**Optogenetic Inhibition of Dopamine Neurons in the VTA at Reward Reception Alters Choice Behaviour**

Jessica Chan

Supervisor: Jonathan Britt1, 2

Acknowledgements: Jesse Mendoza1

1Integrated program in Neuroscience, Montreal Neurological Institute, Montreal, QC, Canada

2 Department of Psychology, McGill University, Montreal, QC, Canada  
page1image3798144

**Abstract**  
 Well-adjusted reward seeking behaviour is essential to the survival of almost all organisms, complicit in necessities such as mating and feeding. Optimal reward seeking behaviour involves the recognition and selection of actions that would result in the highest probability or magnitude of reward, and an underlying mechanism of the valuation of these states is the processing of reward prediction error (RPE) signals in the ventral tegmental area (VTA) and substantia nigra. We hypothesized that altering the RPE would lead to a change in behaviour. To investigate this, we used VTA targeted optogenetic dopamine neuron inhibition at crucial events of the RPE, namely the reward-predicting stimulus and reward. Manipulations were hypothesized to decrease either the magnitude of expectation or the perceived value of the reward. According to various reinforcement learning theories, the change in RPE should cause a change in reward processing and valuation of each state, resulting in a preference in decision making behaviour for the action that leads to a more valuable reward. Whether reward processing is a retrospective or prospective mechanism has yet to be settled, however our findings produce a decent demonstration of the retrospective nature of RPE processing.

Introduction

Prediction errors are thought to be central to associative learning, and the field has produced numerous accounts of the parallels between reward prediction error (RPE) processing and different reinforcement learning models (CITE 1, 2 from brief opto). Dopamine neuron activity in the VTA and the substantia nigra has been implicated in reward prediction error production, a crucial component to associative learning that is used to update values of states based on differences between expected and observed rewards (Schultz neural substrate). RPEs in cue-reward associations appear to be processed in a two-component dopamine signal, in which a first dopamine spike represents cue salience, and the second signal represents novelty of reward (cite). Positive RPEs occur when a reward is better than expected, with a greater dopamine spike at reward reception. Negative RPEs occur when rewards are absent or lesser than expected. No signal occurs when a reward appears that is completely expected, showing that associative learning operates as a function of novelty (CITE SCHULTz). Expected and observed reward values are subtracted, and the cue’s state value is adjusted accordingly. Numerous studies cite temporal difference (TD) learning algorithms, which primarily use past observed errors between temporally relative stimuli to inform present predictions (CITE). Other studies cite the Rescorla-Wagner reinforcement learning model, in which strengths of association between cue and reward are updated as functions of cue salience and previous associative strength among other things (CITE). This paper will use an extended Q-learning model to analyze expected outcomes.

To explore the nature of RPE processing, we used optogenetic inhibition in the VTA to attenuate dopamine neuron activity at crucial events in an FR8 schedule lever press paradigm. Mice were trained to obtain reward by pressing a lever 8 times, after which sucrose would be available at the food bowl. Animals had the option to choose one of two levers, one of which was subject to optogenetic inhibition time locked to either presentation of the lever or reward reception. These events were speculated to be components of the RPE that would eventually update the value of each lever’s presentation, hypothesized to be the cue component of the association. Phasic optogenetic inhibition has been shown to mimic endogenous negative reward prediction errors (CITE). Manipulations were made to change the values of RPEs produced, in turn altering the valuation of the target lever. We speculated that such change would cause an increase in the valuation of the targeted lever when time locked to cue as a result of a forced positive RPE, however no such effect was observed (2). Interestingly, dopamine inhibition to reward reception had the expected effect, resulting in a decreased valuation of the target lever reflected in preference behaviour (Figure 2).

**Materials and Methods**

*Subjects.* Ten TH:Cre mice were used in this experiment. Mice were housed in groups of up to three with same-gendered animals and maintained on a 12-hour dark/light cycle prior to and throughout the manipulations. Over the course of the experiment, mice were food restricted to 80% of their body weight and fed 2.5 grams of grain pellets after each round of experimentation.

*Surgical procedure.* Animals were anesthetized using intraperitoneal injections of ketamine (Ventoquinol, 100 mg/kg) and Xylazine (Bayer, 10mg/kg) cocktail and secured to a stereotaxic frame (Kopf Instruments). Prior to initiation of the procedure, animals were tested for the absence of the hind paw withdrawal reflex and opthalamic ointment (Natural Tears, Alcon) was applied to the corneas to protect the eyes. The scalp was then shaved and sterilized before a midline incision was made to expose the bregma and lambda. The exposed areas were disinfected, and skin was held in place with haemostats. Small holes were then drilled at specified coordinates for viral injections and optic fiber insertion. A 10µm tip diameter pulled glass pipette (Drummond Scientific) was loaded with a given viral vector. Ten mice were injected with the viral vector AAV5-DIO-eArch30-eYFP (UNC VectorCore) bilaterally targeted at the VTA. Injection was administered at a 10-degree angle from the vertical and targeted to the following coordinates relative to bregma: AP (-3.45), ML [+/-] 0.93, DV (-4.37). Injections were delivered by a Nanoject II injector system (Drummond Scientific) and consisted of 36.8 nanolitres every 15 seconds for a total of 20 injections. Mice were then implanted bilaterally with 200µm core fibre optics (0.39NA, ThorLabs) equipped with a 230-240µm core ceramic ferrule (PFP) which was inserted into the VTA just above the site of injection at the following coordinates to bregma: AP (-3.45), ML [+/-] 0.93, DV (-4.30). Three screws (Morris) were then inserted partially into the skull, and dental cement (Lang) was used to cover the skull and secure the optic implants to the screws to form a head cap. The wound was sealed with Vetbond glue and physiological saline (0.9% NaCl) was administered subcutaneously to replenish fluids. Mice were then placed on a warming pad until consciousness was regained before returning to their home cages. Mice were allowed four weeks of rest post-surgery for proper recovery and viral expression.

*Operant boxes and Laser Equipment.* Experiments were conducted in plexiglass operant boxes (Med Associates) equipped with two retracting metal levers and a food port that presented a single bowl. Sucrose solution was administered to the bowl through tubing from a syringe pump (Med Associates). Behavioural data was registered by the operant boxes and recorded into text files for further analysis. Optogenetic inhibition was administered through splitters or patch cords connected to rotary joints (Doric Lenses Inc.) to allow greater freedom of movement. Splitters and patch cords were connected to the optic implants secured to the head-caps of the animals using ceramic sleeves (ThorLabs). Green (532 nm wavelength) lasers supplied the light that was delivered into the brain and output was confirmed to be ~20mW or ~25mW for bilateral and unilateral inhibition respectively before each testing period.

*Training.* Mice were trained on lever pressing for sucrose reward on an FR1 schedule until a threshold of capability was reached that showed sufficient learning of the cue-reward association. After sufficient recognition of the relation, mice were moved onto FR3, FR5 then FR8 in respectively depending on level of performance. The experiment was not initiated until subjects had reached a level of performance on the FR8 schedule that suggested sufficient reward seeking operant behaviour. Throughout training stages and manipulation, mice were tethered by cables connected to their implants to maintain a consistent state. Operant boxes used for experimentation were used for training stages for the same purpose.

*Behavioural Paradigms and Stimulation Protocols.* Mice were placed in an operant box and subject to an hour-long FR8 behavioural paradigm in which the presentation of the lever gave mice the opportunity to press eight times for reward administration, consisting of 30 microlitres of 15% sucrose solution. Trials consisted of choice or forced trials, where choice trials presented both levers and forced trials presented just one. Levers were presented for 60 seconds, with and an average ITI of 30 seconds. Trial probabilities were determined in blocks of 6, in which 1/6 trials would be choice and 5/6 trials were forced. Half of the forced trials have a 1/2 chance of being left or right (Figure 1). Left or right levers were designated to each mouse for inhibition such that half the subjects were on either lever to counterbalance the results. Both optogenetic manipulations utilized the same behavioural paradigm.

Ten mice with Arch expression in the VTA were tested for 6 days in the first experimental protocol, with four baseline days without inhibition, then two days of optogenetic inhibition to cue. Administration of constant green (532nm) wavelength light for 2 seconds was time locked to the presentation of a designated lever during the two final days of the experiment.

The same ten mice with Arch expression in the VTA were then tested for 12 days in a second protocol, with 4 days on baseline and 8 days of optogenetic inhibition time-locked to reward reception. Reward reception in this protocol was defined as the first lick bout after a successful rewarding series of lever presses. The lick bout would not trigger inhibition in the program unless it was within 10 seconds of the final lever press. If conditions were satisfied for reward reception, constant green (532 nm) wavelength light was administered for 2 seconds.

*Statistical Analysis.* Custom Matlab script (Mathworks) was used to analyze and graph behavioural data collected from operant boxes (Med Associates), with a focus on choice and forced behaviour. Latencies analyzed were taken from forced trials of the designated inhibition lever and comparisons were made between baseline and inhibition days. Two-way ANOVAs were performed between baseline and manipulation days, and between stimulated and non-stimulated levers averaged across the tested mouse population. Sidik’s multiple comparison test was used to test for significance (p<0.05). The two-way ANOVAs and significance testing was computed using Prism 7 (Graphpad). Animals in which optic implants had improper contact to the VTA were excluded from analysis.

**Results**

To test the hypothesis that dopamine inhibition at crucial events during the learning of operant behaviour would cause an alteration in the valuation of lever press action, Archaerhodopsin, an inhibitory opsin, was delivered to the VTA via a cell body infecting virus (Figure 2). After training stages, light was shone into the VTA during experimental days, time locked to certain events to depress dopamine neuron activity. Events were postulated to be one of the two components in RPE processing, a cue related signal for expectation of value or a reward related signal for observed value.

*Dopamine inhibition to cue.* No effect was observed when inhibiting dopamine neurons to the presentation of the lever. Both choice and forced trials were unaffected by the manipulation (Figure 2; forced trials p=0.5946, choice trials 0.2456). Latency to lever press and head entry also remained unaffected (Figure 2; p=0.2569, p=0.2841 respectively). This may suggest that the DA spike in response to the cue does not have a significant influence on decision making in reward seeking behaviour. However, the lack of effect may be due to a failure to associate the presentation of the lever to the reward, preventing the event from being incorporated into the RPE.

*Dopamine inhibition to reward reception.* Inhibition to bout produced a significant effect in both forced and choice trials, with a decrease in lever presses on the inhibited lever (Figure 3; p<0.001, t=7.249 and p<0.01, t=5.11). This suggests that altering the dopamine spike caused by the recognition of reward reception is sufficient to reduce reward seeking behaviour directed to the inhibited lever relative to the non-inhibited lever. Behaviour may be explained by decreased valuation of the target lever due to repeated production of negative RPEs (1). In this case, an attenuated dopamine spike in the second component of the RPE would have caused a negative value to be incorporated into processing, reflected in preference behaviour (Figure 2). The value of pressing the targeted lever is lowered as a result.

1. RPE = actual value – expected value
2. Qt+1 (st) = Qt (st) + RPE

*Where Qt represents the current state, and Qt+1 represents the value of the next expectation.*

A significant increase in latency to lever press was also observed to the forced trials of the inhibited lever, however latency to head entry in the same trials did not deviate in comparison to baseline (Figure 3; p<0.05, t=3.45 and p=0.977 respectively).

**Discussion**

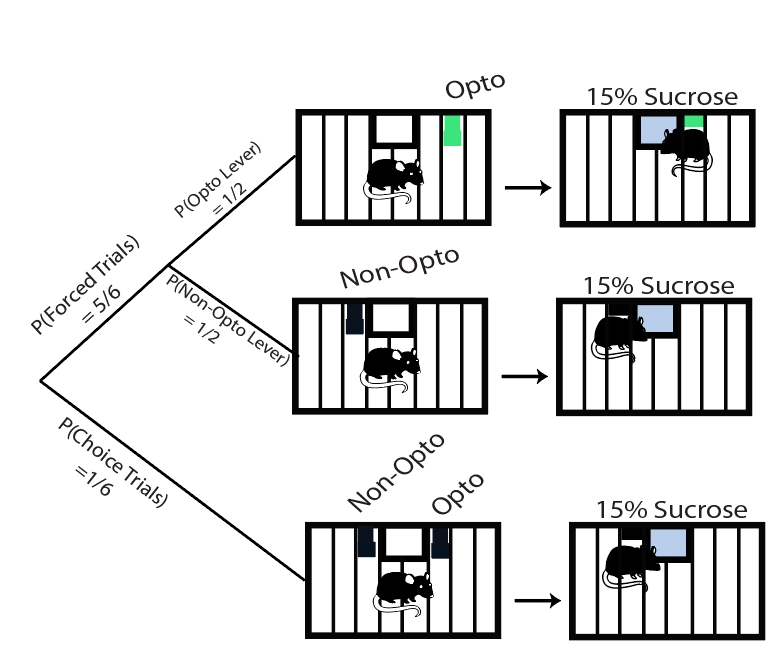
Dopamine inhibition to reward reception had a significant decrease in choice and forced lever presses on the inhibited lever (Figure 3). These results align with our hypothesis and with general models of processing RPEs. Inhibition at reward reception was hypothesized to reduce the value of the perceived reward and cause a negative or attenuated RPE relative to baseline. This change in the value of the RPE should have reduced the valuation of actions and cues related to the reward state according to standard reinforcement learning algorithms, and in turn make the related action less desirable **(cite).** Latency to head entry was unaffected by the manipulation, which could be interpreted as a lack of difference in motivation to obtain reward. Interestingly, the latency to lever pressing significantly increased with inhibition, suggesting that the change in valuation had some effect on motivation to perform the actions (Figure 3). These results when interpreted through retrospective RPE processing align with the proposed hypothesis. RPE processing changed the valuation of the state in which animals were able to make actions based on the differences between expected and perceived reward (cite). As the altered RPEs are used to update the expectations of the related states, values relaying information about expectation are lessened, resulting in decreased expectations, and thus lowered motivation to action. Head entry latency being unaffected is expected in the retrospective model, as the reward state is simply a component of the RPE and does not get updated. Previous studies have also shown that reward reception in relation to a cue is time sensitive, such that reward that is received even 1 second later than usual would cause a dopamine depression(CITE). This may explain the consistency of head entry latency between baseline and inhibition days.

We found that optogenetic inhibition time-locked to the presentation of the lever produced little to no effect. However, this does not rule out the possibility that alteration of expectations can affect choice behaviour. This may be due to the unconfirmed participation of the lever presentation in the RPE being manipulated. Schultz showed that recognition of unassociated cues was unselective, and various stimuli could cause dopamine signals that could then be associated after some repetition (cite). Furthermore, if a cue was recognized but no reward was received, over time the dopamine neuron activity reflects that of a neutral stimulus (CITE). In the paradigm described, the mice had 60 seconds from presentation to press the lever 8 times before the reward was administered. Administration of the reward was not always followed by reward reception. Because inhibition was time locked to the lever presentation, it is unsure whether or not this stimulus was actually associated to the reward. It is possible that because of the temporal separation of the events the cue was learned to be neutral. Another possibility allowed by the temporal gap is the presence of intermediary signals that may act as secondary cues. It is not well known how secondary cues are valued.

To work around the issues that may have affected the outcome, future directions for investigating effects of expectation alteration would include inhibition to either the final lever press or an intermediate lever press, in hopes of closing the temporal gap between the cue and the reward. In previous studies, it was shown that actions may play the role of cues as action-states that have their values updated, and these action-states may be more temporally relevant and physically salient to the animal than lever presentation (CITE). In this case, interpretation of the effects should be analyzed through an extended Q-learning model that incorporates actions into states as a form of an action vector that stores values of state-action pairs (1). In these cases, each state (st) would propose the possibility of some actions (at) and the chosen action would be considered as part of a state-action pair whose value would be updated using the RPE. In the previous consideration, the state in question would have proposed a vector of possible actions, the state being the target of alteration. It is difficult to assert that altering an expectation signal to a state with various possible actions can single out just one to update. In the interpretation below, actions are valued and updated separately despite being related to the same state. Were manipulations of expectation enacted on the valuation of action state pairs rather than on an overarching state, the different action possibilities may be disentangled and a more targeted alteration may occur.

1. Q(st+1, at+1) = Q(st, at) + RPE

Conclusion

**Figures**

*Figure 1.* **a.** Probability of forced or choice trials in behavioural paradigm. **b.** Visual representation of optic fibre placement and location of light delivery.



*Figure 2.* Constant 2 second inhibition at the VTA to lever presentation. **a,b.** Average proportion of choice and forced trials responded over all mice and sessions for Arch-VTA (n=10). **c,d.**  Average latency to head entry and lever press in forced trials over all mice and sessions for Arch-VTA (n=10). Force trials restricted to optogenetically inhibited lever only.



*Figure 3.* Constant 2 second inhibition at the VTA to reward reception. **a, b.** Timeline of behavioural paradigm. Points shown are proportion of forced and choice trials averaged across all mice over course of manipulation. 8 days baseline followed by 4 days of inhibition. **b, c.** Proportion of forced and choice trials responded averaged across all mice. Comparison made between basline days and inhibition days. **e, f.** Latencies to lever press and head entry to forced trials averaged across all mice. Comparison made between basline and inhibition days.

**References**

Balasubramani, P. P., Chakravarthy, V. S., Ravindran, B., & Moustafa, A. A. (2014). An extended reinforcement learning model of basal ganglia to understand the contributions of serotonin and dopamine in risk-based decision making, reward prediction, and punishment learning. Frontiers in Computational Neuroscience, 8(47). doi:10.3389/fncom.2014.00047

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron, 47(1), 129-141. doi:10.1016/j.neuron.2005.05.020

Bermudez, M. A., & Schultz, W. (2014). Timing in reward and decision processes. Philos Trans R Soc Lond B Biol Sci, 369(1637), 20120468. doi:10.1098/rstb.2012.0468

Chang, C. Y., Esber, G. R., Marrero-Garcia, Y., Yau, H. J., Bonci, A., & Schoenbaum, G. (2016). Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. Nat Neurosci, 19(1), 111-116. doi:10.1038/nn.4191

Enomoto, K., Matsumoto, N., Nakai, S., Satoh, T., Sato, T. K., Ueda, Y., . . . Kimura, M. (2011). Dopamine neurons learn to encode the long-term value of multiple future rewards. Proc Natl Acad Sci U S A, 108(37), 15462-15467. doi:10.1073/pnas.1014457108

Fiorillo, C. D., Newsome, W. T., & Schultz, W. (2008). The temporal precision of reward prediction in dopamine neurons. Nat Neurosci, 11(8), 966-973. doi:10.1038/nn.2159

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. Science, 299(5614), 1898-1902. doi:10.1126/science.1077349

Kakade, S., & Dayan, P. (2002). Dopamine: generalization and bonuses. Neural Netw, 15(4-6), 549-559.

Lak, A., Stauffer, W. R., & Schultz, W. (2014). Dopamine prediction error responses integrate subjective value from different reward dimensions. Proc Natl Acad Sci U S A, 111(6), 2343-2348. doi:10.1073/pnas.1321596111

Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature, 459(7248), 837-841. doi:10.1038/nature08028

Meck, W. H. (2005). Neuropsychology of timing and time perception. Brain Cogn, 58(1), 1-8. doi:10.1016/j.bandc.2004.09.004

Morris, G., Nevet, A., Arkadir, D., Vaadia, E., & Bergman, H. (2006). Midbrain dopamine neurons encode decisions for future action. Nat Neurosci, 9(8), 1057-1063. doi:10.1038/nn1743

Rao, R. P., & Sejnowski, T. J. (2001). Spike-timing-dependent Hebbian plasticity as temporal difference learning. Neural Comput, 13(10), 2221-2237. doi:10.1162/089976601750541787

Rescorla, R. A., & Solomon, R. L. (1967). Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. Psychol Rev, 74(3), 151-182.

Satoh, T., Nakai, S., Sato, T., & Kimura, M. (2003). Correlated coding of motivation and outcome of decision by dopamine neurons. J Neurosci, 23(30), 9913-9923.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. J Neurophysiol, 80(1), 1-27. doi:10.1152/jn.1998.80.1.1

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. Annu Rev Psychol, 57, 87-115. doi:10.1146/annurev.psych.56.091103.070229

Schultz, W. (2016). Dopamine reward prediction error coding. Dialogues Clin Neurosci, 18(1), 23-32.

Schultz, W. (2017). Reward prediction error. Curr Biol, 27(10), R369-R371. doi:10.1016/j.cub.2017.02.064

Tobler, P. N., Dickinson, A., & Schultz, W. (2003). Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. J Neurosci, 23(32), 10402-10410.