Force tuning explains changes in phasic dopamine signaling during stimulus-reward learning

Authors: Konstantin Bakhurin[1], Ryan N. Hughes[1], Qiaochu Jiang[1], Isabella P. Fallon[1,2], Henry H. Yin[1,2]

Affiliations:

[1]Department of Psychology and Neuroscience, Duke University, Durham, NC, 27708, USA.,

[2]Department of Neurobiology, Duke University School of Medicine, Durham, NC, 27708, USA.

Correspondence to: **hy43@duke.edu**

Keywords: reward; reward prediction error; ventral tegmental area; dopamine; Pavlovian conditioning; force; movement; mesolimbic

Abstract

According to a popular hypothesis, phasic dopamine (DA) activity encodes a reward prediction error (RPE) necessary for reinforcement learning. However, recent work showed that DA neurons are necessary for performance rather than learning. One limitation of previous work on phasic DA signaling and RPE is the limited behavioral measures. Here, we measured subtle force exertion while recording and manipulating DA activity in the ventral tegmental area (VTA) during stimulus-reward learning. We found two major populations of DA neurons that increased firing before forward and backward force exertion. Force tuning is the same regardless of learning, reward predictability, or outcome valence. Changes in the pattern of force exertion can explain results traditionally used to support the RPE hypothesis, such as modulation by reward magnitude, probability, and unpredicted reward delivery or omission. Thus VTA DA neurons are not used to signal RPE but to regulate force exertion during motivated behavior.

Introduction

The ventral tegmental area (VTA) is the source of the mesolimbic dopamine (DA) pathway that has been implicated in reward and motivation [1]. According to the influential reward prediction error (RPE) hypothesis, VTA DA neurons encode the difference between actual and predicted reward, providing a teaching signal for associative learning [2,3]. However, others propose that DA contributes to vigor, incentive salience, or movement kinematics [4-7].

One reason for conflicting opinions on DA function is the lack of precise and continuous behavioral measurements. Many previous studies used head-fixed animals and behavioral measures that are usually limited to arm movements[8-10] or licking [3,11]. However, head restraint does not prevent mice from moving or attempting to move their head. In studies using freely moving animals, behavioral measures are limited to licking or time spent in the reward port [12-16]. When continuous behavioral measures were used with high temporal and spatial resolution, DA activity in the substantia nigra pars compacta was found to be highly correlated with vector components of head velocity, regardless of reward prediction or outcome valence, and selective stimulation of DA neurons generated movements [6]. Recent work using sensitive measures of force also showed that VTA DA neurons represent and regulate force exertion in head-fixed mice [17,18]. These results suggest that previous observations in support of the RPE hypothesis could be explained by the contributions of DA signaling to online behavioral performance rather than learning.

In this study, we used force sensors to measure subtle movements in head-fixed mice during a Pavlovian stimulus-reward learning task previously used to study RPE signaling. Using optogenetics and *in vivo* electrophysiology, we recorded and manipulated the activity of VTA DA neurons.

Results

Similar to previous work on RPE, mice were trained in a Pavlovian conditioning task: a conditioned stimulus (CS: 200 ms tone) is followed 1 second later by an unconditioned stimulus (US: sucrose reward) received just in front of the mouse (**Figure 1A**) [3]. The head-fixation device has load cells that continuously measure forces exerted through the head-bar (**Figure 1B)**[17-20]. Such force exertion was not required to earn reward or to reach the spout, but reflect approach and consummatory behaviors. The reward spout was comfortably positioned ~1 mm away from the lower jaw and was easily accessible with licking alone (**Supplemental Figure 1**).

We could quantify changes in force exertion during stimulus-reward learning (**Supplemental Figure 1**). The mice exerted forward forces in anticipation of reward and applied additional force following reward delivery while they consumed the water (**Figure 1C, Supplemental video 1**). The patterns of force exertion changed systematically during training (**Figure 1D**). Force changes in response to the CS increased with training (**Figure 1D, 1E**), but responses to the US decreased (**Figure 1D, 1E**). Thus, force exertion became more anticipatory during learning. Learning the task specifically involved changes in how force was exerted in response to CS and US. These subtle changes in movements have never been quantified in previous studies. They are not captured by standard measures of the conditioned response (CR) or unconditioned response (UR) such as licking (**Supplemental Figure 1**).

Mice also generated spontaneous movements during the inter-trial-interval (**Figure 1F & G**). While the relative frequency of spontaneous movements decreased during training, their amplitude did not change, even in well-trained mice (**Figure 1H**, **Supplemental Figure 1**).

Force tuning in VTA DA neurons

We recorded single unit activity from the VTA in mice with moveable optrodes and used optogenetic stimulation to confirm cell type (*n* = 1285 single units; n = 630 from putative DA neurons; *n* = 159 from tagged DA neurons; **Figure 2A & Supplemental Figures 2 & 3**)[17,18,20]. Because single units recorded during each session were counted independently, some of the single units may reflect repeated recordings from the same neurons.

We identified two major classes of DA neurons based on their unique direction-specific relationships to spontaneous force exertions (**Supplemental Figure 3**). One class of DA neurons (Forward DA, n = 300) produced a burst just prior to spontaneous forward movements (**Figure 2B, left**). In contrast, during spontaneous backward movements Forward DA neurons decreased their firing (**Figure 2B, right**). A second class of DA neurons (Backward DA, n = 116) showed the opposite pattern of activity, suppressing their firing pattern during spontaneous forward movements (**Figure 2C, left)** and increasing their firing (**Figure 2C, right**) during spontaneous backward movements. These two populations show opponent signaling depending on the direction of force exertion (**Figure 2D**). They made up ~65% of confirmed DA neurons (**Figure 2E**). Many of the remaining neurons recorded in the VTA also showed activation to spontaneous movements (**Supplemental Figure 3**). Some had high firing rates and short-duration spike waveforms, properties associated with GABAergic projection neurons. Others were either not tagged and failed to show direction-specific force tuning. These neurons were not analyzed in this study.

Forward and Backward neurons showed significant correlations with change in force in their preferred direction (**Figure 3A & B**). Neuron representations of force by these populations were "rectified," in that their firing rates had a linear relationship with force in their preferred direction but were uncorrelated with force in the opposite direction (**Figure 3C & D**).

109

110 **Force tuning explains changes in phasic DA activity during learning**

111 To determine whether these DA neurons show similar force tuning during the stimulus-
112 reward task, we next analyzed force after CS and US presentation. Forward DA neurons again
113 showed low correlations with backward movements that mice produced both early and late in
114 learning (**Figure 3E**, Early: $R^2 = 0$; Late: $R^2 = 0.26$) but showed correlation with change in
115 forward force (Early: $R^2 = 0.95$; Late: $R^2 = 0.94$). Backward DA neurons showed linear
116 correlations with Backward movements observed during the task (**Figure 3F**) but were less
117 correlated with forward movements. Together these results show that force tuning in VTA DA
118 neurons was consistent across task-related and spontaneous movements.

119 DA neurons are known to show burst firing following CS presentation [21,22] during
120 stimulus-reward learning and reduce their activity when rewards become better predicted. Such
121 changes have been attributed to signaling an RPE [3,8,22]. But our results suggest that changes in
122 phasic DA may be explained by changes in performance rather than learning. In agreement with
123 previous reports, Forward DA activity steadily increased during learning in response to the CS
124 and decreased slightly after the US (**Figure 4A & B**). We found evidence in support of a shift of
125 phasic DA activity from the US to the CS (**Figure 4C**). But these changes are accompanied by
126 changes in the way force is exerted (**Figure 1, Figure 4D & 4E**). Initially there was more force
127 exertion after reward delivery. With training, this behavior becomes more anticipatory, as the CR
128 increases in magnitude while the magnitude of the UR reduces (**Figure 4F**). The increasing force
129 CR and decreasing force UR can explain the changes in firing of DA neurons during stimulus-
130 reward learning.

131 Several additional features of Forward DA neurons indicated their role in generating
132 movements during the task. For example, regardless of the extent of training, Forward DA
133 neurons were only activated by the CS if the animal made a subsequent CR (**Supplemental**
134 **Figure 4**). Even well-trained mice occasionally did not show anticipatory movement as
135 measured by force exertion. In such cases, Forward DA neurons were not activated
136 (**Supplemental Figure 4**). Sometimes mice were already engaged in spontaneous movements at
137 the start of the trial, as detected by force sensors, so CS presentation interrupted the spontaneous
138 movements. When this occurs, Forward DA neurons also showed less activity.
139 With learning the fraction of trials containing a CR increased, and the latency of the force
140 CR became less variable (**Supplemental Figure 4).** The probability of burst activity in DA
141 neurons was correlated with the fraction of trials containing CRs. Training reduced the variance
142 in burst latency following the CR, as the timing of the movements became more consistent. On
143 the other hand, we did not observe any changes in firing rate or bursting probability of Backward
144 DA neurons as mice learned, likely because training resulted in increased forward movements
145 reflecting anticipatory approach behavior (**Supplemental Figure 5**).
146

147 **Force exertion explains how DA firing is modulated by reward probability or size**

148 Prior work found that phasic DA activity was modulated by reward probability[9]. We
149 manipulated reward probability and compared its effect on force exerted and DA activation
150 (**Figure 5A**). The duration of the UR was extended when mice received rewards only 50% of the
151 time (**Figure 5B**). Mice produced lower CR force when reward probability was 50%, but then
152 generated larger UR force upon reward delivery (**Figure 5C**). These patterns were paralleled by
153 similar changes in firing rate of Forward DA neurons (**Figure 5D**). Firing rates after the CS are

154 reduced when the CS signaled a 50% probability of reward. Once mice received the uncertain
155 reward, however, DA neurons showed higher activity. The opposite pattern was observed when
156 the reward was delivered at 100% probability (**Figure 5E**).

157 Phasic DA signaling also changes with reward size manipulations[9,11]. Previous studies
158 did not quantify the behavioral changes associated with such manipulations other than licking.
159 We found that manipulations of reward size also altered movements generated by mice. The total
160 amount of forward force exerted after both the CS and US were greater for larger rewards
161 (**Supplemental Figure 6**) Increasing reward size resulted in a greater impulse (force over time)
162 both during the CR and the UR (**Supplemental Figure 6**)[18]. Consistent with prior work,
163 Forward DA neurons increased their firing rates to the CS and to the US with a larger reward
164 (**Supplemental Figure 6**). Average firing rates for the Forward DA populations were greater
165 during trials with larger rewards (**Supplemental Figures 6**). On the other hand, Backward DA
166 populations did not respond differently to the changes in reward size (**Supplemental Figure 6**).
167

168 **Dissociating reward prediction error from force tuning**

169 To test whether phasic DA represents reward prediction or force exertion, we
170 manipulated the location of the reward spout while keeping the reward predictability constant.
171 Our head-fixed setup is flexible enough so that the mouse can still access the reward when the
172 spout is moved slightly behind the mouth. Although this change in spout location is only ~ 2
173 mm, it required a change in the direction of movement to access the reward (**Supplemental**
174 **Video 2**). Because the value of the outcome is identical, the RPE hypothesis predicts similar
175 phasic DA regardless of spout location. In contrast, if the DA neurons represent force vectors,

176 DA neurons with different force tuning (e.g. forward and backward) should be differentially
177 recruited when the spout is moved to a new location (**Figure 6A**).
178 Anticipatory licking rates (a conventional measure of the CR on this task) were similar
179 for both spout positions (**Figure 6B**), which reflects similar reward prediction. However, the
180 slight change in spout position resulted in different movements generated by the mice: when the
181 spout was moved backward by only 2 mm, mice generated more backward force and less
182 forward force (**Figure 6C**). As expected based on force tuning, Forward DA neurons showed
183 higher firing rates when the spout was positioned in front than when it was behind the mouth
184 (**Figures 6D & E**). In contrast, Backward DA neurons showed greater activation when mice
185 were required to exert force in the backward direction (**Figures 6F & G**). These opponent
186 relationships were consistent with how these neurons fired during spontaneous movements
187 (**Supplemental Figure 7**) and further demonstrate that the activity of DA neurons in reward
188 learning tasks are related to online performance rather than learning.
189
190 **Phasic DA to aversive stimuli is explained by changes in force direction**
191 If DA activity reflects force exertion, a similar relationship between DA and force should
192 be observed regardless of whether the outcome is rewarding or aversive. According to the RPE
193 hypothesis, an unexpected aversive stimulus should produce a negative prediction error, reflected
194 in a decrease or pause in VTA DA activity [23,24]. To test this possibility, we delivered aversive air
195 puffs on separate sessions (**Figure 7A**) [18]. Unexpected air puffs resulted in a movement backwards
196 away from the source of the air puff, followed by rebound movement forward (**Figure 7B**). This
197 pattern was observed across all mice (**Figure 7C**). Latency measures confirmed the backward
198 component had a shorter latency than the rebound forward component (**Figure 7D**). Contrary to
199 the RPE hypothesis, we found that the direction-selective DA neurons reflected the temporal

200 patterns of force changes during an aversive air puff (**Figure 7E**). The Backward DA population
201 was activated first during the initial backward movement, followed by the Forward DA population
202 that was activated prior to the onset of forward movement (**Figure 7F**). Forward DA neurons and
203 Backward DA neurons did not differ in their response magnitudes (**Figure 7G**). Lastly, we also
204 calculated force tuning in these two populations for movements generated after the air puff.
205 Consistent with our prior results (**Figure 3**), Forward and Backward DA neurons maintained
206 tuning for changes in force in their preferred direction (**Figure 7H**). Our results agree with past
207 research showing activation of DA neurons in response to aversive stimuli as well as rewarding
208 stimuli, [3,6,18,25] but we now show that previous observations could be explained by distinct types
209 of DA neurons signaling force in different directions.

210 We also noticed additional features of DA neuron activity that were inconsistent with
211 RPE interpretations. Some Forward DA neurons exhibited multiple bursts after the reward when
212 mice occasionally generated persistent movements (**Supplemental Figure 8**). We also observed
213 that some Backward neurons exhibit additional 'hump'-like firing during movements backward
214 under 'spout behind' conditions. These observations could only be explained by the pattern of
215 force exertion. However, because they were rare, we did not analyze the results further.

216

217 **Reward omission reveals 'dips' in DA firing and force**

218 A common interpretation of DA's role in learning is reduced firing after reward omission
219 that is thought to signal a negative RPE (actual reward less than predicted) [22]. In some well-
220 trained mice, we also performed experiments in which the reward was sometimes omitted after
221 the CS (**Figure 8A**). Forward DA neuron populations showed reduced activity after reward
222 omission (**Figure 8B & C**). Interestingly, upon reward omission, mice abruptly terminated force

223 exertion, and did not generate additional force which is normally observed after reward delivery
224 (**Figure 8D**). The change in force is therefore negative following reward omission (**Figure 8E**).
225 This shows that the characteristic 'dip' in DA neuron activity after reward omission (**Figure 8F**)
226 coincides with a pronounced 'dip' in force exertion (**Figure 8G, Supplemental Video 3**).
227 Backward DA neurons also showed reductions in firing after reward omission, but they were
228 more modest (**Supplemental Figure 8**), which suggests that mice stop forward force exertion
229 rather than producing a backward movement.
230

231 **Bidirectional optogenetic manipulation altered force exertion without affecting learning**

232 According to the RPE hypothesis, DA serves as a teaching signal to adjust learned
233 associations[2]. We tested whether stimulating DA neurons in place of a sucrose reward after CS
234 presentation would lead to learning of the CS-stimulation association, as reported previously[13].
235 We trained both VTA *DAT-Cre* with ChR2 expression *(n* = 3*)* and *WT* mice (n = 5) by delivering
236 a tone (CS) followed by stimulation instead of sucrose reward (**Figure 9A**). Stimulation did not
237 produce a learned association, as mice did not generate any licking (**Figures 9B**) or anticipatory
238 forward movement to the CS (**Figure 9C**). After 400 trials, we then introduced a sucrose reward
239 in addition to stimulation. When compared with control mice, stimulation did not affect the
240 learning rate when the reward was also delivered (**Figure 9C**). Previous work has shown that
241 activation of DA neurons could produce movement[6,18]. In agreement with such observations,
242 optogenetic stimulation 4 seconds after reward delivery (**Figure 9D**) also produced forward force
243 (**Figure 9E**). Moreover, each light pulse resulted in increased force (**Figure 9F**).

244 To further test whether VTA DA signaling is necessary for learning, we used an
245 inhibitory channelrhodopsin (SIO-*stGtACR2* [26], *N* = 4 DAT-cre mice, control: DIO-GFP in N = 6
246 DAT-Cre mice) to inhibit VTA DA neurons during the CS-US interval (**Figure 9G**). Inhibition

247 produced backward movement, but did not impair learning (**Figure 9H**). In fact, the mice with
248 DA inhibition licked more after inhibition than controls, contrary to what the RPE hypothesis
249 predicts (**Figure 9I**). Inhibition also reduced their latency to backward movement compared to
250 controls (**Figure 9J**). According to the RPE hypothesis, if DA neurons are inhibited immediately
251 after reward delivery, there should be a negative RPE and reduce anticipatory licking on future
252 trials [27]. Contrary to this prediction, when we inhibited DA neurons at reward delivery to mimic
253 a negative prediction error (**Figure 9K**), anticipatory licking (CR) was not reduced (**Figures 9L**
254 **& 9M**). Together our optogenetic results show that phasic DA is neither necessary nor sufficient
255 for stimulus-reward learning.
256

257 **Discussion**

258 Using sensitive measures of force generated by head-fixed mice, we demonstrated
259 systematic changes in behavior and the activity of VTA DA neurons during a stimulus-reward
260 task (**Figure 1**). We found two populations of DA neurons whose activity precedes and predicts
261 force in different directions (**Figures 2 & 3**). This relationship between DA activity and force
262 generated is independent of learning. Although we found the classic pattern of burst firing
263 following reward presentation (US) and the reward-predicting cue (CS), such activity does not
264 signal RPE. Instead, the activity of VTA DA neurons regulate performance continuously during
265 reward-guided behavior (**Figures 4, 5, 7 & 8, Supplemental Figures 4 & 6**) and causally
266 contribute to force generation (**Figure 9**). By keeping reward prediction constant but changing
267 the direction of movement required by moving the spout, we were able to demonstrate
268 systematic changes in phasic DA activity based on direction of force exertion rather than reward
269 prediction (**Figure 6**). Force exertion could also explain modulation of DA activity by reward
270 probability or magnitude (**Figure 5, Supplemental Figure 6)**[11,28].

Our results suggest that phasic DA signaling is critical for regulating the direction and magnitude of force exertion continuously, independent of learning. The force tuning of DA neurons was not altered by learning. It was similar during spontaneous movements outside of the task and during movements in response to aversive air puffs. Finally, although it has been proposed that tonic DA activity is related to response vigor [29], our results suggest that tonic and phasic components of DA activity represent the same variables, and differ only in magnitude (**Figures 2, 3 & 4**).

Outcome valence

We showed that VTA DA neurons can increase firing to rewarding and aversive stimuli, but these activations are determined by subtle differences in characteristic force profiles (**Figures 7, 8, 10**), indicating DA activity is independent of outcome valence [3,25]. A recent study found phasic DA activity in response to aversive events when the probability of receiving a reward was high. In contrast, DA neurons were inhibited when the probability of reward was low [30]. However, in our study, air puff trials were run separately so the mice only received air puffs without having any probability of receiving a reward. If VTA DA neurons signals RPE, they should be inhibited by the air puff. We found, however, that they were excited and modulated by the direction of movements evoked by the air puff (**Figure 7**). Importantly, force tuning remains the same despite the change in outcome valence.

Reward omission

A widely reported observation supporting RPE interpretations of DA is the 'dip' in DA firing when mice do not receive an expected reward [22]. We found that mice abruptly changed the

direction of force exertion after reward omission. This negative change in force appears as a 'dip' that parallels the dip in DA activity at the expected time of reward delivery (**Figure 8**).

Recent work also found that distinct populations of VTA DA neurons that project to different sub-regions of the nucleus accumbens (medial and lateral shell) respond differently to aversive stimuli [31]. However, there was no continuous behavioral measure, and the only behavioral measure during the shock experiments was percent of time spent freezing. It is possible that animals may briefly move in different directions in addition to freezing, as we have found on air puff trials. Thus, VTA DA projections to distinct accumbens regions could be responsible for either forward (approach) or backward (avoidance) movements. In fact, our results suggest that with shock, the activation of opposing force systems may be in conflict and result in no overt movement.

Limitations of RPE models

Determining how learning impacts performance has long been a major challenge, as it is well established that changes in performance cannot be equated with learning [4,32]. In reinforcement learning models, the RPE is assumed to be a teaching signal, though in practice learning is often conflated with performance in such models. Reward prediction is assumed to be translated directly into performance, and the RPE that updates learning is assumed to change response probability [33].

There are two popular interpretations of the DA's role in reinforcement learning. According to the strong version, DA neurons not only signal RPE's, but also directly cause learning. According to the weak version, DA activity does not necessarily *cause* learning [34]. It is only one signal among many in a larger circuit for predictive learning. However, our results

317 are at odds with both accounts. By manipulating reward location while maintaining reward
318 prediction, we tested whether RPE or force generation can better explain DA activity (**Figure 6**).
319 If phasic DA encodes RPE, reversing the parameters of movement required to obtain reward
320 should not alter DA signaling, as reward prediction and the value of the reward remain similar as
321 indicated by conventional measures of the conditioned response like anticipatory licking [3,35].
322 Although this manipulation changed the location of the reward spout by only ~2 mm, it
323 significantly changed the direction of force exerted by the animal, as detected by highly sensitive
324 force sensors. As shown in **Figure 6,** the activity of VTA DA neurons also changed drastically
325 when the reward spout is moved, and this change can be explained by the change in the direction
326 of force exertion. More backward force is associated with increased activity of DA neurons
327 representing backward force, and more forward force increases the activity of DA neurons
328 representing forward force. Neither RPE nor reward prediction can explain this pattern of results.
329 In addition, the RPE hypothesis fails to predict several other observations, such as
330 multiple successive bursts after reward, and a lack of DA activation to the tone on trials without
331 CRS, even when the mice are well-trained (**Figure 2, Supplemental Figure 4, Supplemental**
332 **Figure 8**). These observations are compatible with the role of DA in continuous regulation of
333 performance online, rather than updating stimulus-reward associations.
334 Our optogenetic experiments also indicate that phasic VTA DA is neither necessary nor
335 sufficient for learning (**Figure 9**). According to the RPE hypothesis, if a burst occurs reliably
336 after a neutral stimulus, the stimulus should generate reward prediction. However, optogenetic
337 stimulation in place of reward did not produce any learning: there was no anticipatory licking,
338 and the rate of learning did not change compared to controls when the stimulation was paired
339 with reward (**Figure 9**). Optogenetic excitation and inhibition did however produce bidirectional

force changes. Force exerted during head-fixation is expected to result in changes in velocity during freely moving behavior, as previously reported [6,36].

Vectorial representation of force

RPEs used in traditional reinforcement learning models are scalar quantities. They do not explain force vector tuning. Tabular reinforcement learning models usually define actions as discrete events, leaving out the spatial and temporal dimensions necessary for generating behavior. With many degrees of freedom, the discretization of the action easily leads to explosive growth of action values and an action space that is too large to explore efficiently. Every possible combination of force direction and magnitude must be learned, and their value stored in memory for the appropriate action to be chosen, which represents an unrealistic computational demand for the nervous system.

Although more recent continuous reinforcement learning approaches attempt to address the limitations of tabular models[37,38], they still cannot explain our results. For example, using deep-Q learning, Lillicrap et al attempt to generate continuous control signals by mapping states to a probability distribution of torque values. However, millions of learning trials are needed just to generate a simple movement in simulation, and the learned policy is unlikely to work with unpredictable environmental disturbances. Recently, Lee et al argued that a vector RPE model can explain results showing heterogeneous coding of task variables but uniform reward responses[39]. They suggest that DA heterogeneity reflects a high-dimensional state representation with a distributed RPE code. However, our findings indicate that DA neurons do not show uniform reward responses. Their results fail to rule out whether the encoding of diverse variables, such as response to cues or movement, can be explained by the same underlying

363 variable of force generation, which was not measured. In our results, the phasic DA activity at
364 the time of CS, reward, or spontaneous movements can all be explained by the force vector.
365

366 **Limitations in traditional experimental designs and measures**

367 Our results suggest that traditional experimental design and analysis have obscured the
368 key behavioral changes during learning, resulting in misinterpretations of the role of phasic DA
369 in learning and performance (**Figure 10**). The lack of continuous behavioral measures with high
370 temporal and spatial resolution is a major limitation in previous research on the role of DA in
371 learning. Previous work relied on discrete measures of licking or limb movements [3,8]. More
372 recent work in head-fixed mice attempted to quantify behavior more thoroughly [21,40-42], but still
373 neglected direction-specific torso and head movements, which are critical in understanding the
374 contribution of DA neurons in tasks that are organized around feeding behaviors.

375 Earlier work attempted to monitor EMG signals from the arm in restrained monkeys and
376 found no relationships between arm muscle and DA activity [10]. We demonstrate that head-
377 restrained mice attempt to move their heads when attempting to obtain water with the mouth.
378 Measurements of arm muscle activity are inadequate if the key movements are generated by the
379 head and torso. Lacking adequate behavioral measures, previous work could not rule out
380 confounding factors like force exertion in different directions during performance and changes in
381 such performance variables during learning.
382

383 **Conclusions**

384 By using precise and continuous force measurements, we have shown that VTA DA
385 neurons can carry signals that specify both the magnitude and direction of behavior. They

386 suggest that previous results used to support the RPE hypothesis can be explained by a
387 performance account if the force exerted is precisely measured.
388 During stimulus-reward learning, there was a significant increase in anticipatory force
389 exertion, even though the force tuning of DA neurons remains the same. This observation
390 suggests that the learning mechanism takes place upstream of DA neurons, and the excitatory
391 drive to DA neurons in response to the CS increases with training. In turn, DA signaling may be
392 recruited to modulate the gain of corticostriatal transmission in generating approach behavior.
393 DA concentration in the striatal targets of the mesolimbic DA pathway is expected to reflect
394 force, which is roughly proportional to the integral of phasic DA activity[18].
395 Our results are in accord with previous research that has implicated VTA DA in effort,
396 vigor, or incentive salience [4,5,43-45]. However, these previous concepts lack detailed vectorial
397 representations needed to explain the spatiotemporal features of continuous behavior. Based on
398 our results, it is expected that DA would produce changes in movement velocity in unrestrained
399 mice [6,46]. Since force generation is an analog of approach and retreat behavior, and dissociable
400 from licking, our results are also in accord with previous findings that VTA DA contributes more
401 to approach rather than consummatory behaviors [47,48]. We hypothesize that, by acting as an
402 adaptive gain signal that modulates striatal output, VTA DA regulates the continuous
403 performance of anticipatory approach and avoidance [49]. This hypothesis can be tested by
404 determining how specific populations of DA neurons exert their effects on target striatal
405 projection neurons.
406
407
408

References

1 Morales, M. & Margolis, E. B. Ventral tegmental area: cellular heterogeneity, connectivity and behaviour. *Nature reviews. Neuroscience* **18**, 73-85, doi:10.1038/nrn.2016.165 (2017).

2 Schultz, W. Predictive reward signal of dopamine neurons. *J Neurophysiol* **80**, 1-27. (1998).

3 Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B. & Uchida, N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* **482**, 85-88 (2012).

4 Berridge, K. C. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology (Berl)* **191**, 391-431 (2007).

5 Salamone, J. D. & Correa, M. The mysterious motivational functions of mesolimbic dopamine. *Neuron* **76**, 470-485 (2012).

6 Barter, J. *et al.* Beyond reward prediction errors: the role of dopamine in movement kinematics. *Frontiers in Integrative Neuroscience* **9**, 39 (2015).

7 Panigrahi, B. *et al.* Dopamine Is Required for the Neural Representation and Control of Movement Vigor. *Cell* **162**, 1418-1430, doi:10.1016/j.cell.2015.08.014 (2015).

8 Waelti, P., Dickinson, A. & Schultz, W. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* **412**, 43-48. (2001).

9 Fiorillo, C. D., Tobler, P. N. & Schultz, W. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* **299**, 1898-1902 (2003).

10 Romo, R. & Schultz, W. Dopamine neurons of the monkey midbrain: contingencies of responses to active touch during self-initiated arm movements. *J Neurophysiol* **63**, 592-606 (1990).

11 Eshel, N. *et al.* Arithmetic and local circuitry underlying dopamine prediction errors. *Nature* **525**, 243-246, doi:10.1038/nature14855 (2015).

12 Mohebi, A. *et al.* Dissociable dopamine dynamics for learning and motivation. *Nature* **570**, 65-70 (2019).

13 Saunders, B. T., Richard, J. M., Margolis, E. B. & Janak, P. H. Dopamine neurons create Pavlovian conditioned stimuli with circuit-defined motivational properties. *Nature neuroscience* **21**, 1072 (2018).

14 Sharpe, M. J. *et al.* Dopamine transients do not act as model-free prediction errors during associative learning. *Nature communications* **11**, 1-10 (2020).

15 Steinberg, E. E. *et al.* A causal link between prediction errors, dopamine neurons and learning. *Nature neuroscience* **16**, 966-973 (2013).

16 Keiflin, R., Pribut, H. J., Shah, N. B. & Janak, P. H. Ventral tegmental dopamine neurons participate in reward identity predictions. *Current Biology* **29**, 93-103. e103 (2019).

17 Bakhurin, K. I., Hughes, R. N., Barter, J. W., Zhang, J. & Yin, H. H. Protocol for Recording from Ventral Tegmental Area Dopamine Neurons in Mice while Measuring Force during Head-Fixation. *STAR Protocols*, 100091 (2020).

18 Hughes, R. N. *et al.* Ventral Tegmental Dopamine Neurons Control the Impulse Vector during Motivated Behavior. *Current Biology* **30**, 1-14 (2020).

19 Hughes, R. N., Bakhurin, K. I., Barter, J. W., Zhang, J. & Yin, H. H. A head-fixation system for continuous monitoring of force generated during behavior. *Frontiers in integrative neuroscience* **14**, 11 (2020).

20 Madisen, L. *et al.* A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nat Neurosci* **15**, 793-802, doi:nn.3078 [pii]

10.1038/nn.3078 (2012).

21 Coddington, L. T. & Dudman, J. T. The timing of action determines reward prediction signals in identified midbrain dopamine neurons. *Nature neuroscience* **21**, 1563 (2018).

22 Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593-1599 (1997).
23 Matsumoto, M. & Hikosaka, O. Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature* **459**, 837-841 (2009).
24 Fiorillo, C. D. Two dimensions of value: dopamine neurons represent reward but not aversiveness. *Science* **341**, 546-549 (2013).
25 Horvitz, J. C. Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience* **96**, 651-656, doi:S0306452200000191 [pii] (2000).
26 Mahn, M. *et al.* High-efficiency optogenetic silencing with soma-targeted anion-conducting channelrhodopsins. *Nat Commun* **9**, 4125, doi:10.1038/s41467-018-06511-8 (2018).
27 Lee, K. *et al.* Temporally restricted dopaminergic control of reward-conditioned movements. *Nature Neuroscience* **23**, 209-216 (2020).
28 Tobler, P. N., Fiorillo, C. D. & Schultz, W. Adaptive coding of reward value by dopamine neurons. *Science* **307**, 1642-1645 (2005).
29 Niv, Y. Cost, benefit, tonic, phasic: what do response rates tell us about dopamine and motivation? *Ann N Y Acad Sci* **1104**, 357-376, doi:10.1196/annals.1390.018 (2007).
30 Matsumoto, H., Tian, J., Uchida, N. & Watabe-Uchida, M. Midbrain dopamine neurons signal aversion in a reward-context-dependent manner. *Elife* **5**, e17328 (2016).
31 de Jong, J. W. *et al.* A neural circuit mechanism for encoding aversive stimuli in the mesolimbic dopamine system. *Neuron* **101**, 133-151. e137 (2019).
32 Rescorla, R. A. *Pavlovian second-order conditioning: studies in associative learning*. (1980).
33 McClure, S. M., Daw, N. D. & Read Montague, P. A computational substrate for incentive salience. *Trends in neurosciences* **26**, 423-428 (2003).
34 Schultz, W. Dopamine reward prediction-error signalling: a two-component response. *Nature reviews. Neuroscience* **17**, 183-195, doi:10.1038/nrn.2015.26 (2016).
35 Tobler, P. N., Dickinson, A. & Schultz, W. Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. *J Neurosci* **23**, 10402-10410 (2003).
36 Dodson, P. D. *et al.* Representation of spontaneous movement by dopaminergic neurons is cell-type selective and disrupted in parkinsonism. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E2180-2188, doi:10.1073/pnas.1515941113 (2016).
37 Lillicrap, T. P. *et al.* Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971* (2015).
38 Van Hasselt, H. & Wiering, M. A. in *2007 IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning.* 272-279 (IEEE).
39 Lee, R. S., Engelhard, B., Witten, I. B. & Daw, N. D. A vector reward prediction error model explains dopaminergic heterogeneity. *bioRxiv*, 2022.2002. 2028.482379 (2022).
40 Engelhard, B. *et al.* Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature*, 1 (2019).
41 Howe, M. W. & Dombeck, D. A. Rapid signalling in distinct dopaminergic axons during locomotion and reward. *Nature*, doi:10.1038/nature18942 (2016).
42 Hamid, A. A. *et al.* Mesolimbic dopamine signals the value of work. *Nat Neurosci* **19**, 117-126, doi:10.1038/nn.4173 (2016).
43 Vancraeyenest, P. *et al.* Selective Mesoaccumbal Pathway Inactivation Affects Motivation but Not Reinforcement-Based Learning in Macaques. *Neuron* **108**, 568-581. e566 (2020).
44 Cagniard, B. *et al.* Dopamine scales performance in the absence of new learning. *Neuron* **51**, 541-547 (2006).
45 da Silva, J. A., Tecuapetla, F., Paixão, V. & Costa, R. M. Dopamine neuron activity before action initiation gates and invigorates future movements. *Nature* **554**, 244 (2018).

46 Yin, H. H. The Basal Ganglia in Action. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry* **23**, 299-313, doi:10.1177/1073858416654115 (2017).
47 Ikemoto, S. & Panksepp, J. Dissociations between appetitive and consummatory responses by pharmacological manipulations of reward-relevant brain regions. *Behavioral neuroscience* **110**, 331 (1996).
48 Ikemoto, S. & Panksepp, J. The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. *Brain Research Reviews* **31**, 6-41 (1999).
49 Lahiri, A. K. & Bevan, M. D. Dopaminergic Transmission Rapidly and Persistently Enhances Excitability of D1 Receptor-Expressing Striatal Projection Neurons. *Neuron* (2020).

Acknowledgments: We would like to thank Fengxia Allen, Guozhong Yu, Joe Barter, and Jinyong Zhang for technical assistance.

Funding: This work was supported by NIH grants NS094754 and MH112883 to HHY.

Author contributions: KB., RNH., and HHY designed the experiments. RNH and KB performed surgeries, *in vivo* electrophysiological and optogenetic experiments, and histological analysis. KB, RNH, IPF, and QJ performed data analysis. KB and HHY wrote the manuscript. All authors have read and approved the manuscript.

Competing Interests: The authors report no financial interests or potential conflicts of interest.

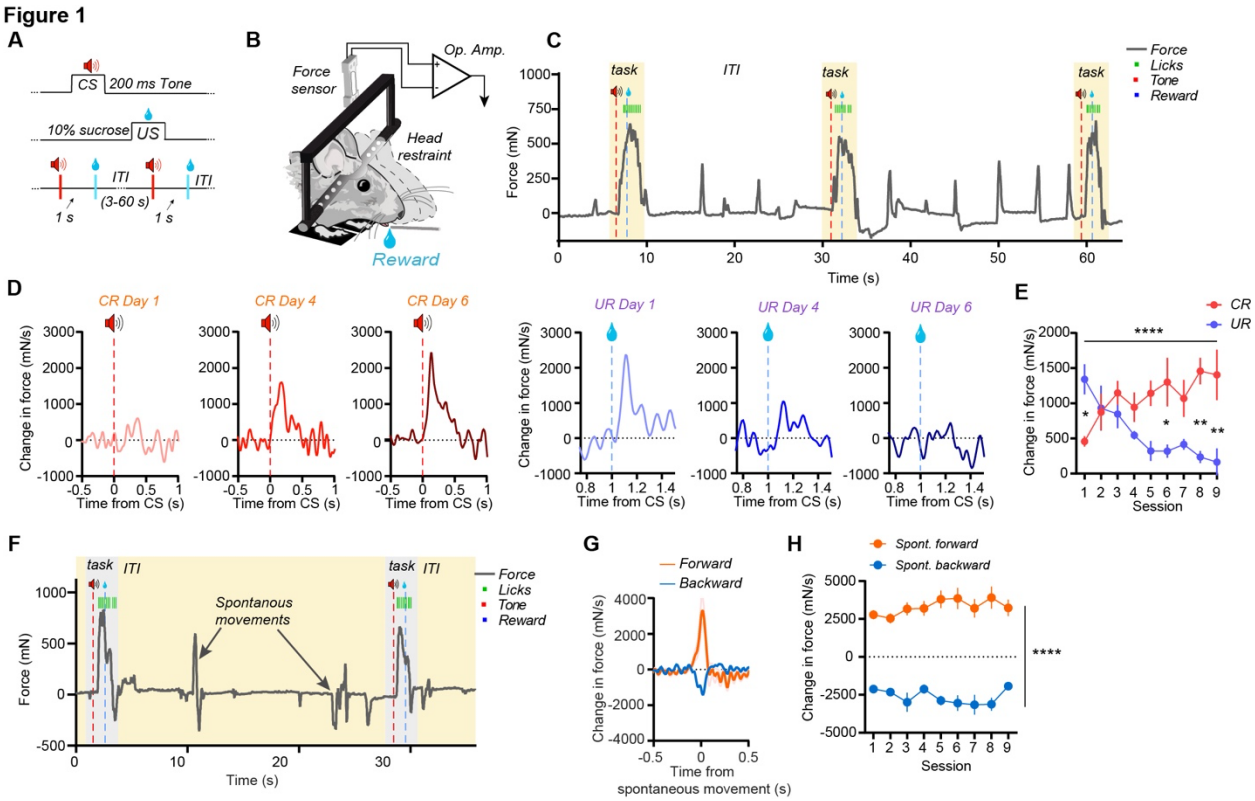


Figure 1. Force is a previously neglected behavioral measure in Pavlovian conditioning

A) Behavioral task design for stimulus-reward learning. **B)** Mice were restrained in a head-fixation device that continuously measures force exerted in the forward and backward directions. **C)** Representative force data obtained from a mouse well trained on a Pavlovian conditioning task. Note changes in force following CS and US presentations (**Supplemental Video 1**). **D)** Changes in force exertion following CS and US presentation across training in a representative mouse. **E**) Mice show a shift of peak force exertion from the UR to the CR during learning (Two-way ANOVA, significant interaction, $F(8,28) = 8.0$; Post-hoc tests found significant differences in force on days 1 ($p<0.05$), 6 ($p<0.05$), 8 ($p<0.01$) & 9 ($p<0.01$); n = 5 mice). **F)** Example force signal recorded between trials in a well-trained mouse. Note the presence of spontaneous backward and forward force exertion events during the ITI. **G**) Average spontaneous forward and backward change in force. **H**) No changes occurred in spontaneous movement amplitude over training (Two-way ANOVA, significant main effect of movement direction, $F(1,18) = 171.4$, significant interaction between training and direction on spontaneous force, $F(8,134) = 2.65$. Post-hoc tests showed that each session had significant differences between forward and backward movements ($p < 0.0001$), indicating the interaction reflected slight variability in differences in movement magnitude across sessions). Data points represent mean +/- SEM. * $p<0.05$, ** $p < 0.01$, **** $p < 0.0001$.

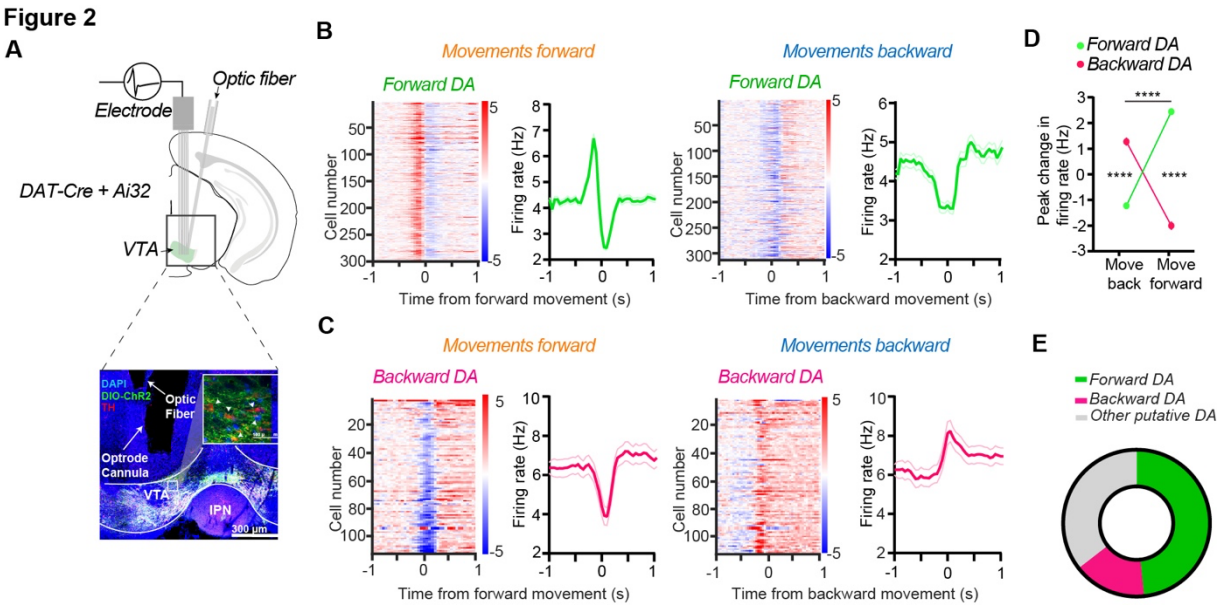


Figure 2. Opponent signaling of force direction by distinct DA neuron populations

A) *Top***,** Identification of VTA dopamine (DA) neurons by optically tagging neurons that express *ChR2* only in DAT+ neurons. *Bottom*, co-registration of GFP with TH-expressing neurons. **B)** *Left*, Forward DA neurons are activated by spontaneous forward movements (n = 300 neurons). *Right*, the same Forward DA neurons show suppression in activity during spontaneous backward movement. C**)** *Left,* Backward DA neurons are suppressed by spontaneous forward movements (n = 116 neurons). *Right,* the same Backward DA neurons are activated by spontaneous movements backward. **D**) There was a significant interaction between the change in firing rate and the DA neuron class (Two-way ANOVA, $F(1,408) = 930.13$, $p < 0.0001$). Post-hoc tests found that peak firing of Forward DA neurons was greater than Backward DA during movements forward ($p < 0.0001$) and that peak firing of Backward DA was greater than Forward DA during movements backward ($p < 0.0001$). **E**) Forward and backward DA populations of the VTA make up approximately 65% of all DA neurons. Data represent mean +/- SEM. **** $p < 0.0001$.

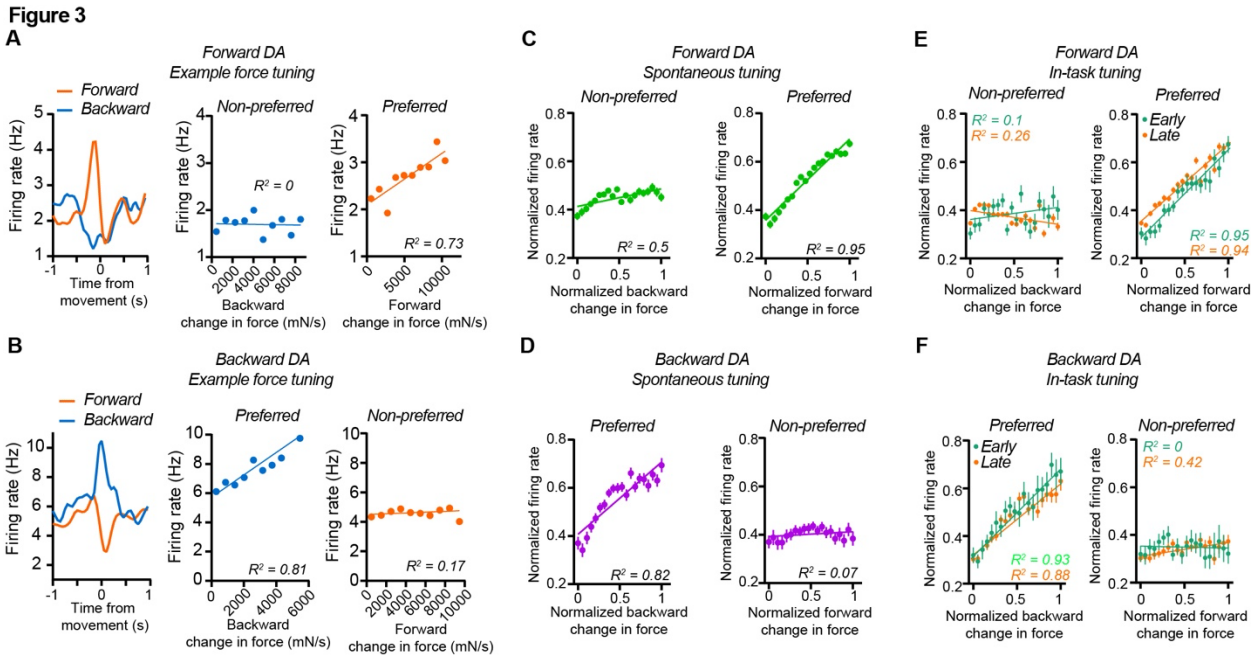


Figure 3. Force tuning by dopamine neurons is independent of learning

A) *Left*, Example Forward DA neuron showing preferred activity for movements forward. *Right*, tuning profiles for the same Forward DA neuron, indicating a linear tuning for only forward movements (R^2 for forward component = 0.73, $p < 0.01$, R^2 for backward component: 0, $p > 0.05$). **B)** *Left*, Example Backward DA neuron showing preferred activity for spontaneous movements backward. *Right*, tuning profile for the same Backward DA neuron (R^2 for forward component = 0.17, $p > 0.05$, R^2 for backward component: 0.81, $p < 0.001$). **C**) Forward DA neurons show a linear relationship with change in force magnitude during spontaneous forward movements but less so to spontaneous backward movements (R^2 for forward = 0.94, $p < 0.0001$, R^2 for backward = 0.5, $p < 0.001$, n = 300). **D**) Backward DA neurons show linear relationship with change in force magnitude during spontaneous backward movements but not to spontaneous forward movements (R^2 for forward = 0.07, $p > 0.05$, R^2 for backward = 0.82, $p < 0.0001$, n = 116). **E**) Forward DA neurons are preferentially tuned for forward movements during the task, regardless of learning (Preferred direction: R^2 for Early = 0.95, $p < 0.0001$, n = 46; R^2 for late = 0.94, $p < 0.0001$, n = 257. Non-preferred direction: R^2 for Early = 0.1, $p > 0.05$, n = 46; R^2 for Late = 0.26, $p < 0.05$, n = 257). **F**) Backward DA neurons are preferentially tuned for backward movements during the task regardless of learning (Preferred direction: R^2 for Early = 0.93, $p < 0.0001$, n = 36; R^2 for late = 0.88, $p < 0.0001$, n = 80. Non-preferred direction: R^2 for Early = 0, $p > 0.05$, n = 36; R^2 for Late = 0.42, $p < 0.01$, n = 80). Data represents mean +/- SEM.

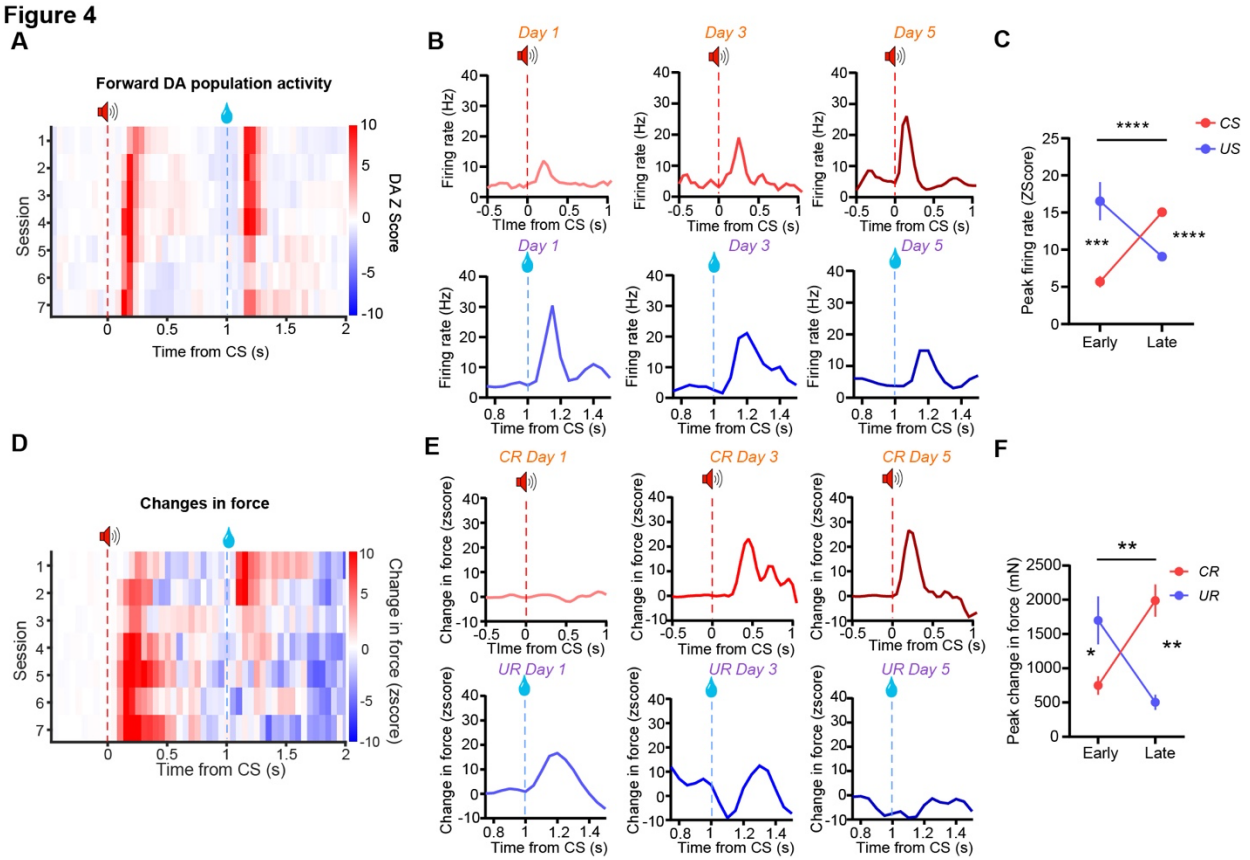
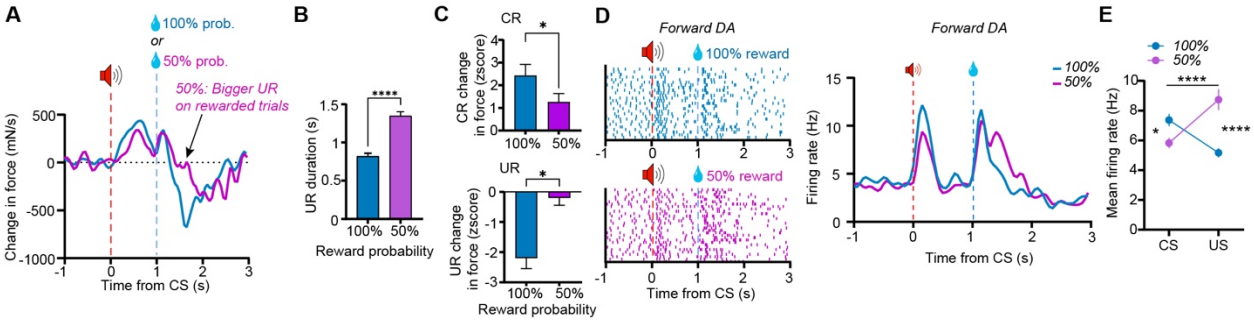


Figure 4. Force exertion explains shift in the timing of phasic DA activity from the US to the CS during learning

A) Example population of Forward DA neurons recorded during learning. **B**) Example Forward DA population activity during learning on sequential sessions. Firing rate increases in response to the CS (*top row*) and decreases to the US (*bottom row*). **C**) Forward DA neurons show increases in rate following CS delivery and decreases in rate following US delivery after training. A two-way ANOVA found a significant interaction between training (Early, Late) and stimulus (CS, US) in their effect on peak neural responding ($F(1,147) = 87.65$, $p < 0.0001$, n = 153 units from 5 mice). Post-hoc tests found that Early in training, activity after the US was greater than activity after the CS ($p < 0.001$) and that Late in training, activity after CS was greater than activity after the US ($p < 0.001$). **D**) Example change in force of a representative animal as function of training session. **E**) Transitions in CR and UR force from the same sessions as the neural recordings shown in ***B***. **F**) CR force increases as a function of training while the UR amplitude decreases. A two-way ANOVA showed a significant interaction between training (Early, Late) and response type (CR, UR) on peak change in force ($F(1,5) = 57.04$, $p < 0.001$, n = 5 mice). Post-hoc tests found that Early in training, UR force was greater than CR force ($p < 0.05$) and that Late in training, CR force was greater than UR force ($p < 0.01$). Data represent mean +/- SEM. * $p < 0.05$, ** $p<0.01$, *** $p < 0.001$, **** $p < 0.0001$.

Figure 5



606

607 **Figure 5. Larger DA responses to greater reward uncertainty can be explained by**
608 **increased force exertion**

609 **A**) Example change in force on rewarded trials under different levels of uncertainty. **B**) UR duration is
610 extended on rewarded trials that are less certain (paired t-test, $p < 0.0001$, $n = 6$). **C**) Top, mice exert
611 greater force during the CR when rewards are more probable (paired t-test, $p < 0.05$, $n = 6$). Bottom,
612 mice show greater changes in force on rewarded trials when reward is less likely (paired t-test, $p < 0.05$,
613 $n = 6$). **D**) Left, Raster plots showing firing activity of an example Forward DA neuron recorded while
614 rewards were delivered with 100% probability and 50% probability. Right, mean firing rate for the
615 neuron shown at left. Extra firing after reward promotes greater force exertion on trials with lower
616 reward probability. **E**) While reward delivery is uncertain, forward DA neurons show lower mean firing
617 after the CS and greater firing after the US. When reward is more reliable, Forward DA neurons fire
618 more after CS, but are not as activated after the US (Two-way ANOVA, significant interaction between
619 trial phase and reward probability, $F(1,56) = 55.37$, $p < 0.0001$. Post-hoc tests revealed that firing rate
620 was higher after CS in 100% condition than 50% condition ($p < 0.05$) but was higher after US in 50%
621 condition than 100% condition, $p < 0.0001$, $n = 30$ neurons). Data represent mean +/- SEM. * $p < 0.05$,
622 **** $p < 0.0001$.

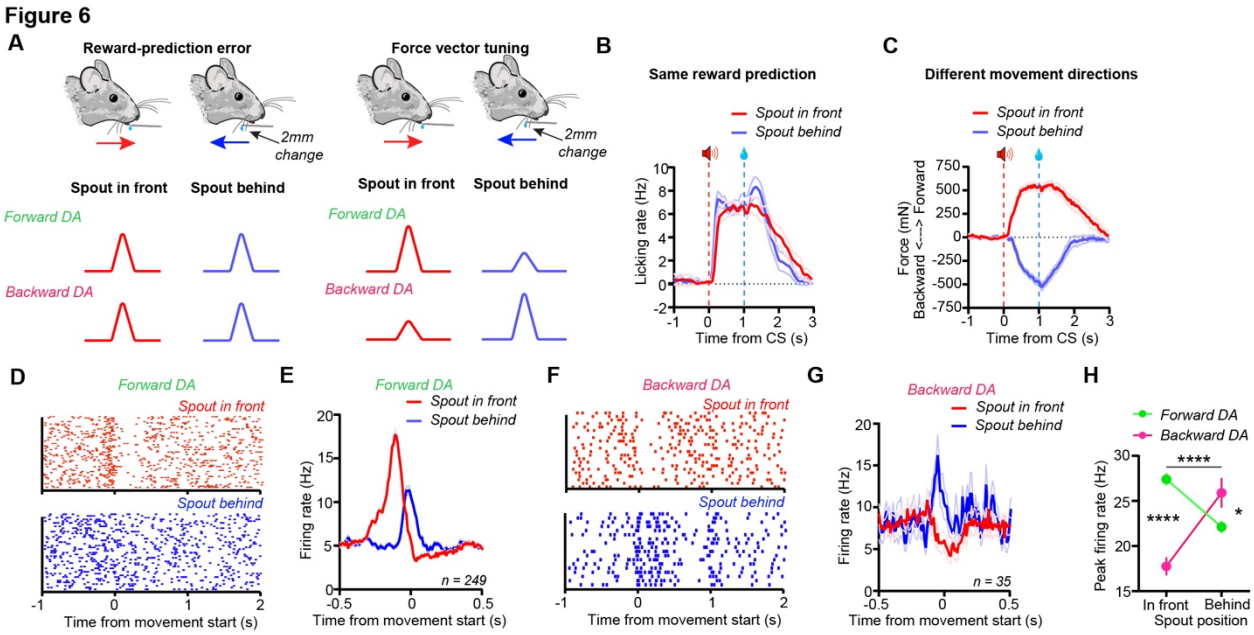


Figure 6. Distinct populations of VTA DA neurons represent force in different directions, despite identical reward prediction

A) Force sensors are slightly flexible and allow mice to move small distances forward or backward. The spout was positioned either in front of or slightly under the mice' chin to encourage distinct movement directions to obtain reward of the same magnitude and value (**Supplemental Video 2**). Diagrams show that firing patterns of Forward and Backward DA neurons are distinguished between the two conditions because of their opponent force tuning. **B**) Anticipatory and consummatory licking rates are identical between 'spout in front' and 'spout behind' conditions. **C**) Mice make movements in opposed directions between the 'spout in front' and 'spout behind' conditions. **D**) Example raster plots showing firing of Forward DA neuron aligned to within-task movement under the two spout position conditions (in front vs behind). **E**) Forward DA neurons showed greater responding to movements forward than backward (n = 249). **F**) Example raster plots showing firing of Backward DA neuron aligned to within-task movement under the two spout position conditions (in front vs behind). **G**) Backward DA neurons showed greater firing to movements backward than forward (n = 35). **H**) Opponent activity of Forward and Backward DA neurons depending on reward location (Two-way ANOVA, significant interaction $F(1,408) = 83.74$, $p < 0.0001$. Post hoc tests revealed that Forward neurons had firing rates higher than Backward neurons when the spout was in front ($p < 0.0001$), but Backward neurons had higher responses than Forward neurons when the spout was behind ($p < 0.05$). Data represent mean +/- SEM. * $p < 0.05$, **** $p < 0.0001$.

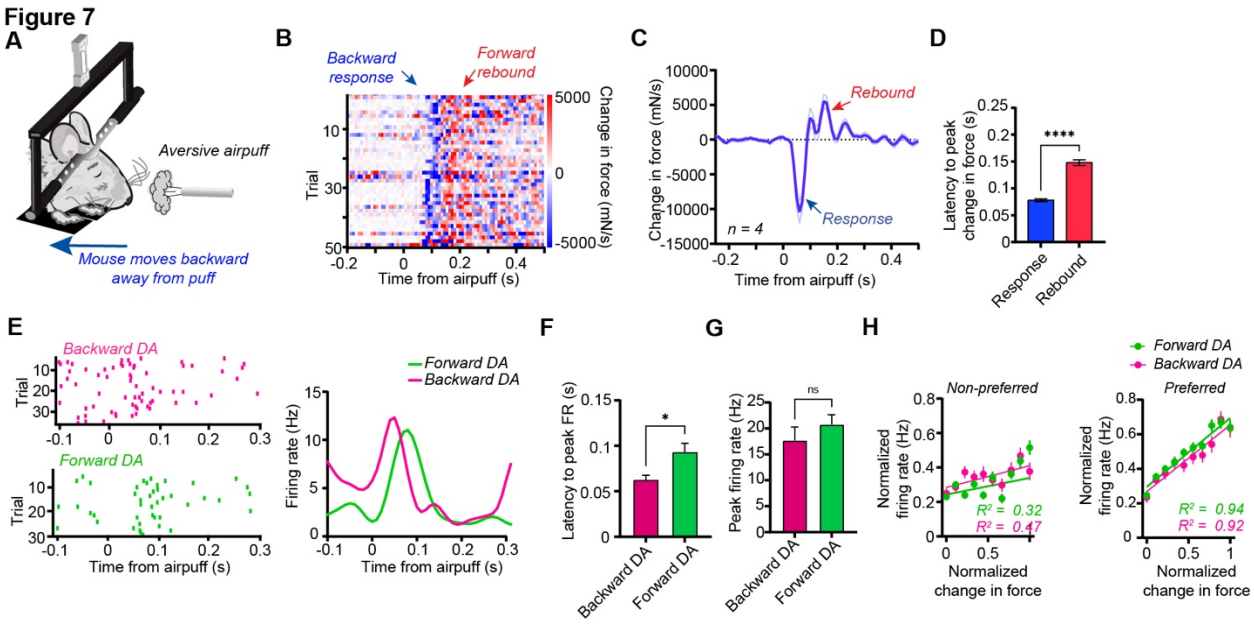


Figure 7. DA activation to aversive air puff produces is explained by bidirectional force changes

A) We exposed mice to a mild un-cued air puff delivery to the face in a session separate from Pavlovian conditioning **B**) Mice respond to air puffs with a backward change in force followed by a forward change in force (rebound). **C**) Average change in force trace across mice (n = 4). **D**) Backward responses occurred before the rebound during the air puff trials (paired t-test, $p < 0.0001$, n = 24 sessions from 4 mice). **E**) *Left,* Rasters for a representative Backward DA neuron and a representative Forward DA neuron showing activation at different latencies. *Right*, average firing rate traces for the neurons whose rasters are shown to the left, note latency differences. **F**) Latency for Backward DA neuron firing was lower than Forward DA (unpaired t-test, $p < 0.05$, n = 41 Forward DA, n = 35 Backward DA). **G**) Forward DA and Backward DA peak firing rates were not different (unpaired t-test, $p > 0.05$, n = 41 Forward DA, n = 35 Backward DA). **H**) Forward and Backward DA neuron firing activity was tuned for each population's preferred force direction during air puff trials (*Left*, Non-preferred, Forward DA, $R^2 = 0.32$, $p > 0.05$, Backward DA, $R^2 = 0.47$, $p < 0.05$; *Right*, Preferred, Forward DA, $R^2 = 0.94$, $p < 0.0001$, Backward DA, $R^2 = 0.92$, $p < 0.0001$). Data represent mean +/- SEM. * $p < 0.01$, **** $p < 0.0001$.

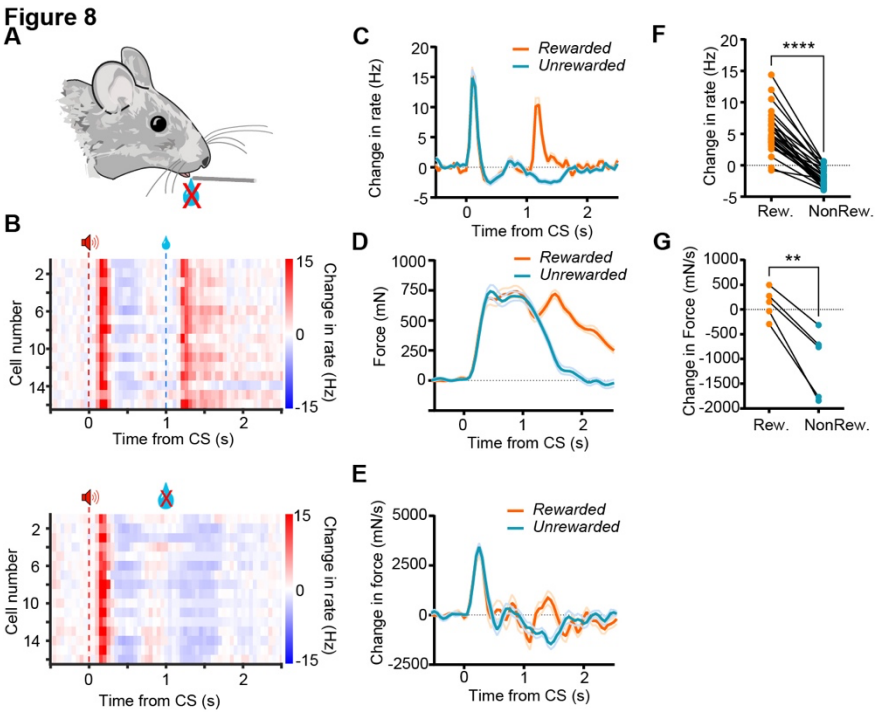


Figure 8. Parallel 'dips' in force and DA firing following omission of expected reward
A) Schematic of reward omission experiment, whereby expected reward was omitted on 80% of trials. **B**) Simultaneously recorded population of Forward DA neurons during trials with expected reward delivery (*top*) and omission (*bottom*). Note characteristic 'dip' in DA neurons during reward omission. **C**) Mean change in firing rate for the population in B during rewarded and unrewarded trials shows a robust dip in activity following omission. **D**) Mice abruptly withhold forward force exertions upon reward omission (**Supplemental Video 3**). **E**) Reward omission results in parallel 'dips' in change in force and DA activity. **F**) Reduction in firing rate after reward omission (paired t-test, $p < 0.0001$, $n = 28$ neurons from 3 mice). **G**) Omission resulted in a reduction in mean force in the period following reward omission (paired t-test, $p < 0.01$, $n = 5$ mice). Data represent mean +/- SEM. ** $p < 0.01$, **** $p < 0.0001$.

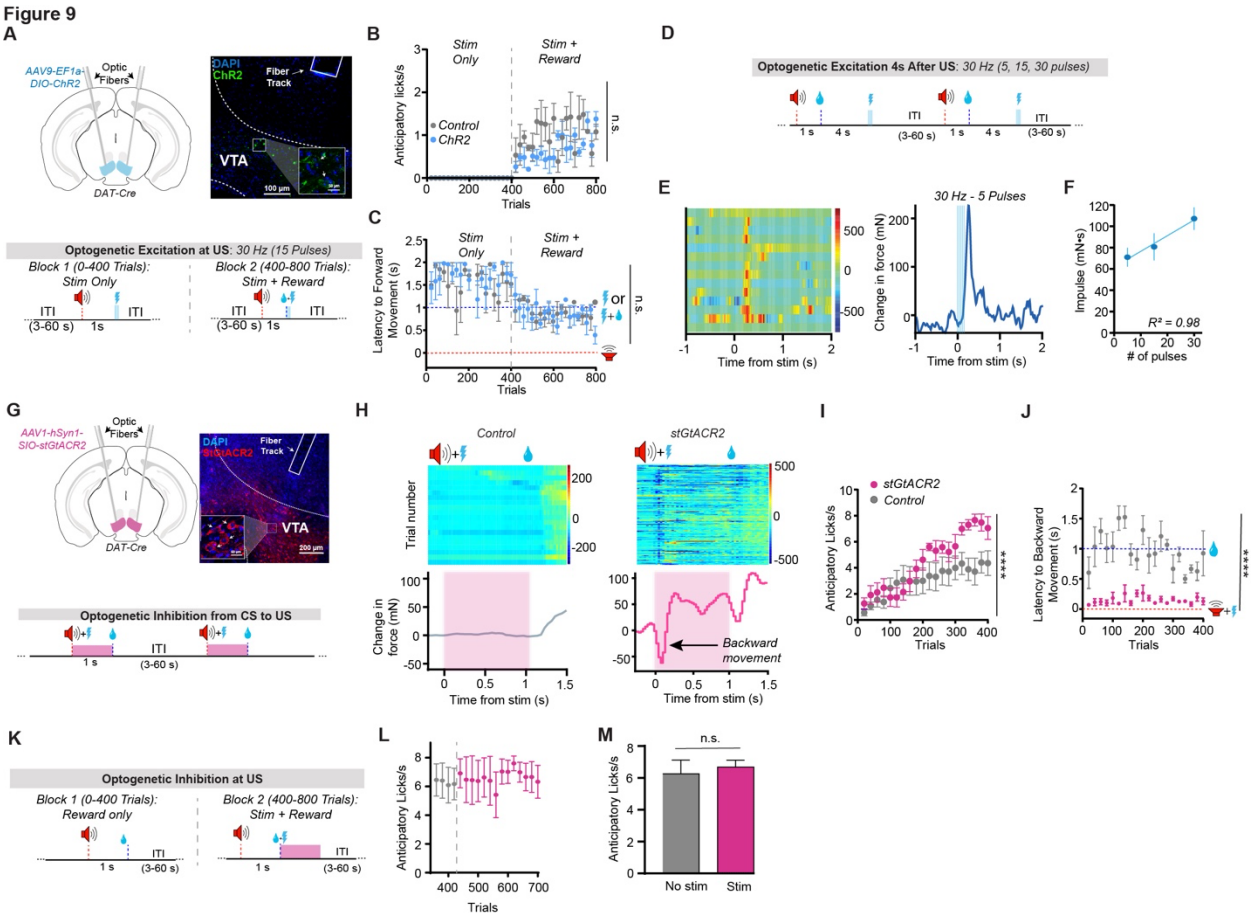


Figure 9. VTA DA is neither necessary nor sufficient for learning the CS-US association **A**) Cre-dependent ChR2 or DIO-eYFP was injected into the VTA of DAT-Cre mice (N = 3) to excite VTA DA neurons (Top). To test whether VTA DA is sufficient for learning, brief high frequency stimulation (30 Hz – 15 pulses) was used in place of reward (*Bottom*). After 400 trials, a sucrose reward was delivered together with stimulation. **B**) Stimulation did not produce a CS-US association with stimulation alone compared to controls (N = 5), as indicated by anticipatory licking (Two-way ANOVA, significant interaction, $F(38,152) = 1.752$, $p < 0.01$, but post-hoc analyses revealed no significant effects between any of the groups ($p > 0.05$)). **C**) The same was found for latency to move to the CS (Two-way ANOVA, significant main effect of Trial Number, $F(38, 152) = 5.80$, $p < 0.0001$, no effect of Stimulation $F(1,4) = 0.1$, no interaction, $F(38,152) = 0.9$, $p > 0.05$). **D**) DA neurons were stimulated 4 seconds after reward during a time period when mice are least likely to generate spontaneous movements. **E**) Example effect of DA activation on forward force generation. **F**) Varying the number of pulses delivered at 30 Hz resulted in a linear relationship between pulse number and impulse ($R^2 = 0.98$). **G**) We used the Cre-dependent inhibitory opsin stGtACR2 (N = 4) to inhibit VTA DA neurons during learning or DIO-eYFP (N = 6) for controls (Top). Bottom, we inhibited DA activity during the CS-US interval try to suppress learning. **H**) Example backward movement resulting from inhibition of DA neurons compared to a control. Rows of the heatmap indicate stimulation trials and the curves below indicate average across all trials with stimulation. **I**) Optogenetic inhibition did not

prevent learning of the CS-US association, as indicated by anticipatory licking compared to controls. However, it did cause an increase in anticipatory licking, in contradiction with the RPE hypothesis (Two-way ANOVA significant main effects of Stimulation, $F(1, 160) = 25.34$, $p < 0.0001$, and of Trial Number, $F(19, 160) = 6.775$, $p < 0.0001$, no interaction, $F(19,160) = 1.3$, $p > 0.05$). **J**) Optogenetic inhibition significantly reduced latency to backward movement initiation compared to control stimulation (Two-way ANOVA, significant main effect of Stimulation, $F(1, 7) = 339.5$, $p < 0.0001$, no main effect of Trial number, $F(19, 133) = 0.8$, $p > 0.05$, no interaction $F(19,133) = 0.86$). **K**) To test whether VTA DA inhibition causes a negative prediction error, VTA DA neurons were inhibited at the time of reward for 1 second in trained mice. **L**) Inhibition at reward did not significantly impair learning. **M**) There was no change in anticipatory licking (paired t-test, $p = 0.46$). Data represent mean +/- SEM.

Figure 10

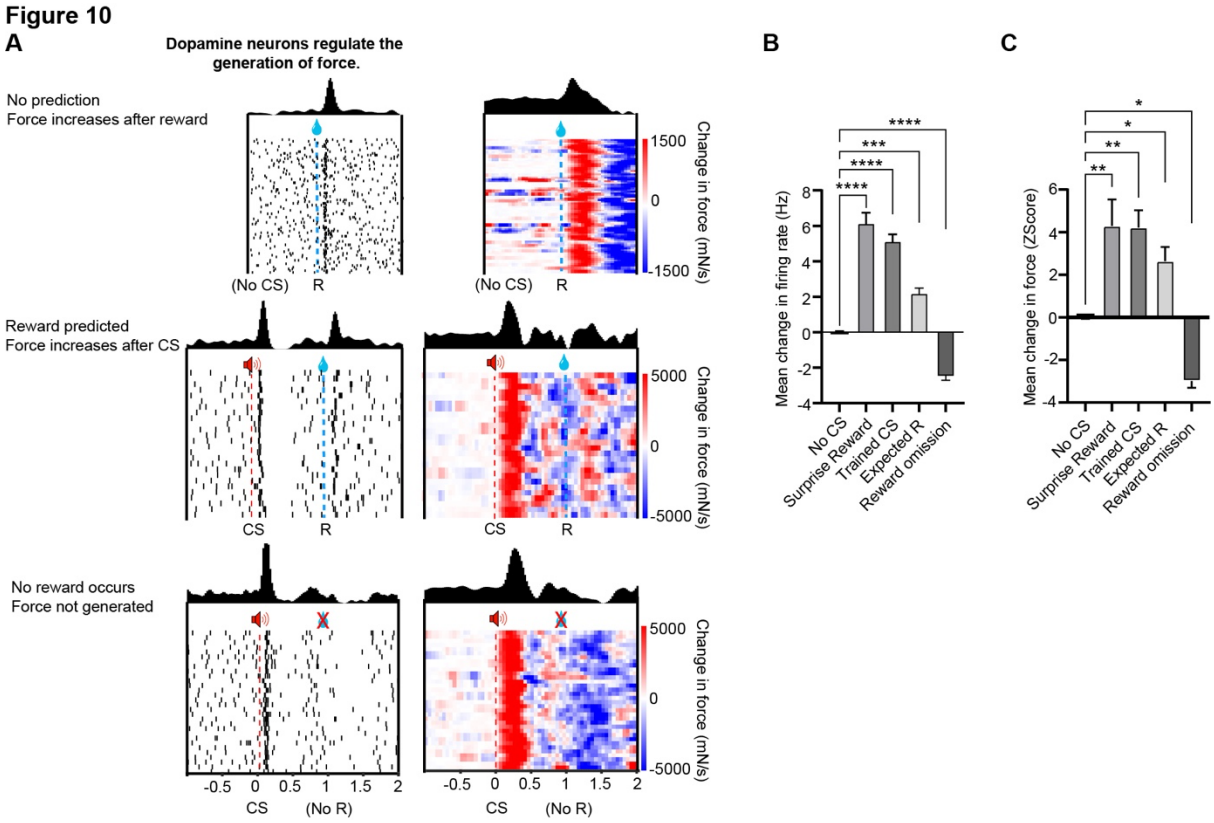


Figure 10. Changes in DA activity explain changes in force exertion rather than reward prediction

A) Raster plots of example neurons from mice depicting classically observed DA firing patterns to surprise reward, reward predicted by tone, and omitted reward. On the right are changes in force occurring in parallel with DA responses in the same session. Mice push forward to surprise rewards, to cues predicting reward, and after receiving expected reward. Mice terminate force generation upon reward omission, producing a negative change in force. **B**) Bidirectional changes in firing rate under conditions depicted in A (One-way ANOVA, $F(4,131) = 7.971$, $p < 0.0001$; Post hoc tests using No CS condition as the control showed increased firing after Surprise reward, $p < 0.0001$, after CS in well-trained mice, $p < 0.0001$, in response to expected reward, $p < 0.001$, and significant decreases after omission of expected reward, $p < 0.0001$; n = 84 neurons). **C**) Mean changes in force for the same sessions (One-way ANOVA, $F(4,24) = 2.31$, $p < 0.0001$; Post hoc tests using No CS condition as the control showed increases in force after Surprise reward, $p < 0.01$, after CS in well-trained mice, $p < 0.01$, in response to expected reward, $p < 0.05$, and significant negative changes in force after omission of expected reward, $p < 0.05$; n = 10 mice). Note that patterns of force changes closely parallel changes in DA activity shown in B. Data represent mean +/- SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.