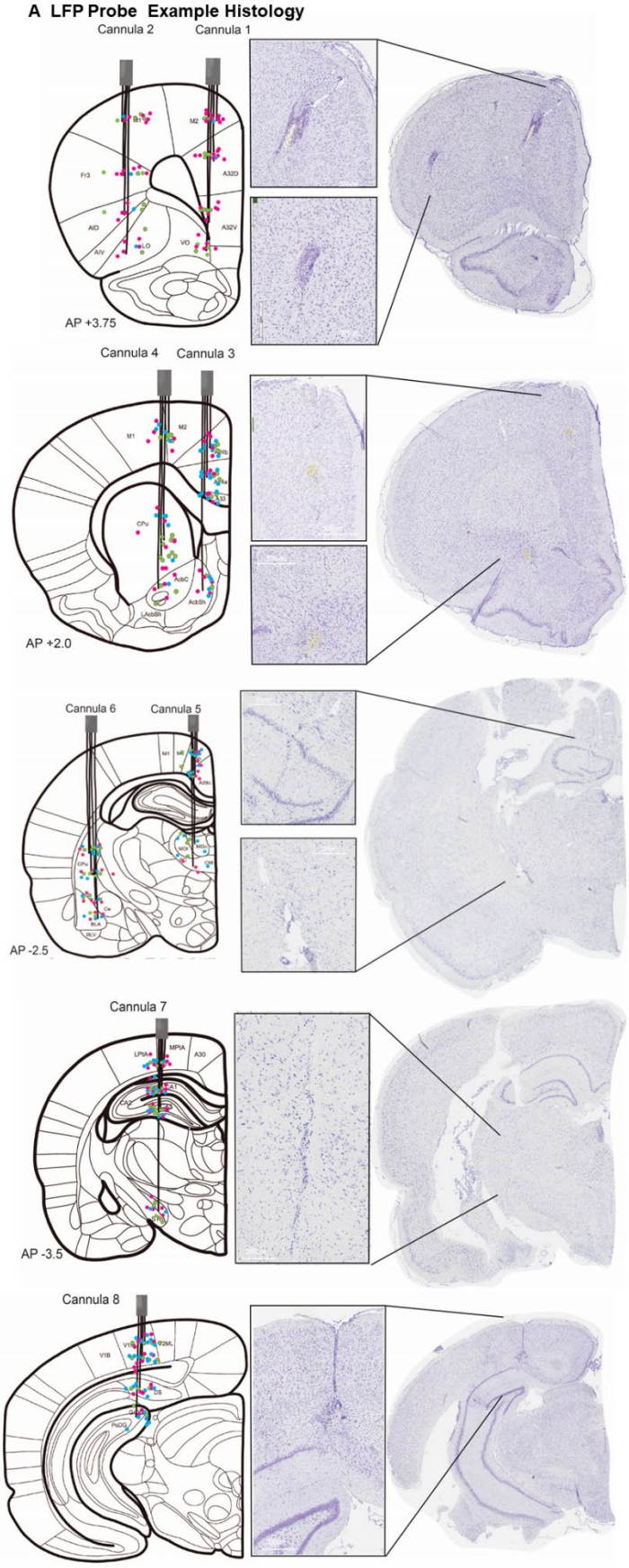


Figure 5: Likelihood of Reward Outcome Represented on the PRL Task.

(A) Trial structure of the probabilistic reversal learning (PRL) task. Rats choose between two response ports: the target choice delivered reward 80% of the time and the non-target choice delivered reward 20% of the time. Reward contingencies reversed after 8 out of the last 10 trials were target choices (regardless of reward outcome). (B) Behavior on the PRL task measured by number of reversals per session and proportion of win-stay trials. The number of reversals is compared in the first and last session of individual rats to show improvement over time (N=36 sessions). The average and SEM

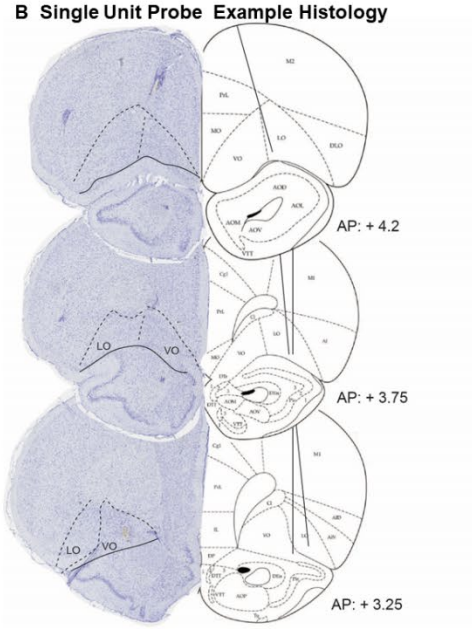
proportion of win-stay trials/ session is plotted with black dots visualizing the average for each individual rat. (C) Average power across delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (15-30 Hz), low gamma (50-70 Hz) and high-gamma (70-150Hz) frequencies on the lOFC electrode during reward-feedback (500-2500ms). We used a two-way ANOVA to compare power across frequencies and trial type (target vs. non-target rewards). (D) The shaded error bar plot shows mean and SEM traces of lOFC normalized beta power for high-probability rewards (blue) and low-probability rewards (red). Traces are time-locked to response. (E) LOFC beta frequency power based on reward outcome. Average and SEM are shown for rewarded outcomes (solid) and non-rewarded outcomes (striped) for both choices (target and non-target). (F) Beta power during reward-feedback (averaged activity from 5000ms-2500ms after response) across the 12 brain regions. The average and SEM are shown separately for target (solid) and non-target (striped) rewarded choices in each brain region. Brain regions are organized from 1. anterior to posterior and 2. dorsal to ventral. (F) Similarly, we also show the average and SEM beta power activity on win-stay (solid) compared to win-go (striped) trials in the 12 brain regions. (H) Beta power (from all 12 brain regions) is compared before and after a reversal on target choices. The average and SEM are shown separately for rewarded trials pre-reversal (light blue), rewarded trials post-reversal (dark blue), non-rewarded trials pre-reversal (light orange) and non-rewarded trials post-reversal (dark orange).

A LFP Probe Example Histology



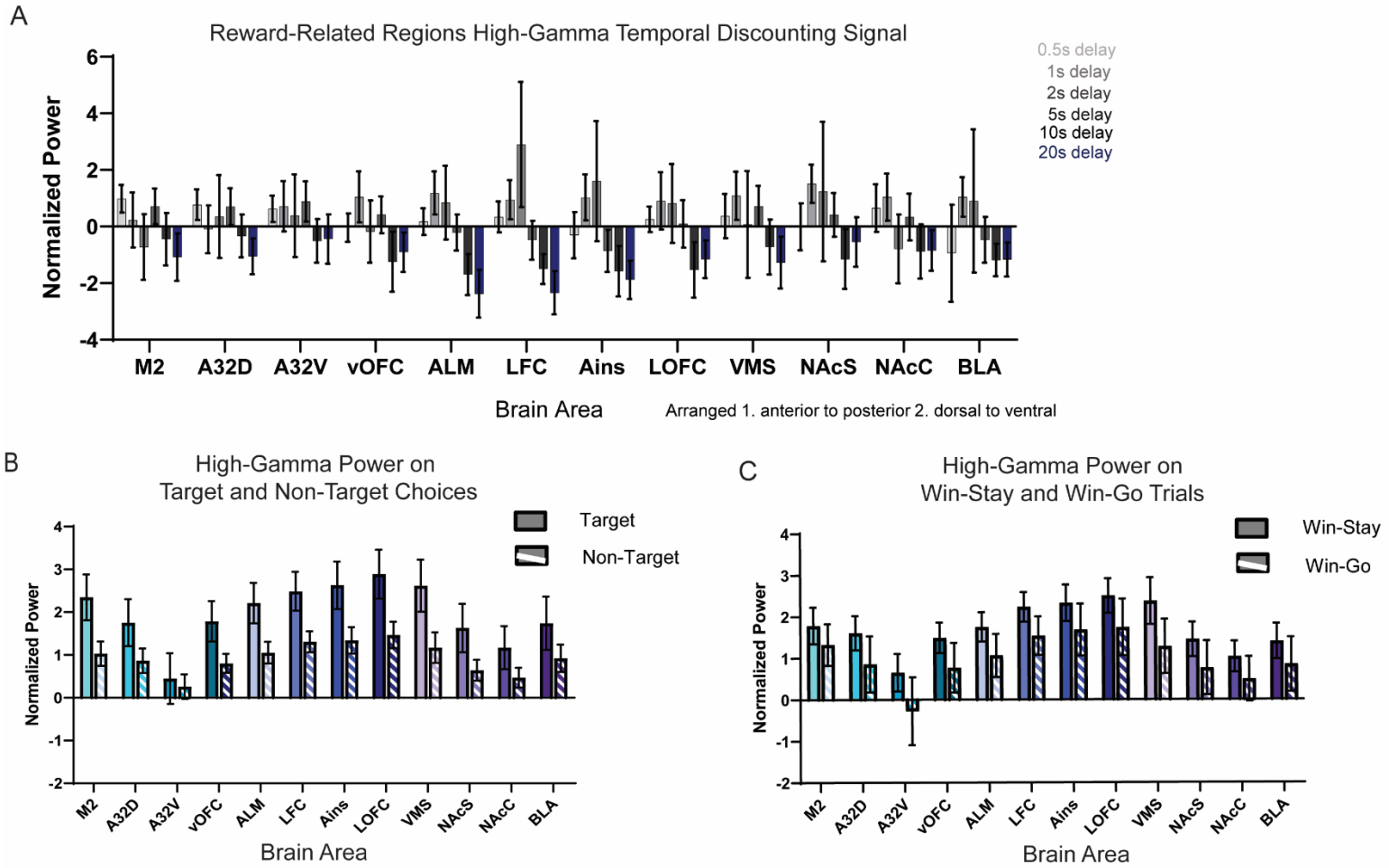
AP	ML	DV	Target Area
Cannula 1			
3.75	0.8	0.8	M2
3.75	0.8	3.2	A32D
3.75	0.8	4.8	A32V
3.75	0.8	5.8	ventral orbitofrontal cortex
Cannula 2			
3.75	3.2	1.0	anterolateral motor cortex
3.75	3.2	3.6	lateral frontal cortex
3.75	3.2	4.8	anterior insula
3.75	3.2	5.8	lateral orbitofrontal cortex
Cannula 3			
2.0	0.6	2.0	A24b
2.0	0.6	3.0	A24a
2.0	0.6	3.5	A33
2.0	0.6	6.6	nucleus accumbens shell
Cannula 4			
2.0	1.8	1.5	M2
2.0	1.8	4.5	dorsomedial striatum
2.0	1.8	5.7	ventromedial striatum
2.0	1.8	6.9	nucleus accumbens core
Cannula 5			
-2.5	0.7	1.3	A30
-2.5	0.7	2.3	A29
-2.5	0.7	4.7	mediodorsal thalamus
-2.5	0.7	5.7	centro-median thalamus
Cannula 6			
-2.5	4.9	5.1	dorsolateral striatum
-2.5	4.9	6.1	dorsolateral striatum
-2.5	4.9	7.1	central amygdala
-2.5	4.9	8.1	basolateral amygdala
Cannula 7			
-3.5	2.5	1.4	posterior parietal cortex
-3.5	2.5	2.5	CA1
-3.5	2.5	3.5	CA3
-3.5	2.5	8.0	subthalamic nucleus
Cannula 8			
-6.0	3.5	1.0	V1
-6.0	3.5	1.7	V1
-6.0	3.5	2.8	dorsal subiculum
-6.0	3.5	3.7	dentate gyrus

B Single Unit Probe Example Histology



Supplemental Figure 1: **Histological Verification of Recording Sites.**

(A) LFP implants histology identifying locations of the 32 electrodes (8 cannula). A graphical representation of the placement of each cannula is plotted on a coronal section of a modified rat brain atlas (Paxinos & Watson, 2013) Each cannula contains four wires each targeting a unique DV location. Multiple cannulas may be implanted on the same coronal plane (same AP coordinates). The identified centers of each electrode are shown as dots color coded based on task to (green: go/wait N= 6/11; pink: temporal discounting N= 9/10; blue: PRL N=7/7). An example thionine-stained coronal slice at the corresponding AP location is also shown for each cannula placement with magnification of each track in the brain (white bar provides scale). The table includes the AP, ML, and DV coordinates for all 32 electrodes and their corresponding nomenclature. (B) Single unit implants histology for electrode tracks in OFC. Example coronal sections are shown from +4.2 AP through +3.25 AP relative to bregma. The LO/VO subdivisions are outlined on example thionine-stained slices and the center of the electrode tracks are marked on sections of a modified rat brain atlas (Paxinos & Watson, 2013) (N=5/8 rats).



Supplemental Figure 2: **High-gamma Activity Related to Reward Feedback.**

High-gamma (70-150 Hz) activity in the cortico-striatal network during temporal discounting (A) and PRL tasks (B,C). (A) The difference in power on high reward and low reward choices (high-value choice – low-value choice) across 12 brain regions. High-gamma power during reward-feedback (averaged activity one second after response) is plotted at each delay condition (0.5, 1, 2, 5, 10, 20s) for each brain region. Brain regions are organized from 1. anterior to posterior and 2. dorsal to ventral. (B) High-gamma power during reward-feedback (averaged activity from 5000ms-2500ms after response) during the PRL task across the 12 brain regions. The average and SEM

are shown separately for target (solid) and non-target (striped) rewarded choices in each brain region. (C) Similarly, we also show the average and SEM high-gamma power activity on win-stay (solid) compared to win-go (striped) trials in the 12 brain regions during reward-feedback of the PRL task.

1356 **Table 1. Table of Reward-Related Brain Regions**

Abbreviation	Brain Area
M2	Secondary Motor Cortex
A32D	Dorsomedial Prefrontal Cortex
A32V	Ventromedial Prefrontal Cortex
vOFC	Ventral Orbitofrontal Cortex
ALM	Anterolateral Motor Cortex
LFC	Lateral Frontal Cortex
Ains	Anterior Insula
lOFC	Lateral Orbitofrontal Cortex
VMS	Ventromedial Striatum
NAcS	Nucleus Accumbens Shell
NAcC	Nucleus Accumbens Core
BLA	Basolateral Amygdala

1357 *Listed in order from 1. Anterior to posterior 2. Dorsal to ventral*