Dorsal striatum coding for the timely execution of action sequences

María Cecilia Martinez[1,2], Camila Lidia Zold[1], Marcos Antonio Coletti[1], Mario Gustavo Murer[1]*, Mariano Andrés Belluscio[1]*.

1- Universidad de Buenos Aires y Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Instituto de Fisiología y Biofísica "Bernardo Houssay" (IFIBIO-Houssay), Grupo de Neurociencia de Sistemas, Buenos Aires C1121ABG, Argentina.

2- Departamento de Fisiología, Biología Molecular y Celular "Dr. Héctor Maldonado", Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, CONICET, Ciudad Universitaria Pabellón 2, AR1428EHA Buenos Aires, Argentina.

Author list footnotes*. Co-senior authors

* Correspondence: mbellu@fmed.uba.ar, gmurer@fmed.uba.ar.

Summary

The automatic initiation of actions can be highly functional. But occasionally these actions cannot be withheld and are released at inappropriate times, impulsively. Striatal activity has been shown to participate in the timing of action sequence initiation and it has been linked to impulsivity. Using a self-initiated task, we trained adult male rats to withhold a rewarded action sequence until a waiting time interval has elapsed. By analyzing neuronal activity we show that the striatal response preceding the initiation of the learned sequence is strongly modulated by the time subjects wait before eliciting the sequence. Interestingly, the modulation is steeper in adolescent rats, which show a strong prevalence of impulsive responses compared to adults. We hypothesize this anticipatory striatal activity reflects the

animals' subjective reward expectation, based on the elapsed waiting time, while the steeper waiting modulation in adolescence reflects age-related differences in temporal discounting, internal urgency states or explore-exploit balance.

Keywords: Dorsal Striatum, Rat, Reward, Action sequence, Waiting time, Impulsivity, Adolescence, Reinforcement learning.

Introduction

The striatum is involved in the acquisition and execution of action sequences (Costa, 2011; Graybiel, 1998; Hikosaka et al., 1999). It has also been linked to temporal information processing (Bakhurin et al., 2017; Emmons et al., 2017; Gouvêa et al., 2015; Matell et al., 2003; Mello et al., 2015) and to the use of temporal information for the action initiation timing (Kunimatsu et al., 2018; Thura and Cisek, 2017; Yau et al., 2020). Sometimes, well-learned action sequences need to be initiated at precise times to obtain the desired outcome. However, these actions may be difficult to withhold when triggering cues are present, and also difficult to stop once they have been initiated (Dalley and Robbins, 2017; Gillan et al., 2016a; Graybiel, 2008; Knowlton and Patterson, 2016; Robbins and Costa, 2017). Moreover, in neuropsychiatric conditions involving malfunctioning of cortico-basal ganglia circuits, such as attention deficit hyperactivity disorder, Tourette syndrome, obsessive compulsive disorder and drug addiction (Dalley and Robbins, 2017; Gillan et al., 2016a; Singer, 2016) action sequences might be started at inappropriate contexts and timings. However, how the striatum contributes to action sequence initiation timing remains poorly understood.

Interestingly, impulsivity has been identified as a vulnerability factor for compulsive drug seeking habits (Belin and Everitt, 2008). Recent studies link impulsivity to automaticity in behavior (Ersche et al.,

2019; Gillan et al., 2016b; Hogarth et al., 2012) and a preponderance of habitual over goal-directed behavioral control (Everitt et al., 2008; Voon et al., 2015). An influential theory in the field postulates a dual control system for behavior, where dorsolateral striatal (DLS) circuits support habitual stimulus-response control whereas dorsomedial striatal (DMS) circuits mediate cognitive-based deliberative control (Balleine and Dickinson, 1998; Daw et al., 2005; Graybiel, 2008; Yin and Knowlton, 2006). During action sequence learning, neuronal activity in the DLS rapidly evolves to mark the initiation and termination of the acquired sequence (Jin and Costa, 2010; Jog, 1999), possibly contributing to its release as a behavioral unit or chunk (Graybiel, 2008). In contrast, the DMS encodes reward expectancy, reward prediction errors and trial outcomes even after extensive training (Kim et al., 2009; Kubota et al., 2009; Rueda-Orozco and Robbe, 2015; Samejima et al., 2005; Thorn et al., 2010; Vandaele et al., 2019). All these features likely contribute to the regulation of explore-exploit (Barnes et al., 2005), cost-benefit (Floresco et al., 2008; Schultz, 2015) and speed-accuracy (Thura and Cisek, 2017) trade-offs during decision making. A bias toward exploration and risk-taking (Addicott et al., 2017), low tolerance for delayed rewards (Dalley and Robbins, 2017; Monterosso and Ainslie, 1999; Wittmann and Paulus, 2008) and elevated internal urgency states (Carland et al., 2019), may all contribute to impulsivity traits. However, how neuronal activity in the dorsal striatum informs about enhanced premature responding in conditions of high impulsivity, remains unknown.

Here we studied striatal activity during a self-paced task where rats have to withhold a rewarded action sequence until a waiting interval has elapsed (Zold and Hussain Shuler, 2015). Prematurely initiated sequences were penalized by re-initiating the waiting interval. Despite the negative effect this timer re-initialization may have, the animals showed premature responding even after extensive training. Moreover, they often failed to interrupt the execution of the action sequence although there was sensory evidence of its untimely initiation. Thus, the task allowed comparing striatal activity during behaviorally indistinguishable prematurely and timely executed learned action sequences. We found a peak of striatal

activity preceding trial initiation that was modulated by the time waited before responding. In addition, this modulation grew at a faster rate in adolescent rats. This likely reflects a steeper growing increase in reward expectancy during waiting that could underlie their more impulsive behavior compared to adult rats.

Results

Rats learn to make timely action sequences to obtain a water reward

Water-deprived rats were trained to obtain water from a lick tube located within a nose-poke by emitting a sequence of eight licks following a visual cue (Figure 1a and b; task modified from Zold and Hussain Shuler,2015). Animals self-initiated the trials by entering into the nose-poke. In those trials initiated 2.5 s after the end of the previous trial (timely trials), a 100 ms duration visual cue (two symmetrical green LEDs located at the nose-poke sides) reported that there was a 0.5 probability of receiving a water reward. In contrast, prematurely initiated trials (<2.5 s waiting time) were penalized by re-initiating the timer and had no visual cue associated with them. Spike discharges and local field potentials (LFP) were recorded from the dorsal striatum using custom-made tetrodes (Vandecasteele et al., 2012) (representative localization Figure 1c, for detailed localization see Supplementary Figure 10).

Adult rats learned to make timely nose-poke entries followed by an 8-lick sequence (Figure 1d). This was evidenced by a twofold increase in the reward rate in sessions late in training (after three consecutive sessions with >70% correct trials) compared to early in training (Figure 1e, $p=6.10 \times 10^{-5}$, Wilcoxon matched pairs test, n pairs=15). Performance in 8-lick trials became faster with training (Figure 1f-h): trial duration (learning stage: $F_{(1, 14)}= 19.85$, $p=5.43 \times 10^{-4}$, trial: $F_{(1, 14)}= 19.91$, $p=5.37 \times 10^{-4}$), latency to the first lick (learning stage: $F_{(1, 14)}= 12.59$, $p=3.21 \times 10^{-4}$) and time to complete the 8-lick sequence (learning stage: $F_{(1, 14)}= 18.33$ $p=7.60 \times 10^{-4}$), diminished with training for both rewarded and unrewarded

timely trials (significant effect of learning stage, non-significant interaction, two-way RM ANOVAs). A representative raster plot showing licking bouts during a whole session is shown in Fig. 1i and a representative video recording of behavior displayed during premature and timely trials is shown in Supplementary video 1.

Premature nose-poke entries delayed the opportunity to get the reward as evidenced by a negative correlation between the relative number of premature trials and reward rate across sessions (Figure 2a, R square=0.090, $F_{(1,60)}$=5.96, p=0.018, n=62). Premature trials diminished with training from >30% of all trials at the beginning of training to ~15% at the end of training. However, premature trials followed by an 8-lick sequence rose from 40% to 70% of all premature trials with training, paralleling the relative increase of 8-lick sequences observed in timely trials (Figure 2b, $F_{(1, 14)}$=32.87, p=5.20 $x10^{-5}$ for learning stage and $F_{(1, 14)}$=121.74, p=2.73 $x10^{-8}$ for trial timing, no interaction, two-way RM ANOVA). These data suggest that behavior during premature trials was modified by learning, just like in timely trials. Further supporting this presumption, time to complete the 8-lick sequence ($F_{(1, 14)}$=15.63, p=0.001), latency to the first lick of the sequence ($F_{(1, 14)}$=14.30, p=0.002) and variation coefficient of the inter-lick intervals (an index of the regularity of such intervals; $F_{(1, 14)}$=6.78, p=0.021), diminished with training both for timely and premature trials (Figure 2c-e; significant effect of learning stage, no effect of trial type, no interaction, two-way RM ANOVA). Remarkably, even though premature trials were detrimental to reward rate (Figure 2a), there was a significant positive correlation between the percentage of premature trials followed by an 8-lick sequence and reward rate (Figure 2f, R square=0.478, $F_{(1,60)}$=54.98, p=1.03 $x10^{-9}$, n=62).

To characterize the timing of trial initiations in adult rats, plots showing the frequency distribution of all trial initiation times were built (Figure 2g-h). The probability of a trial including an 8-lick sequence sharply increased at the end of the 2.5 s waiting interval, peaked immediately after its finalization, and then diminished gradually. Similar results were observed in a separate group of rats trained with a longer waiting interval (Supplementary Figure 1a-h). Finally, rats trained with the long waiting interval (5 s)

quickly learned to adjust trial initiations to a shorter waiting interval (2.5 s) (Figure 2i), suggesting that premature trials with 8-lick sequences served to adapt behavior to changes in the waiting time requirements of the task that otherwise would have passed unnoticed to the rats.

An additional group of adult rats was trained with a modified version of the task that required initiating trials not before 2.5 s and no later than 5 s after exiting the port in the previous trial. These animals showed similar behavior with a more marked decrease of trial initiations after the 2.5 s was passed (see below; Supplementary Figure 2).

Altogether, the data show that adult rats improved their reward rate by waiting the least possible time between trials. Noteworthy, premature execution of the learned behavioral response was (relatively) more common late in training than early in training, suggesting that with training behavior became less sensitive to the absence of the reward predictive visual cue.

Task-sensitive striatal activity concentrates at the boundaries of the learned behavioral response

To determine if striatal activity marks the boundaries of the learned action sequence in our task (Jin and Costa, 2010; Jog, 1999), dorsal striatum activity recorded from adult rats was analyzed by aligning the activity to port entry and port exit (Figure 3a-c). Visual inspection of neuronal raster plots and peri-event time histograms (PETH) showed strong modulations of striatal activity preceding port entry and/or at the time of port exit, during timely trials (Figure 3a-b). Overall, ~50% of the recorded units (473 out of 867) showed higher activity (>1 SD over baseline) preceding port entry ("anticipatory activity"; 19%, Figure 3d), at the time of port exit (18%, Figure 3e) or both before port entry and at port exit (12%, Figure 3f). On average, these neurons showed lower than baseline firing rates during the execution of the learned action sequence (Figure 3d-f). There were also neurons (n=108, 12% of all recorded neurons) showing higher activity when the animal was inside the port than during the waiting period or at the initiation and

finalization of the action sequence (Figure 3g). Finally, 82 neurons classified as non-task responsive were tonically active during the waiting period regardless of the waited time (Supplementary Figure 3a). All main types of task-related activity emerged early during training (Supplementary Figure 3b).

Thus, although striatal activity was continuously modulated during the present task, modulations at the boundaries of the behavioral response accounted for about 50% of all task related activity and more than 30% of the recorded neurons showed a peak of activity anticipating trial initiation.

Waiting time modulates anticipatory activity

Striatal activity that anticipates a learned behavioral response could specifically mark the initiation of a previously rewarded action sequence (Jin and Costa, 2010; Jog, 1999; Martiros et al., 2018). Likewise, it could relate to additional factors, such as reward anticipation or the vigor and value of an upcoming action (Lauwereyns et al., 2002; Samejima et al., 2005; Wang et al., 2013). Moreover, it has been proposed that changes in striatal activity preceding the initiation of a prepotent action may predict premature responding (Buckholtz et al., 2010; Donnelly et al., 2014; Wu et al., 2018). Because in the present task the learned action sequence is often prematurely released, we asked if the observed anticipatory activity could specifically predict its release, and if, additionally, encodes its timing. When all port entry responsive neurons were considered (i.e., port entry only plus port entry/port exit neurons), the average firing rate modulation anticipating trial initiation was higher for timely than for premature trials irrespective of the upcoming action including the 8-lick sequence or not (Figure 4a-b, $F_{(1, 330)}=30.84$, $p=7.01 \times 10^{-8}$, significant main effect of trial initiation timing, no effect of action sequence-structure $F_{(1, 330)}=0.089$, $p=0.765$, no interaction, $F_{(1, 330)}=9.64 \times 10^{-5}$, $p=0.992$, two-way RM ANOVA). On average, this activity began 1 s before and peaked 0.5 s before the animal crossed the infrared beam located at port entry, both in premature and timely trials (Figure 4a). Since this anticipatory activity closely preceded approaching movements towards the nose-poke, we analyzed accelerometer recordings of head

movements available from two adult rats. The accelerometer recordings did not show differences in movement initiation time between premature and timely 8-lick trials (0.357 ± 0.027 and 0.382 ± 0.023 seconds -mean and SEM- for premature and timely trials respectively, $t_{(8)}$=0.636, p=0.543, paired t test; Figure 4c). Furthermore, the 8-lick sequences emitted during premature and timely trials lasted the same and had the same latency and inter-lick interval regularity (Figure 2c-e), suggesting similar action vigor during premature and timely 8-lick trials. Thus, although head acceleration data may not include differences in movements of other body parts, our data support that in this task, the firing rate modulation preceding trial initiation discriminates between premature and timely trials and does not predict the speed, regularity, structure, value or vigor of the subsequently released action sequence.

To further investigate this anticipatory activity, we plotted its amplitude at increasing waiting times observing that it increased with a steep slope as time waited surpassed the learned waiting interval and then plateaued (Figure 4d-e). Similar results were obtained in rats trained with a longer waiting interval (Supplementary Figure 4). In contrast, there was no modulation of this neuronal activity by time waited in the following trial (Figure 4f), nor of port exit related activity by the previous waiting time (Figure 4g). The steep slope of the curve at the criterion waiting interval suggested that the neuronal activity does not linearly report elapsed time but is rather related to changes in reward anticipation as waiting progressed. To explore this possibility, we analyzed striatal activity of rats trained with a modified version of the task requiring initiating trials not before 2.5 s and no later than 5 s after exiting the port in the previous trial (Supplementary Figure 2). Waiting less than 2.5 s (premature trials) or more than 5 s (late trials) was penalized by resetting the waiting timer (Figure 4h). We reasoned that if this activity provides a reward anticipation signal for the upcoming action, it should decrease after 5 s of waiting in the modified version of the task, considering that trials with >5 s waiting-time are not rewarded. As in the standard version of the task, the animals learned to wait the less possible time between trials, and also noticed the effect of the cutoff time on reward probability, as evidenced by a reduced number of late trials with training (Figure 4i,

($q_{(20)}$=10.64, p=1.10 x10^{-9} versus timely trials, Tukey post hoc test after significant one-way RM ANOVA ($F_{(1.70,\ 34.05)}$= 33.07, p=1.81 x10^{-6}); Supplementary Figure 2). The firing rate modulation preceding trial initiation increased with a steep slope at 2.5 s, but instead of plateauing, it decreased after surpassing the 5 s cutoff time (Figure 4j). A comparison of the effects of waiting time on anticipatory activity in the tasks with and without the 5 s cutoff yielded a significant interaction in a two-way RM ANOVA (Figure 4k, significant effect of task, $F_{(1,\ 334)}$=4.54, p=0.034; trials, $F_{(1.96,\ 656.34)}$=67.73, p=1.01 x10^{-15}; interaction $F_{(2,\ 668)}$=29.27, p=6.53 x10^{-13}, two-way RM ANOVA).

In summary, the marked modulation of striatal activity preceding trial initiation probably reflects subjective changes in reward anticipation as waiting progressed.

Trial initiation timing modulates striatal activity at predicted outcome time

Striatal activity can be modulated by reward-predictive sensory cues (Schultz, 2015). In the present task, a small population of neurons whose activity was modulated at the time of the visual cue (n= 27, 3% of all recorded neurons) showed lower activity during premature trials, when the visual cue was not presented ($q_{(26)}$=6.64, p=2.15 x10^{-4} versus timely trials, Tukey post hoc test after significant one-way RM ANOVA ($F_{(1.75,\ 45.49)}$=17.80, p=5.00 x10^{-6}) (Supplementary Figure 5). This modulation by trial initiation timing was similar to that observed in port entry neurons and may represent the same kind of wait time-based reward anticipation activity that extends until when sensory feedback discloses if reward could be obtained or not. In contrast, neurons showing increased activity during licking did not show any modulation by trial initiation timing (Figure 5a and 5b). Furthermore, when we studied the neuronal activity around the individual licks, we found largely overlapping populations of positively modulated neurons with broad activity peaks encompassing many inter-lick intervals (Figure 5b). The lick-related modulation lasted longer during rewarded trials,where licking also persisted for longer, than in timely unrewarded (>2.5 s of waiting) and premature 8-lick trials (<2.5 s of waiting). However, it showed a

similar amplitude irrespective of reward delivery or omission (Figure 5b, activity centered at the eighth lick), suggesting that these neurons were modulated by licking and not by the delivery of the reward itself.

Neuronal activity in the dorsal and ventral striatum is also modulated by the trial outcome and distinguishes between rewarded and non-rewarded trials in probabilistic tasks (Atallah et al., 2014; Histed et al., 2009; Nonomura et al., 2018; Shin et al., 2018; Yamada et al., 2011). In the present task, there are two types of non-rewarded, yet correctly performed trials: i) the premature 8-lick trials, which seem to be necessary to maintain an internal representation of the criterion waiting time, but should be minimized to improve the reward rate; and ii) the 50% of timely unrewarded trials that still need to be maintained to sustain the reward rate. Importantly, the animal cannot predict if a timely trial will be rewarded or not but they should be able to predict that no reward will be received after initiating a premature trial as the absence of the light cue unequivocally indicates that no reward will be delivered (Figure 1b). To determine if the negative outcome associated with premature 8-lick trials is reported by striatal neurons, and particularly, if this report differs from the report of an unrewarded outcome during timely trials, we looked for outcome-related activity modulations at the time of the eight lick during timely rewarded, timely unrewarded and premature trials with 8-lick sequences (Figure 5c-g). Overall, ~15% of the neurons recorded in adult rats (125 out of 867) showed responses at the time of the 8-lick that differed between these trial types. We found neurons that responded preferentially to: i) reward delivery (Figure 5c, “reward-responsive neurons”), ii) to non-rewarded conditions (Figure 5d, “no reward-responsive”), and iii) the absence of reward during premature 8-lick trials (Figure 5e, “expected no reward-responsive”). More specifically, 5 percent of all recorded neurons (n= 49) showed a higher activity modulation when a reward was obtained than in either no-reward conditions (Figure 5f top: reward-responsive neurons, Rewarded vs. Unrewarded trials, $q_{(48)}=11.18$, $p=5.77 \times 10^{-15}$, Rewarded vs. Premature trials, $q_{(48)}=12.52$, $p=9.82 \times 10^{-17}$, Unrewarded vs. Premature trials, $q_{(48)}=4.89$, $p=1.18 \times 10^{-5}$, Tukey test after significant one-way RM ANOVA ($F_{(1.27, 61.14)}=66.01$, $p=2.69 \times 10^{-11}$). This difference is observed just before lick rates began

to diverge by trial type (Figure 5h). Five percent of the striatal neurons showed activation at the time of the no-reward outcome in timely unrewarded trials and were unresponsive to reward delivery (Figure 5d). In addition, these neurons showed a marked activation at the time of the eighth lick during premature 8-lick trials (Figure 5f middle: no reward-responsive neurons, Rewarded vs. Unrewarded trials, $q_{(45)}$=12.61, p=2.23 x10^{-16}, Rewarded vs. Premature trials $q_{(45)}$=9.77, p=1.08 x10^{-12}, Unrewarded vs. Premature trials $q_{(45)}$=3.85, p=3.69 x10^{-4},Tukey test after significant one-way RM ANOVA ($F_{(1.28,\ 57.7)}$=57.11, p=3.75 x10^{-10})). Strikingly, there were 33 neurons (4% of all recorded neurons) that showed a two-fold higher activation at the time of the eighth lick in premature trials than in timely unrewarded trials (Figure 5f bottom: Rewarded vs. Unrewarded trials $q_{(30)}$=5.51, p=5.39 x10^{-6}, Rewarded vs. Premature trials $q_{(30)}$=12.07, p=4.81 x10^{-13}, Unrewarded vs. Premature trials $q_{(30)}$=10.72, p=8.86 x10^{-12}, Tukey test after significant one-way RM ANOVA ($F_{(1.71,\ 51.39)}$=51.34, p=2.99 x10^{-9})). This difference cannot be explained by differences in lick rates because licking decreased at similar rates after the eighth lick in both types of non-rewarded trials (Figure 5h, bottom). Importantly, the no-reward outcome is certain in premature trials as indicated by the absence of the light cue. Further data analysis showed that, indeed, the three populations of outcome-responsive neurons (reward-responsive, no reward-responsive and “expected no-reward” -responsive) distinguished between those trials when the timing was right but there was no reward from those where the animal entered the nose-poke prematurely, which are never rewarded. However, those neurons showing a higher activation during premature trials discriminated better between the two no-reward conditions (Rewarded vs. Unrewarded trials $q_{(45)}$=0.39, p=0.694; Rewarded vs. Premature trials $q_{(30)}$=5.01, p=2.23 x10^{-5}; Unrewarded vs. Premature trials $q_{(30)}$=5.43, p=6.89 x10^{-6}; Mixed-effects analysis after significant one-way RM ANOVA ($F_{(1.62,\ 99.89)}$=11.80, p=8.65 x10^{-4}), Figure 5g). Overall, the data support that activity at the time of expected trial outcome reflects differences in reward prediction between premature and timely trials.

Activity at port exit reports whether the animal has performed the learned action sequence or not

Although striatal activity predominated at the boundaries of the learned behavioral response, the activity preceding response initiation was not invariably connected to the execution of the learned action sequence. However, when all port exit neurons were considered (i.e., port exit only plus port entry/port exit neurons), a higher activity was observed for 8-lick trials than for trials with an incomplete licking sequence, independently of whether the trials were timely initiated or not (Figure 6a-b; $F_{(1,\ 101)}$=36.28, p=2.79 x$10^{-8}$for lick sequence structure, no effect of trial timing $F_{(1,\ 101)}$=1.86, p=0.176, interaction $F_{(1,\ 101)}$=4.07, p=0.046, two-way RM ANOVA). Moreover, there was no difference between trial types in the accelerometer data at the time animals exited the port (Figure 6c). Thus, port-exit related activity seems to tell if the learned response was emitted or not regardless of its timing and outcome.

Adolescent rats make more premature trials

Adolescence is a high impulsivity period more prone to make high-risk decisions, while neurodevelopmental processes are still taking place (Romer, 2010). In order to better understand how adolescent animals behave in this self-initiated task where an internal representation of time should be kept to withhold a prepotent response, we recorded striatal activity from five 35-46 days-old rats and trained them in the 2.5 s waiting-interval task. The rats successfully improved their performance in the task as shown by a higher percentage of correct trials (Figure 7a), a higher reward rate (Figure 7b, p=2.00 x10^{-6}, Wilcoxon matched-pairs signed-rank test n pairs=20), and a faster performance with training (Supplementary Figure 6). Trial duration, latency to first lick and 8-lick sequence duration did not differ between 8-lick trials of adolescent and adult rats (Supplementary Figure 6a-c), nor between premature 8-lick and timely 8-lick trials in adolescent rats (Supplementary Figure 6e-g). Accelerometer recordings of head movements available from two adolescent rats did not show differences of movement initiation time between premature and timely 8-lick trials (0.591 ± 0.014 and 0.637 ± 0.022 -mean and SEM- seconds for

premature and timely trials respectively, $t_{(9)}$=1.537, p=0.1586, paired t test; Supplementary Figure 8a). Finally, the proportion of premature 8-lick sequence trials increased with training in adolescent rats (Figure 7c; significant effects of learning stage ($F_{(1, 19)}$=93.97, p=8.66 x10^{-9}) and trial timing ($F_{(1, 19)}$=211.67, p=9.40 x10^{-12}), no interaction ($F_{(1, 19)}$=5.01, p=0.037), two-way RM ANOVA), in parallel with reward rate (Figure 7d, R square=0.067, $F_{(1,84)}$=170.98, p=5.88 x10^{-22}, n=86), and the frequency distribution of trial initiation times showed that adolescent rats learned the waiting interval (Figure 7e). In summary, adolescent rats also released the learned action sequence prematurely despite extensive training, and indeed, they did it more frequently than adult rats (Figures 7f and 7g). The number of premature trials per reward obtained was twofold higher in adolescent than adult rats even after training (Figure 7f; significant effects of age group ($F_{(1, 34)}$=28.51, p=6.22 x10^{-6}), training ($F_{(1, 34)}$=50.19, p=3.49 x10^{-8}) and interaction ($F_{(1, 34)}$=10.43, p=2.74 x10^{-3}), two-way RM ANOVA), and this was due to preferential retention of the premature 8-lick trials through training (Figure 7g; significant effect of age group ($F_{(1, 34)}$=17.88, p=1.67 x10^{-4}), no effect of training ($F_{(1, 34)}$=0.19, p=0.662), no interaction ($F_{(1, 34)}$=1.58, p=0.217), two-way RM ANOVA). Altogether, adolescent rats achieved the same reward rate as adult rats but at the expense of a higher cost as evidenced by the excess of premature responses per reward obtained.

Steepest reward anticipation signal preceding trial initiation in adolescent rats

Overall, task-related striatal activity was qualitatively similar in adolescent and adult rats (Supplementary Figure 7). Of the registered units in adolescent rats trained in the 2.5 s waiting interval task (n=876), the proportion of neurons showing port entry, port exit and lick-related activity was similar to that found in adult rats (Figure 7h; Chi-square$_{(4)}$=1.87, p=0.759, Chi-square test). Also like in adult rats, the activations preceding port entry were higher for timely than for premature trials regardless of whether the 8-lick sequence was completed or not (Figures 7i and j; significant main effect of trial initiation timing $F_{(1,362)}$=148, p=3.72 x10^{-27}, no effect of action sequence structure, $F_{(1,362)}$=1.04, p=0.309, no interaction

$F_{(1,362)}$=0.12, p=0.723, Two-way RM ANOVA). Outcome-related activations discriminating reward delivery from reward omission, and selective activations at expected reward time during premature 8-lick trials, not attributable to differences in lick rate, were also present (Supplementary Figure 8).

Therefore, we looked for quantitative differences in task-related activities that could account for the more impulsive behavior of adolescent rats. Data from six adolescent and five adult rats trained in the 2.5 s waiting interval task were used for the following analysis. Activations at the time of expected reward (Figure 7k, effect on trial initiation timing ($F_{(1.95,\ 76.18)}$=8.53, p=4.64 x10-3), no effect of age ($F_{(1,\ 50)}$=0.42, p=0.518), no interaction ($F_{(2,\ 78)}$=0.95, p=0.390), Restricted Maximum Likelihood test, REML) and activity at port exit -probably reflecting that the correct action sequence was emitted- (Figure 7l, sequence structure, $F_{(1,\ 356)}$=124.14, p=6.18 x10-25, timing $F_{(1,\ 356)}$=4.11, p=0.043, significant interaction between trial initiation timing and sequence structure $F_{(1,\ 356)}$=21.21, p=5.74 $x10^{-6}$, three-way ANOVA), did not differ between adolescent and adult rats (effect of age $F_{(1,\ 356)}$=0.12, p=0.727).

On the other hand, the data displaying the amplitude modulation of striatal activity preceding trial initiations at increasing waiting times (n= 235 and 124 striatal neurons for adolescent and adult rats respectively) differed between adolescent and adult rats (Figure 7m). To substantiate the dynamics of the age-related differential modulation of anticipatory activity, we adjusted a logistic function to the binned neuronal activity data of each age group (Figure 7m). Despite the relatively low number of 8-lick premature trials, which forced us to use relatively long bins for the shortest waiting times, the equation fitted the data convincingly (R square=0.904 and 0.945 for adolescent and adult rats respectively). Importantly, the adolescent rats curve showed a significantly steeper increase than the adult rats curve (p2 coefficient of the logistic function, profile likelihood estimation of the 95% confidence interval: Adolescent rats=1.080 (95% CI: 0.533 to 2.160) and Adult rats=0.4525 (95% CI: 0.273 to 0.747); $F_{(1,16)}$= 4.86, p=0.042). Interestingly, the difference between the curves was more evident during the first seconds that follow the criterion wait time than for shorter waiting times (Figure 7m, inset). To understand the

behavioral correlate of the observed neuronal activity differences, the same logistic function previously used to fit the neuronal data was adjusted to the cumulative frequency distribution of 8-lick trial initiation times of each age group (Figure 7n, R square=0.9307 and 0.8874, for adolescent and adult rats respectively). As expected, the curve of the adolescent group showed a significantly steeper growth rate (p2 coefficient of the logistic function, profile likelihood estimation of 95% confidence interval: Adolescent rats=0.703 (95% CI: 0.683 to 0.724) and Adult rats=0.392 (95% CI: 0.380 to 0.406) $F_{(1, 9960)}$=658.8, p=9.60 $x10^{-14}$). Similar behavioral results were obtained in an independent sample of 6 adolescent and 6 adult rats without electrode implants (Supplementary Figure 9). Remarkably, in addition to a relative excess of premature 8-lick trials, the adolescent rats showed relatively more timely trials than adult rats at short times after the criterion wait time was surpassed (Figure 7n), suggesting that adult rats are more tolerant to reward delay. Indeed, for a waiting time of 5.1 s, adult rats had accumulated 50% of 8-lick trials, while adolescent rats had already accumulated 70% of the trials. Finally, we adjusted the logistic function to the data of each rat separately. We then computed a correlation between the p2 parameters obtained for the waiting modulation of neuronal activity and trial initiation times, and found a significant correlation (Figure 7o, R square=0.833, p=2 $x10^{-4}$, n=10).

In sum, the data show that the waiting modulation of reward anticipation grows at a faster rate in adolescent rats, paralleling a higher rate of accumulation of initiated trials across time in this age group.

Discussion

In the present task, water-deprived rats learn to withhold a prepotent response to a water port to avoid a negative contingency, the re-initiation of the waiting interval, which delays the opportunity to get the next reward. The task resembles differential reinforcement of low rates of responding protocols (DRL) that are used to assess response inhibition and timing factors associated with impulsivity (Monterosso and Ainslie, 1999; Neill, 1976). Factors that increase premature responding in humans, such as

psychostimulant drugs (Evenden, 1998), psychostimulant withdrawal (Peterson et al., 2003), D2-type receptor agonists (Engeln et al., 2016), maternal separation and social isolation (Lovic et al., 2011), sleep restriction (Kamphuis et al., 2017) and adolescence (Andrzejewski et al., 2011) also increase premature responding in rodents trained with DRL procedures. Furthermore, these procedures have been used in educational and clinically meaningful contexts to reduce impulsive behavior (Bonner and Borrero, 2018; Lennox et al., 1987). Unlike classical DRL procedures, the present task requires responding with an action sequence that becomes highly automatized with training and could not always be stopped despite sensory feedback telling that it will not be rewarded. Interestingly, although premature responding showed an overall decrease with training, the premature releases of the learned action sequence were selectively preserved, suggesting that these responses are necessary to learn the duration of the waiting interval and to adapt to its changes. As long as the waiting interval remains predictable, investigation of the waiting interval through premature responding should be minimized to improve reward rate. Importantly, adolescent rats learn the task but make more premature responses per reward obtained. In addition, they initiate their timely trials closer to the criterion wait time than the adult rats, supporting that they are less patient. Overall, the adolescent rats seem to take a higher risk of releasing the learned action sequence prematurely, which makes them more impulsive than adult rats in the present task

Recent studies link premature responding in anticipation of reinforcement to a higher tendency to automatize behavior and form habits (Everitt et al., 2008; Voon et al., 2015). Striatal activity marking the boundaries of automatized action chains has been perceived as a sign of "packaged behavioral sequences" (Graybiel, 2008) that would be difficult to stop after their release regardless of whether their initiation was more or less goal-directed (Geddes et al., 2018; Robbins and Costa, 2017). While this “bracketing activity” prevails in the DLS, according to some studies, the representation of task events contributing to goal-directed behavior persists in the DMS even after extensive training (Kubota et al., 2009; Thorn et al., 2010). These parallel representations may allow switching between automatic and deliberation-based task-

solving strategies when outcomes change (Balleine and Dickinson, 1998; Daw et al., 2005; Yin and Knowlton, 2006) or interventions that modify neural activity impair behavioral control by one of the circuits involved (Gremel and Costa, 2013; Smith and Graybiel, 2013; Yin et al., 2004). The remarkably similar behavior observed in 8-lick prematurely released and timely unrewarded trials led us to expect that a stronger boundary activity, and/or a less precise coding of task events, could explain the higher rates of premature responding observed in adolescent rats. Indeed, a recent study suggested that the stronger the activity at the action initiation, the lower the deliberation at the turn-choice site in a T-maze task (Smith and Graybiel, 2013). Here, the strength of anticipatory activity increased with the time waited before response release and was higher in the more impulsive adolescent rats. However, it did not predict that the rewarded 8-lick sequence would be included in the behavioral response. On the other hand, a similar activity followed all responses containing the learned 8-lick sequence regardless of trial outcome and despite the markedly different lick rates observed between rewarded and unrewarded responses after outcome disclosure. A recent study showed that hierarchical control during action sequence execution may allow the selective removal of intermediate sequence elements (Geddes et al., 2018), which could be a mechanism through which licks could be deleted from behavioral responses preceded by similar wait time-based reward expectations in the present task. Moreover, since our recordings were mainly obtained from the medial and central dorsal striatal regions, we cannot rule out that a different anticipatory activity specific for the learned sequence emerges in the DLS in our task as previously reported for other tasks (Jin and Costa, 2010; Jog, 1999; Martiros et al., 2018; Thorn et al., 2010). Importantly, (Jin and Costa, 2010; Jog, 1999), the activity of many striatal neurons seemed to continuously follow the expression of the lick sequence as has also been reported previously by others (Jin and Costa, 2010; Jin et al., 2014; Rueda-Orozco and Robbe, 2015) and a recent study showed a continuous modulation of striatal neuronal activity across the initiation, execution and termination of cue-guided locomotion (Sales-Carbonell et al., 2018). Overall, our data show that the concentration of striatal activity at the beginning and end of a learned task is widely expressed in the dorsal striatum even after skilled performance is reached, as reported in a recent

study with a different task (Vandaele et al., 2019), and suggest that, this activity is more flexible regarding the properties of the behavioral response it bounds, and carries information about its pertinent timing.

We speculated that the more impulsive behavior of adolescent rats observed in the present task could relate to changes in wait time-based modulation of reward anticipation and/or outcome evaluation signals. Striatal signals at the time of expected outcome not only discriminated between rewarded and unrewarded correct trials as observed by others (Atallah et al., 2014; Nonomura et al., 2018; Shin et al., 2018), but also premature from timely responses. A specific outcome evaluation signal after premature 8-lick trials may serve to inhibit premature responding and minimize the exploration of the waiting interval. To compute such outcome signal, the animal has to retain information regarding the time waited before responding (or regarding visual feedback on proper trial initiation timing) while monitoring the execution of the lick sequence. Such kind of integration could be implemented by retaining information of early task events along the sequential activation of striatal ensembles (Her et al., 2016; Nonomura et al., 2018). However, although the effect of waiting time on task-related striatal activity extended until the time of the visual cue, it was absent in ensembles with lick-related activity. Thus, the sustained lick-related signal may serve to estimate outcome timing (Zold and Hussain Shuler, 2015), but information about trial initiation timing or the visual cue should arrive at outcome-sensitive ensembles through other mechanisms likely involving inputs from prefrontal and orbitofrontal cortex areas (Asaad et al., 2017; Hamid et al., 2021; Wassum et al., 2011). Nonetheless, we found no differences in outcome evaluation signals that could explain the more impulsive behavior of adolescent rats. On the other hand, although premature responding has classically been linked to poor behavioral inhibition (Bari and Robbins, 2013), alternative views equates it to preferring a smaller more immediate reward over a larger delayed one (Dalley and Robbins, 2017; Monterosso and Ainslie, 1999), and to deficient gathering or evaluation of evidence while waiting (Dalley and Robbins, 2017). Interestingly, influential models propose that time estimates derived from accumulation of pacemaker counts can be compared to a memorized time interval to decide whether a

target time has been reached or not (Buhusi and Meck, 2005; Gibbon, 1977; Namboodiri and Hussain Shuler, 2016). Previous studies showed that dorsal striatum ensembles can provide time estimations (Bakhurin et al., 2017; Emmons et al., 2017; Gouvêa et al., 2015; Matell et al., 2003; Mello et al., 2015; Zhou et al., 2020) and we also found sequential and tonic ensemble activation during the waiting period (Figure 3b and Supplementary Figure 3a) that could serve to track time while waiting in the present task. By comparing the time accumulated while waiting against the memorized waiting interval, the animal could anticipate how likely is the upcoming action to be rewarded. How fast reward expectancy steps up when time waited approaches the reference time interval would depend on several factors, including temporal discounting effects on reward value (Monterosso and Ainslie, 1999; Namboodiri and Hussain Shuler, 2016; Wittmann and Paulus, 2008). In this sense, the anticipatory signal we observe could be interpreted as a reading of the temporal discount function at the time chosen to release the learned action sequence. A recent theory proposes that animals time their decisions by estimating if they can improve the reward rate experienced in the recent past; according to it, the longer the time window over which reinforcement history is estimated, the higher the tolerance to delays of future rewards (i.e., the less steep the temporal discount effect) (Namboodiri et al., 2014). Adults should be able to integrate information about past reinforcement history over longer time windows that adolescents (Namboodiri et al., 2014), which is consistent with the finding that temporal discount rate decreases throughout childhood and adolescence (Green et al., 1999; Scheres et al., 2006; Steinberg et al., 2009). Alternative theories propose that waiting impulsivity relates to perceiving durations longer than they are, which would be associated to perceiving a higher cost of time and with a steeper temporal discounting (Wittmann and Paulus, 2008). Moreover, time perception depends on several cognitive processes that are modified by adolescence including working memory, attention and mood (Baumann and Odum, 2012; Wittmann and Paulus, 2008). It has been also been proposed that impulsivity relates to changes in an internal urgency signal that influences the timing of decisions and may be primary responsible for the build-up of neural activity observed in the striatum and cortex that often precedes action initiation (Carland et al., 2019).

Interestingly, in the present task, adolescent rats achieve a similar reward rate to adult rats and, indeed, they do not show the negative effect of premature trials on reward rate that is observed in adults (Supplementary Figure 6). Importantly, although they achieve this reward rate at the expense of a higher energy cost (more premature trials per reward obtained), by prematurely executing the learned action sequence, adolescents would likely get more precise information regarding the duration of the waiting interval, which, in turn, would refresh more frequently their cognitive map of the task. This view is consistent with the idea that impulsivity is advantageous when it helps to adapt to uncertain environments and that increased exploratory behavior in adolescence allows gaining knowledge that would be useful to guide decisions in adulthood (Addicott et al., 2017; Spear, 2000). In this context, the steepness of the wait time-based reward anticipation signal could be under the influence of a gain factor that, according to reinforcement learning models, reflects the degree of preference for the highest value option and regulates explore-exploit trade off (Addicott et al., 2017). Further studies are needed to disclose which factors influence the wait time-based reward anticipation signal we observe in the present task and which among them are responsible for the differences observed between adult and adolescent rats. Moreover, since the experiments have been conducted in male rats and impulsivity has sexually dimorphic features (Weafer and de Wit, 2014), studies are needed to understand how female animals behave in the present task and whether female and male animals encode wait time-based reward anticipation differently.

In summary, detailed analysis of behavior in a waiting task that promotes premature responding with an automatized action sequence shows a more impulsive behavior in adolescent rats. Unlike other tasks previously used to study neural correlates of waiting impulsivity (e.g., the 5-choice serial reaction time task), the present task is learned rapidly, which makes it suitable for studying premature responding during rodent adolescence. Moreover, by requiring self-paced responding with a stereotyped action sequence, and making uncertain reward obtention during timely trials, the task generates conditions that “clamp” behavior. This allowed identifying a modulation of anticipatory striatal activity by time waited

before responding, which grows with a steeper slope in adolescent rats, likely reflecting age-related changes in temporal discounting, internal urgency states or explore-exploit balance. Translational studies are necessary to understand if similarly designed tasks capture the relationship between impulsivity and automaticity in behavior that has been related to vulnerability to drug addiction.

Materials and Methods

Subjects

Subjects were adult male Long-Evans rats (Adolescents: 28-46 days-old at the beginning of the experiments, weight ~120 g; Adults: 70-128 days old at the beginning of the experiments, weight ~390 g), kept under a 12 hr light/dark cycle (lights on at 7:00 A.M.) at a constant temperature range of 21-23˚C. Rats were housed in groups of 3–4 in regular cages up to the time of surgery, after which they were singly housed with moderate environmental enrichment (toys, tissue paper strips). Two days before the beginning of the behavioral training, subjects entered a water restriction protocol in which they received free access to water for 20 min each day. Once a week, animals were allowed to have water *ad libitum*. This schedule maintained animals at 90% of their pre-deprivation weight. All procedures complied with the National Institutes of Health Guide for Care and Use of Laboratory Animals (Publications No. 80-23, revised 1996) and were approved by the Animal Care and Use Committee of the School of Medicine of the University of Buenos Aires (CICUAL). Striatal recordings were obtained from: 5 adult males that were part of the 2.5 s WT experiments, 3 adult males that were part of the 5 s WT experiments (Fig. 2i and Fig. S1) and another three adult males were included in the 2.5-5 s WT experiments (Fig 4h-k, Fig. S2 -all except S2f, see below-). Six adolescent males were part of 2.5 s WT experiments (Fig. 7, Fig S7-S9). In addition, six adult and six adolescent male rats, not implanted with recording electrodes, were part of the behavioral experiments reported in Supplementary Figure 9.

Electrodes

Eight tetrodes (each made of four 12μm, tungsten wires, California fine wire company, USA, 0.2 MΩ impedance) were attached to a homemade micromanipulator (Vandecasteele et al., 2012). The array of 8 tetrodes was gradually lowered each day before training to find neuronal activity. At the end of the recording session the array was also lowered to change the recording site for the next training session (approximately 80 μm each day).

Surgery

Before surgery, subjects were treated with local anaesthetic (lidocaine) on the scalp. Under deep anesthesia with isoflurane (3-4% for induction, 1-1.5% for maintenance), rats were placed in the stereotaxic frame and chronically implanted with an array of tetrodes aimed at the striatum (AP: -0.12 cm, L: -0.25 cm, DV: -0.35 cm). Body temperature was maintained using a heating pad. Through a small craniotomy performed in the corresponding area the tetrodes were then lowered to a depth of 3.5 mm. The micromanipulator was fixed in place with dental cement. Two stainless steel screws (0-80 X 1/8” Philips Pan head) were inserted above the cerebellum to be used as ground and reference. Three additional steel screws were inserted to anchor the whole implant to the skull. The craniotomy was covered with a sterile 50-50 mixture of mineral oil and paraffin. Towards the end of the surgical procedure, animals were treated with antibiotics (i.m., enrofloxacin 10 mg/kg) and a veterinary ointment (antiseptic, anti-inflammatory and anesthetic) was applied on the skin in contact with the external side of the implant to prevent microbial infections. Subjects were kept under careful observation until awakening. The first recording session followed 7 days of post-surgical recovery.

Behavioral training

Rats were placed in a dark behavioral operant chamber that contained a nose-poke where they could seek a water reward by licking through a slot onto a lick tube (Med associates) (Figure 1a). After a minimum of 2.5 seconds (criterion waiting time), rats could initiate a trial by entering the nose-poke, 200 ms after which they received a 100 ms visual cue through two green leds placed on the sides of the nose-poke. The animal was then required to lick 8 times to gain a water reward (~20 µl of water) in 50% of the trials in a pseudo-random manner. Licks were detected through an infrared beam located at the sides of the lick tube. Trials ended when the animal removed its head from the nose-poke. Sessions ended when the subject completed 300 trials or when the session reached 120 min. Premature trials are those in which the criterion waiting time was violated (Figure 1b). Once the animal reached 80% of timely trials with a complete 8-lick sequence for two consecutive days, sessions were considered “late sessions”. “Early” sessions were considered from the first day of training and the number of early sessions was determined to match those classified as “late” for each animal, for comparison purposes. The remaining sessions between “early” and “late” were classified as “intermediate”.

Variants of the rewarded task

We performed two additional versions of the task: one in which the criterion waiting time was duplicated (5 s waiting time task) and another in which the animals had to enter the nose-poke within a waiting interval of 2.5 s-5 s (2.5-5 s waiting time task). For the 5 s waiting time task, subjects were trained just like in the 2.5 s task, being the minimum waiting interval of 5 s. After reaching a stable performance with the 5 s waiting-interval, two of the animals were switched to the 2.5 s protocol (data shown in Figure 2i).

For the 2.5-5 s waiting time task, rats were trained with a “shaping protocol”. Briefly, they were trained for 3 consecutive sessions with a 2.5 s task and afterwards we incorporated an upper limit to the waiting interval: any trial started with a waiting-interval longer than the upper limit was not rewarded. The upper limits for shaping were 15 s (one session), 10 s (two sessions) and 5 s (one to three sessions). One day after the end of shaping, tetrodes were chronically implanted with the surgical procedure described above. Trials in the 2.5-5 s waiting time task were classified as premature (if animals entered the nose-poke before the lower limit of 2.5 waiting), timely (if they entered the nose-poke between 2.5 and 5 s of waiting) and late (when they entered the nose-poke after 5 s of waiting). Premature and late trials were not rewarded and timely trials were rewarded with a probability of 0.5 (and reported to the animal with the visual cue).

Accelerometer recordings

Accelerometer data were obtained from two adult and two adolescent rats. Data from adult rats were collected with a MPU-6050 accelerometer (sampling rate of 25 Hz) and registered using an Arduino UNO board connected to a PC via PLX-DAQ software (Parallax Inc.). Accelerometer data from adolescent rats were obtained with an ADXL335 accelerometer (sampling rate of 125 Hz) and registered using an Arduino UNO board and a custom-made script in Processing software (https://processing.org/).

The 3-axis acceleration data were then band-pass filtered (0.25-6 Hz cutoff, fourth-order Butterworth filter) after which the Euclidean norm was taken. This acceleration signal was segmented into 10-sample windows without overlapping for the adolescent data and into 10-sample windows with 80% overlapping for the adult data (in both cases the time resolution was of 80 ms). The standard deviation for all data points in each window was calculated. Windows with SD <3.5 were considered as a no-movement window. We limited the analysis to movement initiation times detected up to 1.2 seconds before port entry.

In addition, only trials with a waiting time below 5 seconds were considered in order to include trials where animals are clearly engaged in the task.

Histology

Animals were given a lethal dose of ketamine xilazine and transcardially perfused with cold PBS containing heparin (500 U/L) followed by 4% PFA in PBS. The brain was quickly removed, postfixed in PFA at 4°C for 2–12 h, and placed in a 30% sucrose solution in PBS. Frozen coronal sections (50 μm) were collected with a sliding microtome, and histological verification of the electrode endpoints and recording tracks was done in microscopy fluorescent pictures. Later, sections were safranin-stained to confirm the visualization of the tracks.

QUANTIFICATION AND STATISTICAL ANALYSIS

Neural recording and data analysis

Brain activity and behavioral data were collected during the training sessions using commercially available hardware and software (sampling rate 32.5 kHz, Cheetah, Neuralynx). Neurophysiological and behavioral data were explored using NeuroScope (http://neuroscope.sourceforge.net; Hazan et al., 2006). Spike sorting was performed automatically, using KlustaKwik (http://klustawik.sourceforge.net), followed by a manual adjustment of the clusters (using “Klusters” software package; http://klusters.sourceforge.net, Hazan et al., 2006). After the spike sorting procedure, all data were analyzed with MATLAB software using custom-built scripts and “FMAtoolbox” (http://fmatoolbox.sourceforge.net/). We registered a total of 867 units in adult rats: 244 were task responsive and 158 non-responsive in the 2.5 s waiting time task, 47 task-responsive and 45 non-responsive in the 5 s waiting time task, and 294 task-responsive and 79

non-responsive in the 2.5-5 s waiting time experiments. In adolescent rats, we registered a total of 876 units, 385 with no response and 491 task-responsive, always in the standard 2.5 s waiting time task. All subjects and their corresponding experiment and units are detailed in Table 1.

Peri-event time histograms (PETH) were created using 80 ms bins. PETH were normalized using a z-score. For each cell its mean firing rate and SD were calculated using 20 s before and after port entry. All data are presented as normalized firing rate except for PETH of single neurons reported in Figure 3a. Only for display purposes, PETH were smoothed using a 5-points moving window. A neuron was classified as responsive to a given task event if its average firing rate during the time window of interest (-640 to -80 ms before port entry, -240 to +320 ms around port exit, -200 to +200 ms around a lick) was above 1SD of its mean firing rate. A 2SD threshold and a time window between +160 to +400 ms relative to port entry was used to define visual cue responding cells. A neuron was considered to be active during the execution of the lick sequence (inside port cells, Figure 3g) if its average firing rate between 400 and 1520 ms after port entry was above 1SD of its mean firing rate during the 40s time window centered at port entry. Neurons whose activity during the whole waiting time period, irrespective of its duration, was above 1SD of their mean firing rate during the 40 s time window, were considered tonically active during the waiting period (Figure S2). To be classified as outcome responding a neuron should met the following criteria: 1) their average firing rate between +160 to +560 ms after the eighth lick should be above 0.5 SD of their average firing rate between -640 to -160 ms prior to the eighth lick; 2) their average firing rate between +160 to +560 ms after the 8th lick should be above 2 SD of their mean activity during the 40s time window. We further divided these neurons into three groups: a) reward responsive neurons (Figure 5c), b) no-reward responsive neurons (Figure 5d); c) expected no-reward responsive neurons (Figure 5e), if their activity in either no rewarded condition differed from activity in rewarded trials by >1.5 SD between +160 to +560 ms after the eighth lick.

To analyze the activity of port entry-responsive neurons at different waiting times (e.g., Figure 4d-

g), PETH were constructed using a similar number of trials for each waiting time interval, which required using waiting intervals of different duration because trial initiations were more frequent at waiting times close to the criterion wait time. Trials with waiting time shorter than 450 ms were not included in the analysis. The intervals used were (beginning, end; in ms): [451,1915; 1916,2499; 2500,3131; 3132,3803; 3804,4999; 5000,6084; 6085,8421; 8422,12621; 12622,22013; 22014,1710755]. For the data from the 2.5 s waiting time experiments presented in Figure 7m, the PETH were also constructed with a different set of time intervals to increase the number of data points for shorter waiting time [451:1699, 1700:2199, 2200:2699, 2700:3199, 3200:3799, 3800:4999, 5000:6084, 6085:8421, 8422:12621, 12622:22013, 22014:1710755].

Behavioral analysis

All data were analyzed with MATLAB software using custom-built scripts. Histograms depicting trial initiation timing at different waiting times (e.g., Figure 2g-i) were calculated using 100 ms bins and normalized so that the peak of the curve corresponded to a value of 1. Curves were smoothed with a 4 bin span. Licks PETHs (e.g., Figure 5h) were calculated centered on port entry using a 10 ms bin and normalized to the mean and standard deviation of licks per bin for each session.

Statistical analysis

Statistical analysis of behavioral data was performed with Prism 8, Graphpad Software. All figures were made colorblind safe using palettes from ColorBrewer 2.0 (http://colorbrewer2.org). All early vs. late comparisons were done with repeated measures two-way ANOVA, except for the reward rate in which

Wilcoxon matched pairs test was used. To compare neuronal activity between Rewarded, Unrewarded and Premature trials, repeated measures one-way ANOVA was used. To compare licking activity between Rewarded, Unrewarded and Premature trials a Mixed-effects model (Restricted Maximum Likelihood test, REML) was used. The performance of adults trained with a 5 s criterion time was analyzed at the late stage of training using Mann Whitney Test. Linear regression was used to correlate premature trials and the reward rate. Behavioral and neuronal data were fitted with a logistic function (y ~ p1 / (1 + exp(-p2 * (x-p3)))). The coefficient p1 is related to the bottom part of the curve, the coefficient p3 is related to the top part of the curve and the coefficient p2 is related to the slope, where higher values describe steeper curves.

Acknowledgments

We thank Dr. Juan Belforte for helpful discussions, Yamila Páez for her technical assistance with animal welfare, Graciela Ortega, Bárbara Giugovaz and Germán La Iacona for technical assistance with the experiments (tetrode making and microscope image acquisition for the revised version of the manuscript). This project was funded with research grants from FONCYT (PICT 2016-0396; PICT 2017-0520; PICT 2017-2465) and UBACYT (2018-2020 305BA).

Author contributions

Conceptualization G.M., M.B./ Methodology M.B., C.M., C.Z./ Software M.B., C.M., M.C./ Formal Analysis M.B., C.M., G.M.,C.M./ Investigation C.M., M.B./ Resources G.M.,M.B./ Data Curation M.B./ Visualization C.M., M.B. / Writing –Original Draft G.M., C.Z./ Writing –Review & Editing C.Z., M.B., C.M./ Project Administration G.M., M.B./ Funding Acquisition G.M., M.B., C.M.

Declaration of interests

The authors declare no competing interests.

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Mariano Belluscio. Behavioral and electrophysiological data and code used for analysis are reported in https://doi.org/10.5061/dryad.8kprr4xpv.

Table 1

Rat ID	Age: group/ training start	Training Protocol (waiting time)	Experiments			Recording sessions	Neurons
7	Adult (p118)	2.5 s	Behavior	Electrophysiology		15	32
11	Adult (p85)	2.5 s	Behavior	Electrophysiology		7	62
12	Adult (p82)	2.5 s	Behavior	Electrophysiology		16	69
19	Adult (p82)	2.5 s	Behavior	Electrophysiology		25	106
29	Adult (p128)	2.5 s	Behavior	Electrophysiology		16	108
21	Adolescent (p36)	2.5 s	Behavior	Electrophysiology		19	79
25	Adolescent (p35)	2.5 s	Behavior	Electrophysiology		18	166
26	Adolescent (p37)	2.5 s	Behavior	Electrophysiology		13	307

45	**Adolescent (p42)**	**2.5 s**	**Behavior**	**Electrophysiology**		**10**	**-**
47	**Adolescent (p46)**	**2.5 s**	**Behavior**	**Electrophysiology**	**Accelerometer**	**14**	**95**
49	**Adolescent (p39)**	**2.5 s**	**Behavior**	**Electrophysiology**	**Accelerometer**	**15**	**229**
31	**Adult (p96)**	**5 s**	**Behavior**	**Electrophysiology**		**6**	**41**
33	**Adult (p60)**	**5 s**	**Behavior**	**Electrophysiology**		**6**	**51**
		2.5 s	**Behavior**	**Electrophysiology**		**2**	**25**
34	**Adult (p99)**	**5 s**	**Behavior**	**Electrophysiology**		**7**	**-**
35	**Adult (p70)**	**2.5- 5 s**	**Behavior**	**Electrophysiology**		**22**	**279**
37	**Adult (p80)**	**2.5- 5 s**	**Behavior**	**Electrophysiology**		**10**	**-**
41	**Adult (p91)**	**2.5- 5 s**	**Behavior**	**Electrophysiology**		**9**	**94**
42	**Adult (p89)**	**2.5 s**	**Behavior**		**Accelerometer**	**5**	**-**
43	**Adult (p89)**	**2.5 s**	**Behavior**		**Accelerometer**	**6**	**-**
16	**Adult (p83)**	**2.5 s**	**Behavior**				
17	**Adult (p83)**	**2.5 s**	**Behavior**				
601	**Adult (p60)**	**2.5 s**	**Behavior**				
602	**Adult (p60)**	**2.5 s**	**Behavior**				
603	**Adult (p70)**	**2.5 s**	**Behavior**				
604	**Adult (p70)**	**2.5 s**	**Behavior**				
22	**Adolescent (p35)**	**2.5 s**	**Behavior**				
24	**Adolescent (p33)**	**2.5 s**	**Behavior**				
352	**Adolescent (p36)**	**2.5 s**	**Behavior**				
353	**Adolescent**	**2.5 s**	**Behavior**				

	(p28)						
354	**Adolescent (p30)**	**2.5 s**	**Behavior**				
355	**Adolescent (p30)**	**2.5 s**	**Behavior**				

Figure Legends

Figure 1. Rats become skilled in the task. (a) Behavioral chamber with the nose-poke. Animals' entries / exits from the nose-poke and licks are detected with infra-red (IR) beams. A 100 ms visual cue is presented through a pair of green LEDs placed on the side of the nose-poke to indicate a timely entry. **(b)** Schematic representation of the different types of trials. Timely trials require a minimum waiting time of 2.5 s and premature trials are those in which the minimum waiting time is not met. After that, trials are classified by whether animals performed an 8-lick sequence or not. **(c)** Top: Representative diagram of the electrodes' positioning, aimed at the dorsal striatum. Bottom: Histological section (AP=0.24 cm from bregma) with electrodes traces. **(d)** Percentage of the different trial types per session, timely trials with an 8-lick sequence (Tx8L) or not (T<8L) and premature trials with an 8-lick sequence (Px8L) or not (P<8L). **(e)** Reward rates for early and late training stages. Trial duration **(f)**, latency to the first lick during correct trials **(g)** and time to complete the 8-lick sequence **(h)** for the two types of correct trials (rewarded and unrewarded), at each training stage. e-h: Data are expressed as mean ± SEM, n=15 early and 15 late sessions, n=5 rats performing the standard 2.5 s waiting time task. **(i)** Raster plot of the licks from 100 trials at a late training session of one of the adult rats. Colored dots show the first 8 licks of the trial (timely trials, blue and premature trials, red).

Supplementary Video 1: Behavior in a late training session of an adolescent rat.

Figure 2. Training does not suppress premature initiations of the learned behavioral response. Data correspond to 5 adult animals performing the standard 2.5 s waiting time task. **a)** Correlation between all prematurely initiated trials and reward rate (Y = -0.023*X + 1.921, slope significantly different from zero p=0.018). The proportion of trials **(b)**, time to complete the 8-lick sequence **(c)**, latency to first lick and **(d)** coefficient of variation **(e)** of the 8-lick sequence inter-lick intervals for prematurely and timely initiated trials, at early and late learning stages. **b-e):** Data are expressed as mean ± SEM, n=15 early and 15 late sessions. **f)** Correlation between percentage of prematurely initiated trials followed by an 8-lick sequence and reward rate (Y= 0.024*X – 0.0408, slope significantly different from zero, p=1.03 $x10^{-9}$). **g-i)** Normalized frequency distributions of the trial initiation times (waiting time), separated for trials with (gray solid line) and without (dotted line) 8-lick sequences and for early **(g)** and late **(h)** training stages (n=15 sessions). Insets: percentage of the trials with (gray solid line) and without (dotted line) 8-lick sequences for each bin, zoomed around the criterion time. **i)** Rats were trained with a 5 s criterion time period (blue) and afterwards were switched to a 2.5 s criterion time for the following two sessions (48 h after the last 5 s-waiting time session, lilac). Red dashed lines: criterion time; bin size: 100 ms; reference for normalization: bin with highest value = 1).

Figure 3. Striatal activity marks transitions between behavioral states of the task. Data come from 5 adult animals performing the standard 2.5 s waiting time task and 2 adult animals performing the 2.5-5 s cutoff version of the task. **a)** Representative raster plots (above) and PETH (below) of striatal units showing firing rate modulations related to port entry and/or port exit**.** L1: first lick, L8: eighth lick, Last L: last lick. **b)** Individual PETHs of striatal neurons during correct trials, aligned to port entry (left) or port exit (right).Below: Average PETH (solid line) and SEM (shaded area) for all recorded neurons. Red dashed line: port entry, orange dashed line: port exit, vertical green line: led on. **c)** Proportion of neurons showing task-relate d firing rate modulations. **d-g)** PETH for all individual neurons and population

average PETH aligned to port entry (left panels) and to port exit (right panels) for **(d)** striatal neurons showing only port entry related activity, **(e)** striatal neurons showing only port exit related activity, **(f)** striatal neurons showing activity modulations at both port entry and port exit and **(g)** striatal neurons showing higher activity while animals are inside the port. From **d-g)** data are the mean (solid lines) and SEM (shaded area). Colored bars over the x-axis show the interval used to detect firing rate modulations (red: entry, orange: exit, light-blue: inside port)**.**

Figure 4. Prematurely initiated trials are preceded by low striatal anticipatory activity. Data come from 5 adult animals performing the standard 2.5 s waiting time task except for (h-j), where they correspond to 3 rats trained in the 2.5-5 s cutoff task. **(a)** Average PETH of entry-related neurons, during premature and timely trials, for trials with or without an 8-lick sequence (diagram on top shows the types of trials analyzed). **(b)** Mean striatal activity during -640 to -80 ms before port entry for all PETH shown in (a). **(c)** Accelerometer recordings of head movements around port entry time for premature and timely trials. On the left, data are the mean (solid lines) and 95 CI (shaded area) of the accelerometer Y axis. Vertical lines represent the mean (solid) and 95 CI (dashed) of movement initiation for timely and premature trials. On the right, time from port entry to movement initiation for timely and premature trials. Data were obtained from a total of 9 training sessions of two animals. **(d)** Average PETH of striatal neurons showing entry-related activity sorted by trial waiting time duration. The color code for the intervals is shown on the right. **(e)** Mean normalized firing rate for each of the waiting time segments. The 2.5 s criterion time is shown with a blue dashed line. **(f)** Average PETH of the same striatal neurons segmented according to next trial waiting time. **(g)** Average PETH of striatal neurons showing activation at port exit segmented as in **(d)**. **(h)** Schematic representation of the different trial types in a modified version of the task incorporating a 5 s cutoff time. Timely trials: waiting time between 2.5 – 5 s, Premature trials: waiting time <2.5 s, Late trials: waiting time >5 s. **(i)** Proportion of trials followed by 8-lick sequences for each type of trial n=21 late sessions, **(j)** Average PETH of striatal neurons showing activity

preceding port entry according to the waiting time duration. The color code for the intervals is shown on the right. **(k)** Mean normalized firing rate for each of the waiting time groups for the 2.5 s (left) and 2.5-5 s (right) waiting times tasks, relative to the mean firing rate of timely trials.

Figure 5. Reward-responsive neurons discriminate prematurely from timely initiated trials. Data come from 5 adult animals performing the standard 2.5 s waiting time task and 2 adult animals performing the 2.5-5 s cutoff version of the task. **(a)** Schematic representation of the different types of trials analyzed. **(b)** Mean (solid lines) and SEM of the average PETH of striatal units showing firing rate modulations related to licking activity (centered to the 1st, 3rd and 8th lick). **(c-e)** PETH for all individual neurons and population average PETH of striatal units, aligned to the 8th lick (time 0 s), showing positive firing rate modulations during reward delivery (reward-responsive neurons) **(c)**, reward omission during timely trials (no reward-responsive neurons) **(d)** or expected reward omission during 8-lick premature trials (expected no reward-responsive neurons)**(e)**. **(f)** Mean normalized firing rate for the different trial conditions (timely rewarded, timely unrewarded, premature) for the three types of neurons depicted in c, d and e. From top to bottom: reward-responsive neurons, no-reward-responsive neurons and expected no reward-responsive neurons. **(g)** Discrimination index (DI= ABS(normalized firing rate in timely unrewarded trials - normalized firing rate in premature trials) for each of the groups of neurons shown in **(c-e)**. **(h)** PETH of the licking activity for the three trial conditions, centered at the 8th lick, with its corresponding average at the bottom.

Figure 6. Striatal activity at port exit reports the performance of the action sequence. a-b) Data come from 5 adult animals performing the standard 2.5 s waiting time task. **(a)** Average PETH of neurons responding to port exit, during premature and timely trials, for 8-lick and <8-lick sequence trials (diagram on top shows the types of trials analyzed). **(b)** Average striatal activity corresponding to the PETH shown

in a. Colored bars over the x-axis show the interval used to detect firing rate modulations. **(c)** Mean (solid lines) and 95 CI (shaded area) of head acceleration around port exit for premature and timely trials. Data were obtained from a total of 9 training sessions of two animals.

Figure 7. Adolescent rats are more impulsive and show a steeper wait time modulation of striatal activity. Data come from 6 adolescent and 5 adult rats performing the standard 2.5 s waiting time task. **(a to e)** Behavior of adolescent rats in the 2.5 s waiting time task. **(a)** Percentage of the different trial types per session. **(b)** Reward rates for each training stage. **(c)** Proportion of trials followed by 8-lick sequences at early and late learning stages. **b-c:** Data are expressed as mean ± SEM, n=20 early and 20 late sessions. **(d)** Correlation between percentage of prematurely initiated trials followed by an 8-lick sequence and reward rate (Y=0.026*X +0.112, Slope significantly different from zero, $p=5.88 \times 10^{-22}$). **(e)** Normalized frequency distributions of trial initiation times (waiting time), separated for trials with (dark gray) and without (light gray dotted line) 8-lick sequences, in late stages of training (n=20 sessions). Inset: percentage of trials with and without 8-lick sequences for each bin, zoomed around the criterion time. **(f)** Ratio between the number of premature trials and obtained rewards for adolescent and adult rats. **(g)** Proportion of premature trials with 8-lick sequence per obtained rewards for adolescent and adult rats. **(h)** Percentage of neurons showing task-related firing rate modulations in adolescent rats. **(i)** Average PETH of entry-related neurons, during premature and timely trials, for trials with (x8L) or without an 8-lick sequence (<8L), in adolescent rats. **(j)** Average striatal activity corresponding to the PETH shown in **(i)**. **(k)** Outcome-related neuronal activity: Discrimination index between 8-lick timely unrewarded and premature trials for different types of outcome responsive neurons from adolescent and adult rats. **(l)** Port exit related activity: mean normalized firing rate at the time of port exit for trials with or without 8-lick sequence for both ages. **k-l)**: dashed lines correspond to adolescents and full lines to adults. **(m)** Waiting time-modulation of anticipatory activity: mean normalized firing rate for each of the waiting time

segments for each age group. Dashed lines show the logistic function fitted to the binned neuronal activity data. Inset: detail around 2.5 sec waiting time. **(n)** Cumulative frequency distribution of 8-lick trial initiation times for adolescent and adult rats. Each light color line shows an individual session. Dashed lines show the logistic function fitted to the behavioral data. Inset: detail around 2.5 s waiting time. **(o)** Correlation between p2 slope coefficients of the logistic functions fitting the behavioral and neurophysiological data of each rat.

Supplementary Figure 1. Rats trained with a longer waiting time also emit premature learned responses. (a) Reward rates for each training stage of rats trained with a 5 s waiting time requirement Wilcoxon matched-pairs signed rank test, ** p=7.80 x10-3 n=9 pairs. **(b)** Proportion of trials followed by 8-lick sequences at early and late learning stages. Two-way RM ANOVA: Training stage, $F_{(1, 8)}$=5.119, p=0.053; Type of trial, $F_{(1, 8)}$=198.63 *** p=6.24 $x10^{-7}$; Interaction Training stage x Type of trial, $F_{(1, 8)}$=8.52, * p=0.019. **(c)** Time to complete the 8-lick sequence, Paired t-Test, $t_{(8)}$=0.67. **(d)** Latency to first lick for prematurely and timely initiated trials, at early and late learning stages, Paired t-Test, $t_{(8)}$=0.42, p=0.683. **(e)** Coefficient of variation of the 8-lick sequence inter-lick intervals for prematurely and timely initiated trials, at early and late learning stages, Paired t-Test, $t_{(8)}$=0.13, p=0.896. Data are expressed as mean ± SEM, n=9 sessions. **(f)** Correlation between prematurely initiated trials and reward rate., Linear regression Y=-0.02755*X +2.776, slope significantly different from zero: $F_{(1,24)}$=12.05, p=0.020, R square=0.334, n=26 sessions. **(g)** Correlation between percentage of prematurely initiated trials followed by an 8-lick sequence and reward rate, Linear regression Y=0.01321*X +0.5469, Slope significantly different from zero: $F_{(1,24)}$=6.30, p=0.019, R square=0.208, n=26 sessions. **(h)** Normalized frequency distributions of the trial initiation times (waiting time), separated for trials with (gray) and without (dotted line) 8-lick sequences, in the late stages of training. Inset: percentage of the trials with (gray) and without (dotted line) 8-lick sequences for each bin, zoomed around the criterion time n=9 sessions.

Supplementary Figure 2. Rats also learn a task with lower and upper limits in the waiting time. Rats were trained to enter the nose-poke within a minimum waiting time of 2.5 s and a maximum waiting time of 5. **(a)** Reward rates for the 2.5-5 s waiting time task, *** $p=9.53 \times 10^{-7}$ One sample Wilcoxon Signed-rank test. . **(b)** Trial duration for correct trials, Paired t test *** $t_{(33)}=14.41$, $p=8.46 \times 10^{-16}$. **(c)** Latency to first lick for prematurely, timely and late initiated trials, One-way RM ANOVA NS, $F_{(1.21,\ 24.21)}=3.36$, $p=0.072$. **(d)** Time to complete the 8-lick sequence for the different types of trials, One-way RM ANOVA $F_{(1.29,25.87)}= 0.45$, $p=0.559$. **(e)** Variation coefficient of the 8-lick sequence inter-lick intervals for the different types of trials, One-way RM ANOVA NS, $p=0.6407$. **(f)** Normalized frequency distributions of the trial initiation times (waiting time), separated for trials with 8-lick sequences in the third (dark blue) and the seventh (light blue) sessions of training with a 2.5 s waiting time. Inset: area below the curve calculated between 2.5 s and 5 s for sessions 3 and 4 vs. sessions 6 and 7, Mann- Whitney test, NS, $p=0.8857$. **(g)** Normalized frequency distributions of the trial initiation times (waiting time), separated for trials with 8-lick sequences in the third (lilac) and the seventh (purple) sessions of training with a 2.5-5 s waiting time. Inset: area below the curve calculated between 2.5 s and 5 s for sessions 3 and 4 vs. sessions 6 and 7, Mann- Whitney test * $p= 0.024$.

Supplementary Figure 3. Characteristics of the neurons registered in the 2.5 s waiting time task. (a) Average firing rate of striatal neurons showing tonic activity during the waiting period, aligned to port entry or port exit, for different waiting times. The color code for the normalized activity (in units of standard deviation) is shown on the left. **(b)** Proportion of neurons with different responses in both training stages from animals trained with a 2.5 s criterion time.

Supplementary Figure 4. Anticipatory activity in the 5 s waiting time task. (a) Average firing rates of striatal neurons (of rats trained with a 5 s criterion time) showing activity preceding the initiation of the trials according to the waiting time duration. Color code for the intervals shown on the right. **(b)** Mean of the normalized firing rate for each of the waiting time groups, calculated between -640 and -80 ms before the entrance vs. waiting time, criterion time is shown with a blue hatched-line.

Supplementary Figure 5. Neuronal activity at the moment of the visual cue. Population PETH and Average firing rates of striatal units showing positive firing rate modulations related with the visual cue. The color code for the mean normalized FR is shown on the right. The bar graphs show the statistical comparisons between trial conditions. One-way RM ANOVA, $F_{(1.750, 45.49)}$=17.80, *** p=1.17 x10^{-4}; Tukey's multiple comparisons test $q_{(26)}$=6.64, *** p=2.15 x10^{-4}; Premature vs. Rewarded and, $q_{(26)}$=6.79, p=1.62 x10^{-4} Premature vs. Unrewarded.

Supplementary Figure 6. Adolescent rats make more premature trials. (a) Trial duration for correct trials at each training stage, Two-way RM ANOVA, Training stage, $F_{(1, 19)}$= 91.81, *** p=1.04 x10^{-8}; Type of trial, $F_{(1, 19)}$=41.67, *** p=3.46 x10^{-6}; Interaction Training stage x Type of trial, $F_{(1, 19)}$=2.63, p=0.121. **(b)** Latency to the first lick during correct trials, Two-way RM ANOVA, Training stage, $F_{(1, 19)}$= 26.74, *** p=5.43 x10^{-5}, Type of trial, $F_{(1, 19)}$=0.68, p=0.418, Interaction Training stage x Type of trial, $F_{(1, 19)}$=0.01, p=0.905. **(c)** Time to complete the 8-lick sequence, Two-way RM ANOVA, Training stage, $F_{(1, 19)}$=0.68, *** p=2.12 x10^{-7}; Type of trial, $F_{(1, 19)}$= p=0.024, Interaction Training stage x Type of trial p=7.65 x10^{-3}. **(d)** Correlation between prematurely initiated trials and reward rate, Linear regression Y=-0.01307*X + 1.996, Slope significantly different from zero: p=0.131, R square=0.026, $F_{(1,26)}$=2.33. **(e)** Latency to first

lick for prematurely and timely initiated trials, at early and late learning stages, Two-way RM ANOVA, Training stage, $F_{(1,19)}$=16.17, *** p=7.30 x10^{-4}; Type of trial, $F_{(1, 19)}$=3.81, p=0.066.; Interaction Training stage x Type of trial, $F_{(1, 19)}$=0.11, p=0.745. **(f)** Time to complete the 8-lick sequence, Two-way RM ANOVA, Training stage $F_{(1, 19)}$=43.68, *** p=2.52 x10^{-6}, Type of trial, $F_{(1, 19)}$=1.53, p=0.231; Interaction Training stage x Type of trial, $F_{(1, 19)}$=1.54, p=0.230. **(g)** Coefficient of variation of the 8-lick sequence inter-lick intervals for prematurely and timely initiated trials, at early and late learning stages, Two-way RM ANOVA, Training stage, $F_{(1, 19)}$=26.68, ** p=5.50 x10^{-5}; Type of trial, $F_{(1, 19)}$=0.21, p=0.649, Interaction Training stage x Type of trial, $F_{(1, 19)}$=38.55, p=5.78 x10^{-6}. Data are expressed as mean ± SEM, n=20 sessions. **(h)** Raster plot of the licks from 100 trials at a late training session of one of the adult rats. Colored dots show the first 8 licks of the trial (timely trials, blue and premature trials, red).

Supplementary Figure 7. Striatal activity also marks transitions between behavioral states of the task in adolescent animals. (a) Individual PETHs of all the recorded striatal neurons from adolescent rats during correct trials, aligned to port entry (left) or port port exit (right), with their corresponding average PETH (below). The color code for the normalized activity is shown on the right. Red dashed line: criterion time, vertical green line: time of led on. **(b)** Proportion of neurons showing task-related firing rate modulations. **(c-f)** Population PETH and average firing rates aligned to port entry (top panels) and port exit (bottom panels) for **(c)** striatal neurons showing only port entry related activity, **(d)** striatal neurons showing only port exit related activity, **(e)** striatal neurons showing activity modulations at both port entry and port exit and **(f)** striatal neurons showing higher activity during licking behavior. From **(c)** to **(f)**, it is represented the mean (solid lines) and SE (shaded area). Colored bars over the x axis show the interval used to detect firing rate modulations (red: entry, orange: exit, purple: active during the task). The color code for the mean normalized FR is shown on the right **(a)**.

Supplementary Figure 8. Anticipatory, reward-responsive and exit activity in striatal neurons of adolescent rats. (a) Time from port entry of movement initiation for timely and premature trials. $t_{(9)}$=1.54, p=0.159**. (b)** Population PETH and Average firing rates of striatal units showing positive firing rate modulations during reward delivery, reward omission and premature trials. **(c)** Bar graphs showing statistical comparisons between trial conditions. Top: Reward-responsive neurons, Rewarded vs. Unrewarded trials, *** $q_{(28)}$=10.76, p=1.86 $x10^{-11}$; Rewarded vs. Premature trials, *** $q_{(28)}$=13.60, p=7.33 $x10^{-14}$; Unrewarded vs. Premature trials * $q_{(28)}$=3.22, p=0.003, Tukey test after significant one-way RM ANOVA (*** $F_{(1.99,\ 55.62)}$=50.81, p=2.33 $x10^{-9}$). Middle: No reward-responsive neurons, Rewarded vs. Unrewarded trials *** $q_{(25)}$=10.24, p=1.96 $x10^{-10}$, Rewarded vs. Premature trials *** $q_{(25)}$=4.46, p=1.52 $x10^{-4}$, Unrewarded vs. Premature trials * $q_{(25)}$=3.28, p=0.003, Tukey test after significant one-way RM ANOVA (*** $F_{(1.57,39.27)}$=18.33, p=1.17 $x10^{-4}$). Bottom: Expected no reward-responsive neurons, Rewarded vs. Unrewarded trials * $q_{(31)}$=4.19, p=2.14 $x10^{-4}$, *** Rewarded vs. Premature trials $q_{(31)}$=13.49, p=1.63 $x10^{-14}$, Unrewarded vs. Premature trials *** $q_{(31)}$=8.78, p=6.58 $x10^{-10}$, Tukey test after significant one-way RM ANOVA (*** $F_{(1.95,60.58)}$=43.19, p=1.34 $x10^{-8}$). **(d)** Discrimination index for each of the groups of neurons shown in b-d, Rewarded vs. Unrewarded *$q_{(25)}$=0.51, Adjusted p value=0.128, Rewarded vs. Premature $q_{(28)}$=2.84, Adjusted p value=0.128, Unrewarded vs. Premature $q_{(25)}$=2.92, Adjusted p value= 0.118, Tukey test after significant Mixed-effects model (REML), * $F_{(1.93,51.19)}$=3.34, p=0.045. **(e)** PETH showing lick rates during the three trial conditions, centered at the 8th lick, with its corresponding average and statistical analysis at the bottom. The color code for the normalized activity is shown on the right. **(f)** Average firing rate of neurons responding to port exit, during premature and timely trials, for trials with an 8-lick sequence and trials where less than 8 licks were emitted. **(g)** Average striatal activity corresponding to the PETH shown in g, Two-way RM ANOVA, Licks' sequence *** $F_{(1,255)}$=116.04, p=1.51 $x10^{-22}$, Training $F_{(1,255)}$=1.98, p=0.160,*** Interaction Sequence x Training $F_{(1,255)}$=27.57, p=3.19 $x10^{-7}$.

Supplementary Figure 9: Behavioral performance of non-implanted adolescent and adult rats. (a) Percentage of the different trial types per session of adolescent rats (n =6). **(b)** Reward rates for each training stage of adolescent rats, Wilcoxon matched pairs signed-rank test, *** $p=6.12 \times 10^{-26}$, n=16 pairs. **(c)** Proportion of trials followed by 8-lick sequences at early and late learning stages of adolescent rats, *** Two way RM ANOVA, $F_{(1,15)}=68.02$, $p= 5.91 \times 10^{-7}$, n=16 sessions. **(d)** Correlation between percentage of prematurely initiated trials followed by an 8-lick sequence and reward rate in adolescent rats (Linear regression Y=0.038*X -0.554, Slope significantly different from zero, $F_{(1,90)}=82.88$, $p=2.51 \times 10^{-14}$, R square=0.4850, n=92 sessions). **(e)** Percentage of the different trial types per session of adult rats (n =6). **(f)** Reward rates for each training stage of adult rats, * p=0.0156, n=7 pairs. **(g)** Proportion of trials followed by 8-lick sequences at early and late learning stages of adult rats, *** Two way RM ANOVA $F_{(1,6)}=9.81$ p=0.020, n=7 sessions. **(h)** Correlation between percentage of prematurely initiated trials followed by an 8-lick sequence and reward rate in adult rats (Linear regression Y=0.037*X -0.265, Slope significantly different from zero, $F_{(1,69)}=205.78$, $p=5.88 \times 10^{-22}$, R square=0.749). **(i)** Normalized frequency distributions of the trial initiation times (waiting time), separated for trials with (red line: adolescents, black line: adults) and without (red dotted line: adolescents, black dotted line: adults) 8-lick sequenc0es, in the early and late stages of training. **(j)** Ratio between the number of premature trials made and the obtained rewards for adolescents and adults, Two-way RM ANOVA, *** effect of training $F_{(1, 21)}=16.33$, $p=6.00 \times 10^{-4}$, ** effect of age $F_{(1, 21)}= 8.43$, p=0.008, Interaction $F_{(1, 21)}=0.91$, p=0.350. **(k)** Premature trials with 8-lick sequences made per obtained rewards for adolescents and adults. Two-way RM ANOVA effect of training $F_{(1, 21)}= 3.75$, p=0.066, * effect of age $F_{(1, 21)}= 7.46$, p=0.012, Interaction $F_{(1, 21)}=2.96$, p=0.099. **(l)** Accumulated normalized trials vs. Waiting time for adolescents. Light color lines show each individual

session. Hatched lines show the logistic function fitted to the behavioral data. Inset: detail around 2.5 sec waiting time. p2 coefficient of the logistic function, profile likelihood estimation of 95% confidence interval: Adolescents mean=1.262, (95% CI: 1.223 to 1.304), and Adults mean=0.8490, (95% CI: 0.8032 to 0.8984), $F_{(1, 5652)}$=170.3, p=2.71 x10^{-38}).

Supplementary Figure 10. Localization of recording tetrodes. Schematic representation of the electrodes' placement for each of the animals, based on the electrode tips visible in the histological analysis, adapted from Paxinos' Atlas (Paxinos and Watson, 2007). Animals with incorrect electrode placement (gray squares) were excluded from the electrophysiological analysis but not from the behavioral analysis.

References

Addicott, M.A., Pearson, J.M., Sweitzer, M.M., Barack, D.L., and Platt, M.L. (2017). A Primer on Foraging and the Explore/Exploit Trade-Off for Psychiatry Research. Neuropsychopharmacology *42*, 1931–1939. https://doi.org/10.1038/npp.2017.108.

Andrzejewski, M.E., Schochet, T.L., Feit, E.C., Harris, R., McKee, B.L., and Kelley, A.E. (2011). A comparison of adult and adolescent rat behavior in operant learning, extinction, and behavioral inhibition paradigms. Behav. Neurosci. *125*, 93–105. https://doi.org/10.1037/a0022038.

Asaad, W.F., Lauro, P.M., Perge, J.A., and Eskandar, E.N. (2017). Prefrontal Neurons Encode a Solution to the Credit-Assignment Problem. J. Neurosci. *37*, 6995–7007. https://doi.org/10.1523/JNEUROSCI.3311-16.2017.

Atallah, H.E., McCool, A.D., Howe, M.W., and Graybiel, A.M. (2014). Neurons in the ventral striatum exhibit cell-type-specific representations of outcome during learning. Neuron *82*, 1145–1156. https://doi.org/10.1016/j.neuron.2014.04.021.

Bakhurin, K.I., Goudar, V., Shobe, J.L., Claar, L.D., Buonomano, D. V., and Masmanidis, S.C. (2017). Differential Encoding of Time by Prefrontal and Striatal Network Dynamics. J. Neurosci. *37*, 854–870. https://doi.org/10.1523/JNEUROSCI.1789-16.2016.

Balleine, B.W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology *37*, 407–419. https://doi.org/10.1016/S0028-3908(98)00033-1.

Bari, A., and Robbins, T.W. (2013). Inhibition and impulsivity: behavioral and neural basis of response control. Prog. Neurobiol. *108*, 44–79. https://doi.org/10.1016/j.pneurobio.2013.06.005.

Barnes, T.D., Kubota, Y., Hu, D., Jin, D.Z., and Graybiel, A.M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. Nature *437*, 1158–1161. https://doi.org/10.1038/nature04053.

Baumann, A.A., and Odum, A.L. (2012). Impulsivity, risk taking, and timing. Behav. Processes *90*, 408–414. https://doi.org/10.1016/j.beproc.2012.04.005.

Belin, D., and Everitt, B.J. (2008). Cocaine seeking habits depend upon dopamine-dependent serial connectivity linking the ventral with the dorsal striatum. Neuron *57*, 432–441. https://doi.org/10.1016/j.neuron.2007.12.019.

Bonner, A.C., and Borrero, J.C. (2018). Differential Reinforcement of Low Rate Schedules Reduce Severe Problem Behavior. Behav. Modif. *42*, 747–764. https://doi.org/10.1177/0145445517731723.

Buckholtz, J.W., Treadway, M.T., Cowan, R.L., Woodward, N.D., Benning, S.D., Li, R., Ansari, M.S., Baldwin, R.M., Schwartzman, A.N., Shelby, E.S., et al. (2010). Mesolimbic dopamine reward system hypersensitivity in individuals with psychopathic traits. Nat. Neurosci. *13*, 419–421. https://doi.org/10.1038/nn.2510.

Buhusi, C. V, and Meck, W.H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. Nat. Rev. Neurosci. *6*, 755–765. https://doi.org/10.1038/nrn1764.

Carland, M.A., Thura, D., and Cisek, P. (2019). The Urge to Decide and Act: Implications for Brain Function and Dysfunction. Neuroscientist *25*, 491–511. https://doi.org/10.1177/1073858419841553.

Costa, R.M. (2011). A selectionist account of de novo action learning. Curr. Opin. Neurobiol. *21*, 579–586. https://doi.org/10.1016/j.conb.2011.05.004.

Dalley, J.W., and Robbins, T.W. (2017). Fractionating impulsivity: neuropsychiatric implications. Nat. Rev. Neurosci. *18*, 158–171. https://doi.org/10.1038/nrn.2017.8.

Daw, N.D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. *8*, 1704–1711. https://doi.org/10.1038/nn1560.

Donnelly, N.A., Holtzman, T., Rich, P.D., Nevado-Holgado, A.J., Fernando, A.B.P., Van Dijck, G., Holzhammer, T., Paul, O., Ruther, P., Paulsen, O., et al. (2014). Oscillatory activity in the medial prefrontal cortex and nucleus accumbens correlates with impulsivity and reward outcome. PLoS One *9*, e111300. https://doi.org/10.1371/journal.pone.0111300.

Emmons, E.B., De Corte, B.J., Kim, Y., Parker, K.L., Matell, M.S., and Narayanan, N.S. (2017). Rodent Medial Frontal Control of Temporal Processing in the Dorsomedial Striatum. J. Neurosci. *37*, 8718–8733. https://doi.org/10.1523/JNEUROSCI.1376-17.2017.

Engeln, M., Ansquer, S., Dugast, E., Bezard, E., Belin, D., and Fernagut, P.-O. (2016). Multi-facetted impulsivity following nigral degeneration and dopamine replacement therapy. Neuropharmacology *109*, 69–77. https://doi.org/10.1016/j.neuropharm.2016.05.013.

Ersche, K.D., Ward, L.H.E., Lim, T.-V., Lumsden, R.J., Sawiak, S.J., Robbins, T.W., and Stochl, J. (2019). Impulsivity and compulsivity are differentially associated with automaticity and routine on the Creature of Habit Scale. Pers. Individ. Dif. *150*, 109493. https://doi.org/10.1016/j.paid.2019.07.003.

Evenden, J.L. (1998). The pharmacology of impulsive behaviour in rats III: the effects of amphetamine, haloperidol, imipramine, chlordiazepoxide and ethanol on a paced fixed consecutive number schedule. Psychopharmacology (Berl). *138*, 295–304. https://doi.org/10.1007/s002130050674.

Everitt, B.J., Belin, D., Economidou, D., Pelloux, Y., Dalley, J.W., and Robbins, T.W. (2008). Review. Neural mechanisms underlying the vulnerability to develop compulsive drug-seeking habits and addiction. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *363*, 3125–3135. https://doi.org/10.1098/rstb.2008.0089.

Floresco, S.B., St Onge, J.R., Ghods-Sharifi, S., and Winstanley, C.A. (2008). Cortico-limbic-striatal circuits subserving different forms of cost-benefit decision making. Cogn. Affect. Behav. Neurosci. *8*, 375–389. https://doi.org/10.3758/CABN.8.4.375.

Geddes, C.E., Li, H., and Jin, X. (2018). Optogenetic Editing Reveals the Hierarchical Organization of Learned Action Sequences. Cell *174*, 32-43.e15. https://doi.org/10.1016/j.cell.2018.06.012.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. Psychol. Rev. *84*, 279–325. https://doi.org/10.1037/0033-295X.84.3.279.

Gillan, C.M., Robbins, T.W., Sahakian, B.J., van den Heuvel, O.A., and van Wingen, G. (2016a). The role of habit in compulsivity. Eur. Neuropsychopharmacol. *26*, 828–840. https://doi.org/10.1016/j.euroneuro.2015.12.033.

Gillan, C.M., Kosinski, M., Whelan, R., Phelps, E.A., and Daw, N.D. (2016b). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. Elife *5*. https://doi.org/10.7554/eLife.11305.

Gouvêa, T.S., Monteiro, T., Motiwala, A., Soares, S., Machens, C., and Paton, J.J. (2015). Striatal dynamics explain duration judgments. Elife *4*. https://doi.org/10.7554/eLife.11386.

Graybiel, A.M. (1998). The Basal Ganglia and Chunking of Action Repertoires. Neurobiol. Learn. Mem. *70*, 119–136. https://doi.org/10.1006/nlme.1998.3843.

Graybiel, A.M. (2008). Habits, rituals, and the evaluative brain. Annu Rev Neurosci *31*, 359–387. .

Green, L., Myerson, J., and Ostaszewski, P. (1999). Discounting of delayed rewards across the life span: age differences in individual discounting functions. Behav. Processes *46*, 89–96. https://doi.org/10.1016/S0376-6357(99)00021-2.

Gremel, C.M., and Costa, R.M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. Nat. Commun. *4*, 2264. https://doi.org/10.1038/ncomms3264.

Hamid, A.A., Frank, M.J., and Moore, C.I. (2021). Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. Cell *184*, 2733-2749.e16. https://doi.org/10.1016/j.cell.2021.03.046.

Hazan, L., Zugaro, M., and Buzsáki, G. (2006). Klusters, NeuroScope, NDManager: A free software suite for neurophysiological data processing and visualization. J. Neurosci. Methods *155*, 207–216. https://doi.org/10.1016/j.jneumeth.2006.01.017.

Her, E.S., Huh, N., Kim, J., and Jung, M.W. (2016). Neuronal activity in dorsomedial and dorsolateral striatum under the requirement for temporal credit assignment. Sci. Rep. *6*, 27056. https://doi.org/10.1038/srep27056.

Hikosaka, O., Nakahara, H., Rand, M.K., Sakai, K., Lu, X., Nakamura, K., Miyachi, S., and Doya, K. (1999). Parallel neural networks for learning sequential procedures. Trends Neurosci. *22*, 464–471. https://doi.org/10.1016/S0166-2236(99)01439-3.

Histed, M.H., Pasupathy, A., and Miller, E.K. (2009). Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. Neuron *63*, 244–253. https://doi.org/10.1016/j.neuron.2009.06.019.

Hogarth, L., Chase, H.W., and Baess, K. (2012). Impaired goal-directed behavioural control in human impulsivity. Q. J. Exp. Psychol. *65*, 305–316. https://doi.org/10.1080/17470218.2010.518242.

Jin, X., and Costa, R.M. (2010). Start/stop signals emerge in nigrostriatal circuits during sequence learning. Nature *466*, 457–462. https://doi.org/10.1038/nature09263.

Jin, X., Tecuapetla, F., and Costa, R.M. (2014). Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences. Nat. Neurosci. *17*, 423–430. https://doi.org/10.1038/nn.3632.

Jog, M.S. (1999). Building Neural Representations of Habits. Science (80-.). *286*, 1745–1749. https://doi.org/10.1126/science.286.5445.1745.

Kamphuis, J., Baichel, S., Lancel, M., de Boer, S.F., Koolhaas, J.M., and Meerlo, P. (2017). Sleep restriction in rats leads to changes in operant behaviour indicative of reduced prefrontal cortex function. J. Sleep Res. *26*, 5–13. https://doi.org/10.1111/jsr.12455.

Kim, H., Sul, J.H., Huh, N., Lee, D., and Jung, M.W. (2009). Role of Striatum in Updating Values of Chosen Actions. J. Neurosci. *29*, 14701–14712. https://doi.org/10.1523/JNEUROSCI.2728-09.2009.

Knowlton, B.J., and Patterson, T.K. (2016). Habit Formation and the Striatum. Curr. Top. Behav. Neurosci. *37*, 275–295. https://doi.org/10.1007/7854_2016_451.

Kubota, Y., Liu, J., Hu, D., DeCoteau, W.E., Eden, U.T., Smith, A.C., and Graybiel, A.M. (2009). Stable encoding of task structure coexists with flexible coding of task events in sensorimotor striatum. J. Neurophysiol. *102*, 2142–2160. https://doi.org/10.1152/jn.00522.2009.

Kunimatsu, J., Suzuki, T.W., Ohmae, S., and Tanaka, M. (2018). Different contributions of preparatory activity in the basal ganglia and cerebellum for self-timing. Elife *7*. https://doi.org/10.7554/eLife.35676.

Lauwereyns, J., Watanabe, K., Coe, B., and Hikosaka, O. (2002). A neural correlate of response bias in monkey caudate nucleus. Nature *418*, 413–417. https://doi.org/10.1038/nature00892.

Lennox, D.B., Miltenberger, R.G., and Donnelly, D.R. (1987). Response interruption and DRL for the reduction of rapid eating. J. Appl. Behav. Anal. *20*, 279–284. https://doi.org/10.1901/jaba.1987.20-279.

Lovic, V., Keen, D., Fletcher, P.J., and Fleming, A.S. (2011). Early-life maternal separation and social isolation produce an increase in impulsive action but not impulsive choice. Behav. Neurosci. *125*, 481–491. https://doi.org/10.1037/a0024367.

Martiros, N., Burgess, A.A., and Graybiel, A.M. (2018). Inversely Active Striatal Projection Neurons and Interneurons Selectively Delimit Useful Behavioral Sequences. Curr. Biol. *28*, 560-573.e5. https://doi.org/10.1016/j.cub.2018.01.031.

Matell, M.S., Meck, W.H., and Nicolelis, M.A.L. (2003). Interval timing and the encoding of signal duration by ensembles of cortical and striatal neurons. Behav. Neurosci. *117*, 760–773. https://doi.org/10.1037/0735-7044.117.4.760.

Mello, G.B.M., Soares, S., and Paton, J.J. (2015). A scalable population code for time in the striatum. Curr. Biol. *25*, 1113–1122. https://doi.org/10.1016/j.cub.2015.02.036.

Monterosso, J., and Ainslie, G. (1999). Beyond discounting: possible experimental models of impulse control. Psychopharmacology (Berl). *146*, 339–347. https://doi.org/10.1007/pl00005480.

Namboodiri, V.M., and Hussain Shuler, M.G. (2016). The hunt for the perfect discounting function and a reckoning of time perception. Curr. Opin. Neurobiol. *40*, 135–141. https://doi.org/10.1016/j.conb.2016.06.019.

Namboodiri, V.M.K., Mihalas, S., Marton, T.M., and Hussain Shuler, M.G. (2014). A general theory of intertemporal decision-making and the perception of time. Front. Behav. Neurosci. *8*. https://doi.org/10.3389/fnbeh.2014.00061.

Neill, D.B. (1976). Frontal-striatal control of behavioral inhibition in the rat. Brain Res. *105*, 89–103. https://doi.org/10.1016/0006-8993(76)90925-2.

Nonomura, S., Nishizawa, K., Sakai, Y., Kawaguchi, Y., Kato, S., Uchigashima, M., Watanabe, M., Yamanaka, K., Enomoto, K., Chiken, S., et al. (2018). Monitoring and Updating of Action Selection for Goal-Directed Behavior through the Striatal Direct and Indirect Pathways. Neuron *99*, 1302-1314.e5. https://doi.org/10.1016/j.neuron.2018.08.002.

Paxinos, G., and Watson, C. (2007). The Rat Brain in Stereotaxic Coordinates, Sixth Edition (Academic Press).

Peterson, J.D., Wolf, M.E., and White, F.J. (2003). Impaired DRL 30 performance during amphetamine withdrawal. Behav. Brain Res. *143*, 101–108. https://doi.org/10.1016/s0166-4328(03)00035-4.

Robbins, T.W., and Costa, R.M. (2017). Habits. Curr. Biol. *27*, R1200–R1206. https://doi.org/10.1016/j.cub.2017.09.060.

Romer, D. (2010). Adolescent risk taking, impulsivity, and brain development: implications for prevention. Dev. Psychobiol. *52*, 263–276. https://doi.org/10.1002/dev.20442.

Rueda-Orozco, P.E., and Robbe, D. (2015). The striatum multiplexes contextual and kinematic information to constrain motor habits execution. Nat. Neurosci. *18*, 453–460. https://doi.org/10.1038/nn.3924.

Sales-Carbonell, C., Taouali, W., Khalki, L., Pasquet, M.O., Petit, L.F., Moreau, T., Rueda-Orozco, P.E., and Robbe, D. (2018). No Discrete Start/Stop Signals in the Dorsal Striatum of Mice Performing a Learned Action. Curr. Biol. *28*, 3044-3055.e5. https://doi.org/10.1016/j.cub.2018.07.038.

Samejima, K., Ueda, Y., Doya, K., and Kimura, M. (2005). Representation of action-specific reward values in the striatum. Science *310*, 1337–1340. https://doi.org/10.1126/science.1115270.

Scheres, A., Dijkstra, M., Ainslie, E., Balkan, J., Reynolds, B., Sonuga-Barke, E., and Castellanos, F.X. (2006). Temporal and probabilistic discounting of rewards in children and adolescents: effects of age and ADHD symptoms. Neuropsychologia *44*, 2092–2103. https://doi.org/10.1016/j.neuropsychologia.2005.10.012.

Schultz, W. (2015). Neuronal Reward and Decision Signals: From Theories to Data. Physiol. Rev. *95*, 853–951. https://doi.org/10.1152/physrev.00023.2014.

Shin, J.H., Kim, D., and Jung, M.W. (2018). Differential coding of reward and movement information in the dorsomedial striatal direct and indirect pathways. Nat. Commun. *9*, 404. https://doi.org/10.1038/s41467-017-02817-1.

Singer, H.S. (2016). Habitual and goal-directed behaviours and Tourette syndrome. Brain *139*, 312–316. https://doi.org/10.1093/brain/awv378.

Smith, K.S., and Graybiel, A.M. (2013). A dual operator view of habitual behavior reflecting cortical and striatal dynamics. Neuron *79*, 361–374. https://doi.org/10.1016/j.neuron.2013.05.038.

Spear, L.P. (2000). The adolescent brain and age-related behavioral manifestations. Neurosci Biobehav Rev *24*, 417–463. .

Steinberg, L., Graham, S., O'Brien, L., Woolard, J., Cauffman, E., and Banich, M. (2009). Age Differences in Future Orientation and Delay Discounting. Child Dev. *80*, 28–44. https://doi.org/10.1111/j.1467-8624.2008.01244.x.

Thorn, C.A., Atallah, H., Howe, M., and Graybiel, A.M. (2010). Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. Neuron *66*, 781–795. https://doi.org/10.1016/j.neuron.2010.04.036.

Thura, D., and Cisek, P. (2017). The Basal Ganglia Do Not Select Reach Targets but Control the Urgency of Commitment. Neuron *95*, 1160-1170.e5. https://doi.org/10.1016/j.neuron.2017.07.039.

Vandaele, Y., Mahajan, N.R., Ottenheimer, D.J., Richard, J.M., Mysore, S.P., and Janak, P.H. (2019). Distinct recruitment of dorsomedial and dorsolateral striatum erodes with extended training. Elife *8*. https://doi.org/10.7554/eLife.49536.

Vandecasteele, M., M, S., Royer, S., Belluscio, M., Berényi, A., Diba, K., Fujisawa, S., Grosmark, A., Mao, D., Mizuseki, K., et al. (2012). Large-scale recording of neurons by movable silicon probes in behaving rodents. J. Vis. Exp. e3568. https://doi.org/10.3791/3568.

Voon, V., Derbyshire, K., Rück, C., Irvine, M.A., Worbe, Y., Enander, J., Schreiber, L.R.N., Gillan, C., Fineberg, N.A., Sahakian, B.J., et al. (2015). Disorders of compulsivity: a common bias towards learning habits. Mol. Psychiatry *20*, 345–352. https://doi.org/10.1038/mp.2014.44.

Wang, A.Y., Miura, K., and Uchida, N. (2013). The dorsomedial striatum encodes net expected return, critical for energizing performance vigor. Nat. Neurosci. *16*, 639–647. https://doi.org/10.1038/nn.3377.

Wassum, K.M., Cely, I.C., Balleine, B.W., and Maidment, N.T. (2011). Micro-opioid receptor activation in the basolateral amygdala mediates the learning of increases but not decreases in the incentive value of a food reward. J. Neurosci. *31*, 1591–1599. https://doi.org/10.1523/JNEUROSCI.3102-10.2011.

Weafer, J., and de Wit, H. (2014). Sex differences in impulsive action and impulsive choice. Addict. Behav. *39*, 1573–1579. https://doi.org/10.1016/j.addbeh.2013.10.033.

Wittmann, M., and Paulus, M.P. (2008). Decision making, impulsivity and time perception. Trends Cogn. Sci. *12*, 7–12. https://doi.org/10.1016/j.tics.2007.10.004.

Wu, H., Miller, K.J., Blumenfeld, Z., Williams, N.R., Ravikumar, V.K., Lee, K.E., Kakusa, B., Sacchet, M.D., Wintermark, M., Christoffel, D.J., et al. (2018). Closing the loop on impulsivity via nucleus accumbens delta-band activity in mice and man. Proc. Natl. Acad. Sci. U. S. A. *115*, 192–197. https://doi.org/10.1073/pnas.1712214114.

Yamada, H., Inokawa, H., Matsumoto, N., Ueda, Y., and Kimura, M. (2011). Neuronal basis for evaluating selected action in the primate striatum. Eur. J. Neurosci. *34*, 489–506. https://doi.org/10.1111/j.1460-9568.2011.07771.x.

Yau, Y., Dadar, M., Taylor, M., Zeighami, Y., Fellows, L.K., Cisek, P., and Dagher, A. (2020). Neural Correlates of Evidence and Urgency During Human Perceptual Decision-Making in Dynamically Changing Conditions. Cereb. Cortex *30*, 5471–5483. https://doi.org/10.1093/cercor/bhaa129.

Yin, H.H., and Knowlton, B.J. (2006). The role of the basal ganglia in habit formation. Nat Rev Neurosci *7*, 464–476. https://doi.org/10.1038/nrn1919.

Yin, H.H., Knowlton, B.J., and Balleine, B.W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. Eur. J. Neurosci. *19*, 181–189. https://doi.org/10.1111/j.1460-9568.2004.03095.x.

Zhou, S., Masmanidis, S.C., and Buonomano, D. V. (2020). Neural Sequences as an Optimal Dynamical Regime for the Readout of Time. Neuron *108*, 651-658.e5. https://doi.org/10.1016/j.neuron.2020.08.020.

Zold, C.L., and Hussain Shuler, M.G. (2015). Theta Oscillations in Visual Cortex Emerge with Experience to Convey Expected Reward Time and Experienced Reward Rate. J. Neurosci. *35*, 9603–9614. .

Figure 1

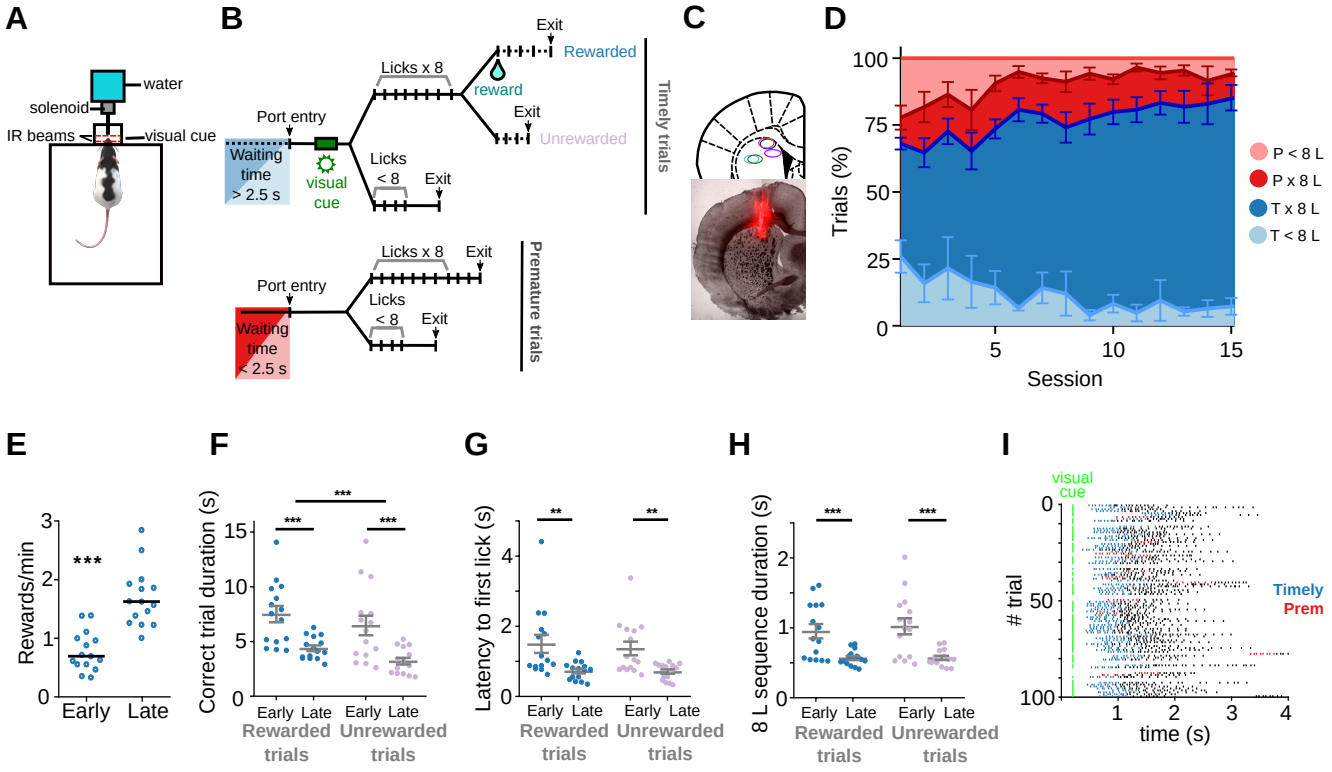

A
water
solenoid
IR beams
visual cue
B
Port entry
Waiting time > 2.5 s
visual cue
Licks x 8
reward
Exit
Rewarded
Licks < 8
Exit
Unrewarded
Timely trials
Port entry
Waiting time < 2.5 s
Licks x 8
Exit
Licks < 8
Exit
Premature trials
C
D
Trials (%)
Session
P < 8 L
P x 8 L
T x 8 L
T < 8 L
E
Rewards/min
Early
Late
F
Correct trial duration (s)
Early Late Early Late
Rewarded trials
Unrewarded trials
G
Latency to first lick (s)
Early Late Early Late
Rewarded trials
Unrewarded trials
H
8 L sequence duration (s)
Early Late Early Late
Rewarded trials
Unrewarded trials
I
visual cue
trial
time (s)
Timely
Prem

Figure 2

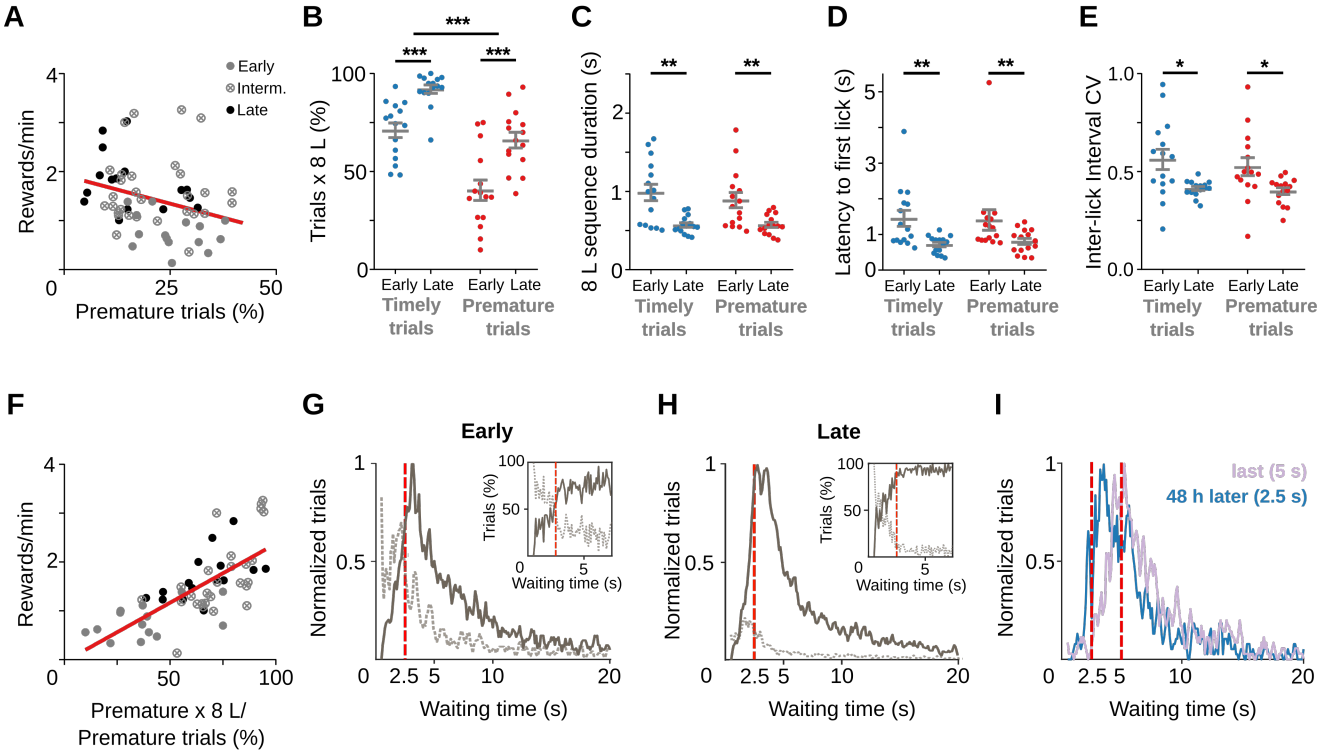

A
Early
Interm.
Late
Rewards/min
Premature trials (%)
B
Trials x 8 L (%)
Early Late Early Late
Timely trials
Premature trials
C
8 L sequence duration (s)
D
Latency to first lick (s)
E
Inter-lick Interval CV
F
Rewards/min
Premature x 8 L/
Premature trials (%)
G
Early
Normalized trials
Trials (%)
Waiting time (s)
H
Late
I
last (5 s)
48 h later (2.5 s)

Figure 3

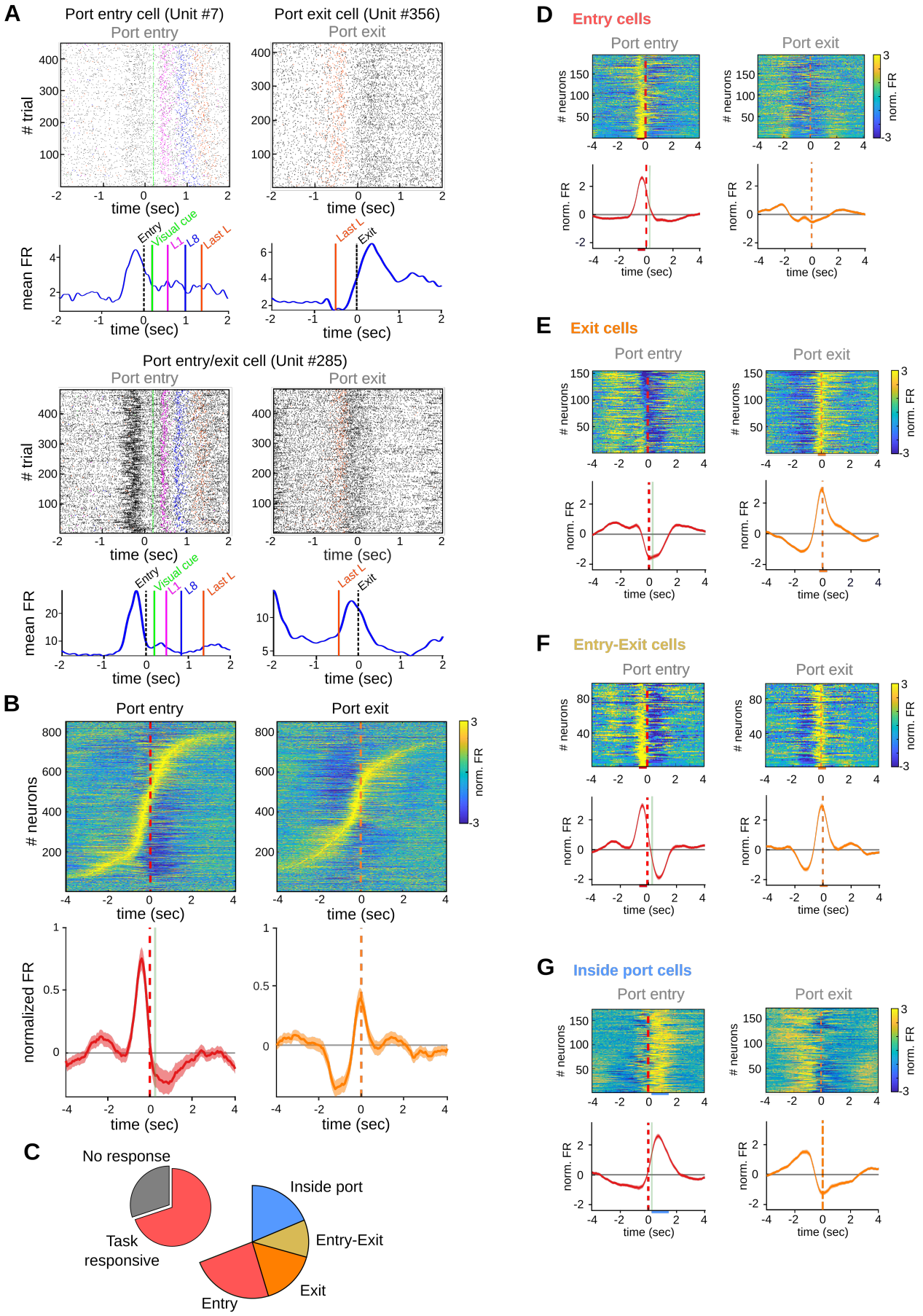

A
Port entry cell (Unit #7)
Port entry
Port exit
trial
time (sec)
mean FR
Entry
Visual cue
L1
L8
Last L
Exit
Port exit cell (Unit #356)
Port entry/exit cell (Unit #285)
B
Port entry
Port exit
neurons
norm. FR
normalized FR
C
No response
Task responsive
Inside port
Entry-Exit
Exit
Entry
D
Entry cells
E
Exit cells
F
Entry-Exit cells
G
Inside port cells

Figure 4

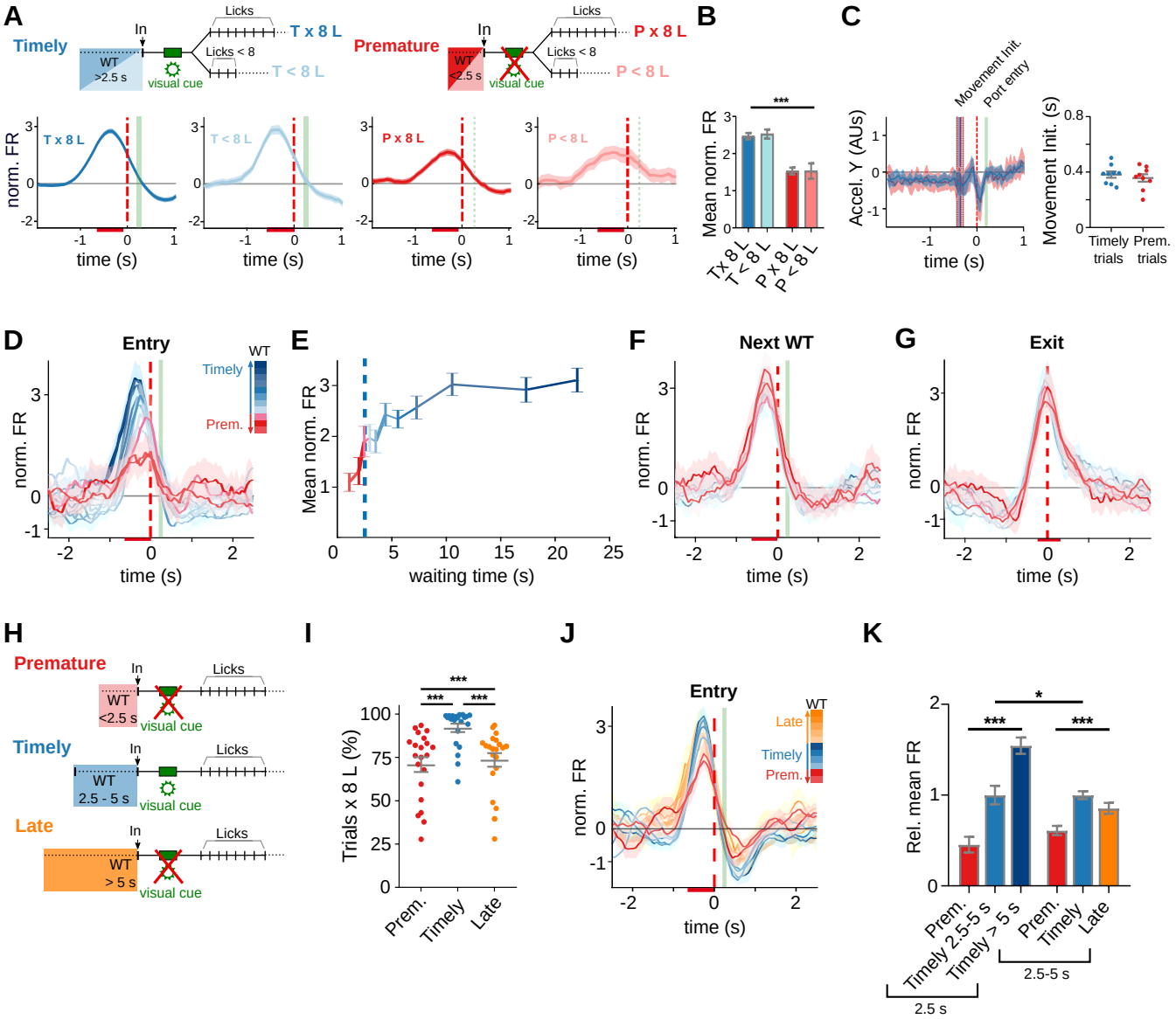

A
Timely
In
Licks
T x 8 L
WT
>2.5 s
Licks < 8
T < 8 L
visual cue
Premature
P x 8 L
P < 8 L
norm. FR
time (s)
B
Mean norm. FR
***
T x 8 L
T < 8 L
P x 8 L
P < 8 L
C
Movement Init.
Port entry
Accel. Y (AUs)
Movement Init. (s)
Timely trials
Prem. trials
D
Entry
WT
Timely
Prem.
E
Mean norm. FR
waiting time (s)
F
Next WT
G
Exit
H
Premature
<2.5 s
Timely
2.5 - 5 s
Late
> 5 s
I
Trials x 8 L (%)
Prem.
Timely
Late
J
Entry
Late
Timely
Prem.
K
Rel. mean FR
Prem.
Timely 2.5-5 s
Timely > 5 s
Prem.
Timely
Late
2.5-5 s
2.5 s

Figure 5

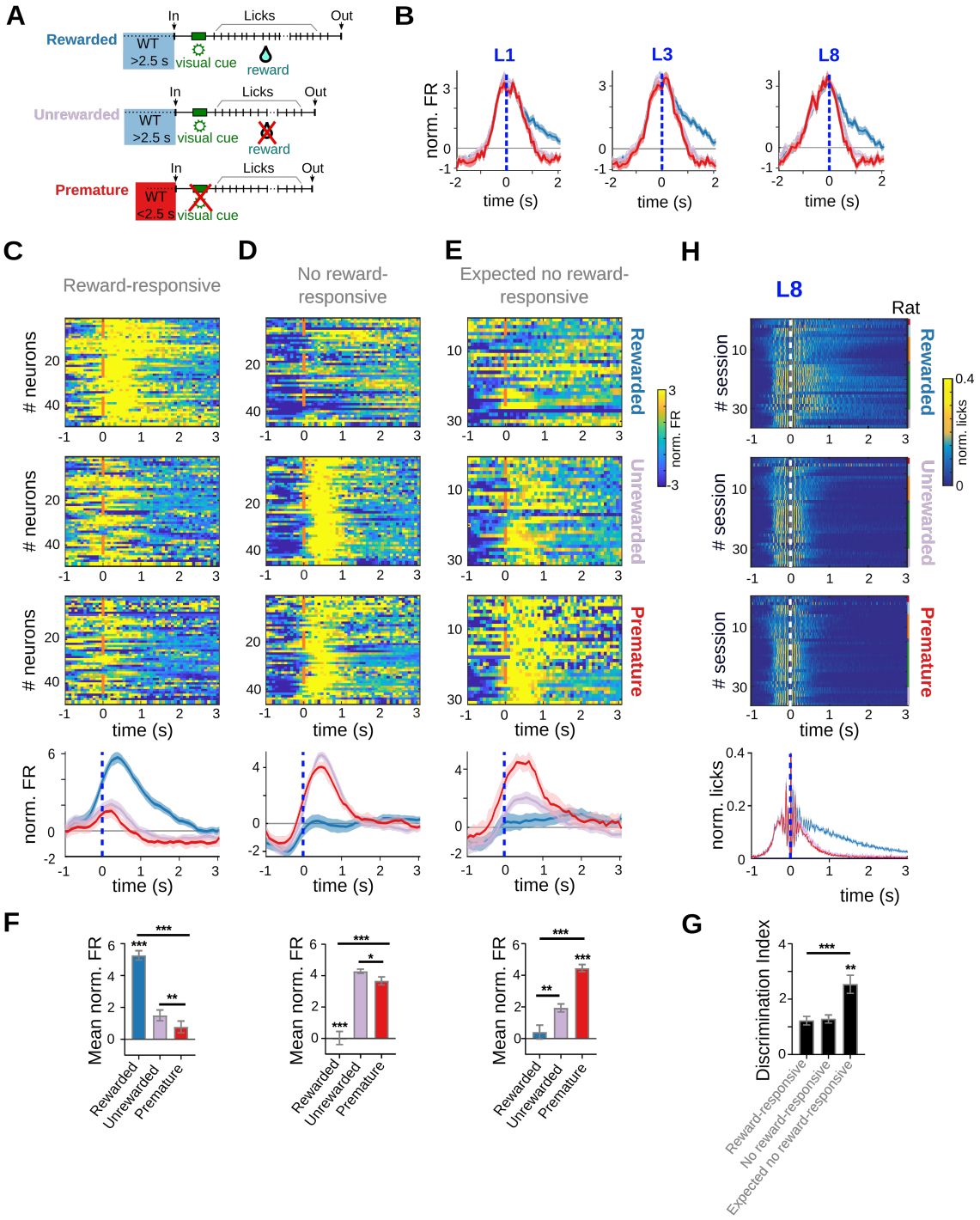

A
Rewarded
WT
>2.5 s
visual cue
reward
In
Licks
Out
Unrewarded
Premature
<2.5 s
B
L1
L3
L8
norm. FR
time (s)
C
D
E
H
Reward-responsive
No reward-responsive
Expected no reward-responsive
neurons
session
Rewarded
Unrewarded
Premature
norm. FR
norm. licks
Rat
F
G
Mean norm. FR
Discrimination Index
Reward-responsive
No reward-responsive
Expected no reward-responsive

Figure 6

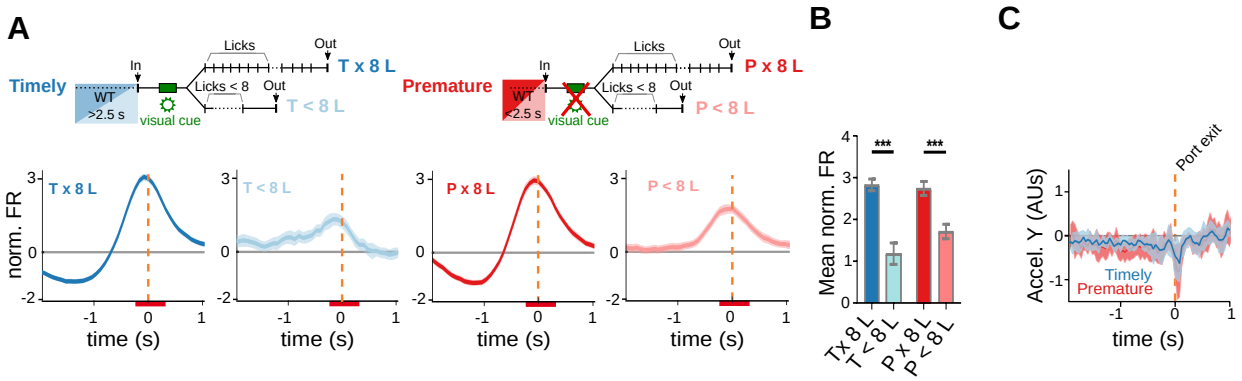

A
Timely
WT
>2.5 s
In
Licks
Out
T x 8 L
Licks < 8
Out
T < 8 L
visual cue
Premature
WT
<2.5 s
In
Licks
Out
P x 8 L
Licks < 8
Out
P < 8 L
visual cue
norm. FR
T x 8 L
T < 8 L
P x 8 L
P < 8 L
time (s)
B
Mean norm. FR
***
***
T x 8 L
T < 8 L
P x 8 L
P < 8 L
C
Accel. Y (AUs)
Port exit
Timely
Premature
time (s)

Figure 7

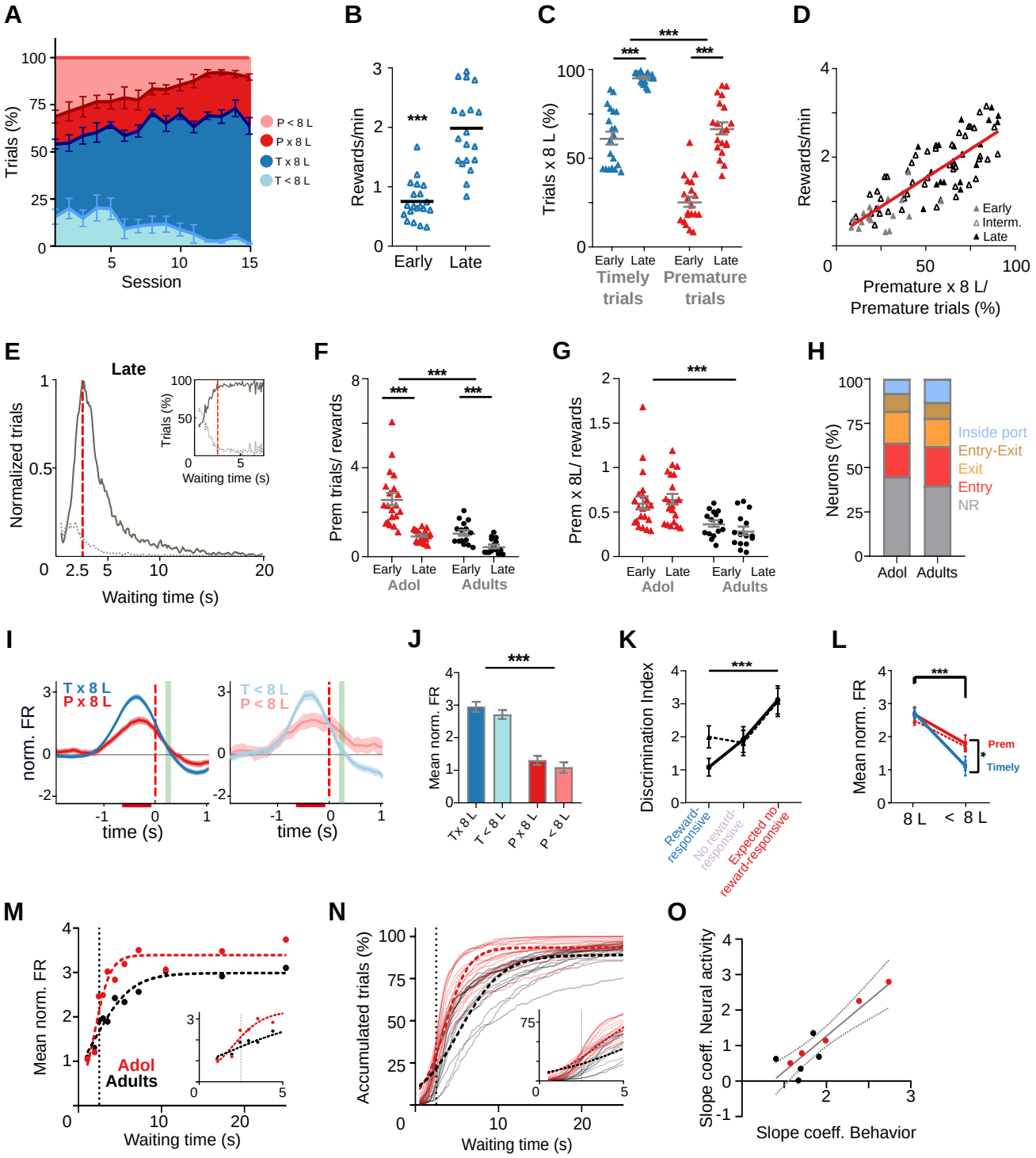

A
Trials (%)
Session
P < 8 L
P x 8 L
T x 8 L
T < 8 L
B
Rewards/min
Early
Late
C
Trials x 8 L (%)
Early Late Early Late
Timely trials
Premature trials
D
Rewards/min
Premature x 8 L/ Premature trials (%)
Early
Interm.
Late
E
Late
Normalized trials
Trials (%)
Waiting time (s)
F
Prem trials/ rewards
Early Late Early Late
Adol
Adults
G
Prem x 8L/ rewards
Early Late Early Late
Adol
Adults
H
Neurons (%)
Inside port
Entry-Exit
Exit
Entry
NR
Adol Adults
I
T x 8 L
P x 8 L
T < 8 L
P < 8 L
norm. FR
time (s)
J
Mean norm. FR
Tx 8 L
T < 8 L
P x 8 L
P < 8 L
K
Discrimination Index
Reward-responsive
No reward-responsive
Expected no reward-responsive
L
Mean norm. FR
8 L
< 8 L
Prem
Timely
M
Mean norm. FR
Adol
Adults
Waiting time (s)
N
Accumulated trials (%)
Waiting time (s)
O
Slope coeff. Neural activity
Slope coeff. Behavior

Figure S1

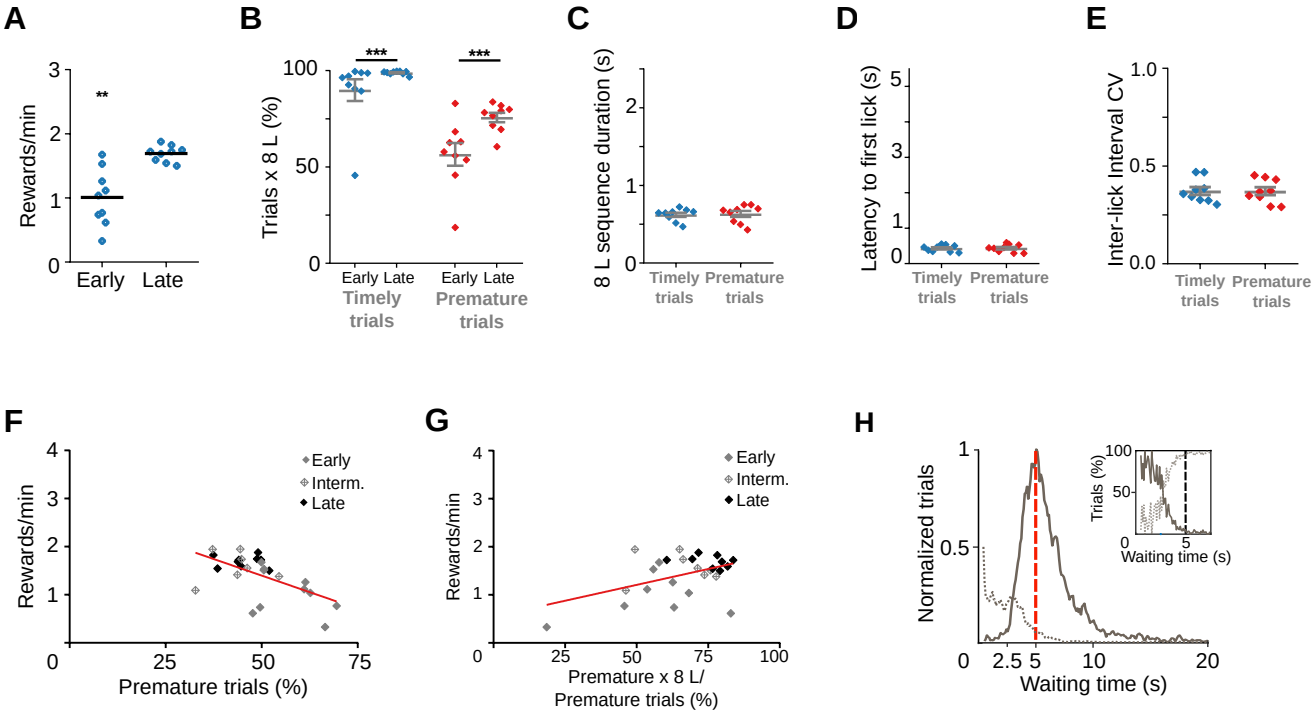

A
Rewards/min
Early
Late
B
Trials x 8 L (%)
Early Late
Timely trials
Early Late
Premature trials
C
8 L sequence duration (s)
Timely trials
Premature trials
D
Latency to first lick (s)
Timely trials
Premature trials
E
Inter-lick Interval CV
Timely trials
Premature trials
F
Rewards/min
Early
Interm.
Late
Premature trials (%)
G
Rewards/min
Early
Interm.
Late
Premature x 8 L/
Premature trials (%)
H
Normalized trials
Trials (%)
Waiting time (s)
Waiting time (s)

Figure S2

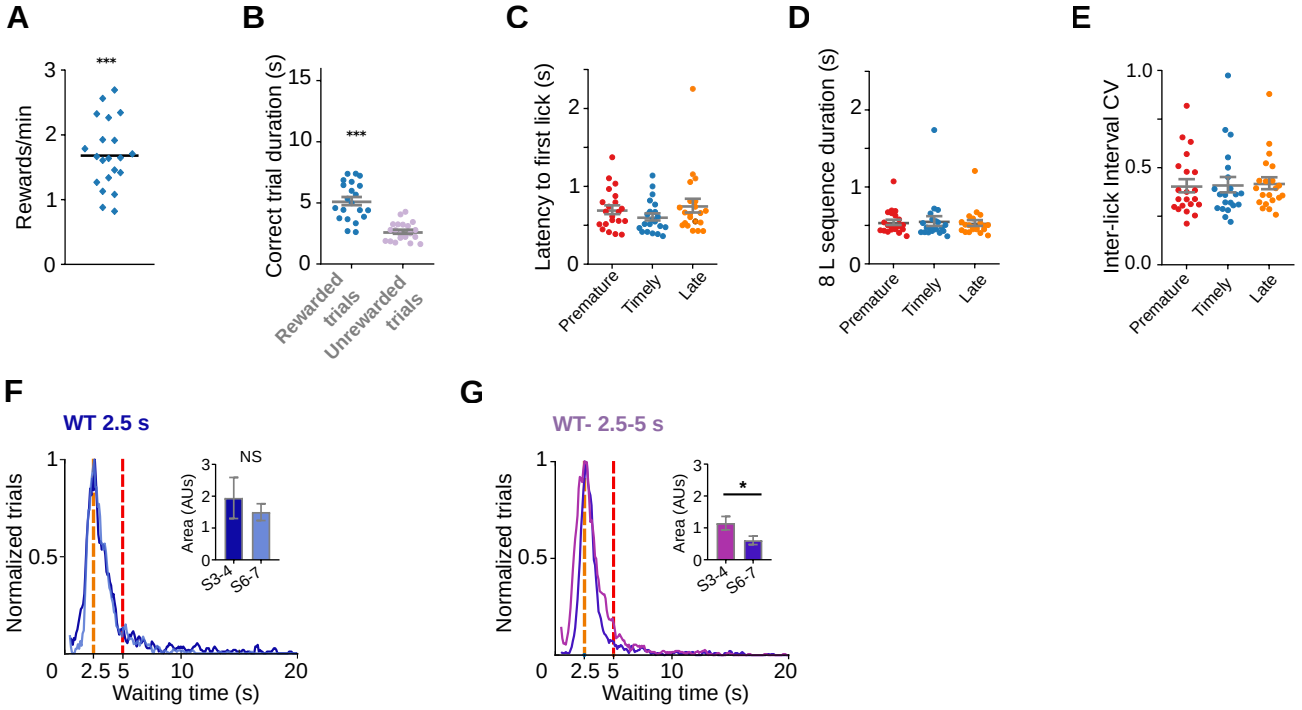
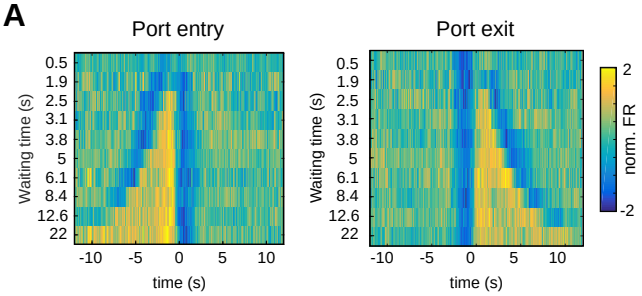

A
Rewards/min
***
B
Correct trial duration (s)
***
Rewarded trials
Unrewarded trials
C
Latency to first lick (s)
Premature
Timely
Late
D
8 L sequence duration (s)
Premature
Timely
Late
E
Inter-lick Interval CV
Premature
Timely
Late
F
WT 2.5 s
Normalized trials
Waiting time (s)
Area (AUs)
NS
S3-4
S6-7
G
WT- 2.5-5 s
Normalized trials
Waiting time (s)
Area (AUs)
*
S3-4
S6-7

Figure S3

A


Port entry
Port exit
Waiting time (s)
0.5
1.9
2.5
3.1
3.8
5
6.1
8.4
12.6
22
-10
-5
0
5
10
time (s)
2
-2
norm. FR

B

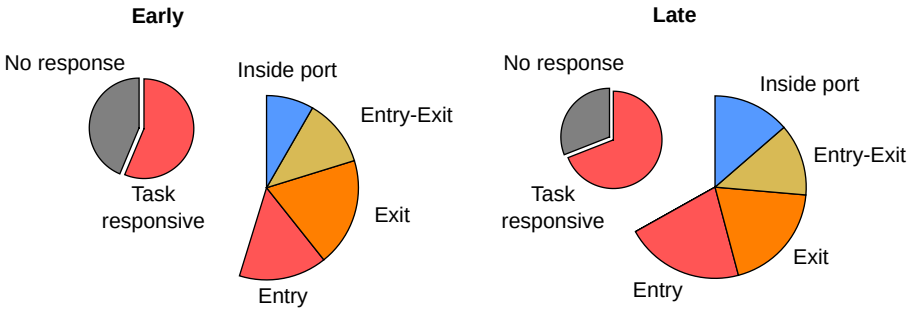
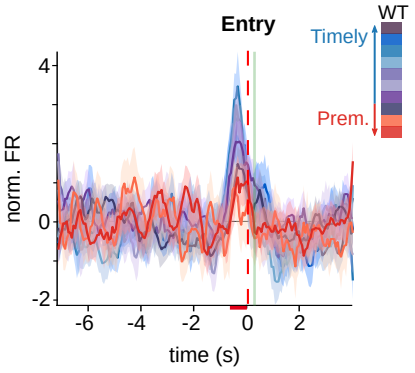

Early
Late
No response
Task responsive
Inside port
Entry-Exit
Exit
Entry

Figure S4

A


Entry
WT
Timely
Prem.
norm. FR
4
0
-2
-6
-4
-2
0
2
time (s)

B

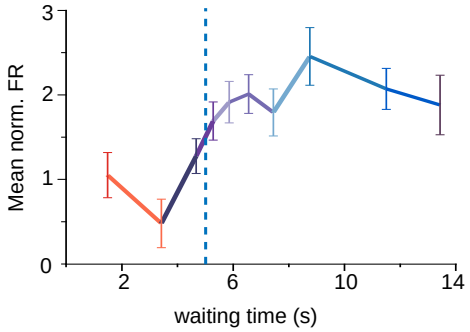

Mean norm. FR
3
2
1
0
2
6
10
14
waiting time (s)

Figure S5

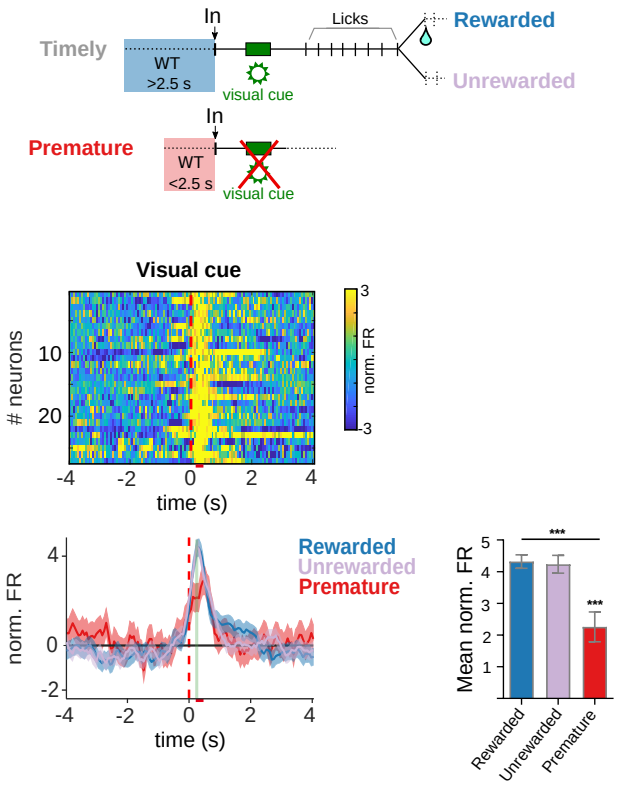

Timely
In
WT
>2.5 s
visual cue
Licks
Rewarded
Unrewarded
Premature
In
WT
<2.5 s
visual cue
Visual cue
neurons
10
20
-4
-2
0
2
4
time (s)
3
norm. FR
-3
norm. FR
4
0
-2
-4
-2
0
2
4
time (s)
Rewarded
Unrewarded
Premature
Mean norm. FR
5
4
3
2
1
***
***
Rewarded
Unrewarded
Premature

Figure S6

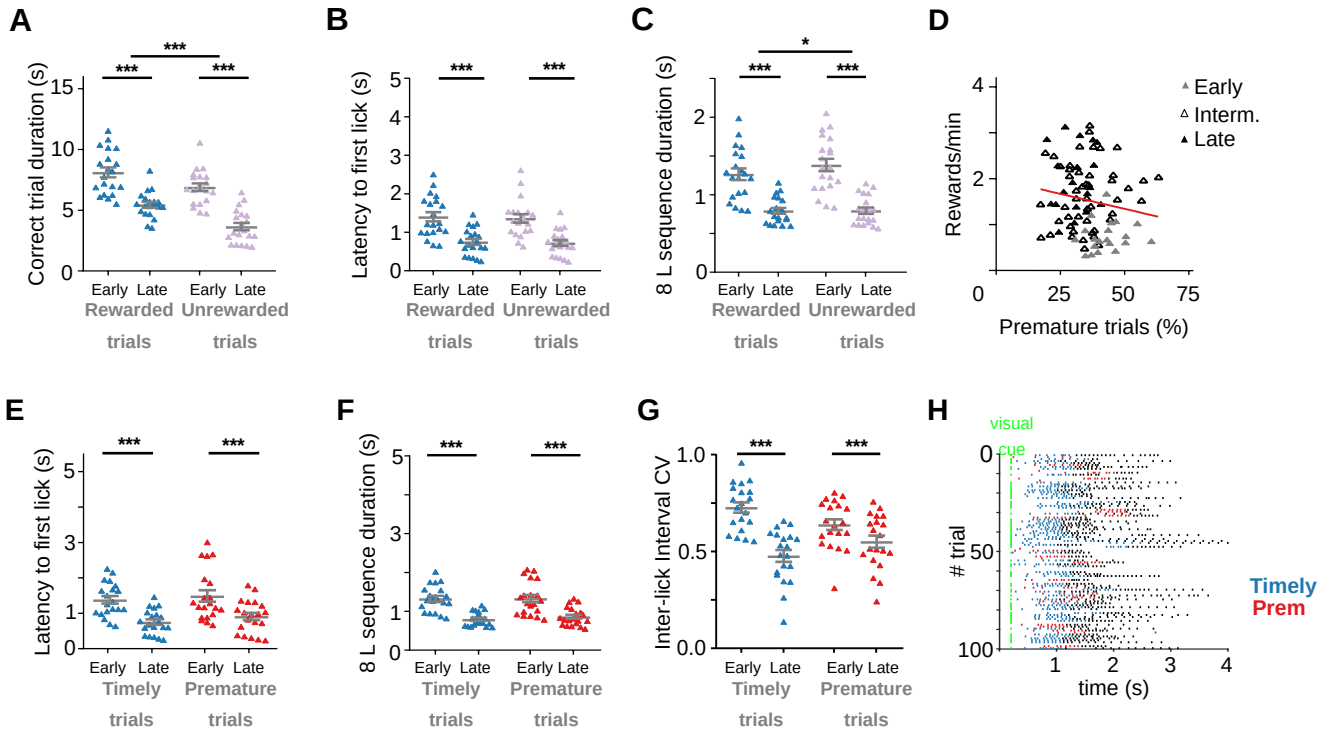

A
Correct trial duration (s)
Early Late Early Late
Rewarded trials
Unrewarded trials
B
Latency to first lick (s)
C
8 L sequence duration (s)
D
Rewards/min
Premature trials (%)
Early
Interm.
Late
E
Latency to first lick (s)
Timely trials
Premature trials
F
8 L sequence duration (s)
G
Inter-lick Interval CV
H
visual cue
trial
time (s)
Timely
Prem

Figure S7

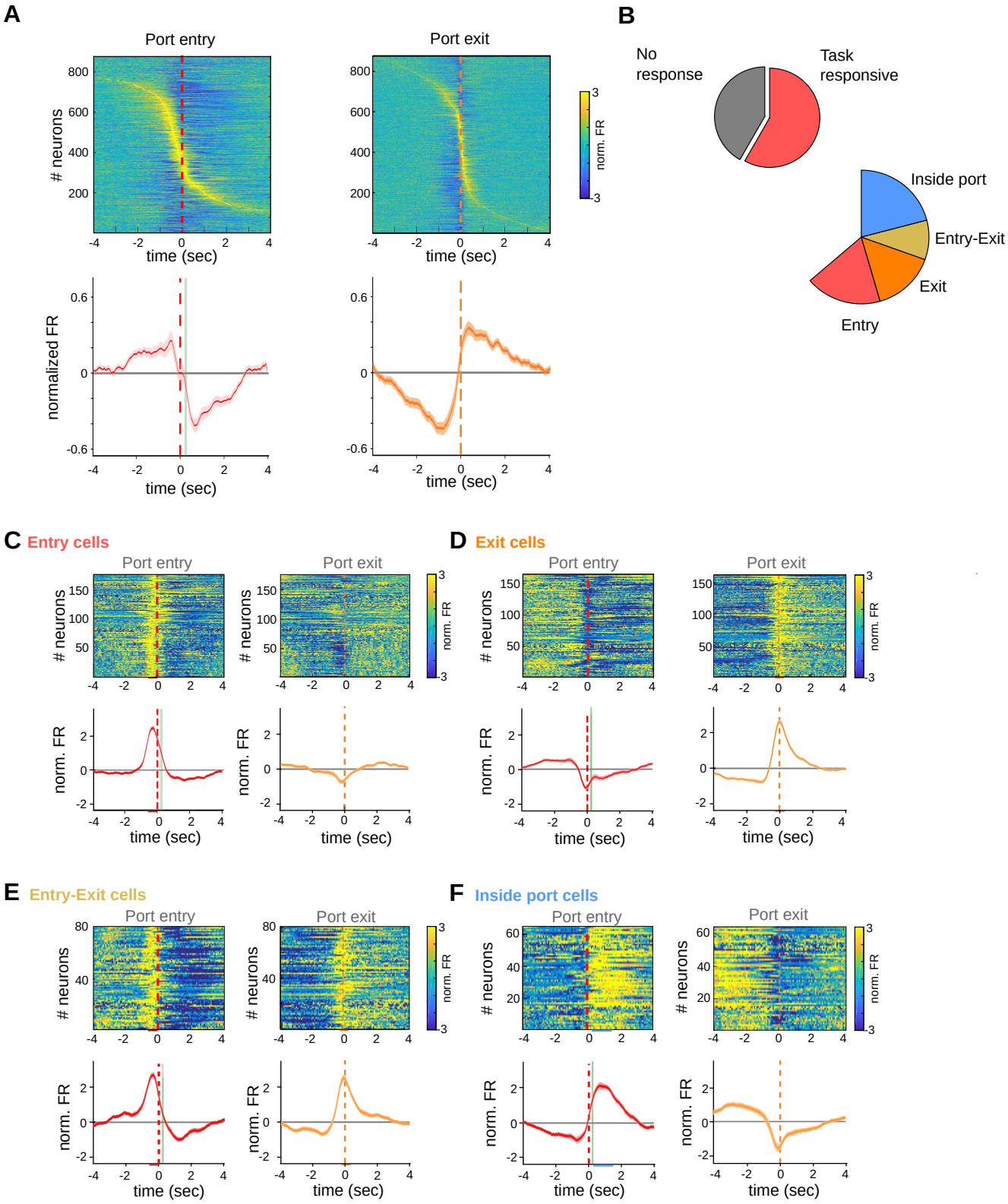

A
Port entry
Port exit
neurons
norm. FR
time (sec)
normalized FR
B
No response
Task responsive
Inside port
Entry-Exit
Exit
Entry
C Entry cells
D Exit cells
E Entry-Exit cells
F Inside port cells

Figure S8

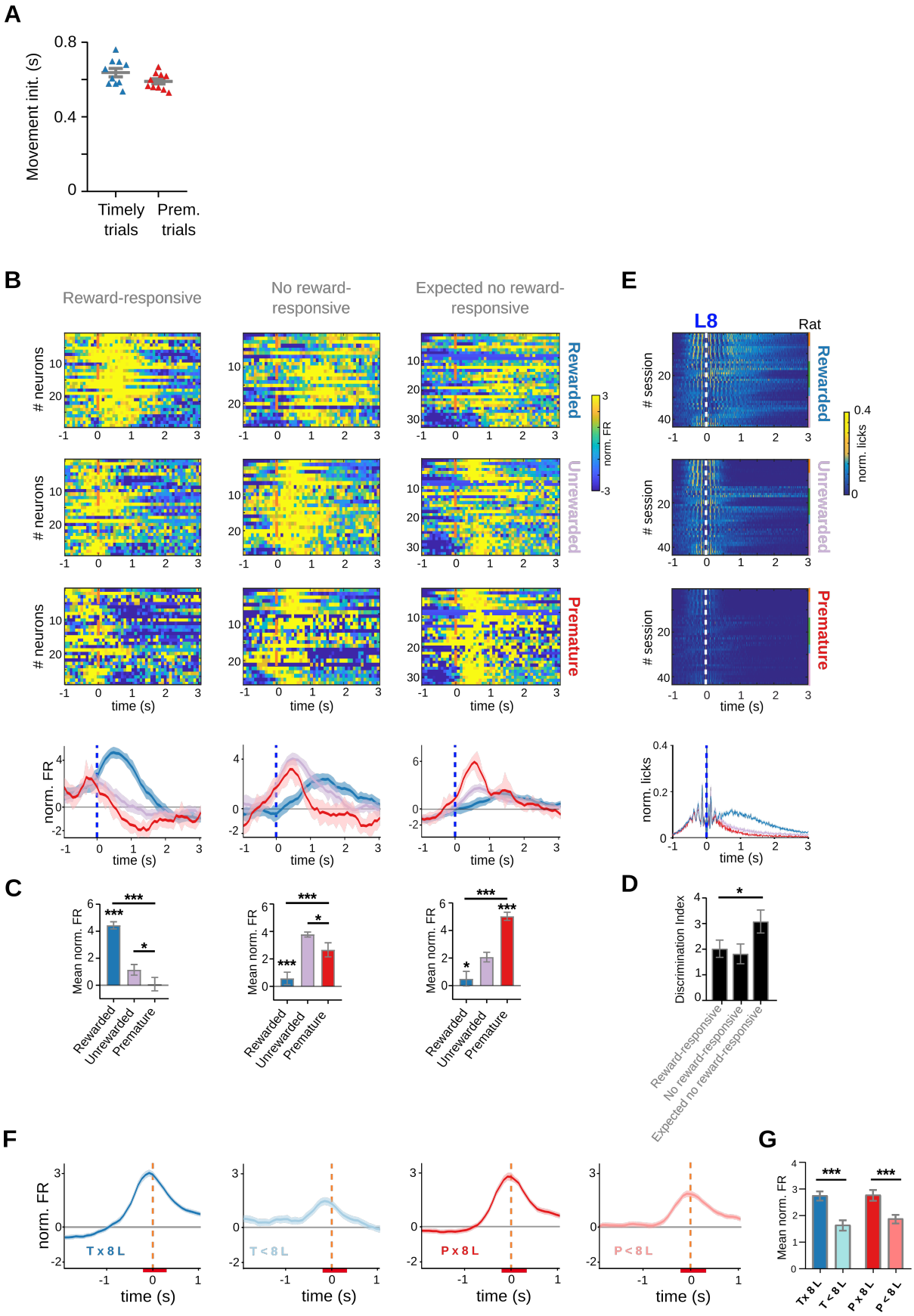

A
Movement init. (s)
Timely trials
Prem. trials
B
Reward-responsive
No reward-responsive
Expected no reward-responsive
neurons
Rewarded
Unrewarded
Premature
norm. FR
time (s)
E
L8
Rat
session
norm. licks
C
Mean norm. FR
Rewarded
Unrewarded
Premature
D
Discrimination Index
Reward-responsive
No reward-responsive
Expected no reward-responsive
F
T x 8 L
T < 8 L
P x 8 L
P < 8 L
G
Mean norm. FR

Figure S9

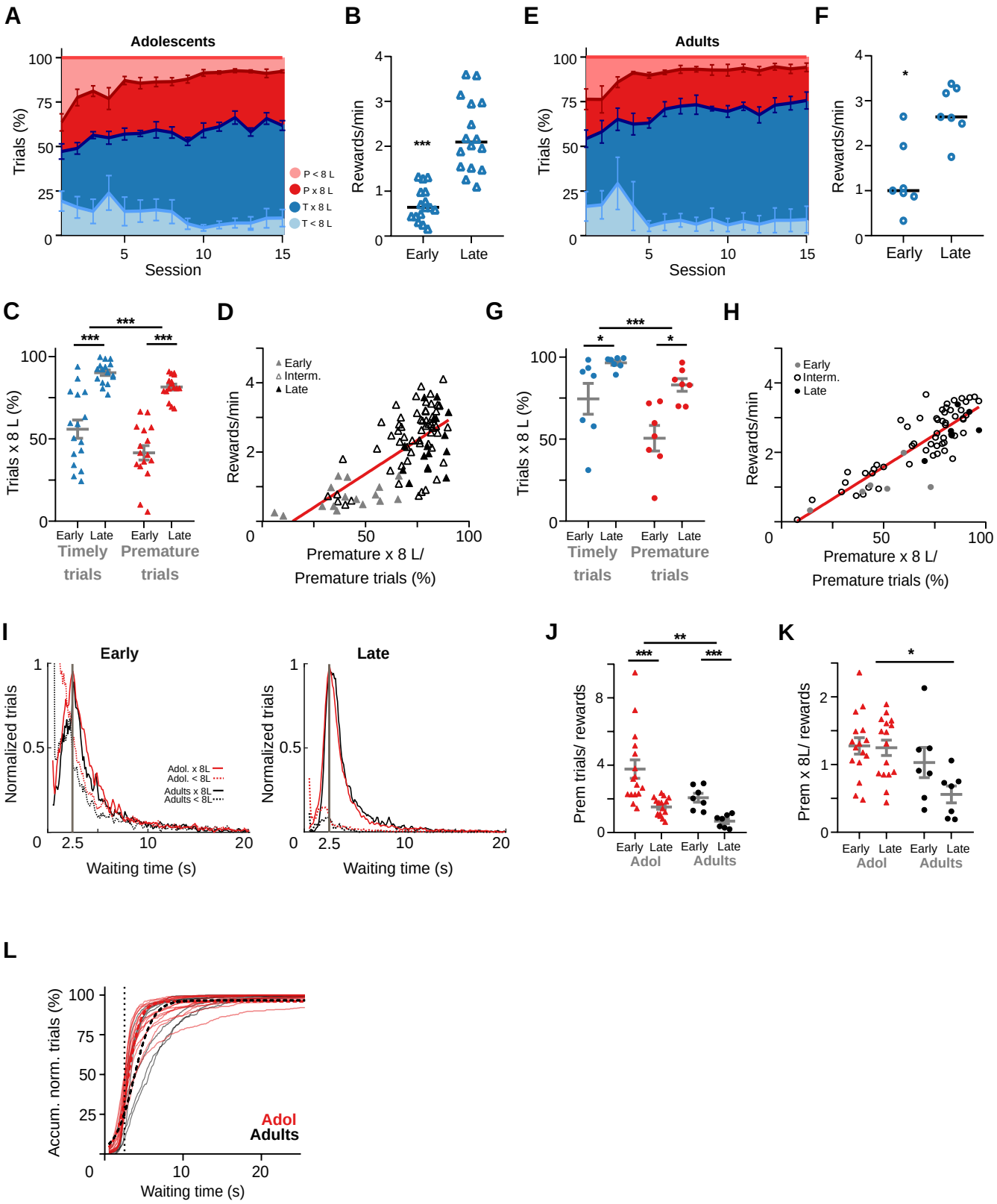
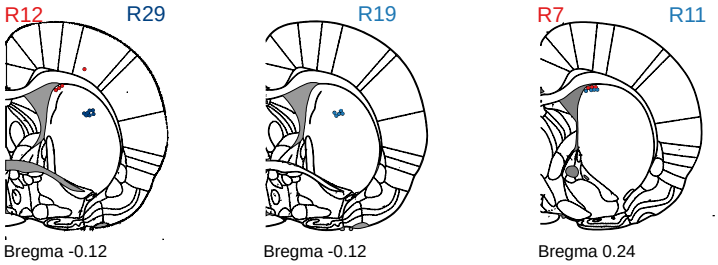
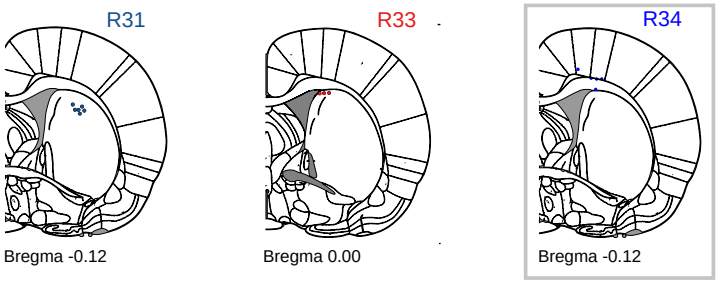

A
Adolescents
Trials (%)
Session
P < 8 L
P x 8 L
T x 8 L
T < 8 L
B
Rewards/min
Early
Late
E
Adults
F
C
Trials x 8 L (%)
Timely trials
Premature trials
D
Early
Interm.
Late
Premature x 8 L/ Premature trials (%)
G
H
I
Early
Late
Normalized trials
Waiting time (s)
Adol. x 8L
Adol. < 8L
Adults x 8L
Adults < 8L
J
Prem trials/ rewards
Adol
Adults
K
Prem x 8L/ rewards
L
Accum. norm. trials (%)
Adol
Adults

Figure S10

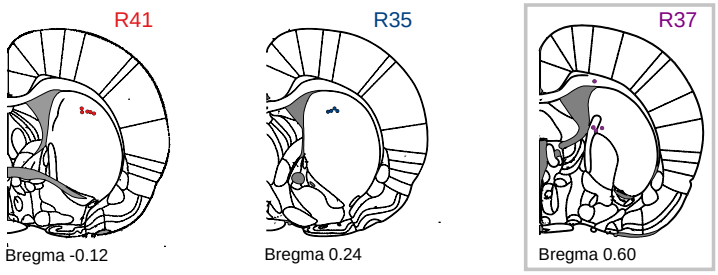
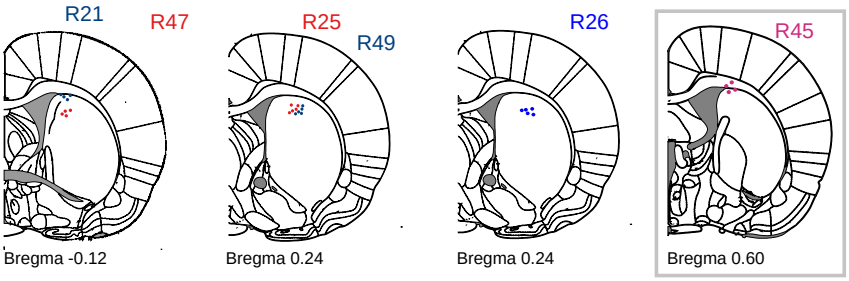
Adults 2.5s WT



Adults 5s WT



Adults 2.5- 5s WT



Adolescents 2.5s WT