Dorsal striatum coding for the timely execution of action sequences

María Cecilia Martinez[1,2], Camila Lidia Zold[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 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 its 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 in decision making (Thura and Cisek, 2017; Yau et al., 2020). Well-learned action sequences may have to be initiated at precise times to obtain the desired outcome, however, they may be difficult to withhold when triggering cues are present, and 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, like 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 to a preponderance of habitual over goal-directed behavioral control (Everitt et al., 2008; Voon et al., 2015). An influential theory 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). By 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), likely contributing 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) tradeoffs during decision making. A bias towards exploration and risk taking (Addicott et al., 2017), low tolerance to 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 encodes action sequence initiation timing, and whether this encoding informs about enhanced premature responding in conditions of high impulsivity, remains to be comprehended.

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; however, the animals showed premature responding even after extensive training and often failed to interrupt sequence execution despite 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 that a peak of striatal activity preceding trial initiation was modulated by time waited before responding. Moreover, this

3

modulation grew at a faster rate in adolescent rats, likely reflecting a steepest reward expectancy increase during waiting that paralleled 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). Trials were self-initiated by the animal 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. Prematurely initiated trials were penalized by re-initiating the waiting interval. 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), as evidenced by a twofold higher reward rate late in training (after three consecutive sessions with >70% correct trials) than early in training (Figure 1e, $p<0.0001$, Wilcoxon matched pairs test). Performance became faster with training (Figure 1f-h): trial duration ($p=0.0005$), latency to the first lick ($p=0.0032$) and time to complete the 8-lick sequence ($p=0.0008$), diminished with training for both rewarded and unrewarded timely trials (significant effect of learning stage, non-significant interaction, two-way RM-ANOVAs).

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).

4

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, $p<0.0001$ for learning stage and trial timing, no interaction, two-way RM-ANOVA). These data suggested that behavior during premature trials was modified by learning. Further supporting this presumption, time to complete the 8-lick sequence ($p=0.0014$), latency to the first lick of the sequence ($p=0.002$) and variation coefficient of the inter-lick intervals (an index of regularity of such intervals; $p=0.021$), 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 had a negative effect on 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).

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 peak rate observed at 2.5 s was passed (see below; Supplementary Figure 2).

5

Altogether, the data show that adult rats optimized 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), DMS 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 DMS activity preceding port entry and/or at the time of port exit, during timely trials (Figure 3a-b). Overall, ~50% of the recorded units (n=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=105, 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 DMS 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.

6

Waiting time modulates anticipatory activity

Striatal activity anticipating a learned behavioral response could specifically mark the initiation of a previously rewarded action sequences (Jin and Costa, 2010; Jog, 1999; Martiros et al., 2018) or relate to additional factors, like reward anticipation and the vigor and value of the 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 an automatized action sequence is often prematurely released, we asked if the observed anticipatory activity specifically predicts the release of the learned action sequence, and if, additionally or alternatively, it 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; $p<0.0001$, significant main effect of trial initiation timing, no effect of action sequence structure, no interaction, 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). Further data analysis showed that this modulation of striatal activity by trial initiation timing was independent of electrode location within the DMS and involved both port entry only and port entry/port exit neurons (Supplementary Figure 4a-b). Since this anticipatory activity closely preceded approaching movements towards the nose-poke, we analyzed accelerometer recordings of head movements performed during the task. The accelerometer recordings did not differ between premature and timely trials (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, in this task, the firing rate modulation preceding trial initiation discriminates between premature and timely trials and does not predict the speed, regularity, structure,

7

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 5). Contrastingly, no modulation of this same neuronal activity by time waited in the following trial (Figure 4f), or of port exit related activity by the preceding waiting time (Figure 4g), was observed. The steep slope of the curve at the criterion waiting interval suggested that the neuronal activity does not linearly report elapsed time but rather changes in reward anticipation as waiting progressed. To explore this possibility further, 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. Waiting less than 2.5 s (premature trials) or more than 5 s (late trials) was penalized by resetting the waiting interval (Figure 4h). We reasoned that if this activity provides a wait time-based reward anticipation signal for the upcoming action, it should decrease (instead of plateauing) after 5 s of waiting in the modified version of the task. 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, $p<0.0001$ versus timely trials, Tukey post hoc test after significant one-way RM-ANOVA; Supplementary Figure 2). The firing rate modulation preceding trial initiation increased with a steep slope at 2.5 s, confirming results obtained with the basic version of the task, but instead of plateauing, it decreased after surpassing the 5 s cutoff time, yielding a significant interaction in a two way ANOVA comparing the effects of waiting time on anticipatory activity (Figure 4j and 4k, significant effect of training, $p=0.033$, trials, $p<0.0001$, and interaction, 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.

8

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 ($p<0.05$ versus timely trials, Tukey post hoc test after significant one-way RM ANOVA) (Supplementary Figure 6). 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). Detection of activations occurring during narrow time windows centered over individual licks yielded largely overlapping populations of positively modulated neurons with broad activity peaks encompassing many inter-lick intervals (Figure 5b). This modulation lasted longer during rewarded trials (where licking also persisted for longer) than during timely unrewarded and premature trials, but was of similar amplitude at the time when reward was expected irrespective of reward delivery or omission (Figure 5b, activity centered at the eight lick), suggesting that these neurons were modulated by licking.

Striatal activity is also modulated by 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, premature 8-lick trials seem to be necessary to maintain an internal representation of the criterion waiting interval. To determine if the negative outcome associated to premature 8-lick trials is reported by striatal neurons, we looked for outcome-related activity modulations at the time of the eight lick during timely rewarded, timely unrewarded and premature trials with complete 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. Five percent of all

9

recorded neurons (n= 45) showed a higher activity modulation when reward was obtained than in the no-reward conditions (Figure 5c, f top – reward responsive neurons, $p<0.0001$, Tukey test after significant one-way RM ANOVA), before lick rates began to diverge by trial type (Figure 5h). Moreover, 5% of the striatal neurons showed activations at the time of the no-reward outcome in timely unrewarded trials and were unresponsive to reward delivery (Figure 5d). Interestingly, these neurons also showed a marked activation at the time of the 8-lick during premature 8-lick trials (Figure 5d, f middle – no-reward responsive neurons, $p<0.0001$ versus rewarded trials, Tukey test after significant one-way RM ANOVA). Strikingly, there were 33 neurons (4% of all recorded neurons) that showed a two-fold higher activation at the time of the eight lick in premature trials than in timely unrewarded trials (Figure 5e, f bottom; $p<0.0001$, Tukey test after significant one-way RM ANOVA). This difference cannot be explained by differences in lick rates because licking decreased at similar rates after the 8-lick in both types of non-rewarded trials (Figure 5h). Further data analysis showed that, indeed, the three populations of outcome responsive neurons discriminated timely unrewarded from premature 8-lick trials; however, those showing a higher activation during premature trials discriminated better between these two no-reward conditions (Figure 5g). Overall, the data supports 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; $p<0.001$ for lick sequence structure, no effect of trial timing, no interaction, two way RM ANOVA). Discrimination of

10

incomplete lick sequence trials was more marked in the port exit only neurons and was independent of striatal recording site (Supplementary Figure 4c-d). Moreover, there was no between trials difference 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

Rats implanted with tetrodes 30-35 days after birth learned the 2.5 s waiting interval task as shown by a higher percentage or correct trials (Figure 7a), a higher reward rate (Figure 7b, *$p<0.0001$, Wilcoxon matched pairs test), and a faster performance with training (Supplementary Figure 7). Moreover, trial duration, latency to first lick and 8-lick sequence duration did not differ between adolescent and adult rats (Supplementary Figure 7a-c), nor between premature 8-lick and timely 8-lick trials in adolescent rats (Supplementary Figure 7e-g). Finally, the proportion of premature 8-lick sequence trials increased with training in adolescent rats (Figure 7c; $p<0.0001$, significant effects of learning stage and trial timing, no interaction, two-way RM ANOVA), in parallel with reward rate (Figure 7d), 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 (Figure 7f and 7g). The number of premature trials per reward obtained was twofold higher in adolescent than adult rats even after training (Figure 7f; $p=0.004$, significant effects of age group, training, and interaction, two-way RM ANOVA), and this was due to a preferential retention of the premature 8-lick trials through training (Figure 7g; $p=0.0024$, significant effect of age group, no effect of training, no interaction, two-way repeated measures ANOVA). Finally, a comparison of the cumulative frequency distributions of 8-lick trial initiation times at late stages of training showed that the relative excess of premature 8-lick trials observed in adolescent rats parallels an excess of late (yet rewarded at $p=0.5$) trials in adult rats (Figure 7h), suggesting that adult rats are more

tolerant to reward delay. 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 8). Of the registered units in adolescents (n=552), the proportion of neurons showing port entry, port exit and lick-related activity was similar to that found in adult rats (Figure 7i; p=0.83, 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 (Supplementary Figure 9). 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 9).

We therefore looked for quantitative differences in task-related activities that could account for the more impulsive behavior of adolescent rats. Data from three adolescent and five adult rats (133 and 103 DMS neurons with anticipatory activity, respectively) trained in the 2.5 s waiting interval task were used for the following analysis. Curves displaying the amplitude modulation of activity preceding trial initiations at increasing waiting times, probably reflecting reward anticipation, showed a steepest increase in adolescent rats as the criterion waiting interval was surpassed (Figure 7j). When all port entry modulated neurons were considered, statistical comparisons based on a general linear model showed a significantly higher activity modulation by waiting time in adolescent rats (significant effect of age, $p<0.001$, and waiting time, <0.001, no interaction). On the other hand, activations at time of expected reward (Figure 7k, $p<0.0001$, effect on trial initiation timing, Restricted Maximum Likelihood test, REML) and activity at port exit, probably reflecting that the correct action sequence was emitted (Figure 7l, $p<0.0001$, significant interaction between trial initiation timing and sequence structure, three-way

12

ANOVA), did not differ between adolescent and adult rats.

In sum, the data show that the waiting modulation of reward anticipation grows at a faster rate in adolescent rats.

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 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, which makes them more impulsive than adults.

13

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 signature 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, 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 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 "opening" activity, 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 closing 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 obtained from the DMS, we cannot rule out that a different opening activity specific for the learned sequence emerges in the DLS in our task as previously

14

reported for other tasks (Jin and Costa, 2010; Jog, 1999; Martiros et al., 2018; Thorn et al., 2010). Importantly, while striatal neurons with opening and/or closing activity were poorly modulated during sequence execution, as shown in other studies (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). Overall, our data show that the DMS expresses opening/closing activity even after skilled performance is reached, as reported in a recent study with a different task (Vandaele et al., 2019), and suggest that, at least in the DMS, 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 reward anticipation signals and/or outcome evaluation signals. Striatal signals at the time of expected outcome not only discriminated between rewarded and unrewarded 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 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 should arrive to 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

15

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 accumulation 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 (Buhusi and Meck, 2005; Gibbon, 1977; Namboodiri and Hussain Shuler, 2016). Previous studies showed that sequentially activated striatal 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

16

durations as longer as they are, which would be associated to perceiving a higher cost of time and to 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 than adult rats and indeed, they do not show the negative effect of premature trials on reward rate that is observed in adults (Supplementary Figure 7). Importantly, although they achieve this reward rate at the expense of a higher energy cost (more premature trials per reward obtained), because most premature trials remaining after reaching skilled performance correspond to the learned action sequence, adolescents would likely get more precise information regarding the duration of the waiting interval and would also 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 to gain 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.

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

17

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 from our own colony (Adolescents: 4 weeks of age at the beginning of the experiments, weight ~120 g; Adults: 8-12 weeks of age at the beginning of the experiments, weight ~390 g), housed on a 12:12 light:dark cycle and experiments were performed during the light phase, 21° C room temperature. Rats were housed in groups of 3–4 in regular cages with wood shavings' bedding and, after surgery, they were singly housed in 45 cm × 35 cm × 22 cm cages with moderate environmental enrichment (toys, tissue paper strips) and cardboard bedding. Two days before the beginning of the behavioral training subjects were deprived of water. All throughout the training schedule subjects had restricted access to water for 20 min each day and a-day rest period each week. This schedule maintained animals at 90% of their predeprivation weight, with any further weight loss being counteracted by increased free water access. 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). Five adult males were part of the ITI 2.5 s experiments, three adult males were part of

18

the ITI 5s experiments (Fig. 2i and Fig. S1) and another three adult males were included in the ITI 2.5-5s experiments (Fig 4h-k, Fig. S2 -all except S2f, see below-). Three adolescent male were part of ITI 2.5 s experiments (Fig. 7, Fig S7-S9). Additionally two adult subjects were included in the accelerometer experiment (Fig. 4 and 6) and two others were only trained with the ITI 2.5 s protocol and did not undergo surgery (Fig. S2f).

Electrodes

Bundles of four microwire electrodes (12µm, tungsten, California fine wire company, USA, 0.2 MΩ impedance) were attached to a homemade micromanipulator (Vandecasteele et al., 2012). Before surgery, the tetrode array was sterilized with 3 % oxygen peroxide solution (AOSept ® PLUS TM). 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). Before being implanted, the upper part of the tetrodes was painted with a DII solution (SCBT, sc-213424) to facilitate the visualization of the electrodes trace before the histological analysis.

Surgery

Before surgery, subjects were treated with local anaesthetic (lidocaine) in 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: +0.06 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

19

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 behavioral operant chamber that contained a 'nose-poke' at which they could seek the reward by licking through a slot onto a lick tube. Breaking an infrared beam in front of the nose-poke ended the inter-trial interval (ITI) and 200 ms after, the nose-poke was then illuminated for 100 ms (visual cue). Following the visual cue, licks were detected by breaking a second infrared beam. Upon licking the required number of times (8 licks), reward (~20 µl of water) was made available on half of the trials on a pseudo-random fashion. Water reward was delivered through a tube with a solenoid attached to it. The behavioral task was controlled using an Arduino Uno board. Trials ended when the animal removed its head from the nose-poke. Sessions ended when the subject completed 300 trials or when it was reached a maximum training time of 120 min. Timely trials are those in which the subject entered the nosepoke after the required minimum waiting time. Premature trials are those in which the entrance beam was interrupted before having reached the minimum waiting time. Training sessions were classified into early or late as follows: having the animal reached a minimum of 80 % of timely trials with the complete 8-licks' sequence for two consecutive sessions, the following sessions were considered as "late" training sessions. After evaluating the total number of late training sessions, the same number of "early" sessions was selected upon the first training session, i.e: having counted 3 late sessions, the first three training

20

sessions were considered as “early” and the rest of the sessions between early and late were classified as “other”. All behavioral data is expressed as mean ± SEM (except for reward rate figures, where the median is indicated, and Waiting time curves). Each dot corresponds to the average value of a session from a single animal.

Accelerometer experiment

Two rats were trained daily in the nose-poke chamber until they reached a stable good performance (equivalent to “late” training sessions). Afterwards, a MPU-6050 accelerometer was chronically attached to their heads following a procedure similar to the electrodes’ implantation. Briefly, four stainless steel screws were inserted in the skull and the accelerometer was glued inside a cap made of a metallic mesh and dental acrylic that was anchored to the skull with dental cement. Subjects were left for one week of recovery before starting again with the behavioral training. Accelerometer data (sampling rate 25 Hz) was collected during each training session (9 sessions in total), using an Arduino UNO board connected to a PC via PLX-DAQ software (Parallax Inc.).

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 and later, sections were safranin-stained to confirm the visualization of the tracks.

Electrode positioning was considered “medial” when the tracks were found between 0 and 0.28 cm

21

relative to bregma and “lateral” from 0.28 cm onwards, considering bregma as starting point from center to right (Fig. S10).

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 was analyzed with MATLAB software using custom-built scripts. We registered a total of 867 units in adult rats: from the 2.5 s WT experiments there were 158 with no response and 244 were task responsive, from the 5 s WT experiments there were 45 with no response and 47 were task responsive and from the 2.5-5 s WT experiments there were 79 with no response and 294 were task responsive. In adolescents we registered a total of 552 units, 233 with no response and 319 task responsive (2.5 s WT).

Peri-event time histograms (PETH) were created using an 80 ms bin. In order to compare between animals and cells each PETH was normalized using the mean and standard deviation ((Xi – Xmean) / Xsd) calculated from a PETH of each cell centered on port entry +/- 20 s. All data is presented as normalized firing rate except for the peri-event examples. Only for display purposes, PETH were smoothed using a 5 points moving window.

In order to analyze the activity of each neuron for the different waiting times (WT), PETH were

22

constructed using trials with similar WT. The WT intervals limits were determined to include a similar number of port entries in each segment. Only 2.5 s and 5 s were fixed. 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]. Trials with WT shorter than 450 ms were not included in the analysis.

A neuron was considered to be responsive to the port entry if the mean of the normalized firing rate (mnFR) was greater by 1SD to the mean. The mnFR was computed between 640 and 80 ms prior to the port entry. Similarly, a cell was considered to be responsive to the port exit if its mnFR between 240 ms prior and 320 ms after port exit was greater than 1SD. A neuron was considered to be active during sequence execution (inside port cells) if its mnFR was > 1SD between 400 and 1520 ms after port entry. Neurons with an increase above 1 SD from port exit to 1 s before port entry, calculated for every waiting time interval, were labelled as tonically active. Visual cue responding cells were those whose mnFR was > 2SD between 160-400ms post port entry in those trials that had the visual cue. A neuron was included as active around lick number 3 or 8 if their mnFR was > 1 SD around the event (+/- 250ms). Neurons with activity related to the reward were selected as follow: 1) their mnFR between 160-560 ms after the 8th lick should be greater to the activity during licking (-640 to -160 ms prior to 8th lick) by 0.5 SD and 2) their mnFR during 160-560 ms after the 8th lick was above 2 SD or rewarded timely trials or during unrewarded timely trials. We divided these neurons in three groups: a) with response in trials with reward (Fig. 5c), b) with response in timely trials with no reward, and activity higher in those trials than in timely rewarded trials by 1.5 SD (Fig. 5d), c) with response in premature trials not included in b (Fig. 5e). Relative mean normalized FR was calculated for each neuron considering average FR equal to 1 for the timely trials between 2,5 and 5 s and timely trials for the modified version of the task (Fig. 4k). Discrimination index for each selection of neurons was calculated as: absolute value of (mnFR in Unrewarded trials – mnFR in Premature trials, Figure 6g).

23

Behavioral analysis

All data was analyzed with MATLAB software using custom-built scripts. Waiting time histograms were calculated from the total number of entries (either early or late in training, with or without the 8-licks sequence) using 100 ms bins, normalized by its maximum so that the peak of the curve corresponded to a value of 1. Curves were smoothed with a 4 bin span. For the accelerometer recordings PETH were created using a 40 ms bin. Data collected from both animals was pooled and the normalized mean and 95 CI were calculated around the entry or exit port event +/- 2.5 s. Licks PETHs were constructed centered on port entry using a 10 ms bin and normalized to the mean and standard deviation of each session. All data is presented as normalized number of licks +/- SEM.

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 reward rate in which Wilcoxon matched pairs test was used. To compare neuronal activity between Rewarded, Unrewarded and Premature groups, repeated measures one-way ANOVA was used. To compare licking activity between Rewarded, Unrewarded and Premature trials 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. Localization of the neurons and their activity at the entry or exit was analyzed using Three-way ANOVA. Linear regression was used to analyze premature trials and the reward rate. A general linear model was used to compare neuronal activity from adult and adolescent rats. Model: Neuronal response ~ intercept + waiting time x age (as factor), using a Gaussian distribution.

24

Acknowledgments

We thank Dr. Juan Belforte for helpful discussions and also Yamila Páez for her technical assistance with animal welfare. 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./ Formal Analysis M.B., C.M., G.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. All the code used for data analysis in this paper will be shared by the lead contact upon written request. Data is reported in

Figure Legends

25

Figure 1. Rats become skilled in the task. (a) Training cage with the nose-poke. Animals' entries and exits from the nose-poke are detected with two infra-red beams. A visual cue of 100 ms is given with a pair of green LEDs placed inside 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 the traces of the electrodes painted with DII. **(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.

Figure 2. Training does not suppress premature initiations of the learned behavioral response. a) Correlation between prematurely initiated trials and reward rate ($Y = -0.023*X + 1.921$, slope significantly different from zero $p=0.0176$, R square 0.09031.). Proportion of trials **(b)**, time to complete the 8-lick sequence **(c)**, latency to first lick **(d)** variation coefficient **(e)** of the 8-lick sequence inter-lick intervals for prematurely and timely initiated trials, at early and late learning stages. **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, $p<0.0001$, R square 0.5016.). **g-i)** Normalized frequency distributions of the trial initiation times (waiting time), separated for trials with (gray) and without (dotted line) 8-lick sequences, and for early **(g)** and late **(h)** training stages. Insets: percentage of the trials with (gray) 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

26

criterion time for the following two sessions (48 h after the last 5s-WT session, light blue). Dotted 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. a) Representative raster plots and PETH of striatal units showing firing rate modulations related to port entry and/or port exit. **b)** Individual PETHs of all the recorded striatal neurons during correct trials, aligned to port entry (left) or port exit (right), with corresponding average PETH (below). Color code for the normalized activity is shown on the right. **c)** Proportion of neurons showing task-related firing rate modulations. **d-g)** Individual and average PETH and average firing rates 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. In d to g, 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).

Figure 4. Prematurely initiated trials are preceded by low anticipatory activity. (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)** Average striatal activity corresponding to the PETH shown in a. **(c)** Accelerometer recordings of head movements around port entry for premature and timely trials. It is represented the mean (solid lines) and 95 CI (shaded area). Data was obtained from a total of 9 training sessions of two animals. **(d)** Average PETH of striatal neurons showing entry-related activity around port entry sorted by waiting time duration. Color code for the intervals shown on the right. **(e)** Mean normalized firing rate for each of the waiting time segments, 2.5 s criterion time is shown with a blue hatched-line. **(f)** Average PETH of the same striatal neurons segmented according to next trial waiting

27

time. **(g)** Average PETH of striatal neurons showing activation at port exit segmented as in **(d)**. **(h)** Schematic representation of the different types of trials in a modified version of the task. Timely trials: waiting time 2.5 – 5 s, Premature: waiting time <2.5 s, Late: waiting time >5 s. **(i)** Proportion of trials followed by 8-lick sequences for each type of trial. **(j)** Average PETH of striatal neurons showing activity preceding port entry according to the waiting time duration. Color code for the intervals shown on the right. **(k)** Relative mean normalized firing rate for each of the waiting time groups for both variants of the task.

Figure 5. Reward responsive neurons discriminate prematurely from timely initiated trials. (a) Schematic representation of the different types of trials analyzed. **(b)** Average PETH of striatal units showing firing rate modulations related to the licking activity (1st, 3rd and 8th lick). It is represented the mean (solid lines) and SE. **(c-e)** individual and average PETHs of striatal units, aligned to 8th lick (time 0 s), showing positive firing rate modulations during reward delivery **(c)**, reward omission **(d)** and premature trials **(e)**. **(f)** Mean normalized firing rate for the different trial conditions, from top to bottom: reward delivery, reward omission and premature trials. **(g)** Discrimination index for each of the groups of neurons shown in **(c-e)**. **(h)** PETH showing lick rates during the three trial conditions, centered at the 8th lick, with its corresponding average at the bottom. Color code for the normalized activity is shown on the right.

Figure 6. Striatal activity at the exit reports the performance of the action sequence. (a) Average PETH 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 (diagram on top shows the types of trials analyzed). **(b)** Average striatal activity corresponding to the PETH shown in a. **(c)** PETH of head acceleration around port exit in premature and timely trials. It is represented the mean (solid lines) and 95

28

CI (shaded area).

Figure 7. Adolescent rats become skilled in the task and show a higher modulatory response. (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. **(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<0.0001, R square 0.7133.). **(e)** Normalized frequency distributions of the trial initiation times (waiting time), separated for trials with (dark gray) and without (light gray dotted line) 8-lick sequences, in the late of training. Inset: percentage of the trials with and without 8-lick sequences for each bin, zoomed around the criterion time. **(f)** Ratio between premature trials made and obtained rewards for adolescents and adults. **(g)** Premature trials with 8-lick sequence made per obtained rewards for adolescents and adults. **(h)** Accumulated normalized trials vs. Waiting time for adolescents (mean: red, individual session: light red) and adults (mean: black, individual session: gray). **(i)** Percentage of neurons showing task-related firing rate modulations. **(j)** Mean normalized FR vs. waiting time for both ages. 2.5 s criterion time is shown with a blue hatched-line. (**k)** Discrimination index calculated at the time of **r**eward delivery. **(l)** Mean normalized FR at the time of exit for trials with or without 8-lick sequence for both ages. j-l: hatched lines correspond to adolescents and full lines to adults.

Supplementary Figure 1. Rats trained with a longer ITI also emit premature learned responses. (a) Reward rates for each training stage of rats trained with a 5s ITI requirement Wilcoxon matched-pairs signed rank test ** p=0.0078. **(b)** Proportion of trials followed by 8-lick sequences at early and late learning stages, Two-way RM ANOVA, Type of trial *** p<0.0001, Interaction Training stage x Type of trial * p=0.0535. **(c)** Time to complete the 8-lick sequence, Mann Whitney Test p=0.4894. **(d)** Latency to first lick for prematurely and timely initiated trials, at early and late learning stages, Mann Whitney Test

29

p=0.8633. **(e)** Variation coefficient of the 8-lick sequence inter-lick intervals for prematurely and timely initiated trials, at early and late learning stages, Mann Whitney Test NS p=0.8633. **(f)** Correlation between prematurely initiated trials and reward rate, Linear regression Y=-0.02755*X +2.776, Slope significantly different from zero: p=0.0020, R-square=0.3343. **(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: p=0.0192, R-square=0.2080. **(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 of training. Inset: percentage of the trials with (gray) and without (dotted line) 8-lick sequences for each bin, zoomed around the criterion time.

Supplementary Figure 2. Rats also learn a task with lower and upper limits in the WT. Rats were trained to enter the nosepoke within a minimum WT of 2.5 s and a maximum WT of 5. **(a)** Reward rates for the 2.5-5 s WT task, *** p<0.0001 Wilcoxon Signed rank test. **(b)** Trial duration for correct trials, Paired t test ***p<0.0001. **(c)** Latency to first lick for prematurely, timely and late initiated trials, One-way RM ANOVA NS, p=0.6053. **(d)** Time to complete the 8-lick sequence for the different types of trials, One-way RM ANOVA NS, p=0.5593. **(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 WT. 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 WT. 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=0.0238.

30

Supplementary Figure 3. Characteristics of the neurons registered in the 2.5 s WT 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. Color code for the normalized (in units of standard deviation) activity is shown on the left. **(b)** Proportion of different responses in both training stages from animals trained with a 2.5 s criterion time.

Supplementary Figure 4. Striatal activity at the boundaries of the trials. (a) Anticipatory activity of entry responsive neurons and neuronal localization. Average striatal activity shown in the right panel. Three-way ANOVA, Localization p=0.0005, Training stage *** p<0.0001, Licks' sequence p=0.0851, Interaction Localization x Training stage *** p<0.0001, Interaction Localization x Licks' sequence p=0.1999, Interaction Training stage x Licks' sequence p=0.2842, Interaction Training stage x Licks' sequence x Localization *** p=0.0006. **(b)** Anticipatory activity of entry-exit responsive neurons and neuronal localization. Average striatal activity shown in the right panel. Three-way ANOVA, Localization p=0.1824, Training stage *** p<0.0001 , Licks' sequence p=0.1751, Interaction Localization x Training stage *** p<0.0001, Interaction Localization x Licks' sequence p=0.0501, Interaction Training stage x Licks' sequence p=0.5239, Interaction Training stage x Licks' sequence x Localization p=0.0178. **(c)** Activity at port exit of exit responsive neurons and neuronal localization. Average striatal activity shown in the right panel. Three-way ANOVA, Localization p=0.9256, Training stage *** p=0.0007, Licks' sequence *** p<0.0001, Interaction Localization x Training stage p=0.0511, Interaction Localization x Licks' sequence p=0.5469, Interaction Training stage x Licks' sequence *** p<0.0001, Interaction Training stage x Licks' sequence x Localization ** p=0.0022. **(d)** Activity at port exit of entry-exit responsive neurons and neuronal localization. Average striatal activity shown in the right panel. Three-way ANOVA, Localization p=0.6139, Training stage p=0.1512, Licks' sequence *** p=0.0005, Interaction

31

Localization x Training stage p=0.9700, Interaction Localization x Licks' sequence p=0.6973, Interaction Training stage x Licks' sequence p=0.5116, Interaction Training stage x Licks' sequence x Localization p=0.6400.

Supplementary Figure 5. Anticipatory activity in the 5s WT 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 6. 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. Color code for the mnFR is shown on the right. Below, bar graphs showing statistical comparisons between trial conditions. One-way RM ANOVA *** p<0.0001, Tukey's multiple comparisons test *** p=0.0002 Premature vs. the other groups.

Supplementary Figure 7. Adolescent rats make more premature trials. (a) Trial duration for correct trials at each training stage, Two-way RM ANOVA, Training stage *** p=0.0002, Type of trial ** p=0.0072. **(b)** Latency to the first lick during correct trials, Two-way RM ANOVA, Training stage *** p=0.0007. **(c)** Time to complete the 8-lick sequence, Two-way RM ANOVA, Training stage *** p<0.0001. **(d)** Correlation between prematurely initiated trials and reward rate, Linear regression Y=-0.002263 *X +1.504, Slope significantly different from zero: p=0.8492, R-square= 0.0077. **(e)** Time to complete the 8-lick sequence, Two-way RM ANOVA, Training stage *** p=0.0005. **(f)** Latency to first lick for

32

prematurely and timely initiated trials, at early and late learning stages, Two-way RM ANOVA, Training stage ** p=0.0045. **(g)** Variation coefficient 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 **p=0.0034, Interaction Training stage x Type of trial p=0.0013.

Supplementary Figure 8. Striatal activity also marks transitions between behavioral states of the task in adolescent animals. (a) Representative raster plots and PETH of striatal units showing firing rate modulations related to port entry and/or port exit. **(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 to 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. In **(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). Color code for the mnFR is shown on the right **(a)**.

Supplementary Figure 9. Anticipatory, reward responsive and exit activity in striatal neurons of adolescent rats. (a) Average firing rate of entry-related neurons responding at port entry, during timely and premature trials, for trials with an 8-lick sequence and trials where less than 8 licks were emitted (diagram on top shows the types of trials analyzed). 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 (-640 to -80 ms). **(b)** Average striatal activity corresponding to the PETH shown in a. **(c)** Population PETH and Average firing rates of striatal units showing positive firing rate modulations during reward delivery, reward omission and premature trials. Below, bar graphs showing statistical comparisons between trial

33

conditions. **(d)** PETH showing lick rates during the three trial conditions, centered at the 8th lick, with its corresponding average and statistical analysis at the bottom. **(e)** Discrimination index for each of the groups of neurons shown in b-d, Mixed-effects model (REML) * p=0.0164. **(f)** PETH showing lick rates during the three trial conditions, centered at the 8th lick, with its corresponding average at the bottom. Color code for the normalized activity is shown on the right. **(g)** 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. **(h)** Average striatal activity corresponding to the PETH shown in g, Two-way RM ANOVA, Licks' sequence *** p<0.0001.

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.

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.

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.

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.

34

763 Bakhurin, K.I., Goudar, V., Shobe, J.L., Claar, L.D., Buonomano, D. V., and Masmanidis, S.C. (2017).
764 Differential Encoding of Time by Prefrontal and Striatal Network Dynamics. J. Neurosci. *37*, 854–870.

765 Balleine, B.W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive
766 learning and their cortical substrates. Neuropharmacology *37*, 407–419.

767 Bari, A., and Robbins, T.W. (2013). Inhibition and impulsivity: behavioral and neural basis of response
768 control. Prog. Neurobiol. *108*, 44–79.

769 Barnes, T.D., Kubota, Y., Hu, D., Jin, D.Z., and Graybiel, A.M. (2005). Activity of striatal neurons reflects
770 dynamic encoding and recoding of procedural memories. Nature *437*, 1158–1161.

771 Baumann, A.A., and Odum, A.L. (2012). Impulsivity, risk taking, and timing. Behav. Processes *90*, 408–
772 414.

773 Belin, D., and Everitt, B.J. (2008). Cocaine seeking habits depend upon dopamine-dependent serial
774 connectivity linking the ventral with the dorsal striatum. Neuron *57*, 432–441.

775 Bonner, A.C., and Borrero, J.C. (2018). Differential Reinforcement of Low Rate Schedules Reduce Severe
776 Problem Behavior. Behav. Modif. *42*, 747–764.

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

780 Buhusi, C. V, and Meck, W.H. (2005). What makes us tick? Functional and neural mechanisms of interval
781 timing. Nat. Rev. Neurosci. *6*, 755–765.

782 Carland, M.A., Thura, D., and Cisek, P. (2019). The Urge to Decide and Act: Implications for Brain
783 Function and Dysfunction. Neuroscientist *25*, 491–511.

784 Costa, R.M. (2011). A selectionist account of de novo action learning. Curr. Opin. Neurobiol. *21*, 579–
785 586.

786 Dalley, J.W., and Robbins, T.W. (2017). Fractionating impulsivity: neuropsychiatric implications. Nat.
787 Rev. Neurosci. *18*, 158–171.

788 Daw, N.D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and
789 dorsolateral striatal systems for behavioral control. Nat. Neurosci. *8*, 1704–1711.

790 Donnelly, N.A., Holtzman, T., Rich, P.D., Nevado-Holgado, A.J., Fernando, A.B.P., Van Dijck, G.,
791 Holzhammer, T., Paul, O., Ruther, P., Paulsen, O., et al. (2014). Oscillatory activity in the medial
792 prefrontal cortex and nucleus accumbens correlates with impulsivity and reward outcome. PLoS One *9*,
793 e111300.

794 Emmons, E.B., De Corte, B.J., Kim, Y., Parker, K.L., Matell, M.S., and Narayanan, N.S. (2017). Rodent
795 Medial Frontal Control of Temporal Processing in the Dorsomedial Striatum. J. Neurosci. *37*, 8718–8733.

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.

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.

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.

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.

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.

Geddes, C.E., Li, H., and Jin, X. (2018). Optogenetic Editing Reveals the Hierarchical Organization of Learned Action Sequences. Cell *174*, 32-43.e15.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. Psychol. Rev. *84*, 279–325.

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.

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*.

Gouvêa, T.S., Monteiro, T., Motiwala, A., Soares, S., Machens, C., and Paton, J.J. (2015). Striatal dynamics explain duration judgments. Elife *4*.

Graybiel, A.M. (1998). The Basal Ganglia and Chunking of Action Repertoires. Neurobiol. Learn. Mem. *70*, 119–136.

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.

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.

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.

36

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.

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.

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.

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.

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

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

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.

Jog, M.S. (1999). Building Neural Representations of Habits. Science (80-.). *286*, 1745–1749.

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.

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.

Knowlton, B.J., and Patterson, T.K. (2016). Habit Formation and the Striatum. Curr. Top. Behav. Neurosci. *37*, 275–295.

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.

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

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.

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.

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.

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.

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.

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

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.

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*.

Neill, D.B. (1976). Frontal-striatal control of behavioral inhibition in the rat. Brain Res. *105*, 89–103.

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.

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.

Robbins, T.W., and Costa, R.M. (2017). Habits. Curr. Biol. *27*, R1200–R1206.

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.

Samejima, K., Ueda, Y., Doya, K., and Kimura, M. (2005). Representation of action-specific reward values in the striatum. Science *310*, 1337–1340.

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.

Schultz, W. (2015). Neuronal Reward and Decision Signals: From Theories to Data. Physiol. Rev. *95*, 853–951.

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.

Singer, H.S. (2016). Habitual and goal-directed behaviours and Tourette syndrome. Brain *139*, 312–316.

Smith, K.S., and Graybiel, A.M. (2013). A dual operator view of habitual behavior reflecting cortical and striatal dynamics. Neuron *79*, 361–374.

38

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.

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.

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.

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*.

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.

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.

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.

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.

Wittmann, M., and Paulus, M.P. (2008). Decision making, impulsivity and time perception. Trends Cogn. Sci. *12*, 7–12.

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.

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.

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.

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

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.

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.

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.

40

Figure 1

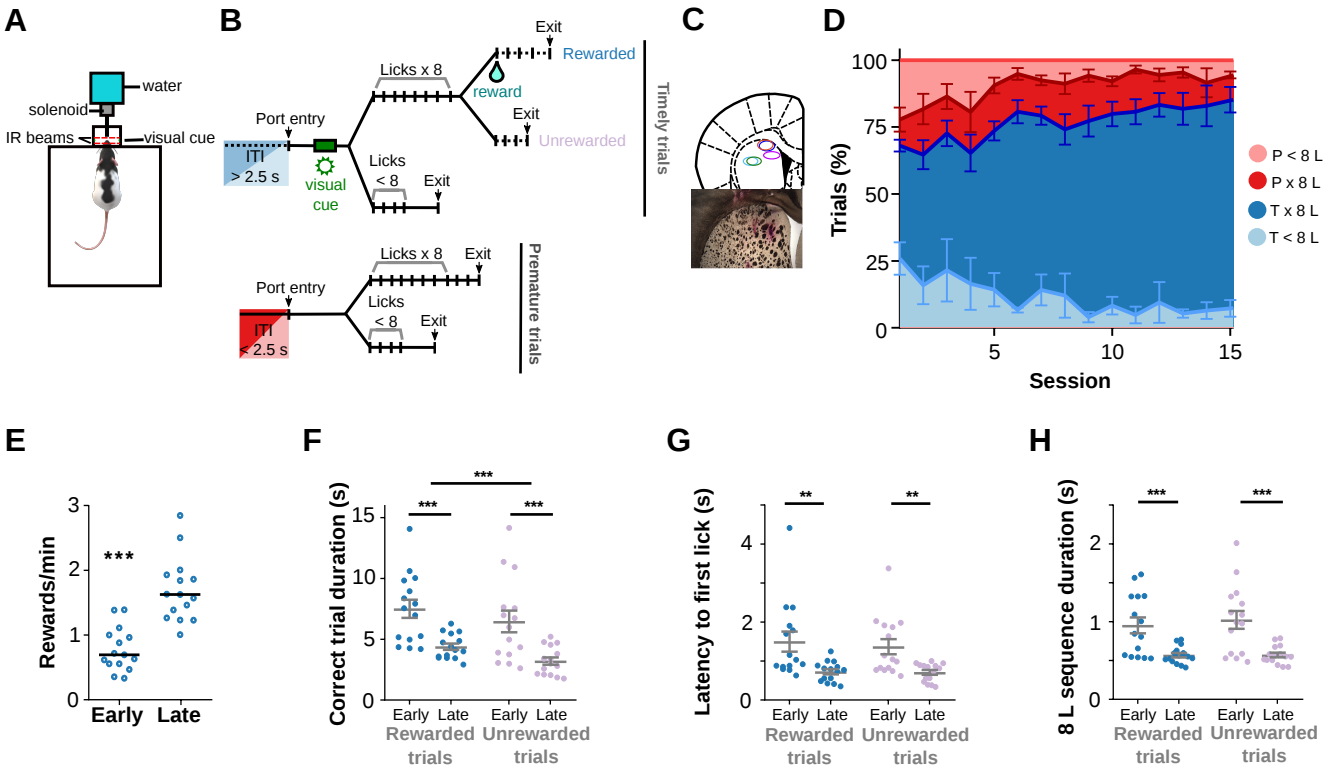


Figure 2

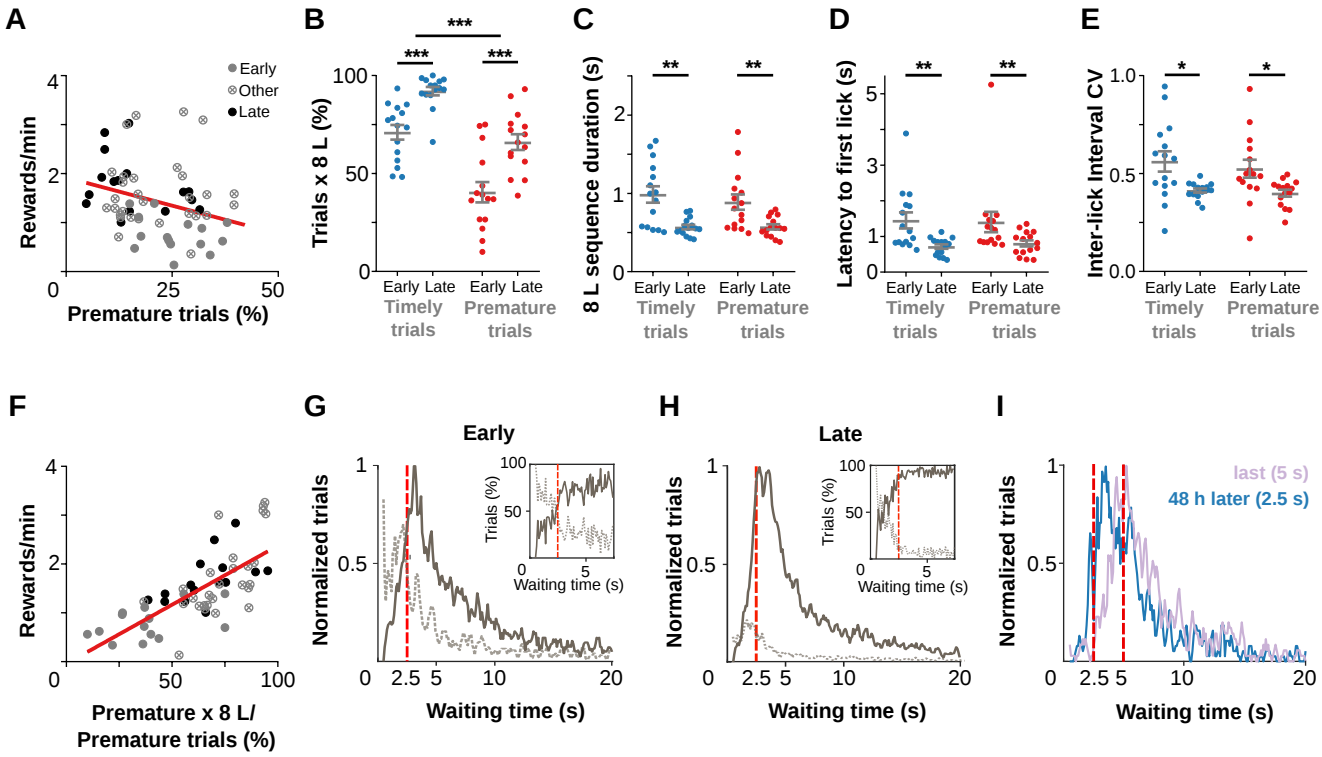


Figure 3

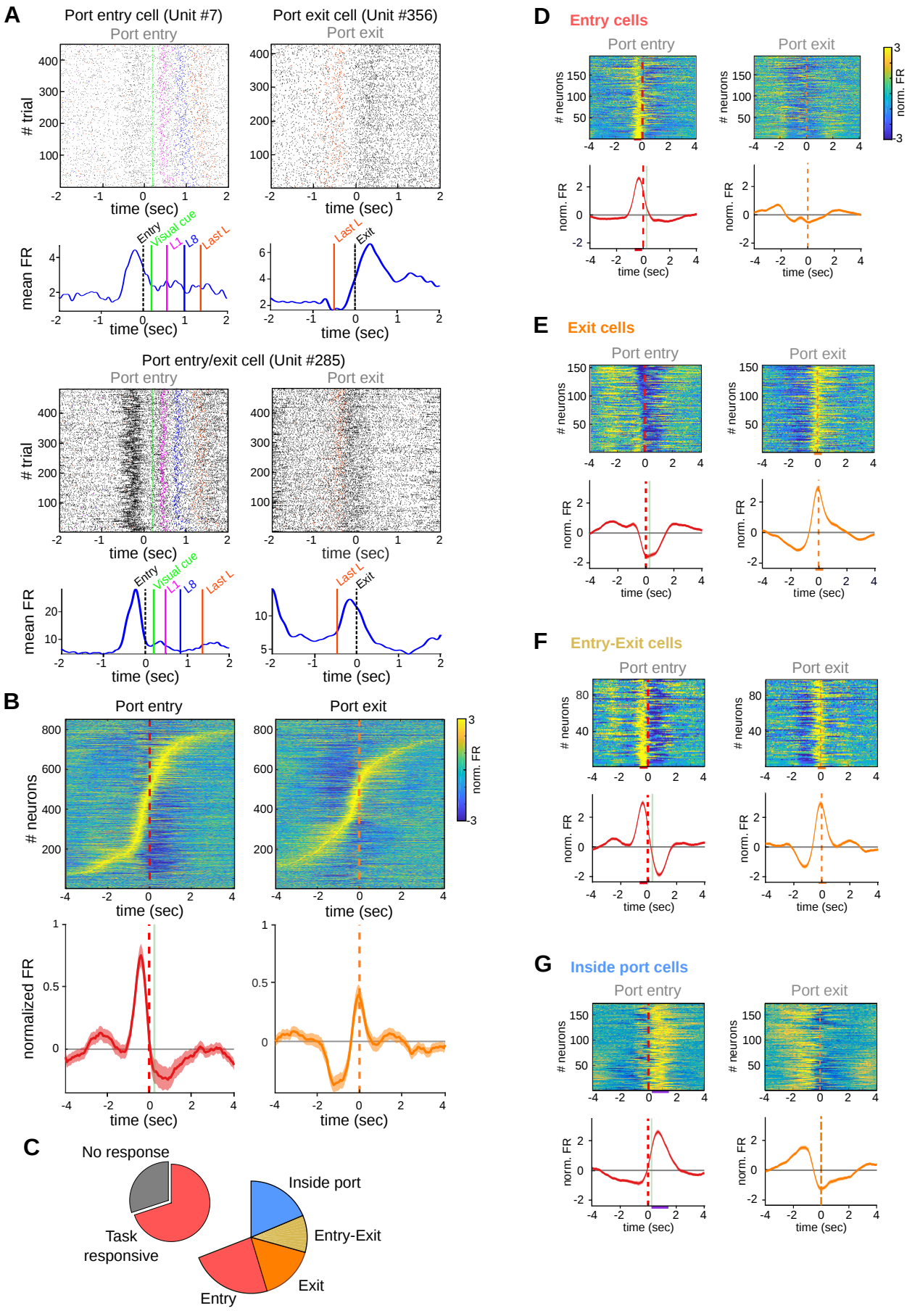


Figure 4

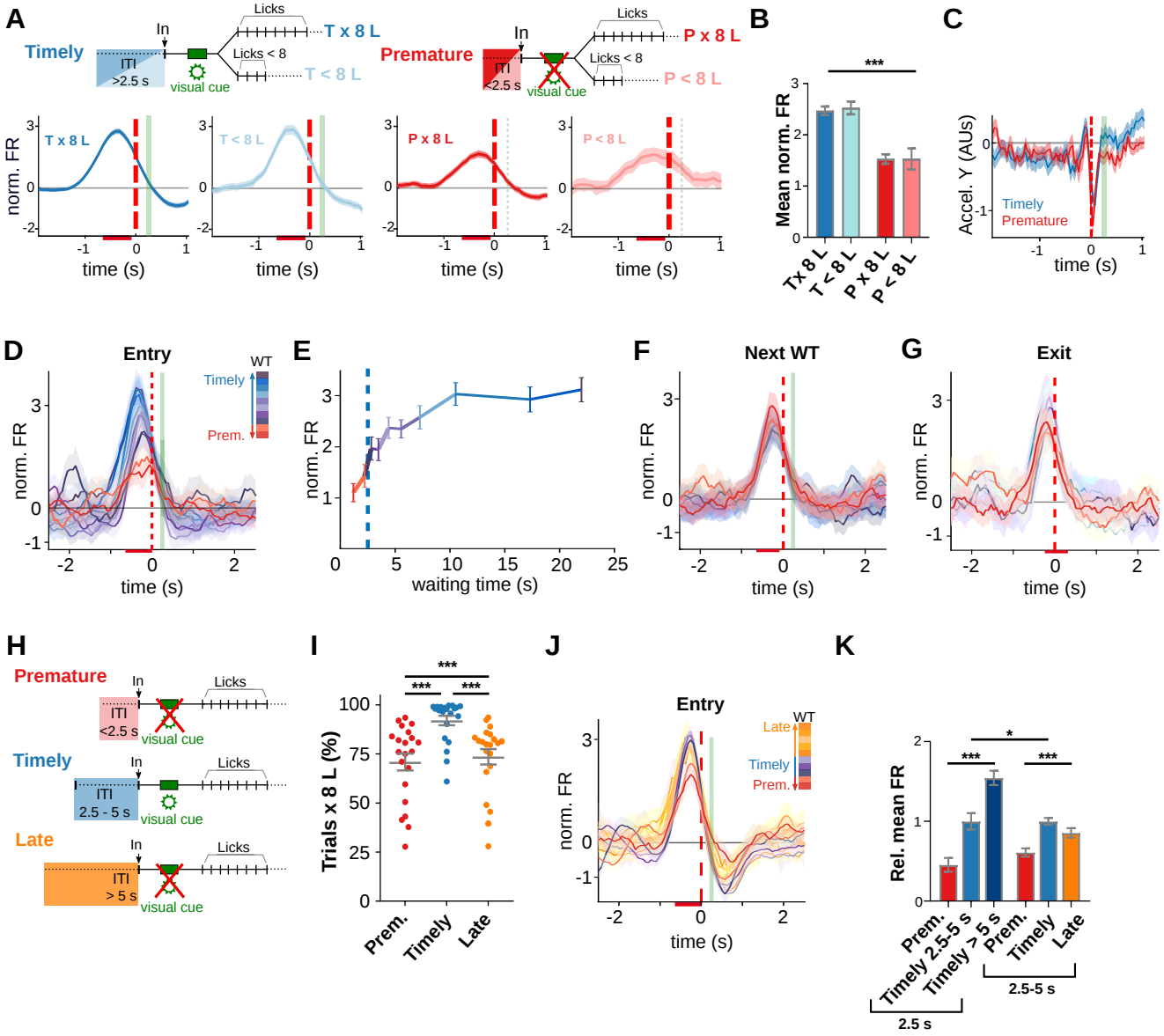


Figure 5

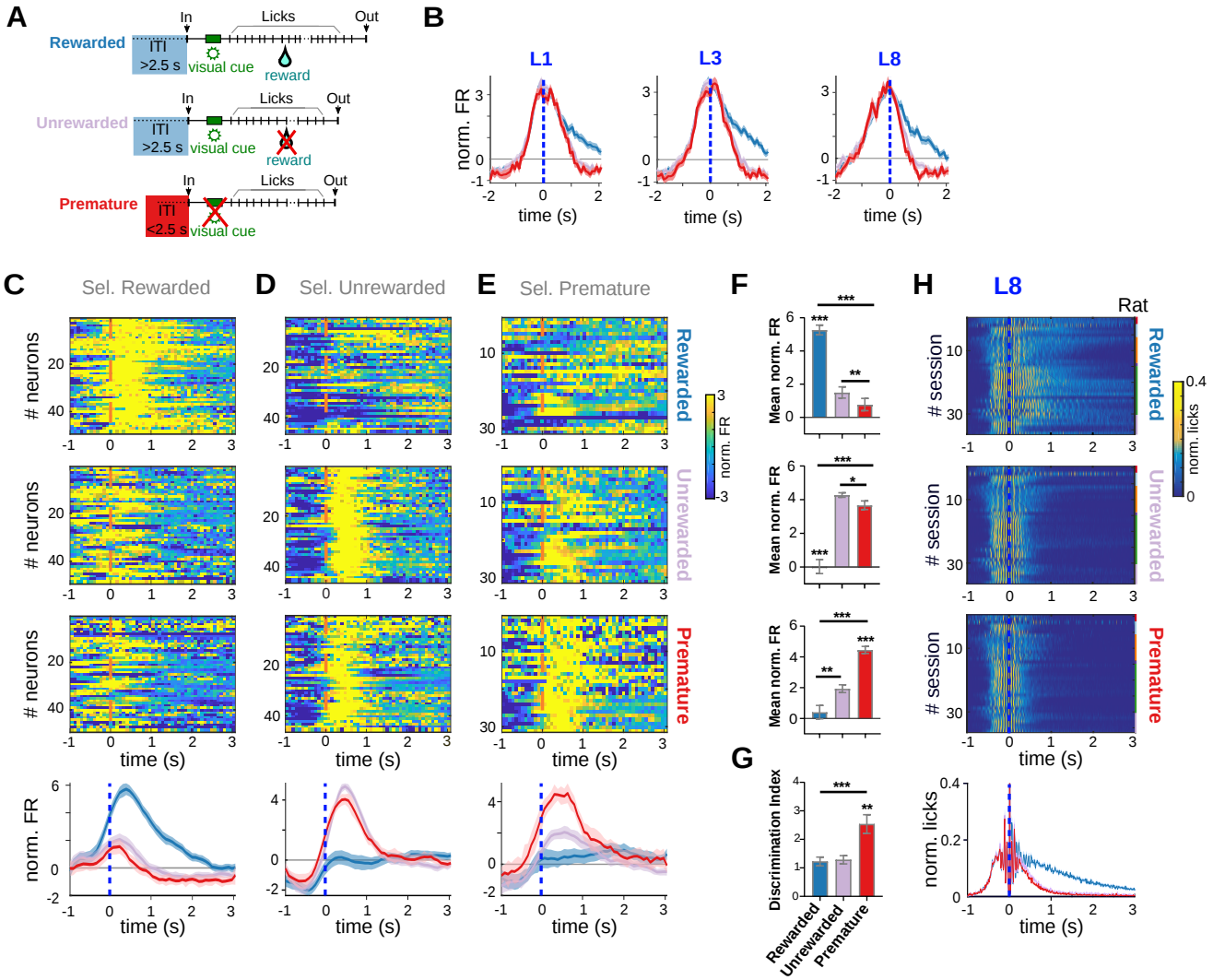


Figure 6

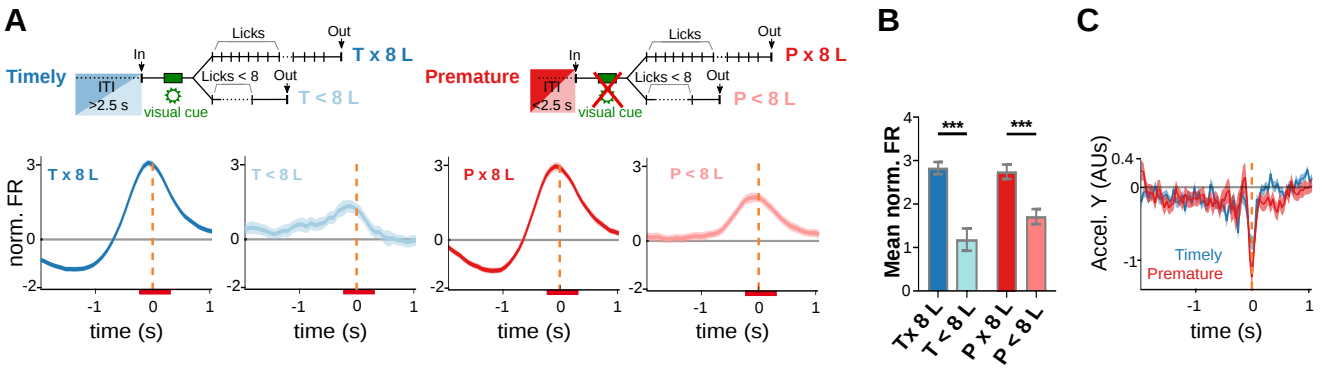


Figure 7

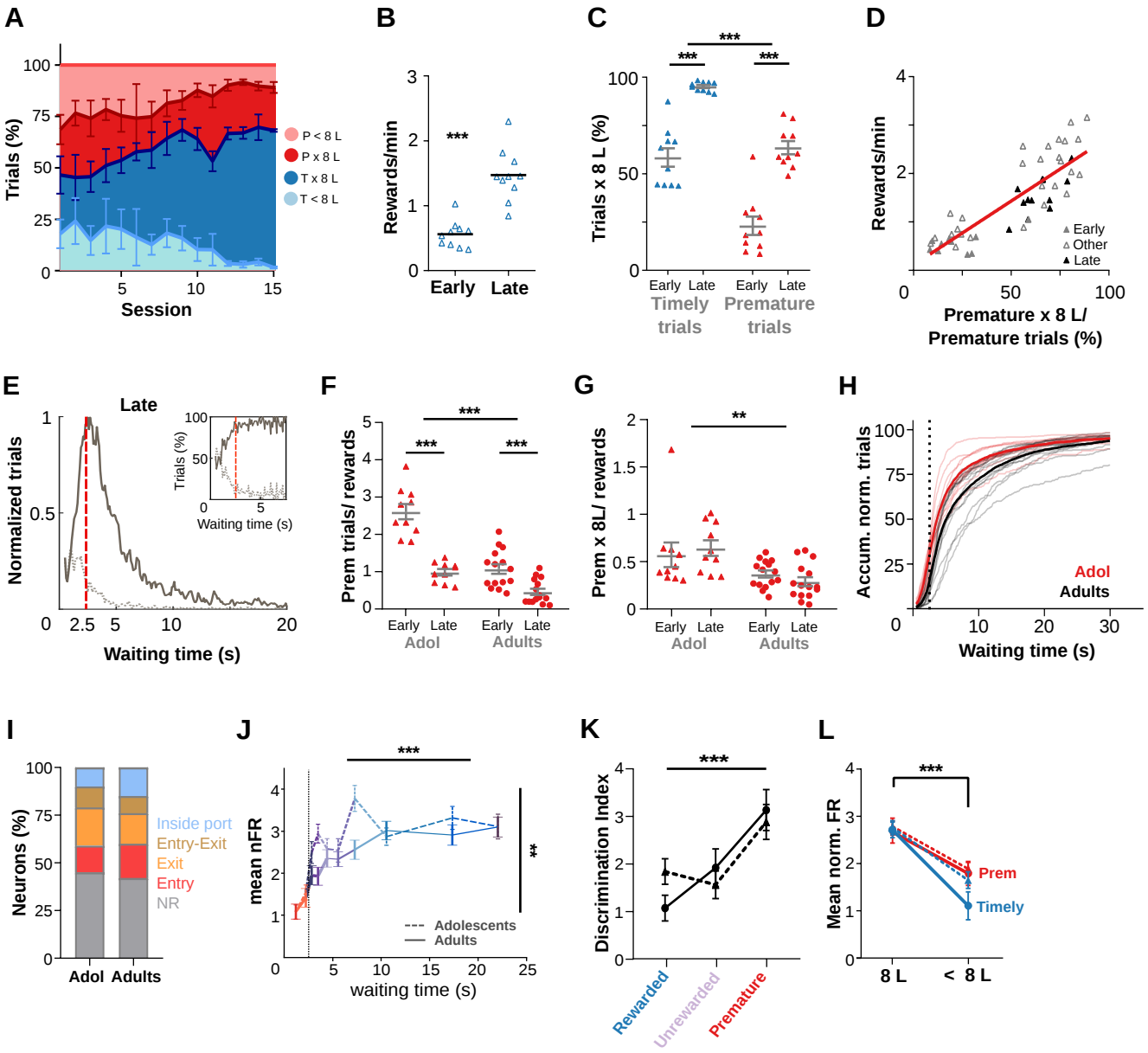


Figure S1

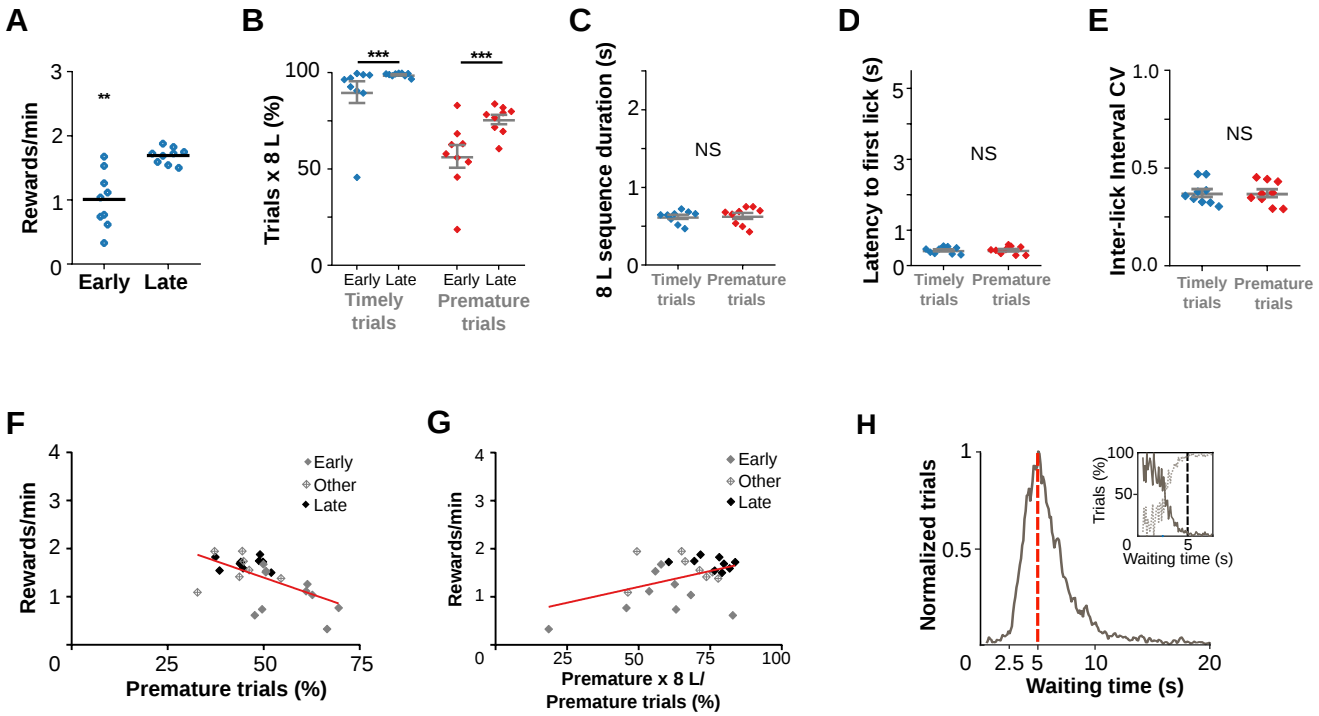


Figure S2

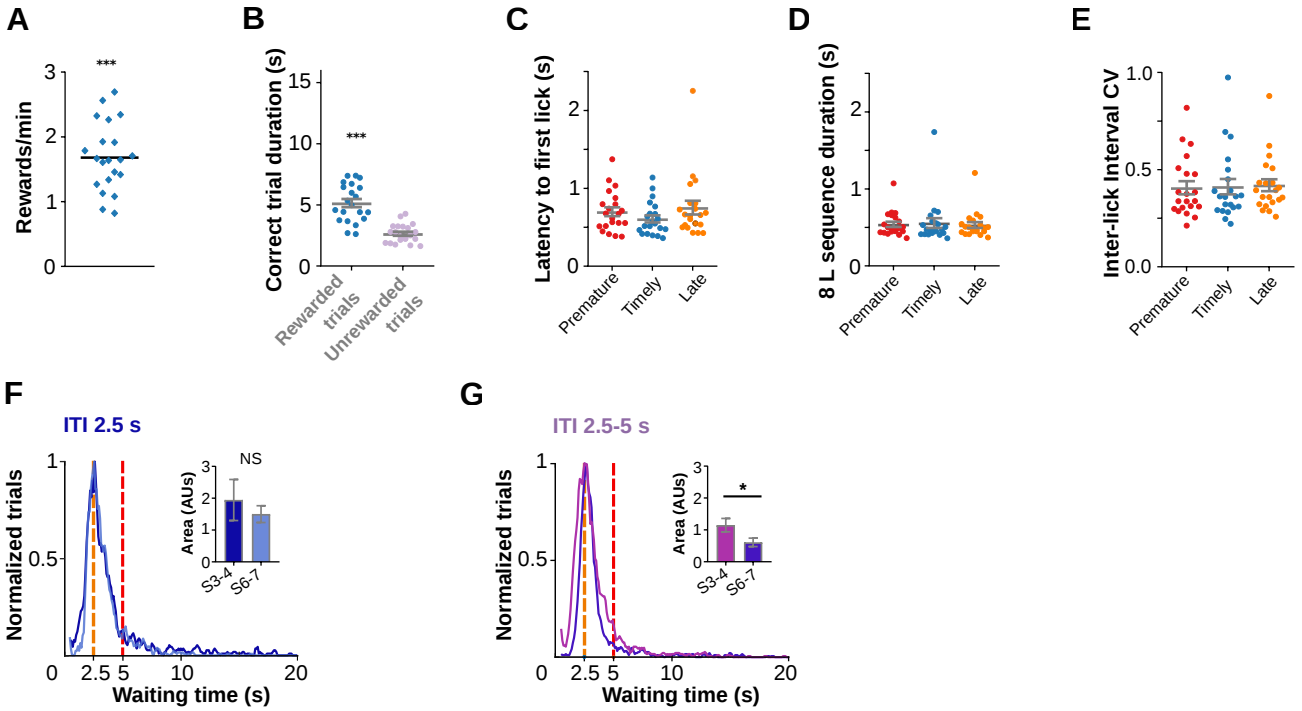
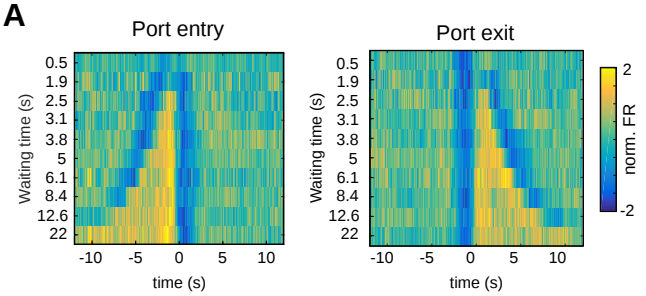


Figure S3

A



B

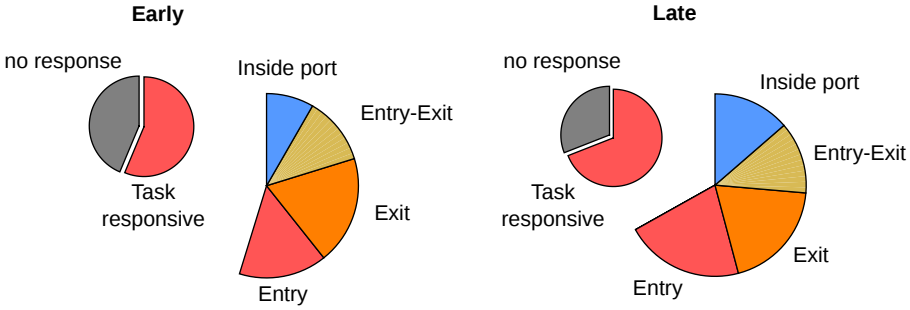


Figure S4

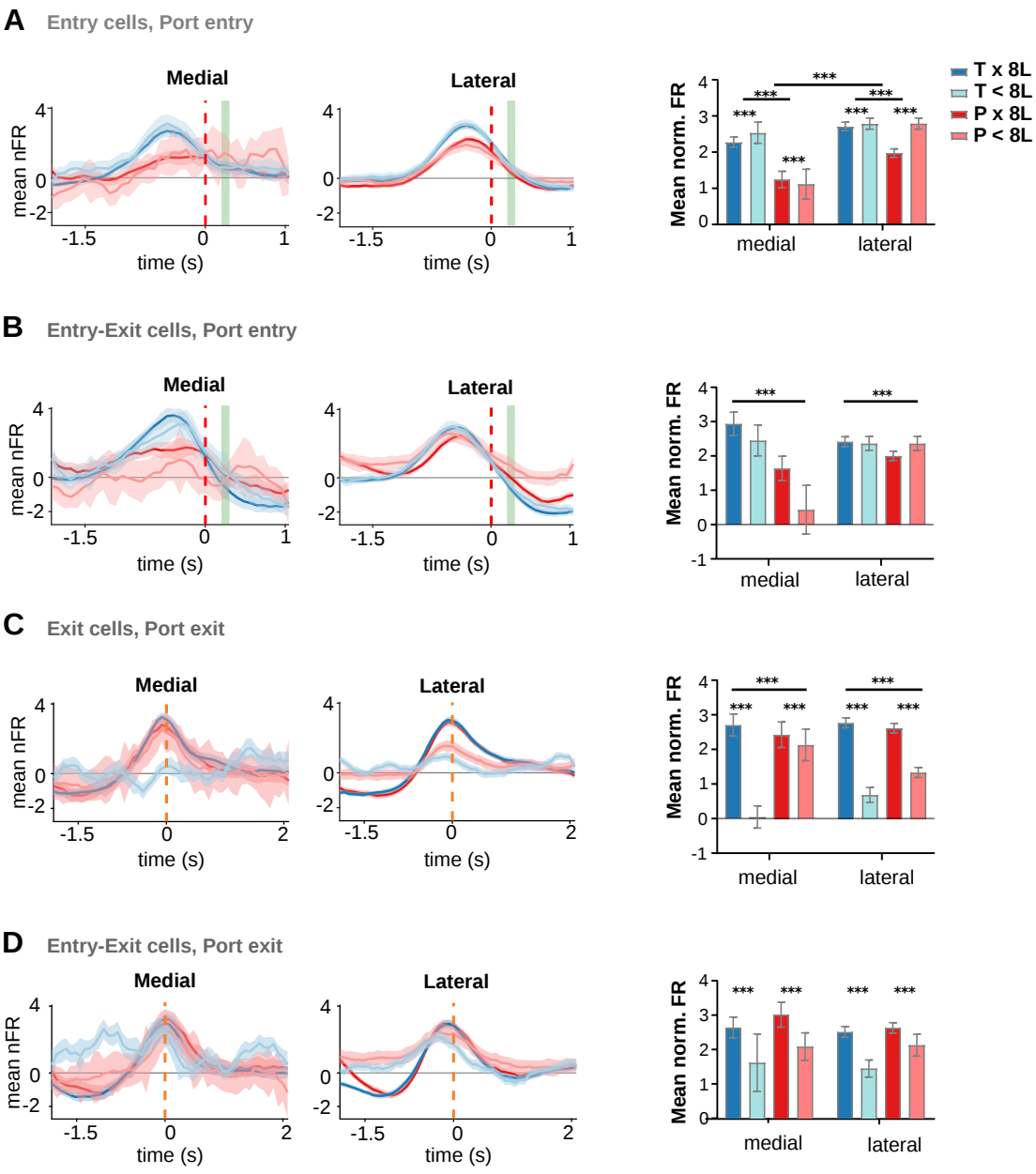


Figure S5

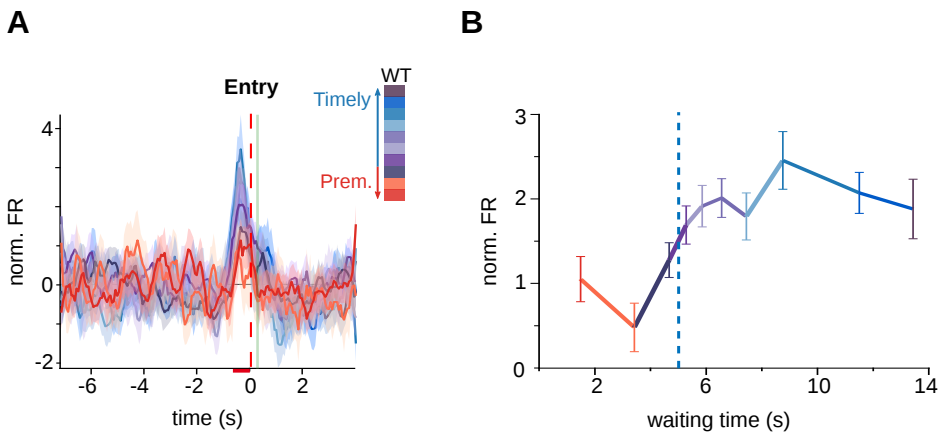


Figure S6

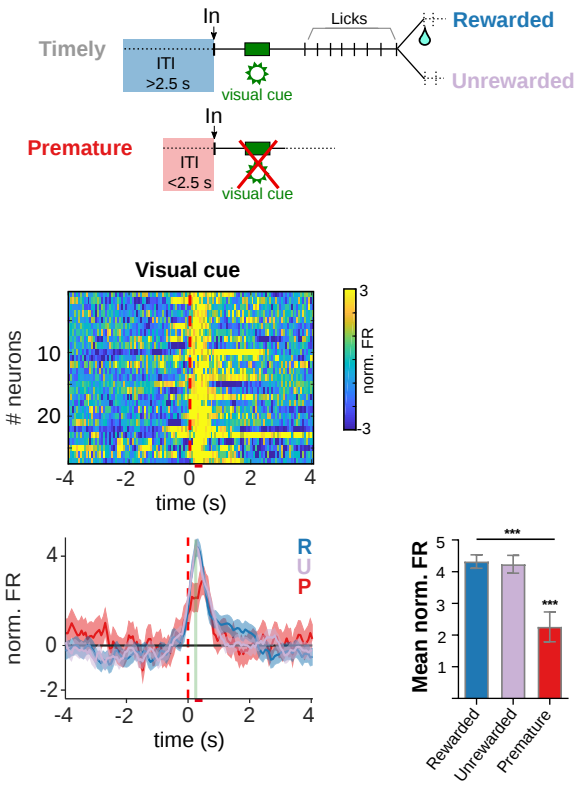


Figure S7

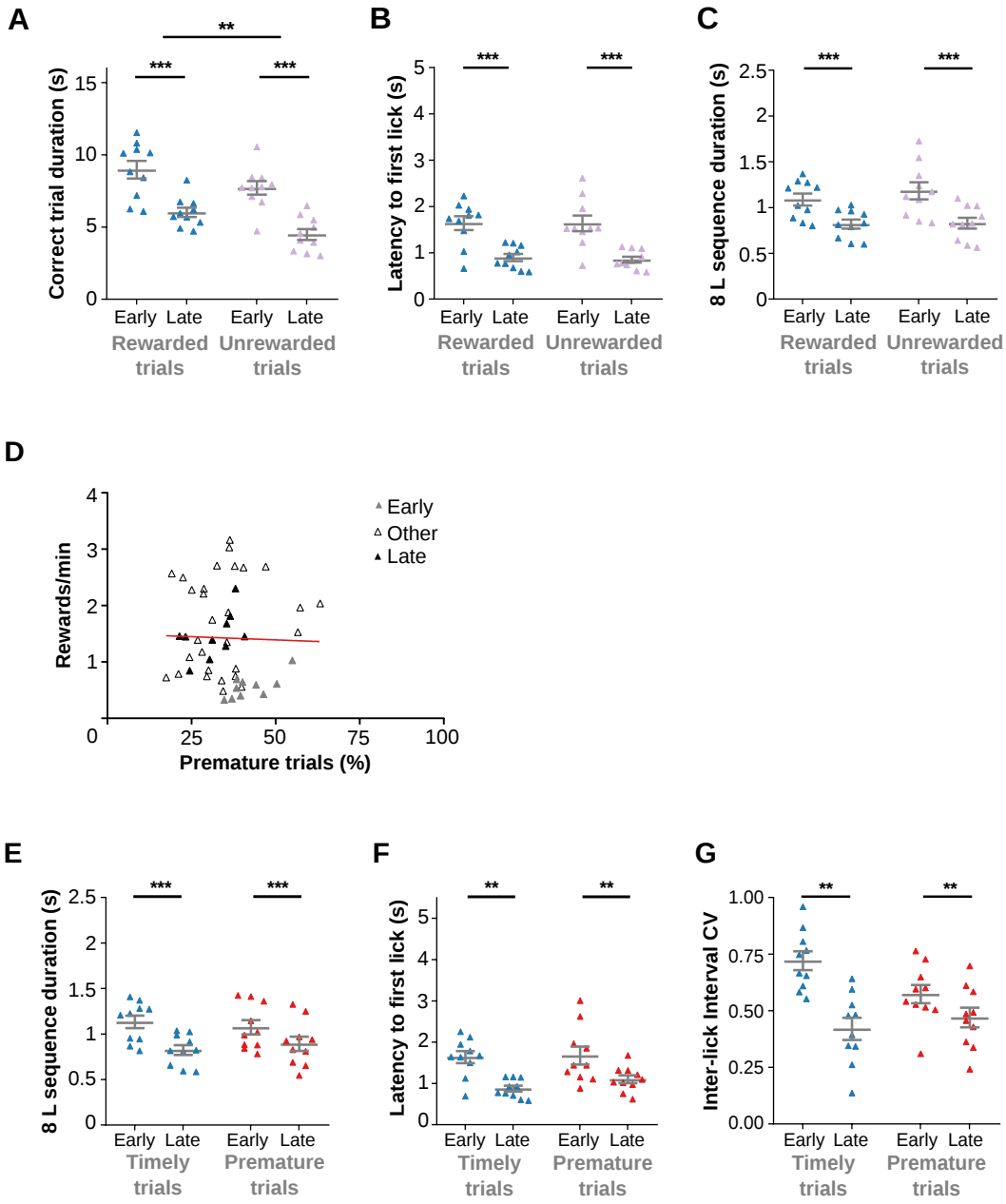


Figure S8

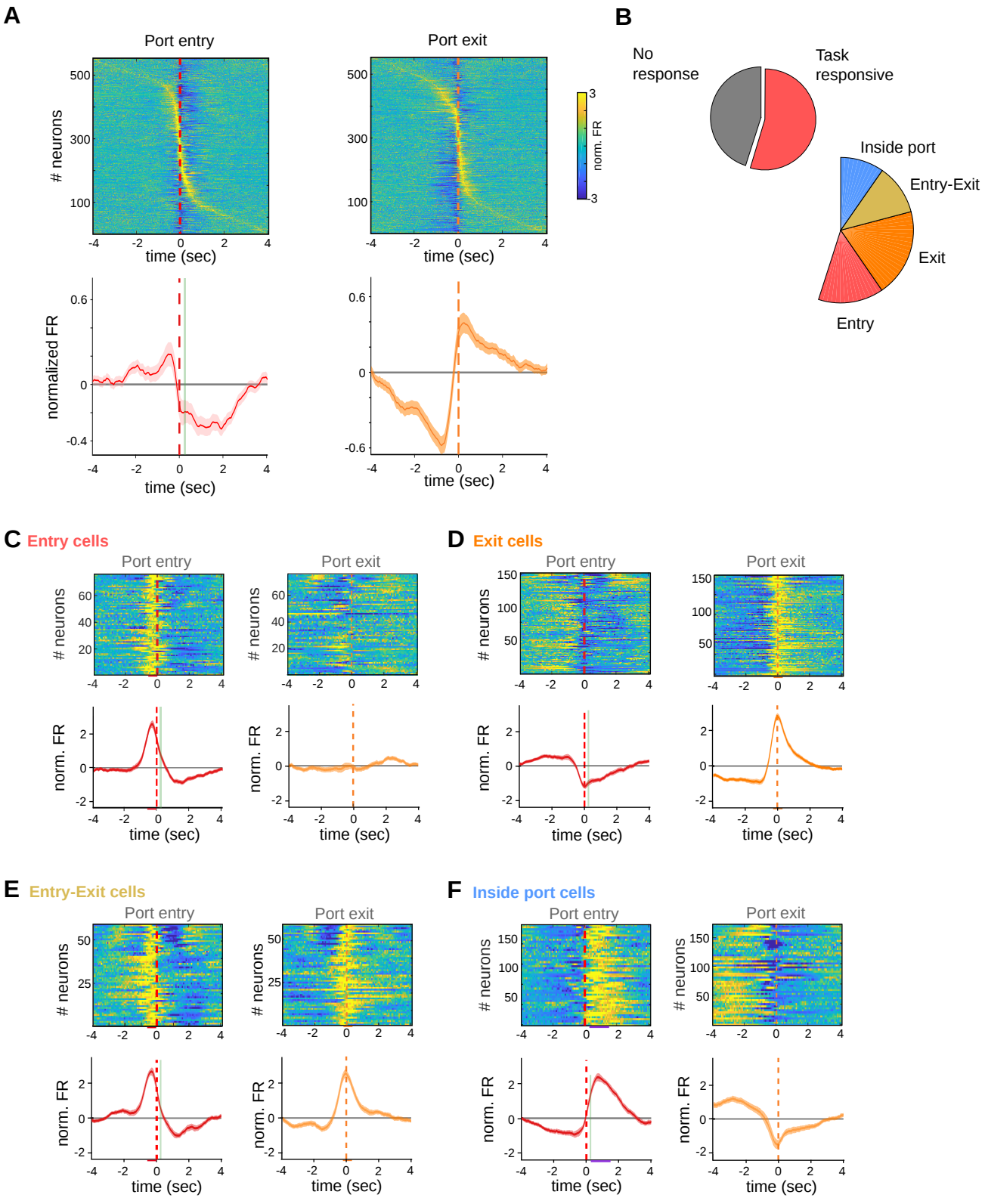


Figure S9

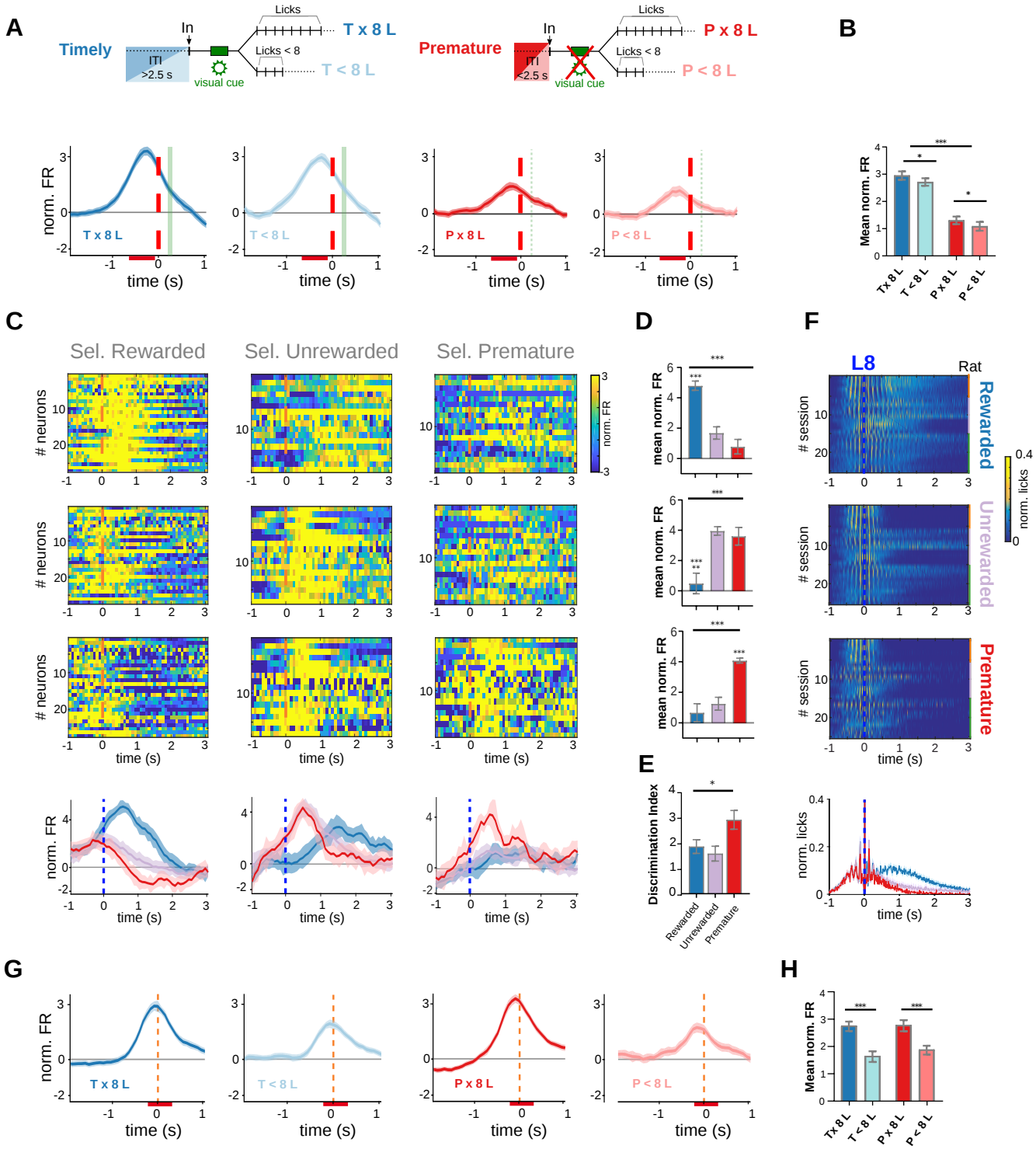
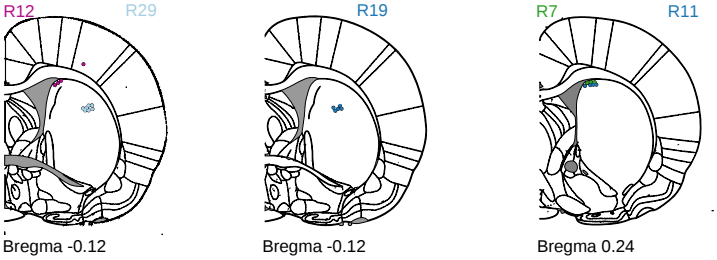
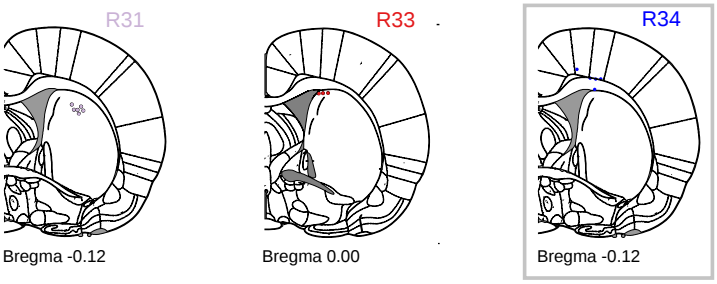


Figure S10

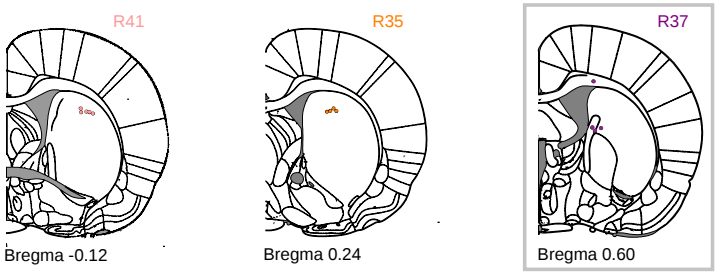
Adults 2.5s WT



Adults 5s WT



Adults 2.5- 5s WT



Adolescents 2.5s WT