Anatomical and computational models of the role of phasic dopamine signaling in intracranial self-stimulation: psychophysical and electrochemical tests

Marie-Pierre Cossette

A Thesis
In the Department
of
Psychology

Presented in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy (Psychology) at
Concordia University
Montreal, Quebec, Canada

February 2019

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Marie-Pierre Cossette

Entitled: Anatomical and computational models of the role of phasic dopamine signaling in intracranial self-stimulation: psychophysical and electrochemical tests

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Psychology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____Chair
Dr. Ian Ferguson

_____External Examiner
Dr. Mark Walton

_____Examiner to Program
Dr. Christopher Brett

_____Examiner
Dr. Mihaela Iordanova

_____Examiner
Dr. Uri Shalev

_____Thesis Supervisor
Dr. Peter Shizgal

Approved by _____
Dr. Andrew Chapman, Graduate Program Director

March 25, 2019

_____
Dr. Andre Roy, Dean
Faculty of Arts and Science

ABSTRACT

Anatomical and computational models of the role of phasic dopamine signaling in intracranial self-stimulation: psychophysical and electrochemical tests

Marie-Pierre Cossette, Ph.D.
Concordia University, 2019

Dopamine (DA) neurotransmission is heavily implicated in electrical intracranial self-stimulation (eICSS), operant performance aimed at triggering stimulation of certain brain regions. To study the underpinnings of reward seeking, we combined ICSS with fast-scan cyclic voltammetry (FSCV), a means of monitoring stimulation-induced DA transients.

Chapter one examines the circuitry linking midbrain DA neurons to the non-DA neurons recruited at the tip of a medial forebrain bundle (MFB) eICSS electrode. We found that unilateral, electrical, MFB stimulation evoked bilateral DA transients and that DA activation occurred, in large part, through polysynaptic circuitry.

The series-circuit hypothesis is the focus of chapter two. On that view, the signal representing the intensity of the stimulation-induced rewarding effect must pass obligatorily through midbrain DA neurons. We found that the the DAergic projections from the ventral tegmental area (VTA) to the nucleus accumbens (NAc) cannot serve as a unique link that relays the reward signal underlying eICSS of the MFB to later stages of the circuitry underlying reward seeking.

The last chapter addresses the dominant theory linking phasic DA signalling to learning. On that view, DA transients signal the discrepancy between expected and experienced rewards and adjust synaptic weights to update reward predictions and bias action selection. Several of our findings are difficult to reconcile with the DA-RPE hypothesis.

Recent technological advances have provided tools for studying the neural bases of reward seeking that are unprecedented in their power and number. It is important to extend this level of sophistication to our electrochemical and behavioural methods.

ACKNOWLEDGMENTS

I have felt deeply grateful many times in my life. Every year brings opportunities, and as some projects turn into amazing journeys it is always striking to realize the power of collective intelligence, creativity and sensitivity. Now, at the end of my graduate studies, I am looking back feeling overwhelmed with gratitude. I must begin by thanking my supervisor Peter Shizgal, it was an absolute privilege to be guided by you. Dave Munro and Steve Cabilio, without you, exciting research is simply not possible. Your dedication is admirable, and what we went through will always stay with me. Members of the Shizgal lab, you introduced me to the fascinating world of psychophysics and reward computations and encouraged me to take on ambitious projects. You believed in me and nourished my desire to think deeply. Members of the CSBN, you supported me through countless lows with smiles, nerdy jokes, the Matlab crush, the sharing of ideas and new friendships, thank you for it all. A special mention to Tricia John who taught me so much about molecular biology when we really needed to develop that expertise, and Ian Moreau-Debord and Pavel Solís for your wonderful views of the world. Heshmat Rajabi, you were my pillar, my expert in everything, thanks for your patience and calm. Aileen Murray, you gave me peace of mind. Isabelle Bouvier and Kim Breux, you have been wonderful. Always greeting me with a smile, you navigated university administration with me, repeatedly rescuing me from serious confusion. Security guards and cleaning staff, you have been amazing. I got to know so many of you over the years. You have brightened my days through conversations on late nights and long weekends of work. You have helped me with the little things that felt like mountains and told me that it always gets better when I needed to hear it. A special thanks to Kimberley Maxwell, the best editor in the world.

I also want to thank collaborators, that allowed me to ask a million questions and experiment with new techniques. Thanks to Dr. Bruce Hope and his team for my internship at the NIH and Dr. Read Montague and his team for an incredible week at Virginia Polytechnic Institute and State University. Thank you to Dr. Peter Dayan and his team for the fruitful conversations about reinforcement learning. Finally, the optogenetic experiments would have not been carried out without the advice of Dr. Antoine Adamantidis and his team as well as Dr. Karl Deisseroth and Dr. Ilana Witten.

Family and friends, you have been extremely patient, compassionate, nonjudgmental and loving. Thanks to all of you for making me a strong enough person to manage my ever-branching trains of thought to finish this Ph.D. Dad, your metaphor of a fragile formula one will stay with me forever. Marwan, you brought me peace, filled me with joy, and fuelled my desire to thrive.

This doctoral thesis was supported by NSERC (Alexander Graham Bell Canada Graduate Scholarships), a Concordia University Graduate Fellowship, a Concordia University Special Entrance Award, a group grant from the "Fonds de recherche Québec—santé" to the "Groupe de Recherche en Neurobiologie Comportementale"/Center for Studies in Behavioural Neurobiology and NSERC Discovery grants (RGPIN-308-11 and RGPIN-2016-06703).

CONTRIBUTION OF AUTHORS

Chapter 2 includes a published manuscript. The contribution of authors is listed below.

I contributed to the experimental design and carried out pilot experiments as well as the main experiment. I performed the surgeries, fast-scan cyclic voltammetry, most of the data analysis, the histology and contributed in writing all versions of the manuscript.

Kent Conover developed and carried out part of the analysis on behavioral data.

Peter Shizgal contributed to the experimental design, developed the data analysis strategy and contributed in writing all versions of the manuscript.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
ChR2	Channelrhodopsin-2
DA	Dopamine
DA-RPE	Dopamine-reward-prediction-error
DMS	Dorsomedial striatum
DS	Dorsal striatum
eICSS	Electrical intracranial self-stimulation
eNpHR	Enhanced halorhodopsin
EYFP	Yellow fluorescent protein
FSCV	Fast scan cyclic voltammetry
GABA	Gamma-Aminobutyric acid
GECIs	Genetically encoded calcium indicators
ICSS	Intracranial self-stimulation
ITI	Inter-trial interval
LDTg	Laterodorsal tegmental nucleus
LH	Lateral hypothalamus
MFB	Medial forebrain bundle
MLE	Maximum likelihood estimate
NAc	Nucleus Accumbens
oICSS	Optical intracranial self-stimulation

PCA	Principal component analysis
PCR	Principal component regression
PVT	Paraventricular nucleus of the thalamus
RL	Reinforcement learning
RNA	Ribonucleic acid
RPE	Reward prediction error
RSS	Residual sum of squares
SN	Substantia nigra
TD	Temporal difference
TDRL	Temporal-difference reinforcement-learning
TH	Tyrosine hydroxylase
VTA	Ventral tegmental area

General Introduction

Reward seeking is fascinating: there can be an infinite number of reasons why we seek rewards and yet only a limited number of behaviours we can execute to obtain them. Even more intriguing is the quest to determine the causes of behaviours. Let us examine a classroom setting as a case in point. When I sign up for a class, I am assiduous and often raise my hand to ask questions. If an observer that frequently visits the nearby park notices my action, he might conclude that I am very engaged when I attend lectures. A more precise analysis of the reasons for my active participation, however, would be much more difficult to perform from the observer's viewpoint. The source of my question might be to express my curiosity, request a clarification, offer my opinion to an ongoing discussion, voice a disagreement, or ask for assistance because I feel unwell. Multiple underlying reasons for my behaviour, yet only one observable outcome—my hand raised to grab the teacher's attention.

When teasing apart the different neural pathways involved in reward seeking, researchers are often akin to the observer in the park. The study of a small number of behaviours is used to infer the computational mechanisms at play when an organism works to harvest rewards. Imposing stringent tests to our existing theories is critical to keep building on the solid foundations established over the years in the field of neuroscience and provide new insightful results to continue our progress. In this thesis, we proposed approaches to extend and challenge theories about one specific type of neurotransmission. More precisely, we examined how midbrain dopamine (DA) neurons are activated by reinforcers and the role of phasic DA release in reward seeking.

DA neurotransmission has been heavily implicated in feeding, drug seeking, reproduction, as well as electrical intracranial self-stimulation (eICSS), a phenomenon in which animals work to harvest stimulation of certain brain regions, particularly when the stimulating electrodes are located along the trajectory of the medial forebrain bundle (MFB) (Olds & Olds, 1963). An increase or decrease in DA neurotransmission potentiates or attenuates lever pressing during eICSS, respectively (Gallistel & Karras, 1984; Wise, 2004; Trujillo-Pisanty, Hernandez, Moreau-Debord et al., 2011; Hernandez, Trujillo-Pisanty, Cossette et al., 2012). In addition,

selective optical stimulation of midbrain DA neurons also elicits ICSS (Witten, Sternberg, Lee et al., 2011). We combined ICSS with fast-scan cyclic voltammetry (FSCV) to monitor stimulation-induced DA transients in the context of reward seeking.

Chapter one examines the circuitry that links midbrain DA neurons to the non-DA neurons recruited at the tip of a MFB self-stimulation electrode. We found that unilateral, electrical, MFB stimulation evoked DA transients in both hemispheres. Given what is known about the physiology and anatomy of the MFB (Yeomans, Kofman & McFarlane, 1985; Yeomans, 1989; Murray & Shizgal, 1994; Shizgal et al., 1980; Wise, 1980; Maeda & Mogenson, 1981; Nieuwenhuys, Geeraedts & Veening, 1982; Veening, Swanson, Cowan, Nieuwenhuys & Geeraedts, 1982; Watabe-Uchida, Zhu, Ogawa et al., 2012; Molochnikov & Cohen, 2014), our finding suggests that action potentials triggered in non-DA neurons are relayed to midbrain DA cell bodies, in large part, through polysynaptic circuitry. Chapter one also bridges two types of studies that have been largely separate: we chose stimulation parameters to empirically close the gap between the methodological differences of experiments reporting the behavioural effects of MFB stimulation and those reporting the neurochemical effects. We provide evidence that data coming from both types of studies can be integrated with confidence.

The idea that non-DA neurons provide inputs to midbrain DA neurons is consistent with a proposition called the series-circuit hypothesis (Shizgal, Bielajew, Corbett et al., 1980; Wise, 1980; Moisan & Rompré, 1998). On that view, the signal representing the intensity of the stimulation-induced rewarding effect must obligatorily pass through midbrain DA neurons, even if it arises in the direct or indirect afferents. The approach presented in chapter two tested this hypothesis and revealed that the DAergic projections from the ventral tegmental area (VTA) to the nucleus accumbens (NAc) cannot serve as a unique link that relays the reward signal underlying eICSS of the MFB to later stages of the circuitry underlying reward seeking.

The last chapter addresses the dominant theory linking phasic DA signalling to learning through the encoding of reward prediction errors (RPEs) (Montague, Dayan & Sejnowski, 1996; Steinberg, Keiflin, Boivin et al., 2013; Colombo, 2014; Schultz, 2015; Keiflin & Janak, 2015; Chang, Esber, Marrero-Garcia et al., 2016; Chang, Gardner, Gonzalez Di Tillio et al., 2017;

Chang, Gardner, Conroy et al., 2018; Parker, Cameron, Taliaferro et al., 2016; Watabe-Uchida, Eshel & Uchida, 2017; Nasser, Calu, Schoenbaum et al., 2017; Sharpe, Marchand, Whitaker et al., 2017; Keiflin, Pribut, Shah et al., 2019). On that view, also called the DA-RPE hypothesis, DA transients signal the discrepancy between expected and experienced rewards (the RPE), and adjust synaptic weights to update reward predictions *and* bias action selection (e.g., lever pressing to harvest a reward). We found several of our findings difficult to reconcile with the dopamine-reward-prediction-error (DA-RPE) hypothesis.

The advent of optogenetic techniques has given researchers tools to selectively stimulate a given type of neuron and precisely control neural activity on a timescale of milliseconds (Deisseroth, 2011; Bernstein & Boyden, 2011; Tischer & Weiner, 2014). Precise real-time monitoring of brain activity (fiber photometry, biosensors, genetically-encoded calcium sensors) is also routinely performed in numerous contemporary neuroscience laboratories (Akerboom, Calderon, Tian et al., 2013; Boyden, 2015; Chernov, Redchuk, Omelina et al., 2017) and is becoming an invaluable approach to probing the effects of optogenetic perturbations on reward seeking. Thus, there is an unprecedentedly large number of new ways to dissect the reward circuitry and examine the role of phasic DA release in reward seeking. To rigorously test theories linking specific neural signals and populations to their precise roles in the evaluation and pursuit of rewards, it is important to extend this level of sophistication to our behavioural and neurochemical methods. This thesis describes several steps taken toward this goal.

Chapter 1

Electrochemical measurement of dopamine release in the nucleus accumbens driven by electrical stimulation of the medial forebrain bundle or optical stimulation of the ventral tegmental area: implications for the structure of brain reward circuitry

Introduction

eICSS has been used for decades to study reward seeking. Rats and other vertebrates will work vigorously to trigger electrical stimulation, particularly when the stimulating electrodes are located along the trajectory of the MFB (Olds & Olds, 1963); vigorous eICSS is observed when animals pursue electrical stimulation of the MFB at the level of the lateral hypothalamus (LH) or at the level of the VTA. Despite a wide literature on the behavioural effects of MFB stimulation and multiple reports that DA neurotransmission is heavily implicated in eICSS, it remains difficult to attribute a precise role in the eICSS substrate to phasic DA release.

Psychophysical methods were developed to quantify the rewarding effects of the stimulation (Gallistel, 1978; Simmons & Gallistel, 1994; Gallistel, Shizgal & Yeomans, 1981) and examine the electrophysiological properties of neurons that produce these rewarding effects (Shizgal, Biejalew, Corbett, Skelton & Yeomans, 1980; Biejalew & Shizgal, 1982, Biejalew & Shizgal, 1986; Rompré & Shizgal, 1986; Shizgal, Schindler & Rompré, 1989, Murray & Shizgal, 1996). Dopaminergic manipulations were shown to have profound effects on performance, enhancing or blunting response vigor. Gradually, these findings were complemented with neurochemical measures to determine the effects of the stimulation on DA activity, first by tissue-punch or microdialysis techniques and later by means of FSCV (Phillips, Jakubovic, & Fibiger, 1987; Garrigues & Cazala, 1983; Nakahara, Ozaki, Kapoor et al., 1989). The timescale of FSCV measurements, with a resolution of 100 ms, is particularly suited to characterize stimulation-evoked DA transients, which made it possible to demonstrate that activation of DA neurons is time-locked to the delivery of individual electrical stimulation trains.

Bridging methodological discrepancies

Although psychophysical and neurochemical approaches appear to provide all the necessary

evidence to interrogate the correspondence between fluctuation in eICSS and DA activity, important methodological differences have limited our ability to fully achieve this goal. Most data collected using psychophysical methods were obtained with LH stimulation whereas FSCV experiments were mainly conducted using VTA stimulation. This methodological dichotomy extends to the choice of stimulation parameters: short pulse durations are commonplace in behavioural experiments whereas neurochemical profiles are obtained using long pulse durations. It would be desirable to pool information from both types of studies, but are the effects of LH and VTA stimulation on DA activity sufficiently comparable to merge the findings? Could the difference in stimulation parameters confound the conclusions drawn from pooled data?

Placing the electrode in different locations along the MFB and using different stimulation parameters could potentially recruit distinct subpopulations of neurons. It has long been argued that MFB stimulation activates midbrain DA neurons trans-synaptically, at least in large part, although the identity and extent of non-DA subpopulations that support this activation remain unknown (Yeomans, Kofman & McFarlane, 1985; Yeomans, 1989; Murray & Shizgal, 1994; Shizgal et al., 1980; Wise, 1980; Maeda & Mogenson, 1981). Moreover, the MFB contains at least 50 different types of fibers of different origin and destination (Nieuwenhuys, Geeraedts & Veening, 1982; Veening, Swanson, Cowan, Nieuwenhuys & Geeraedts, 1982). The heterogeneity of VTA afferents may also be similar (Watabe-Uchida et al., 2012). There are multiple sub-populations that could elicit stimulation-evoked DA transients, and these neurons may or may not be equally recruited by LH and VTA stimulation.

Even when data are acquired from the same brain region, subpopulations of neurons recruited by the stimulation can vary if the stimulation parameters are not held constant. Neuronal subtypes differ in excitability and hence differ in the likelihood they will be fired by a given set of parameters.

Strength-duration curves describe the dependence of the threshold current for firing a neuron on the duration of a stimulation pulse. Figure 1 shows two simulated curves, one of a highly excitable neuron in magenta and the other of a less excitable neuron in blue; the two solid circles represent the pulse durations tested in our experiment, 0.1 and 2 ms. A hyperbolic function is

often used to fit the data (Weiss, 1901).

$$I(d) = I_{r\_b}*(1+(c_b/d))$$

where I_{r_b} is the required current for an infinitely long pulse duration (the rheobase), c_b is the chronaxie, d is the pulse duration, and the subscript b represents a highly excitable axon (such as a **b**ig fiber)

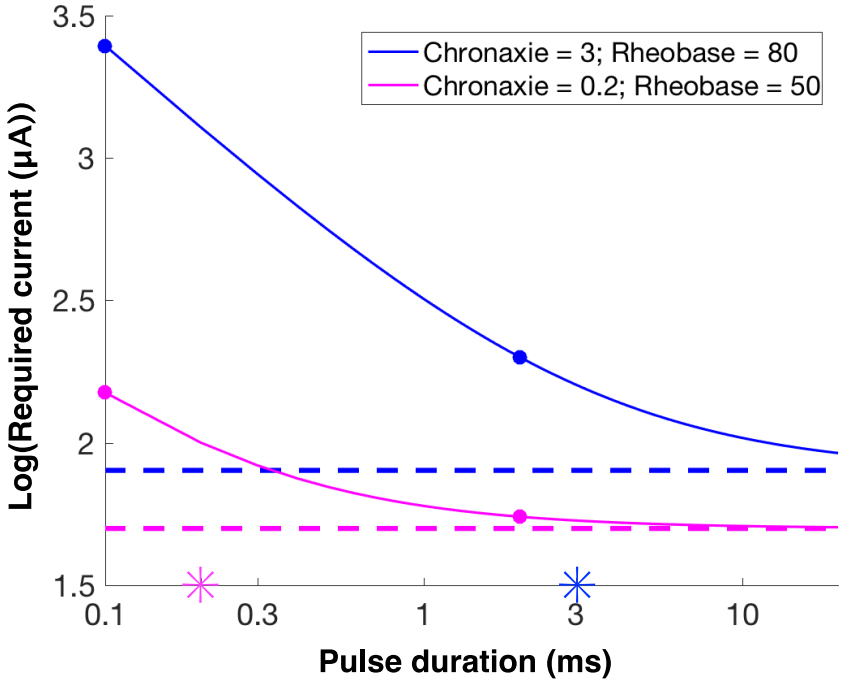


Figure 1. Simulated strength-duration curves for a highly excitable and a less excitable neuron

Strength-duration curves describe the dependence of the threshold current for firing a neuron on the duration of a stimulation pulse. The curves shown trace the hyperbolic function often used to fit strength-duration data (Weiss, 1901: $I(d) = I_r*(1+(C/d))$. The curves for the highly and the less excitable neuron are shown in magenta and in blue,

respectively. The rheobase, I_r, is the lowest current that suffices to fire a neuron in response to a pulse of infinite duration. Thus, the rheobase, represented as dashed lines, determines the horizontal asymptote of the strength-duration curves. The chronaxie, C, determines the curvature of the strength-duration function. It is the pulse duration at which the threshold current is twice the rheobasic value. The chronaxie values are depicted by asterisks in the figure. The solid circles represent the pulse durations used in the experiment, 0.1 and 2 ms. The currents required to fire the two types of neurons are more similar at the longer pulse duration than at the shorter pulse duration. Axes are logarithmically scales so as to facilitate visual comparison of the two curves.

The two parameters, I_r and C, characterize the excitability of neurons. The rheobase is the lowest current required to fire a neuron and thus determines the horizontal asymptote of the strength-duration curve, depicted by the dashed lines in our example. The rheobase depends on physical properties of the neuron, such as size and myelination, as well as on the distance of the neuron from the electrode tip. Highly excitable neurons have lower rheobase compared to less excitable neurons. The chronaxie is the pulse duration at which the threshold current is twice the rheobasic value, shown by the asterisks in our example. Thus, the chronaxie reflects the shape of the strength-duration curve; the lower the chronaxie, the more steeply the threshold declines as a function of pulse duration. Larger, myelinated axons tend to have shorter chronaxies than smaller, unmyelinated axons (West & Wolstencroft, 1983). Strength-duration curves that differ in chronaxie converge as the pulse duration is increased. Thus, short pulse durations will predominantly recruit highly excitable neurons whereas longer pulse durations will recruit a more heterogeneous population that includes both highly excitable and less excitable neurons (Bement & Ranck, 1969; Ranck, 1995, Yeomans, 1989). I have been able to find only one strength-duration curve for a DA neuron in the published literature that spans a range of pulse durations sufficient to obtain a good estimate of the chronaxie (Yeomans, Maidment & Bunney, 1988). Thus, it remains to be determined to what degree long-duration VTA pulses employed in neurochemical studies are more effective in recruiting DA neurons than the short-duration LH pulses used in most eICSS experiments.

It is important to note that even if all types of neurons recruited by LH and VTA stimulation were known and if the exact number of directly activated neurons triggered by any given set of stimulation parameters could be estimated precisely, predicting the overall net effect on phasic DA would still be difficult. Figure 2 illustrates this notion. DA, non-DA excitatory, and non-DA inhibitory neurons are depicted in red, blue, and green, respectively. Even when presenting a simplified circuitry, the complexity is obvious.

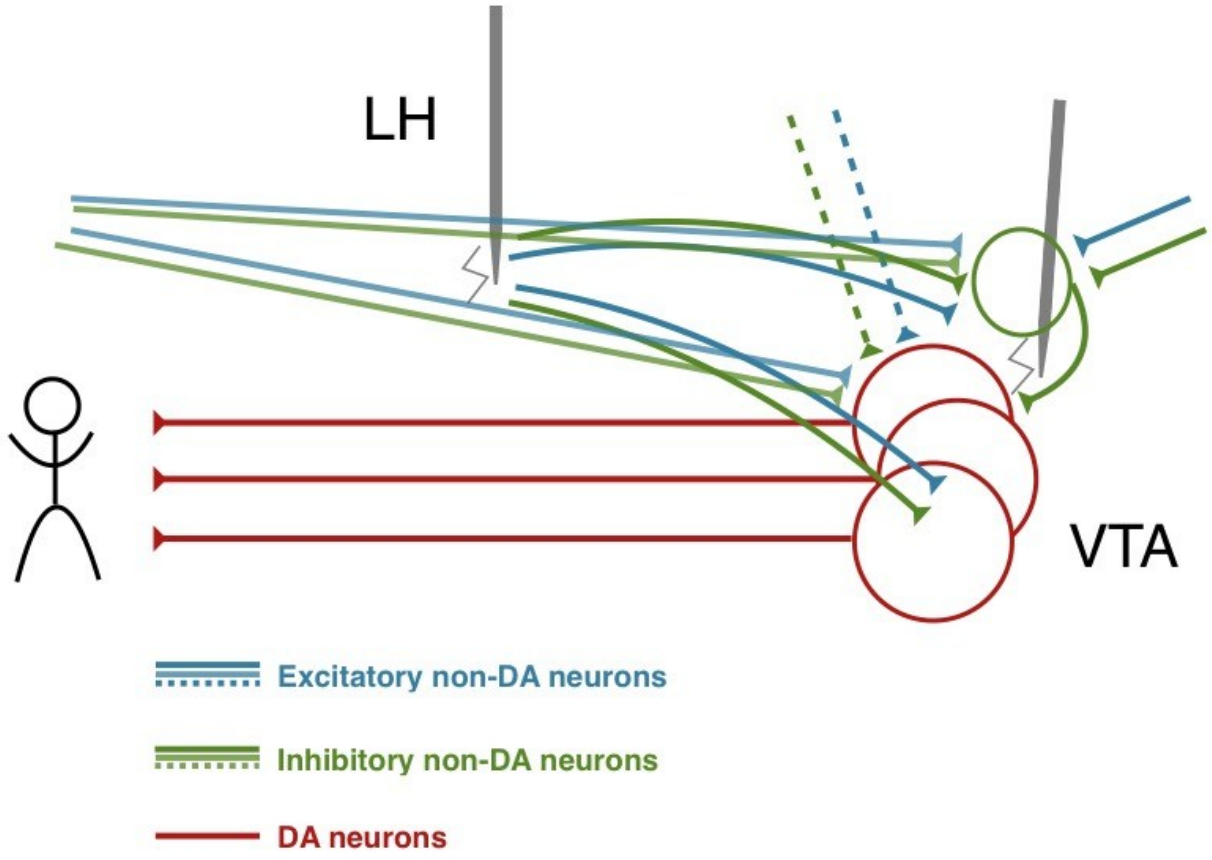


Figure 2. Many interacting neuronal cell types can contribute to the effects of MFB stimulation on phasic DA firing

DA neurons, non-DA excitatory afferents, and non-DA inhibitory neurons are depicted in red, blue, and green, respectively (the varying shades of blue and green and the dashes lines are used strictly to help visualize the overlapping fibers). Both inhibitory and excitatory neurons in the LH or in VTA afferents can be recruited and can have antagonistic effects. The targets of these inputs can be midbrain DA neurons or VTA

GABAergic interneurons (represented as green VTA neurons in the diagram). An inhibitory input can therefore inhibit DA activation or promote DA activation depending on its target, and the same logic applies for excitatory inputs. DA, dopamine; LH, lateral hypothalamus; VTA, ventral tegmental area; GABA, gamma-aminobutyric acid.

Both inhibitory and excitatory neurons in the LH or in VTA afferents can be recruited and can have antagonistic effects. The targets of these inputs can be midbrain DA or VTA gamma-Aminobutyric acid(GABA)ergic interneurons (represented as green VTA neurons in the diagram). An inhibitory input can therefore inhibit DA activation or promote DA activation dependent on its target, and the same logic applies for excitatory inputs. Given these possibilities, precise prediction of DA release by short-duration LH pulses and long-duration VTA pulses would require that all the spatial distributions of all the candidate neuronal populations be known, along with their likelihood of firing (based on strength-duration curves and current-distance curves). This information is not yet available. Thus, the issue must be approached empirically as I have done in the experiments presented in this chapter.

The first goal of this experiment was to bridge the methodological and empirical gap to ensure that behavioural findings obtained using LH stimulation could parallel neurochemical data obtained using VTA stimulation. We tested the hypothesis that the effects of LH and VTA stimulation on phasic DA activity are similar over the ranges of stimulation parameters used in the eICSS and FSCV literature.

The second goal was to gain more insight into the principles that govern input convergence onto midbrain DA neurons. The difficulty of predicting the net outcome of electrical stimulation on phasic DA release stems from a lack of knowledge about how multiple non-DA inputs interact with direct DA activation to produce transients at the terminals. We designed the acquisition of FSCV recordings to begin teasing apart the way non-DA inputs to midbrain DA neurons and direct DA activation combine to influence DA activity during MFB stimulation. More precisely, we questioned the relative contribution of non-DA and DA neurons in producing DA transients in terminal fields during electrical stimulation. To do so, LH and VTA stimulations were

delivered unilaterally and FSCV measurements were acquired in both hemispheres.

Inter-hemispheric comparison to determine the relative roles of MFB non-DA neurons and DA axons in evoking DA transients in the NAc

Unilateral MFB stimulation produces substantial increases in neural activity in both hemispheres some of which is dopaminergic (Phillips, Jakubovic, & Fibiger, 1987; Garrigues & Cazala, 1983; Nakahara, Ishida, Nakamura et al., 1992). The potential mechanisms by which midbrain DA neurons can be activated during MFB stimulation are numerous, but only a subset of these routes could produce contralateral DA transients. Midbrain DA neurons ascend toward striatal regions without crossing the midline, with the exception of a few fibers, estimated to be 1%–3% of projections (reviewed in Molochnikov & Cohen, 2014). This asymmetry provides an ideal way to isolate the contribution of non-DA neurons because these cells must play the principal role in the activation of the contralateral DA population.

The potential routes of DA activation can be grouped into three categories: direct, monosynaptic and polysynaptic. As shown in the simplified schematics in Figure 3, the stimulation electrode can recruit: 1) DA axons ascending directly through the MFB; 2) non-DA, monosynaptic, afferents to DA somata; or 3) non-DA neurons that activate intervening neurons that in turn synapse onto DA neurons. Only the third route is supported by available evidence as a means of producing large bilateral DA transients at the terminal fields.

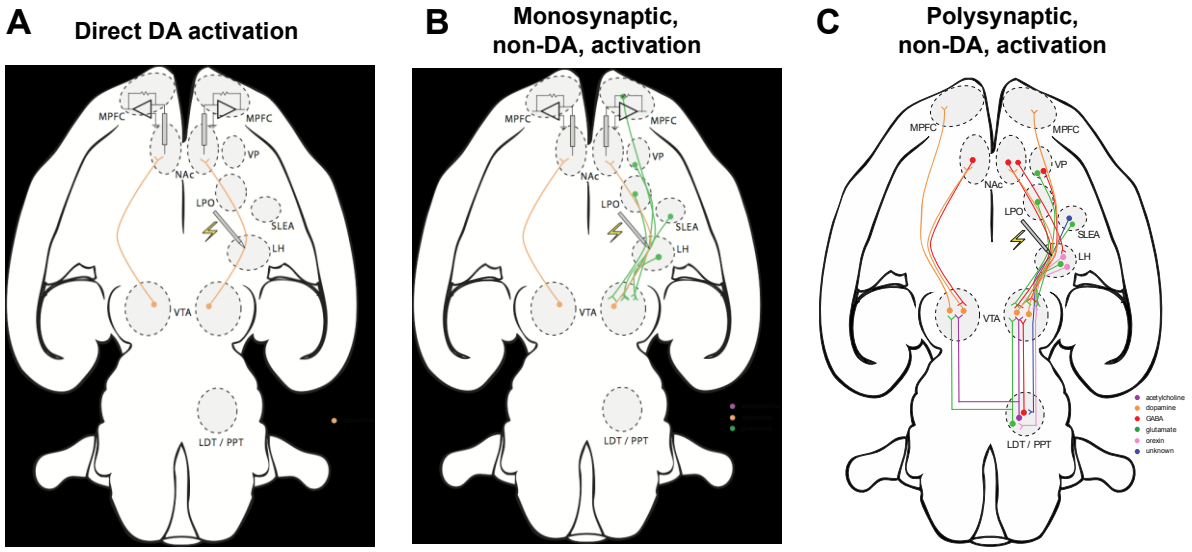


Figure 3. Different ways in which MFB stimulation can trigger DA transients in the NAc

A) Direct activation of the ascending DA axons. B) Monosynaptic activation: Descending, non-DA, MFB fibers are activated directly by the stimulation. These fibers provide excitatory synaptic input to DA neurons in the ipsilateral VTA. C) Polysynaptic activation: Descending, non-DA, MFB fibers are activated directly by the stimulation. These fibers provide excitatory synaptic input to neurons in the laterodorsal tegmental nucleus and pedunculopontine nucleus, which, in turn, provide excitatory synaptic input to DA neurons in both VTAs. Figure prepared by Ivan Trujillo and Peter Shizgal (personal communication). Only subsets of the known monosynaptic and polysynaptic inputs are shown. MFB, medial forebrain bundle; DA, dopamine; NAc, nucleus accumbens; VTA, ventral tegmental area.

Direct route: DA axons ascending directly through the MFB

An overwhelming majority of ascending DA axons project unilaterally to the striatal terminal field. It has been argued that, despite the scarcity of DA fibers that cross the midline, stimulating the ascending DA fibers can produce contralateral DA transients, as measured by FSCV (Fox, Mikhailova, Bass et al., 2016). Fox et al. (2016) examined the effects of unilateral electrical stimulation of the VTA and substantia nigra (SN) on DA release in the NAc and dorsomedial striatum (DMS). We will focus on data recorded in the NAc, which are most germane to our experiment. DA transients were observed in both hemispheres although the amplitude of the contralateral response was much smaller. They also examined the effects of optical stimulation in VTA DA neurons, placing the optical fiber in one hemisphere. They observed bilateral DA transients in that preparation as well, a result that they interpreted as a confirmation that stimulating crossing DA fibers can produce DA transients. They concluded that since stimulation of crossing DA fibers can produce DA transients; a small subset of DA fibers cross the midline; and contralateral DA transients during electrical stimulation were small, activation of crossing DA fibers was likely the source of the electrically-evoked DA transients in the contralateral NAc.

There are two main reasons why the conclusion of Fox et al. about the functionality of crossing DA fibers remains tentative. In their experiment, the expression of channelrhodopsin-2 (ChR2), the light-sensitive opsin, was abundant in VTA DA neurons in both hemispheres (Figure S6 in Fox et al.), and the authors only used single sets of electrical and optical stimulation parameters. To rule out the possibility that contralateral DA transients were produced by direct activation of contralateral VTA neurons, and not crossing DA fibers, control experiments restricting the spread of the virus or the light delivered to one hemisphere must be carried out. We conducted such control experiments. We also compared the amplitude of bilateral DA transients obtained for the set of stimulation parameters used by Fox et al. to the amplitudes produced by a much wider range of parameters and determined whether the contralateral DA transients were systematically smaller than the ipsilateral transients. We found that direct DA activation had negligible effects on phasic DA activity in the contralateral NAc.

Monosynaptic routes: Non-DA neurons or axons synapsing onto DA neurons

Electrophysiological and behavioural studies support the idea that monosynaptic DA activation can contribute to DA transients elicited by MFB stimulation. Optical stimulation restricted to VTA-projecting LH neurons can engage midbrain DA neurons, produce changes in phasic DA in the NAc, and support optical intracranial self-stimulation (oICSS) (Nieh, Matthews, Allsop et al., 2015; Nieh, Weele, Vander et al., 2016, Kempadoo, Tourino, Cho et al., 2013; Stuber & Wise, 2016). Anatomical data suggest that the LH to VTA connections in one hemisphere could impact DA activity in the other hemisphere: a retrograde tracer injected unilaterally into the VTA was found to label LH neurons in both hemispheres (Geisler & Zham, 2005). We reported, based on pilot experiments, that a subset of VTA-projecting neurons crosses the midline and terminates in the contralateral midbrain (Cossette & Shizgal, SfN poster, 2015), consistent with the neuroanatomical data. However, when we transfected wild-type rats to express a ChR2 variant nonspecifically in LH neurons and then optogenetically activated the contralateral LH fibers, only small, non-systematic DA transients were produced. Importantly, the few DA transients obtained did not grow in amplitude with increasing light power. We also found a LH to NAc projection that produced DA transients in the ipsilateral NAc when stimulated optically, consistent with the report that DA transients can be produced by directly stimulating the local circuity at DA terminal fields (e.g. through selective activation of cholinergic interneurons (Threlfell, Lalic, Platt et al., 2012). No transfected fibers, however, were visualized in the contralateral hemisphere. In addition, monosynaptic inputs to midbrain DA neurons originate predominantly from brain regions ipsilateral to their DA target, except for the parabrachial nucleus (Watabe-Uchida et al., 2012) and the neocortex (Menegas, Bergan, Ogawa et al., 2015). Monosynaptic DA activation produced by MFB stimulation is most likely restricted to the ipsilateral hemisphere.

Polysynaptic routes: Non-DA neurons activate intervening neurons that, in turn, synapse onto DA neurons

Anatomical and functional evidence point to at least two putative trans-synaptic routes of DA activation that could drive bilateral DA activity during MFB stimulation. The mesopontine areas (Cornwall, Cooper & Phillipson, 1990; Semba & Fibiger, 1992; Steininger, Rye & Wainer, 1992) and the paraventricular nucleus of the thalamus (PVT) (Otake & Nakamura, 1998 and

reviewed in Kirouak, 2015) receive LH inputs and project bilaterally to midbrain DA neurons (Oakman, Faris, Kerr et al., 1995; Dautan, Huerta-Ocampo). When electrically stimulated, the mesopontine areas and the PVT both produce changes in DA efflux in the NAc, as measured by amperometry (Forster & Blaha, 2000; Forster & Blaha, 2003; Blaha & Phillips, 1996; Yeomans, Forster & Blaha, 2001; Parsons, Li & Kirouac, 2007). These findings demonstrate the existence of trans-synaptic connections between the LH and the midbrain DA neurons that can affect phasic DA at the terminals.

Data presented in this section show that LH and VTA stimulation produce similar increases in phasic DA activity. We also report large contralateral DA transients for both stimulation sites that cannot be explained by direct activation of DA fibers crossing the midline, which are very few in number (reviewed in Molochnikov & Cohen, 2014). We found that the amplitude of DA transients grew systematically as a function of stimulation current, suggesting convergence in the neural activity produced at the tip of the stimulation electrode in the neural substrate that drives changes in phasic DA activity during MFB stimulation.

Data obtained during my Master's thesis and presented in poster format has already demonstrated that bilateral DA transients are detectable during MFB stimulation. As part of my doctoral work, more extensive analyses on the growth of DA transients were carried out in each data set. Following the same procedure, new voltammetric recordings were obtained in one group of animals receiving electrical stimulation in the VTA and in three groups of animals receiving specific optogenetic stimulation of DA VTA neurons. To draw a complete and coherent story, all data from my masters and my doctoral work, are presented as a whole in this chapter.

Methods

Subjects

Subjects were wild-type Long Evans male rats from Charles River (St. Constant, QC, Canada) or Long Evans TH::Cre males bred in-house. Once assigned to their experimental group, animals were housed individually with ad libitum access to food and water and maintained on a 12-h light/dark reverse cycle (lights off from 08:00 to 20:00). The experimental procedures were performed in accordance with the principles outlined by the Canadian Council on Animal Care.

Electrical Stimulation

Surgery

The electrical assembly

Rats were anesthetised with urethane (1.5 g/kg, i.p.), injected with atropine sulphate (0.05 mg/kg) to reduce bronchial secretions, and placed in the stereotaxic frame (Kopf Instrument, Tujunga, CA). Burr holes were made, through which the electrodes were lowered. One stimulating electrode, fashioned from a 0.25 mm stainless steel insect pin and insulated with Formvar enamel to within 0.5 mm of the tip, was aimed at the right MFB at the level of the LH (AP: −2.8 mm; ML: 1.7 mm; DV: 8.9 mm from the skull; all coordinates are referenced to bregma). Another electrode, from which 3 mm of insulation was removed at its tip, was placed in the external plexiform layer of the olfactory bulb (AP: 6.2 mm; ML: −0.1 mm; DV: 5.0 mm from the skull; all coordinates are referenced to bregma) and served as the anode of the stimulation circuit. The location of the anode was chosen so the recording site was mid-way between the cathode and anode of the stimulation circuit. At the mid-way position, the recording site should lie on or near a zero-potential surface, thus reducing interference between the electrical stimulation and the voltammetric measurements. The electrode assembly was secured with dental acrylic and anchored with jeweller screws. The output of the constant-current amplifier was routed to the stimulation electrode to deliver the stimulation trains.

The fast-scan cyclic voltammetry assembly

A carbon-fiber electrode, pre-cleaned with 2-propanol containing activated carbon, was slowly lowered into the NAc shell (AP: 1.7 mm; ML: 1.0 mm; DV: 7 mm from the skull). A sintered

Ag/AgCl reference electrode with a 5-mm conductive surface (In vivo metrics, Healdsburg, CA) was placed in the hemisphere contralateral to the carbon fiber, on or near the zero potential surface of the electrical field around the stimulation current (tip of the electrode roughly at AP: 0.45 mm; ML: 5 mm; DV 6 mm). The carbon-fiber and reference electrodes were then connected to the voltammetric amplifier and data acquisition system.

The density in DA innervation in the NAc varies greatly along the dorsal-ventral axis and can be described as patchy. To optimize the detection of DA transients, the NAc was systematically swept to select the optimal depth at which the carbon-fiber electrode yielded a maximal amplitude response. MFB stimulation trains (2-ms pulse duration, 500-ms train duration, 400 μA, and 60 Hz) were delivered as the carbon-fiber electrode was lowered in small steps. Evoked DA transients were measured (see below) by means of FSCV.

Lowering the same carbon-fiber electrode multiple times often results in attenuated sensitivity with each attempt. A fresh electrode was used for each lowering to avoid introducing such a confounding variable. Responses to all selected stimulations were performed in one hemisphere and then repeated in the other one. The first hemisphere sampled was randomly selected.

The animal's temperature was controlled throughout the procedure with a heating pad. A subcutaneous (s.c.) injection of Ringer's solution (6 mL/kg) was given every hour.

The exact same procedure was followed for the group of animals that received VTA stimulation with the exception that the stimulation electrode was aimed at the VTA (AP: −5.8 mm; ML: 0.8 mm; DV: 8.5 mm from the skull; all coordinates are referenced to bregma).

Stimulation parameters

To best assess the growth of DA transients as a function of current with both the short and long pulse duration, we determined the widest range of currents that could be tested at each sampled NAc site. As briefly explained in the previous section, the depth of the carbon fiber was selected based on voltammograms with the best signal-to-noise ratio in response to MFB stimulation at 2 ms pulse duration, 500 ms train duration, 400 μA, and 60 Hz. Then, stimulating currents were

increased to find the maximal current capable of producing a DA transient without eliciting large jaw or nose movements (such reflexes are commonly reported in urethane-anesthetized rats). The same process was performed to find the lowest current capable of producing a DA transient. In order to determine the ideal step size by which to increment current from the minimal to the maximal value, the next step was to obtain the amplitude in DA while increasing the current in 0.1 log increments. A 0.1 log step was often sufficient to produce significant increases in the size of DA transients at most of the steps. The size of the step could be increased if that requirement was not met. This procedure was repeated for each hemisphere with 0.1-ms and 2-ms pulse widths. All currents were delivered at a fixed pulse frequency of 60 Hz and a train duration of 500 ms. Every combination of parameters, pulse duration and current, was delivered six times, each repeat spaced by 60 s.

Once all necessary data sets were acquired for each pulse duration in both hemispheres, we probed additional sites along the medial-lateral or the anterior-posterior axes and recorded additional voltammetric data if larger DA transients were detected. Only the data sets corresponding to the largest responses were kept for data analysis.

Optogenetic Stimulation

First surgery: viral transfection

For the first surgery, rats were anesthetized with a mixture of ketamine hydrochloride (87 mg/kg) and xylazine hydrochloride (13 mg/kg) (injected i.p.). Rats were given a s.c. injection of atropine sulphate (0.05 mg/kg) to reduce bronchial secretions during surgery, a s.c. injection of buprenorphine to alleviate post-operative pain (0.05 mg/kg), and a s.c. injection of penicillin procaine to prevent infection (0.3 cc/rat). Burr holes were made to lower the injector.

We transfected VTA DA neurons in TH::Cre males with an adeno-associated virus serotype-5 (AAV5) containing an EF1a-DIO-hChR2(H134R)-EYFP-WPRE transcript suspended in phosphate buffered saline. Animals received six 0.5 μL injections in each hemisphere (AP: 5.4, ML: 1.0, DV: 8.3, 7.7, 7.2; AP: 6.2, ML: 1.0, DV: 8.3, 7.7, 7.2). We injected at 0.1 μL per min via a Hamilton syringe connected to a Harvard pump. We waited 10 minutes between injections. The burr holes were covered with bone wax.

Second surgery: optical fiber implantation and stimulation

Prior to the second surgery, 4 to 8 weeks were allowed for virus expression and integration of opsins into the membranes of the targeted neurons. Animals were prepared for voltammetric measurements following the same procedure as for the MFB stimulation group (see method above), with the exception that an optical fiber connected to a 473-nm diode-pumped solid-state (DPSS) laser (Laserglow Technologies) was lowered in the VTA instead of a stimulation electrode in the MFB. The optical implant was fashioned from a 300-µm core optical fiber, stripped of its cladding. One end was inserted in a stainless alloy ferrule glued in place and polished. The optical implants were secured with dental acrylic and anchored with jeweller screws. In four animals, the optical implants were inserted in a bevelled cannula to prevent the light from reaching the contralateral VTA. Prior to surgeries, the light spread emitted at the tips of these optical implants was visually inspected. We confirmed that light was obstructed on the side of the implant that would be placed toward the midline.

The animal's temperature was controlled throughout the surgeries with a heating pad. A s.c. injection of Ringer's solution (6 mL/kg) was given every hour.

Stimulation Parameters

The range of light power for optogenetic stimulation was limited between 1 mW and 40 mW and varied with a spacing of 0.3 log units. One concern with increasing light power is the production of heat capable of altering brain functioning. A 10-mW light stimulation applied continuously elevates brain temperature by 2°C rapidly over a few seconds. This variation drops to 0.2°C when the duty cycle is reduced to 10%. Thermal changes of 1°C to 3°C can lead to drastic perturbations in firing rate, transmembrane ionic transport, membrane properties, probability of transmitter release, and presynaptic uptake (Stujenske, Spellman & Gordon, 2015). Consequently, even if we used a duty cycle of 5%, we limited the highest light power to 40 mW when measured at the output of the optical implant. Moreover, 40 mW supports high rates of self-stimulation for optical VTA DA stimulation over several weeks with no sign of gradual damage, confirming that the highest light power used in our experiment does not present a biological hazard.

First, the depth of the carbon fiber was selected based on voltammograms with the best signal-to-noise ratio in response to optical stimulation at 40 mW and 40 Hz. Then, we found the lowest light power capable of producing DA transients. We tested as many light powers as possible between 1 mW and 40 mW to obtain DA transients significantly different in amplitude. Increasing light power in increments of 0.3 log units was typically sufficient to meet this requirement. All optical trains were delivered at a fixed pulse frequency of 40 Hz, a pulse duration of 5 ms, and a train duration of 1000 ms. Every light power was delivered six times, each repeat spaced by 60 s. This procedure was repeated for each hemisphere in a counterbalanced order.

Once all necessary data sets were acquired, we probed additional sites along the medial-lateral or the anterior-posterior axes and recorded additional voltammetric data if larger DA transients were detected. Only the data sets corresponding to the largest responses were kept for data analysis.

Voltammetry

Electrode fabrication

The carbon fiber was glass-encased and a seal was produced by heating the glass capillary with a pipet puller (PUL-1, WPI, Sarasota, FL). A wire covered with silver paint was inserted in the capillary to make contact with the carbon fiber and secured with shrink tubing coated with epoxy. The exposed portion of the fiber was cut to a length of 100-150 uM.

Data acquisition

Background-subtracted cyclic voltammograms were generated at 10 Hz by applying a 8.5 ms triangular waveform that ramped from −0.4 V to +1.3 V and back to −0.4 V at a scan rate of 400 V/s. The potential was held at −0.4 V between each scan to promote cation absorption at the surface of the FSCV electrode. All potentials were measured with respect to the Ag/AgCl reference electrode. The waveform was generated using LabVIEW (National Instruments, Austin, TX) and a multifunction data acquisition board (PCI-6052E, National Instruments, Austin, TX).

A PCI-6711E board (National Instruments, Austin, TX) was used to synchronize waveform acquisition and data collection. A synchronization signal from the PCI-6711E board was sent to the external input of a multi-channel pulse generator (Master-8, A.M.P.I, Israel) and used to trigger the electrical or optical stimulation 5 s after the start of each recording. In the case of the electrical stimulation, the pulse generator was programmed to prevent overlap between the electrical pulses and the voltammetric scans. This was accomplished by confining pulse generation to the 91.5 ms intervals separating the triangular waveform. Voltages generated by the Master-8 were converted to constant currents via a stimulus isolation unit (AM-2200, AM-Systems, Carlsborg, WA).

Carbon-fiber electrodes were conditioned at 60 Hz for 10 min and allowed to stabilize at 10 Hz for 10 min prior to acquiring a data set in a new location in the NAc.

Statistical Analysis

Averaging

Voltammetric current was first normalized by background subtraction. Backgrounds were manually selected at a time point during the 5 seconds preceding each stimulation delivery. In doing so, we ensured that backgrounds were not distorted by noise artefact (rarely detected in anesthetized animals). The background subtraction was performed by the TarHeel software also used to visualize real-time voltammetric current during the recording session. An average of five scans before and after the manually selected scan was computed and that average was removed from all data points to produce the false-colour voltammogram.

Means and standard errors of the mean were computed for the traces showing the amplitude of DA transients as a function of time, also called voltammetric traces (Figure 4A and 4B), and for the cyclic voltammogram showing changes in voltammetric current as a function of the voltages swept during each application of the triangular waveform (scan), increasing from –0.4 V to 1.3 V and back to –0.4 V (Figure 4A and 4C).

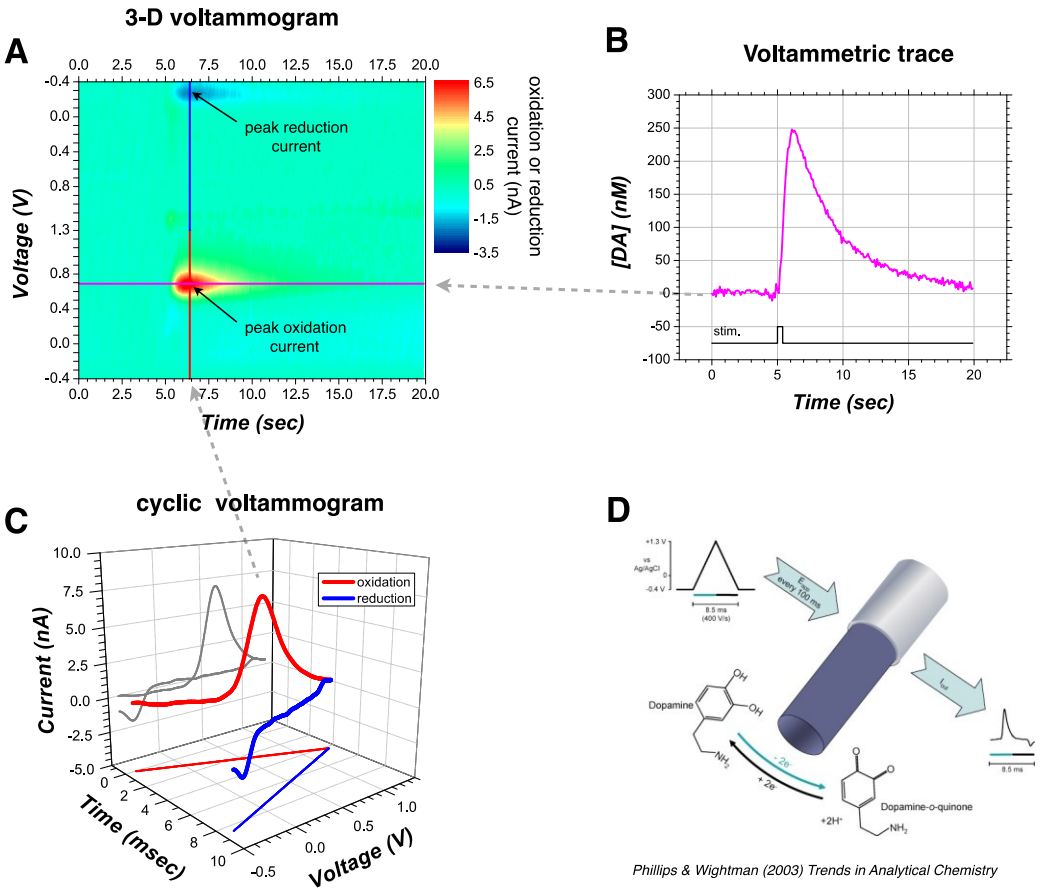


Figure 4. Principles of voltammetry

A) Each stimulation-induced DA transient is represented by a three-dimensional, false-colour voltammogram. Time is represented on the x-axis, whereas the amplitude of a triangular voltage wave applied every 0.1 s to the carbon-fiber electrode is represented on the y-axis. The dimension represented in false colour portrays the background-subtracted currents registered by the carbon-fiber electrode. Backgrounds were manually selected at a time point during the 5 seconds preceding each stimulation delivery. In doing so, we ensured that backgrounds were not distorted by noise artefact (rarely detected in anesthetized animals). The background subtraction was performed by the TarHeel software also used to visualize real-time voltammetric currents during the recording session. An average of 5 scans before and after the manually selected scan was computed and removed from all data points to produce the false-colour voltammogram. B) A voltammetric trace corresponds to the horizontal cross-section of

the false-colour voltammogram. It represents voltammetric changes as a function of time for the potential at which the background-subtracted oxidation current, and hence the DA concentration, was maximal. C) A cyclic voltammograph corresponds to the vertical cross-section of the false-colour voltammograph. It plots the background-subtracted current as a function of potential applied to the carbon fiber at the time when the DA concentration was maximal. D) Each voltammetric scan consists of a 8.5 ms triangular waveform that ramped from −0.4 V to +1.3 V and back to −0.4 V. DA oxidizes into DA-o-quinone optimally at roughly 0.6 V, producing a positive peak in the current around that potential and DA-o-quinone reduces back into DA at roughly –0.2 V, producing a negative peak in current. DA, dopamine.

Peak detection

Peak detection to quantify the amplitude of DA transients at each electrical current or light power was performed using a customized routine in MATLAB (The Mathworks, Natick, MA). The maximal change in oxidation current was extracted from the rising phase of the triangular waveform at a potential around 0.65 V, shown as by the red spike in Figure 4C. For conversion into molar concentrations, these peak currents were compared with an average of *in vitro* pre-calibration measurements taken from several carbon-fiber electrodes.

Polynomial approximations and model selection

The growth in peak DA was assessed by testing polynomial fits of degrees ranging from 0 to 10 for each current or power sweep. The best approximation was selected using the Akaike Information Criterion (AIC). An approach was developed to select the model offering the best combination of accuracy and parsimony for a given empirical data set. This was the model with the lowest AIC value. A customized routine in MATLAB was written based on equations and recommendations for AIC analysis from Burnham and Anderson (2002).

Fits for the different polynomials were performed with the polyfit function in MATLAB, and the residual sum of squares (RSS) was calculated with the polyval function in MATLAB. A maximum likelihood estimate (MLE) was obtained using the following equation:

$$MLE = RSS / n \text{ (Burnham \& Anderson, 2002, p.63)}$$

where n is the number of observations obtained at each current or power sweep.

An AIC for small sample size, often denoted as AICc, was then computed:

$$AICc = -2*\ln(MLE)+2*K(n/(n-K-1)) \text{ (Burnham \& Anderson, 2002, p.66)}$$

where K is the number of parameters in the polynomial, including the intercept.

An AIC value, by itself, is not interpretable; its absolute value carries very little information. AIC values are highly dependent on scales used to obtain data and sample size. AIC values allow for the models to be ranked from the best fit to the worst fit. However, more than one model might have an AIC very close to the lowest AIC. To select of a model that is "convincingly the best" (Burnham & Anderson, 2002, p.77) Burnham and Anderson (2002) recommend computing AIC differences:

$$AIC_{diff} = AIC_i - AIC_{min} \text{ (Burnham \& Anderson, 2002, p.71)}$$

where AIC_i is the AIC for a given model and AIC_{min} is the lowest AIC across all models.

The authors also recommend computing what they call Akaike weights. These weights represent the relative strength of evidence for each model and rely on the principle that the likelihood of the model *i* given the data is proportional to $\exp(-0.5*AIC_{diff})$. This likelihood is normalized to cause the Akaike weights to range from 0 to 1 (Burnham & Anderson, 2002, p.75-78).

$$W_i = \exp(-0.5*AIC_i) / \sum_{0to10}(\exp(-0.5*AIC_{0to10})) \qquad (1)$$

where Wi is the Akaike weight. (Burnham & Anderson, 2002, p.75)

Akaike weights can be interpreted as the weight of evidence in favour of a given model: the closer the 1, the higher the evidence in favour. To determine if models other than the one with the lowest AIC could also be candidates for "convincingly the best," we computed evidence ratios:

$$\text{Evidence ratio} = W_{min} / W_i \quad (2)$$

where W_{min} is the Akaike weight for the model with the lowest AIC and W_i is the Akaike weight for the model under consideration.

Models with small AIC_{diff} have large Akaike weights similar to the weight of the model with the lowest AIC (see equation 1). Their evidence ratio thus tends toward a value of 1. When considering models with increasing AIC_{diff} and smaller weights (less evidence in favour of those models), the evidence ratios grow exponentially (see equation 2 and Burnham & Anderson, 2002, p.77). As a rule of thumb, models associated with AIC_{diff} of more than 10 are not worth considering, and models with AIC_{diff} of 2 or below have strong evidence in their favour. The models with evidence ratios close to 1 are equally likely to be the best model. We applied these rules to select good models with the least number of parameters.

All selected models had AIC_{diff} below 2 and evidence ratios below 3. We inspected plots of the fitted curves and the data points. We found the selected model always appeared to be a good approximation of the data without overfitting.

Growth analysis

The AIC procedure mostly selected linear fits. Consequently, we assessed growth in peak DA using regression lines. We calculated Pearson correlation coefficients using the corr function in MATLAB and displayed the central tendencies between conditions with box plots. We visualized the effect of the conditions (hemispheres and pulse widths) on the slopes of the linear fits by means of box plots.

Histology

Perfusion

After completion of the experiment, a lethal dose of sodium pentobarbital (120 mg/rat, i.p.) was administered to ensure that animals would still be under deep anesthesia before proceeding with the perfusion procedures. The lack of foot and tail responses to a pinch was used to confirm that state. For all groups of animals, a stainless steel rod (diameter: 400 µm) was lowered at each recording site in the NAc prior to perfusion. This step created damage to allow visualization of the placement of the carbon-fiber electrodes using microscopy (see below). In the groups that received MFB or VTA stimulation, a 1-mA anodal current was applied for 15 s to deposit iron ions at the tip of the stimulating electrode, also for visualization. In the optogenetic groups, the optical fibers were large enough to leave damage to confirm their location in the VTA.

Animals were perfused intracardially with 0.9% sodium chloride followed by 4% paraformaldehyde. Then, the brains were removed and stored in a 30% sucrose solution made with 4% paraformaldehyde.

Slicing and microscopy

Sagittal sections, 40 µm thick, were cut with a cryostat (Thermo Scientific). Sections containing neurons transfected with an AAV5-EF1a-DIO-hChR2(H134R)-EYFP-WPRE transcript were coverslipped with Vectashield to prevent quenching of the enhanced yellow fluorescent protein (EYFP) and were examined using epifluorescence microscopy. Sections from animals that received electrical stimulation were examined using light microscopy.

Results

Bilateral changes in phasic DA in response to unilateral MFB stimulation at the level of the LH

The histology revealed that stimulation electrodes were in the targeted region of the MFB at the level of the LH and the recording carbon fibers were in comparable locations in both NAcs (Figure 5).

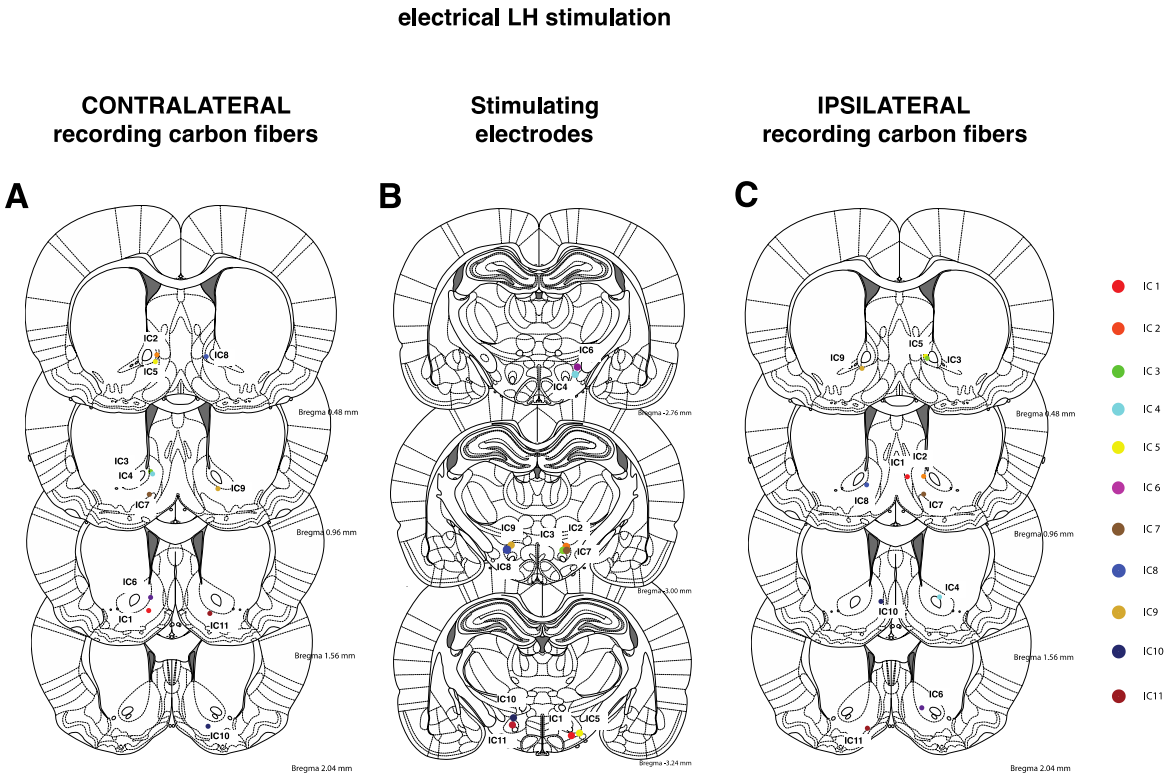


Figure 5. LH stimulation: histological localization of stimulation and recording sites

Location of the tips of A) the contralateral and C) the ipsilateral recording carbon fibers and B) the stimulating electrodes aimed at the LH. LH, lateral hypothalamus.

Unilateral LH stimulation produced changes in phasic DA in both NAcs using a range of stimulation parameters (Figure 6). All DA transients were time locked to the delivery of the stimulation trains and their cyclic voltammographs were characteristic of oxidized and reduced DA.

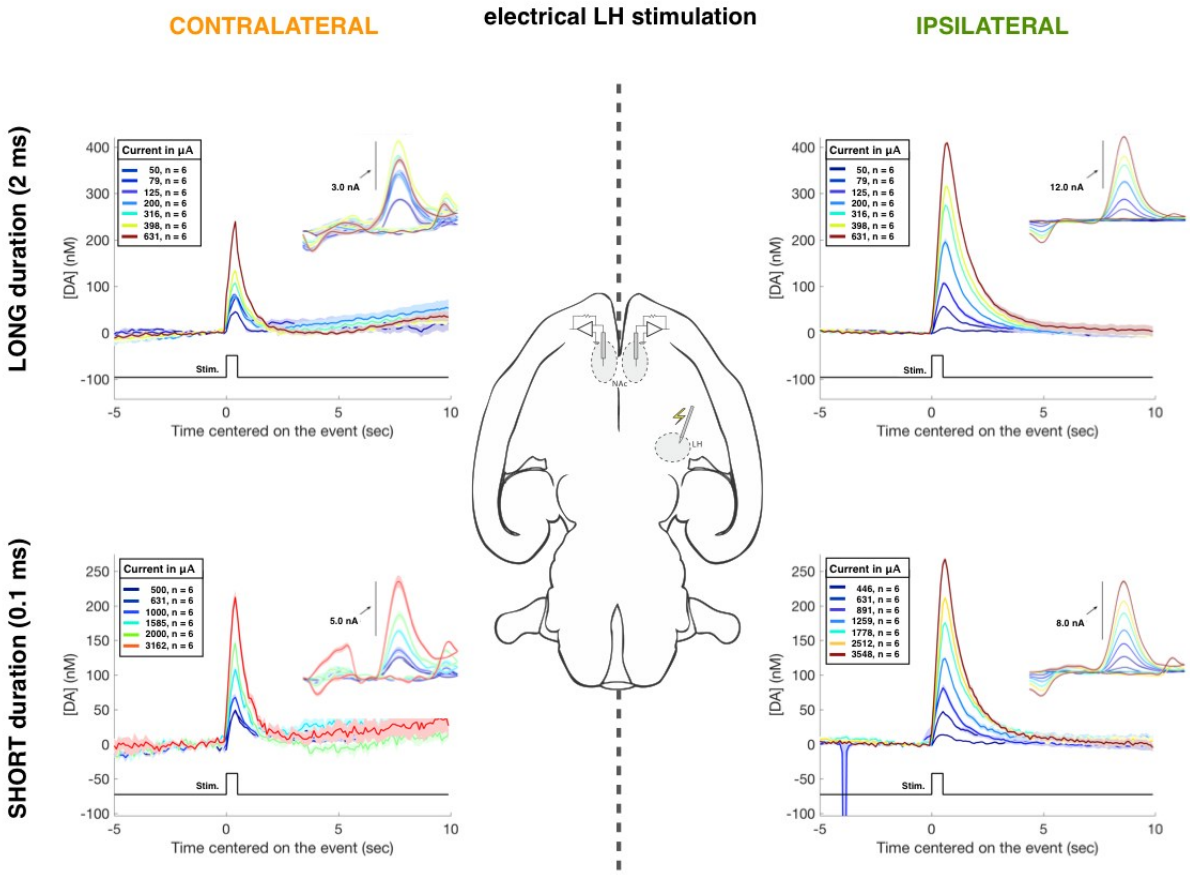


Figure 6. Bilateral responses to unilateral electrical stimulation of the LH

Data from rat IC7. Voltammetric traces and cyclic voltammographs (insets) in response to electrical trains comprised of pulses 2 ms (top row) and 0.1 ms (bottom row) in duration. DA transients recorded in the hemispheres contralateral and ipsilateral to the stimulation electrode are shown in the left and right columns, respectively. The traces are colour-coded according to the intensity of the stimulation current used to evoke the transient. The black line below the traces marks the stimulation delivery at time 0 on the x-axis. The center diagram shows the location of the MFB stimulating electrode and the two NAc carbon fibers. LH, lateral

hypothalamus; MFB, medial forebrain bundle; NAc, nucleus accumbens.

Bilateral changes in phasic DA in response to unilateral MFB stimulation at the level of the VTA

The histology revealed that stimulation electrodes were in the targeted region of the MFB at the level of the VTA and the recording carbon fibers were in comparable locations in both NAcs (Figure 7).

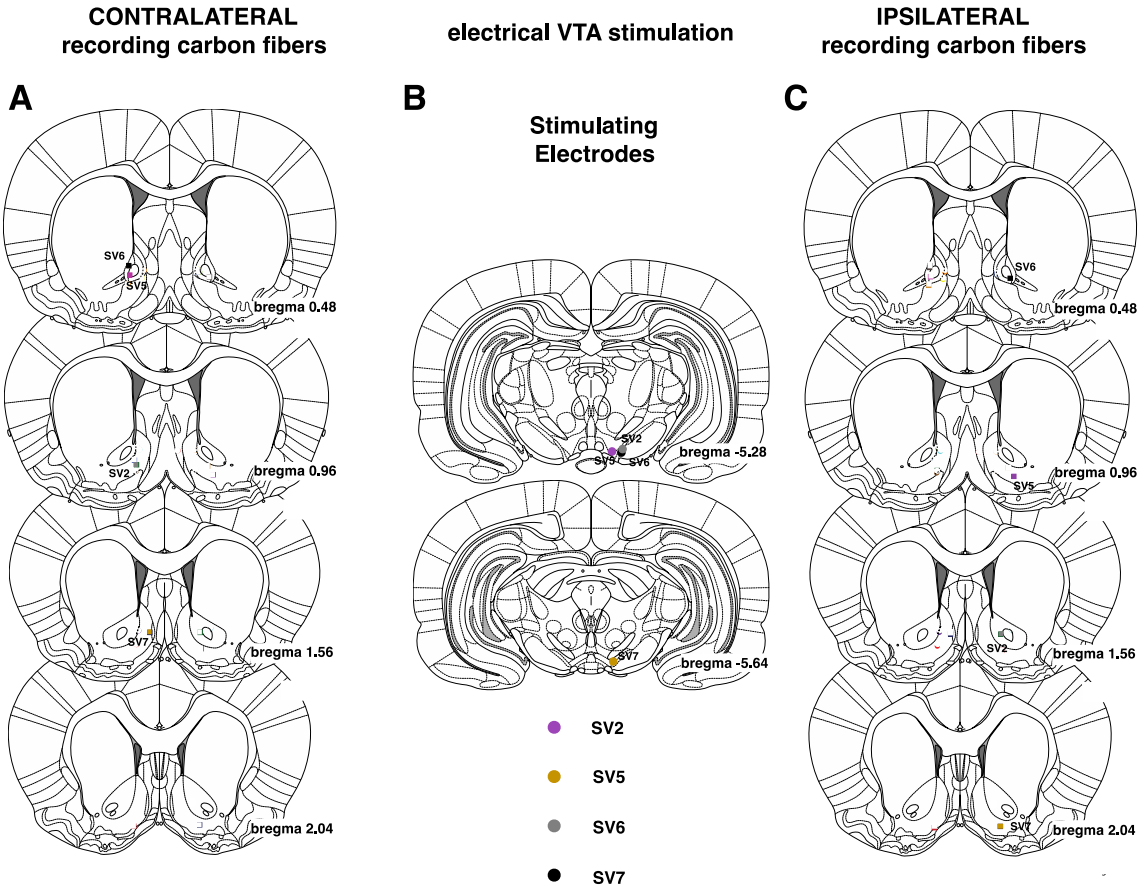


Figure 7. VTA stimulation: histological localization of stimulation and recording sites

Location of the tips of A) the contralateral and C) the ipsilateral recording carbon fibers and B) the stimulating electrodes aimed at the VTA. VTA, ventral tegmental area.

Unilateral VTA stimulation produced changes in phasic DA in both NAcs using a range of stimulation parameters (Figure 8). All DA transients were time locked to the delivery of the stimulation trains and their cyclic voltammographs were characteristic of oxidized and reduced DA.

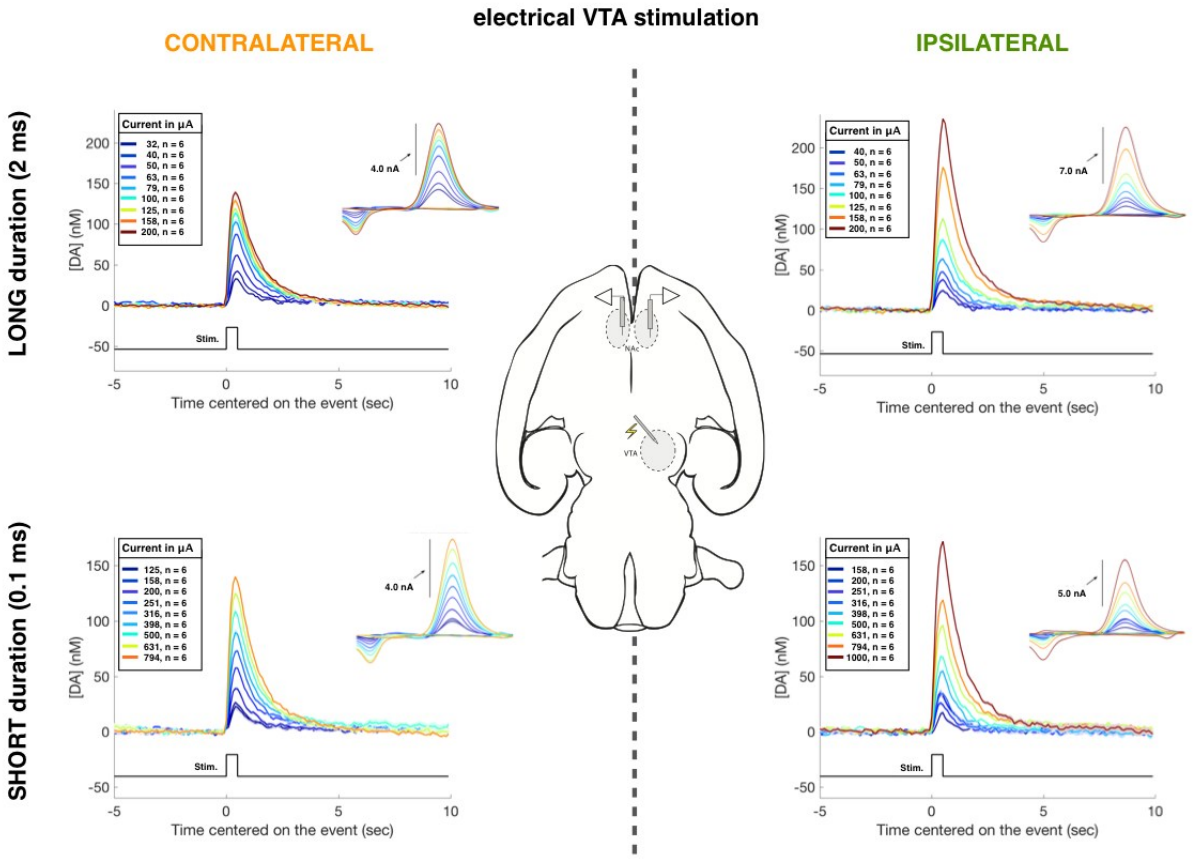


Figure 8. Bilateral responses to unilateral electrical stimulation of the VTA

Data from an example rat, SV1. Voltammetric traces and cyclic voltammographs (insets) in response to electrical trains comprised of pulses 2 ms (top row) and 0.1 ms (bottom row) in duration. DA transients recorded in the hemispheres contralateral and ipsilateral to the stimulation electrode are shown in the left and right columns, respectively. The traces are colour-coded according to the intensity of the stimulation current used to evoke the transient. The black line below the traces marks the

stimulation delivery at time 0 on the x-axis. The centre diagram shows the location of the MFB stimulating electrode and the two NAc carbon fibers. VTA, ventral tegmental area; DA, dopamine; MFB, medial forebrain bundle; NAc, nucleus accumbens.

Comparable threshold currents

As predicted by strength-duration properties of the neural substrate, threshold currents required to obtain detectable DA transients were higher for stimulations delivered at a pulse width of 0.1 ms compared with those delivered at 2 ms (Figure 9, A and B). However, distributions of threshold currents were similar in both hemispheres and when comparing LH to VTA stimulation. Computing the ratios of threshold currents (ipsilateral/contralateral) told the same story: all median ratios were close to 1 and the distributions of ratios were similar in all conditions (2 pulse widths X 2 stimulation sites) (Figure 9, C and D). This finding implies that the amplitudes of DA transients are most similar when using low stimulation currents regardless of pulse duration, hemisphere, and stimulation site and that differences are progressively more pronounced with increasing stimulation currents.

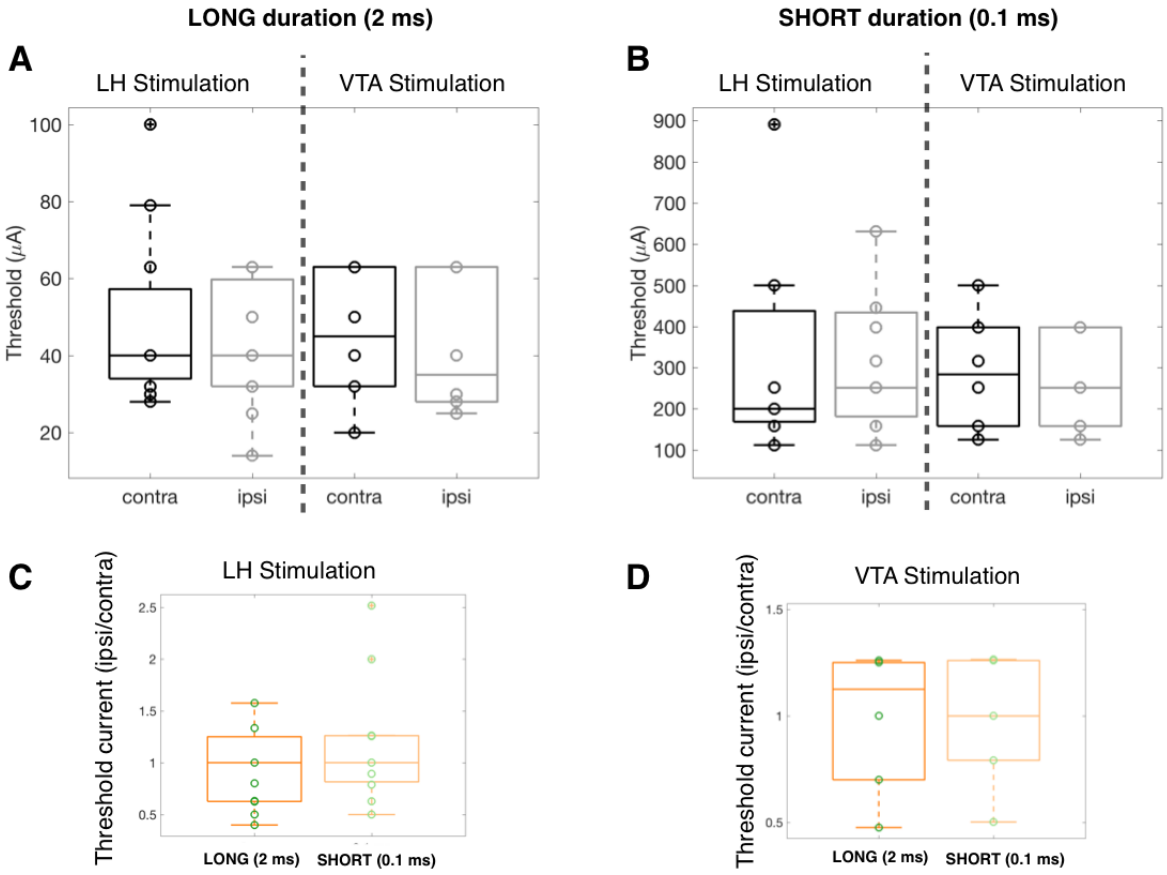


Figure 9. Similar inter-hemispheric threshold currents for LH and VTA stimulation

Threshold stimulation currents required to obtain detectable DA transients are similar between stimulation sites, LH and VTA, and between hemispheres when using the long (A) or the short (B) pulse duration. Ratios of threshold currents (ipsi/contra), computed within animals, are similar between pulse durations for LH (C) and VTA (D) stimulation. Data are presented as box plots. Individual observations are represented by circles, the inter-quartile range by the boxes, and the medians by horizontal lines within the boxes. The whiskers represent the 5th and 95th percentiles. LH, lateral hypothalamus; VTA, ventral tegmental area; DA, dopamine.

Linear approximations best describe the growth in peak DA produced by LH and VTA stimulation

Polynomial fits and AIC model selection for growth in peak DA

To determine to what extent LH and VTA stimulation increase phasic DA activity over a wide range of stimulation currents, we quantified the amplitude, and characterized the growth, of DA transients as a function of increasing stimulation currents.

First, we extracted peak DA concentration at each stimulation current (Figure 10A), and then we examined the growth of DA transients for each condition (2 hemispheres X 2 pulse widths) (Figure 10B). As explained in the introduction, there was no reason to assume the growth of DA transients would be linear. Thus, we evaluated the form of the fits empirically. We described the growth by fitting polynomials with 0 to10 parameters (Figure 10B) and tested for the fit offering the best combination of accuracy and parsimony using the AIC (Figure 10C).

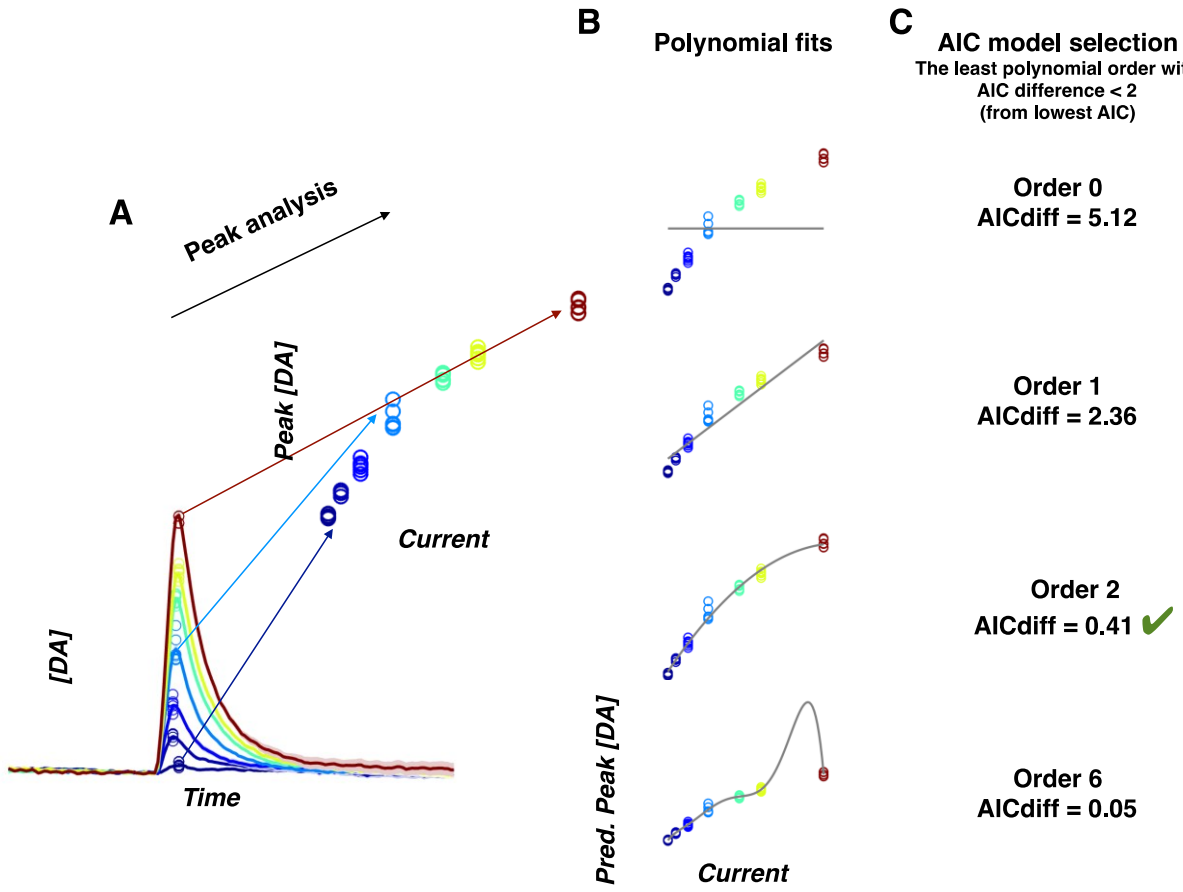


Figure 10. Growth of DA transients as a function of current: polynomial approximations and AIC model selection

A) Estimation of the peak DA concentration for one dataset. The maximal background-subtracted oxidation current was extracted for each voltammetric trace for all 6 repetitions at all stimulation currents or optical powers. B) The growth in the peak DA concentration was assessed by testing polynomial fits of degrees ranging from 0 to 10 for each current or optical-power sweep. For 4 of these fits, the resulting predicted line is plotted in grey. C) We computed AIC differences and selected the model (green checkmark) associated with an AIC difference of 2 or smaller and with the least number of parameters (for more details see the method section and Burnham & Anderson, 2002, p.71). AIC, akaike information criterion; DA, dopamine.

In principle, the fit with lowest AIC should achieve the best trade-off between accuracy and parsimony. However, more than one model might have an AIC very close to the lowest AIC. As recommended by Burnham and Anderson (2002), we used the AIC difference, Akaike weights, and evidence ratios to select the polynomial that was deemed as likely to be the best approximation with the smallest number of parameters (see methods section for more details).

Linear growth in peak DA during MFB stimulation

Growth in the amplitude of DA transients as a function of LH and VTA stimulation current was mostly linear. The percentages of linear best fits were similar when grouping by hemisphere (LH: Figure 11A and VTA: Figure 11C) or by pulse duration (LH: Figure 11B and VTA: Figure 11D). Of the data sets, 68% to 83% were best described by first-order polynomials, 17% to 27% by second-order polynomials, and a few by zero-order or third-order, S-shaped polynomials. All linear approximations had comparable goodness of fit with Pearson coefficients ranging from .83 to .99 (Figure 12).

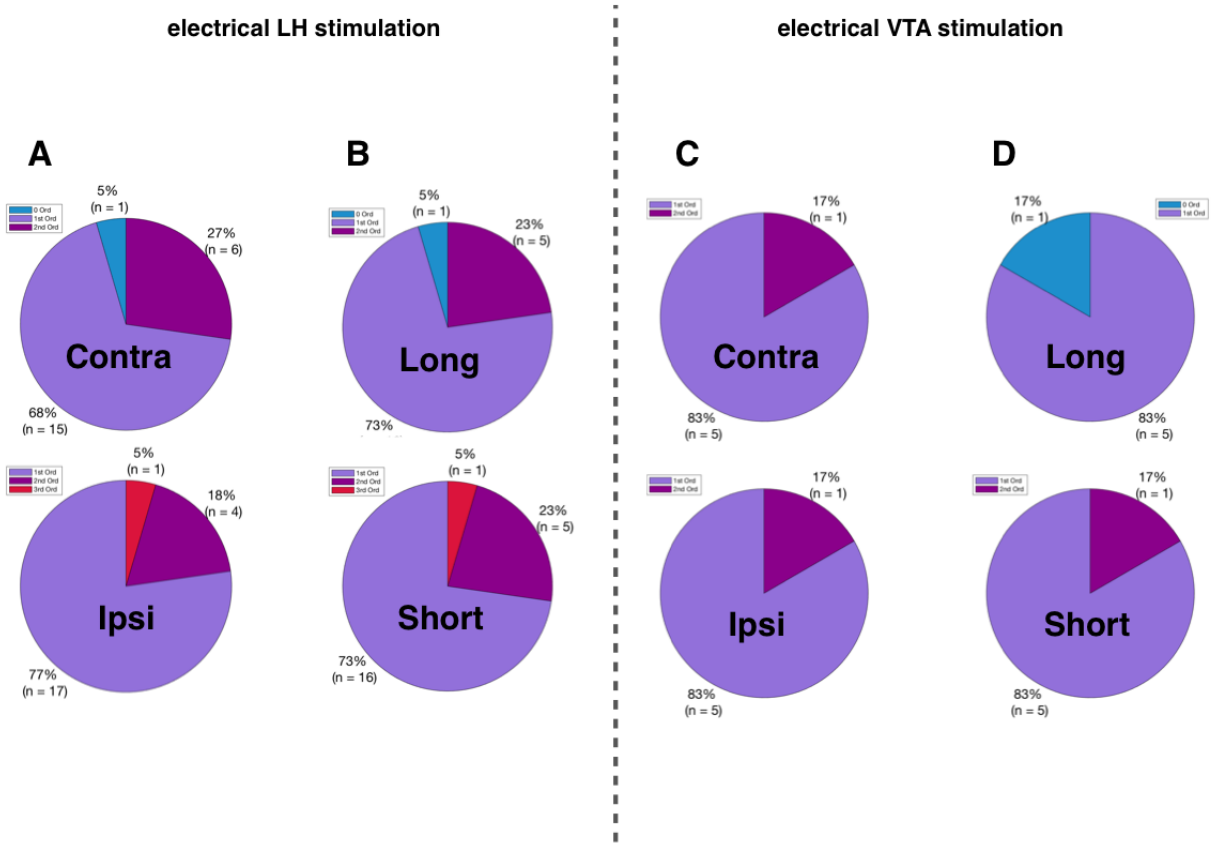


Figure 11. Distributions of the best polynomial fits to datasets obtained with LH and VTA stimulation

Pie charts showing the distribution of polynomial fits deemed as the best approximations. Models associated with an AIC difference of 2 or smaller and with the least number of parameters were selected (for more details see the method section and Burnham & Anderson, 2002, p.71). The distributions are grouped by hemispheres (A and C) and by pulse duration (B and D). Cases in which the best fit was provided by zero, first, second and third order polynomials are shown in red, light purple, dark purple, and blue, respectively. Roughly 75% of the best approximations were first order. Sample sizes varied across conditions. The number of observations is provided below each percentage in the pie charts. LH, lateral hypothalamus; VTA, ventral tegmental area; AIC, akaike information criterion.

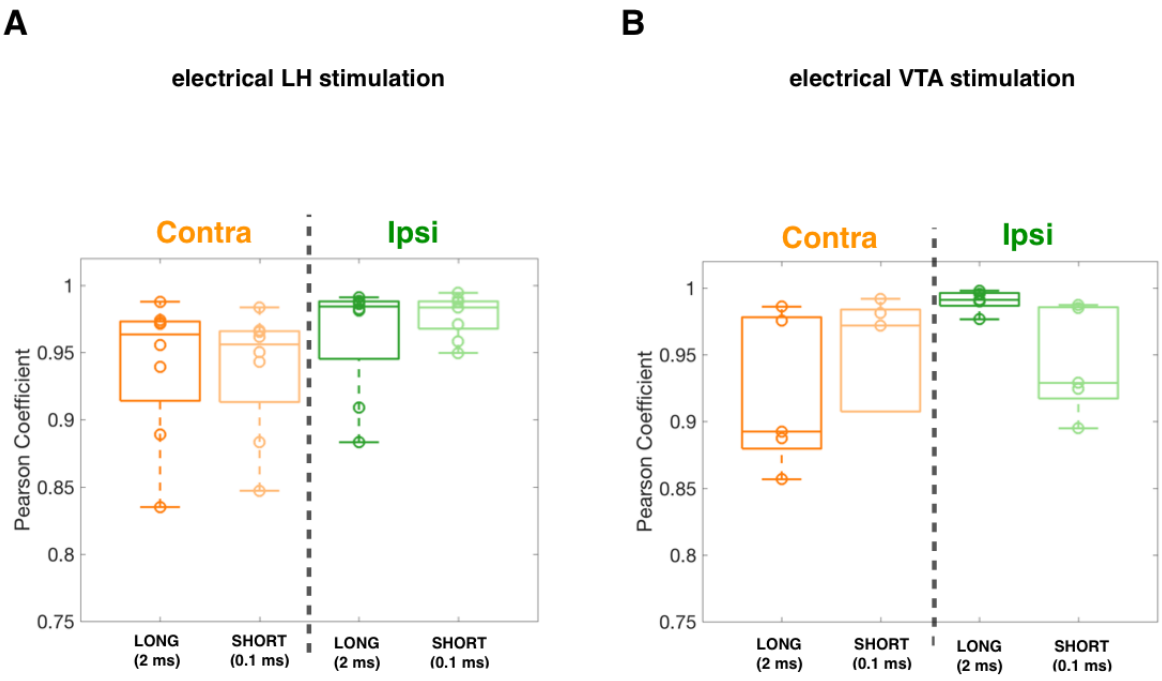


Figure 12. Goodness of linear fits for datasets obtained with LH and VTA stimulation

Pearson coefficients for datasets obtained with LH and VTA stimulation are shown in A and B, respectively. The box plots are grouped by pulse duration and hemisphere. Contralateral coefficients are in orange and ipsilateral coefficients are in green. Data obtained with the long pulse duration (2 ms) are in dark colours, whereas data obtained with the short pulse duration (0.1 ms) are in pale colours. Individual observations are represented by circles, the inter-quartile range by the boxes, and the medians by horizontal lines within the boxes. The whiskers represent the 5th and 95th percentiles. LH, lateral hypothalamus; VTA, ventral tegmental area.

The rest of the analyses only included data sets associated with linear best fits. The degree of similarity between conditions (2 hemispheres X 2 pulse widths) was assessed examining the slopes of these fits and threshold currents required to produce detectable DA transients.

Inter-hemispheric similarities in the amplitude of DA transients evoked by VTA compared to LH stimulation

Steeper growth in peak DA for electrical trains of 2 ms compared to 0.1 ms in pulse duration

Median slopes for linear fits were greater for the long pulse duration compared with the short pulse duration (Figure 13), as expected from the strength-duration curves presented in the introduction. This was true for DA transients obtained in the contralateral and ipsilateral NAc for both LH and VTA stimulations.

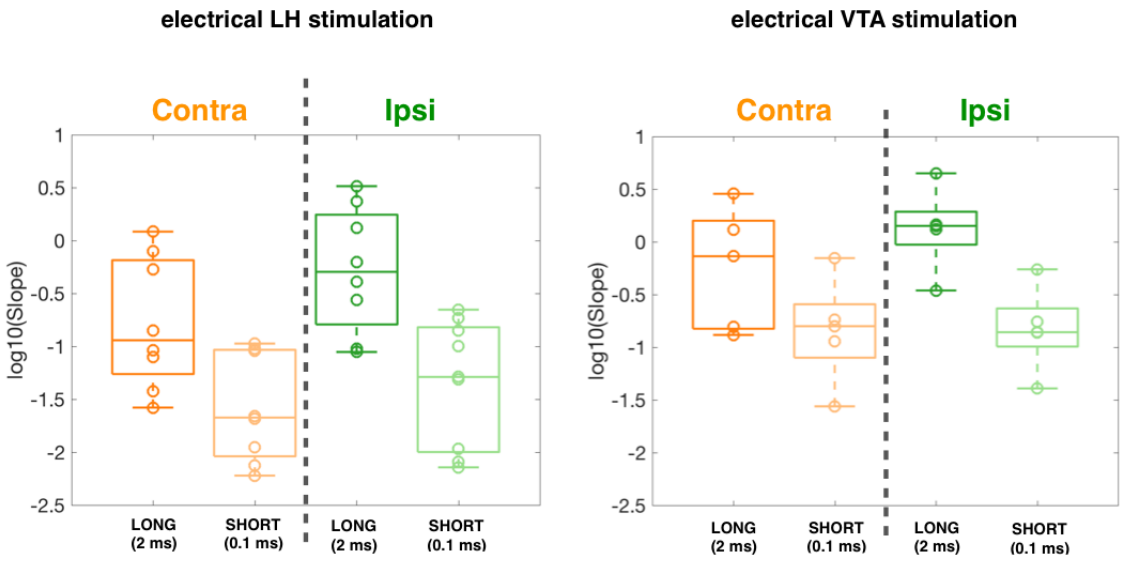


Figure 13. Growth of the peak DA concentration as a function of electrical stimulation current is steeper when 2-ms pulses, rather than 0.1-ms pulses, are employed

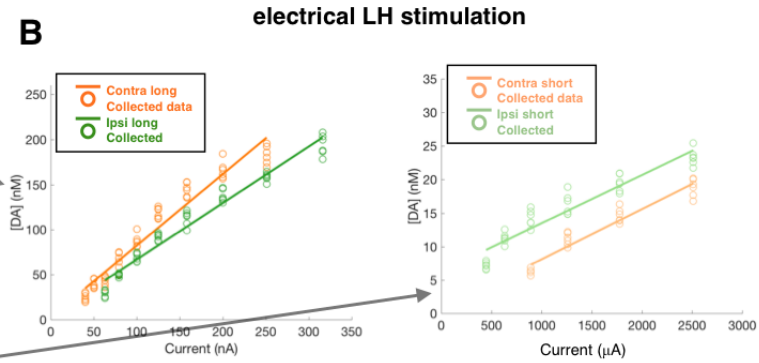
The slopes of the linear fits describing the growth in the amplitude of DA transients as a function of stimulation currents for LH and VTA stimulation are shown in A and B, respectively. Slopes are expressed in log unit to facilitate visual comparison

between pulse durations. The box plots are grouped by pulse duration and by hemisphere. Contralateral coefficients are in orange and ipsilateral coefficients are in green. Data obtained with the long pulse duration (2 ms) are in dark colours, whereas data obtained with the short pulse duration (0.1 ms) are in pale colours. Individual observations are represented by circles, the inter-quartile range by the boxes, and the medians by horizontal lines within the boxes. The whiskers represent the 5th and 95th percentiles. DA, dopamine; LH, lateral hypothalamus; VTA, ventral tegmental area.

If animals had linear fits in both hemispheres for a given pulse duration the slopes were compared. Figure 14 lists all the cases and examples of raw data (A and C) and shows examples of selected cases with superimposed linear fits (B and D). There was no systematic effect of hemisphere on the slopes of DA growth; all combinations of slower, faster, and similar rates of DA growth were found when contrasting contralateral to ipsilateral DA transients.

A

LH (values = slopes of linear fits)		
long		
contra	ipsi	contra DA grows slower?
0.09	0.09	SAME
0.14	0.27	SLOWER
0.79	0.63	FASTER
0.54	2.34	SLOWER
1.21	3.26	SLOWER
short		
contra	ipsi	contra DA grow slower?
0.01	0.01	SAME
0.02	0.05	SLOWER
0.01	0.01	SAME
0.01	0.01	SAME
0.02	0.10	SLOWER
0.11	0.19	SLOWER
0.09	0.14	SLOWER



C

VTA (values = slopes of linear fits)		
long		
contra	ipsi	contra DA grows slower?
0.73	4.47	SLOWER
2.86	1.42	FASTER
0.16	0.35	SLOWER
1.31	1.46	SLOWER
short		
contra	ipsi	contra DA grow slower?
0.18	0.18	SAME
0.11	0.55	SLOWER
0.70	0.14	FASTER
0.03	0.04	SLOWER
0.16	0.14	FASTER

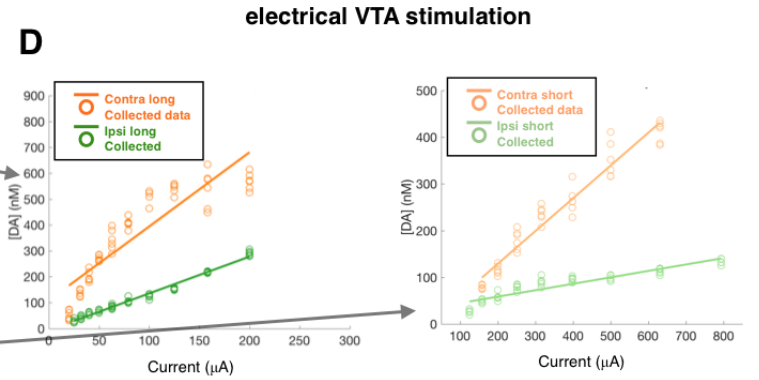


Figure 14. The rate at which the peak DA concentration grows as a function of stimulation current varies across hemispheres

The growth rate of the peak DA concentration as a function of stimulation current was compared across hemispheres in all cases in which linear functions provided the best fit to the data obtained in both hemispheres (LH: 12 out of 22 cases; VTA: 9 out of 12 cases). The comparisons for data obtained with LH and VTA stimulation are shown in the tables in A and C, respectively. The last column of each table classifies the contralateral growth rate as slower, faster, or similar to the ipsilateral growth rate. Raw data with superimposed linear fits for selected cases (arrows) are shown in B and D. The orange and green colours designate data acquired in the contralateral or ipsilateral hemisphere, respectively, and the dark or pale colours designate the long

or short pulse durations, respectively. DA, dopamine; VTA, ventral tegmental area; LH, lateral hypothalamus.

Careful control of virus infection and restricted light delivery prevent or markedly reduce the detection of contralateral DA transients

The histology revealed that optical implants were just above the VTA, at the targeted position, and the recording carbon fibers were in comparable locations in both NAcs for all groups receiving optical stimulation (Figures 15, 16, and 17). The expression of ChR2 in midbrain DA cell bodies was bilateral when the virus was injected only 1 mm away from the midline (Figure 15 and 17). In contrast, the expression of ChR2 was restricted to one hemisphere when the distance from the midline was increased to 1.7 mm (Figure 16). The expression is weaker in Figure 17 although the amplitude of the stimulation-evoked DA transients in the ipsilateral hemisphere was similar in all optogenetic groups. The weaker expression might be due to the longer storage at 4 °C for these brains.

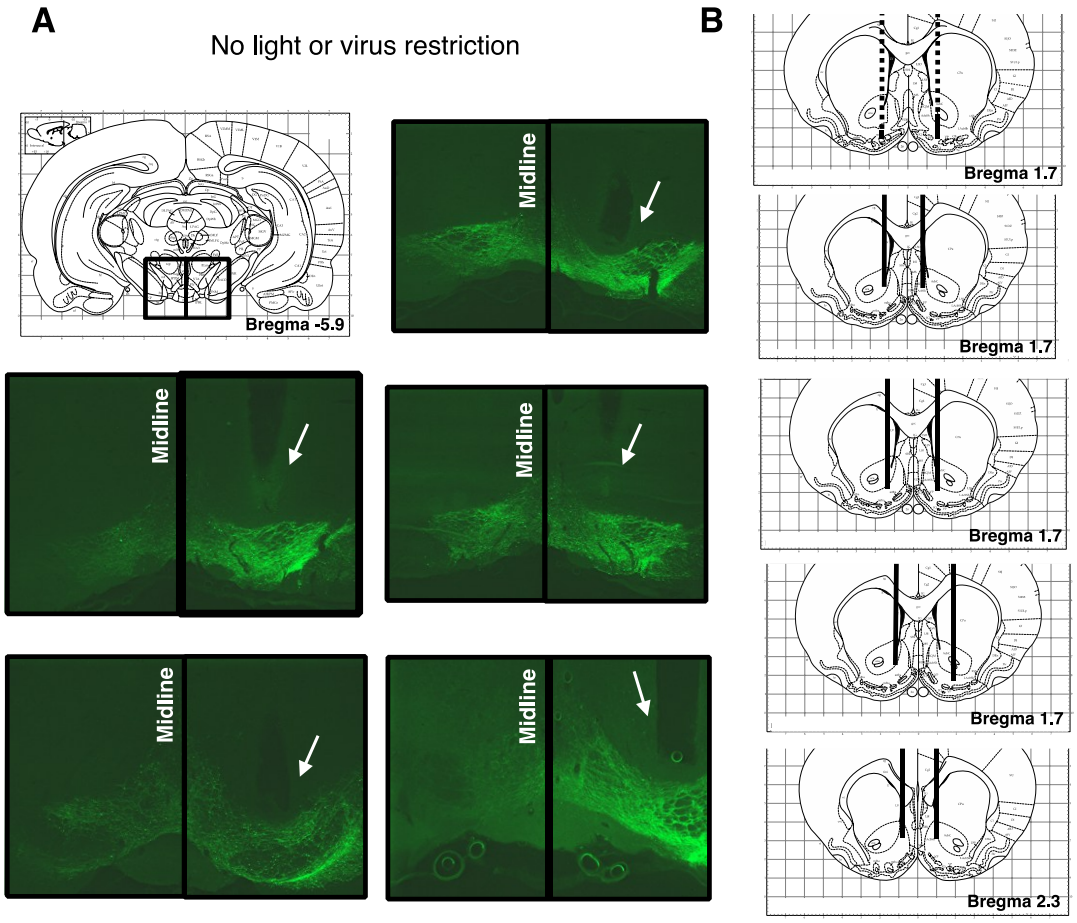


Figure 15. Unrestricted optical stimulation on VTA DA neurons: histological localization of optical stimulation and FSCV recording sites

Location of the tips of A) optical implants and B) the carbon-fiber FSCV electrodes. The track left by the optical implants is marked with a white arrow and ChR2-EYFP expression is shown in green. VTA, ventral tegmental area; DA, dopamine; FSCV, fast scan cyclic voltammetry; ChR2, channelrhodopsin-2; EYFP, yellow fluorescent protein.

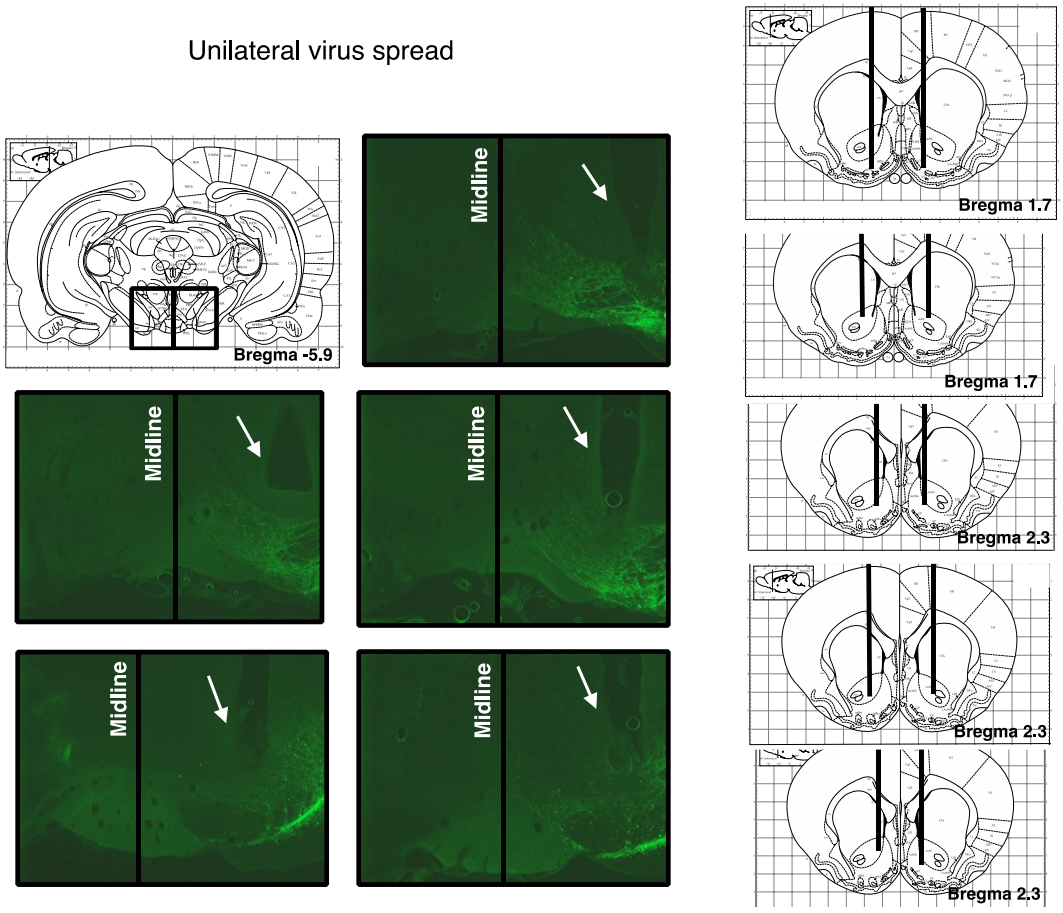


Figure 16. Laterally displaced viral injection: histological localization of optical stimulation and FSCV recording sites

Location of the tips of A) optical implants and B) the FSCV electrodes. The track left by the optical implants is marked with a white arrow and ChR2-EYFP expression is shown in green. FSCV, fast scan cyclic voltammetry; ChR2, channelrhodopsin-2; EYFP, yellow fluorescent protein.

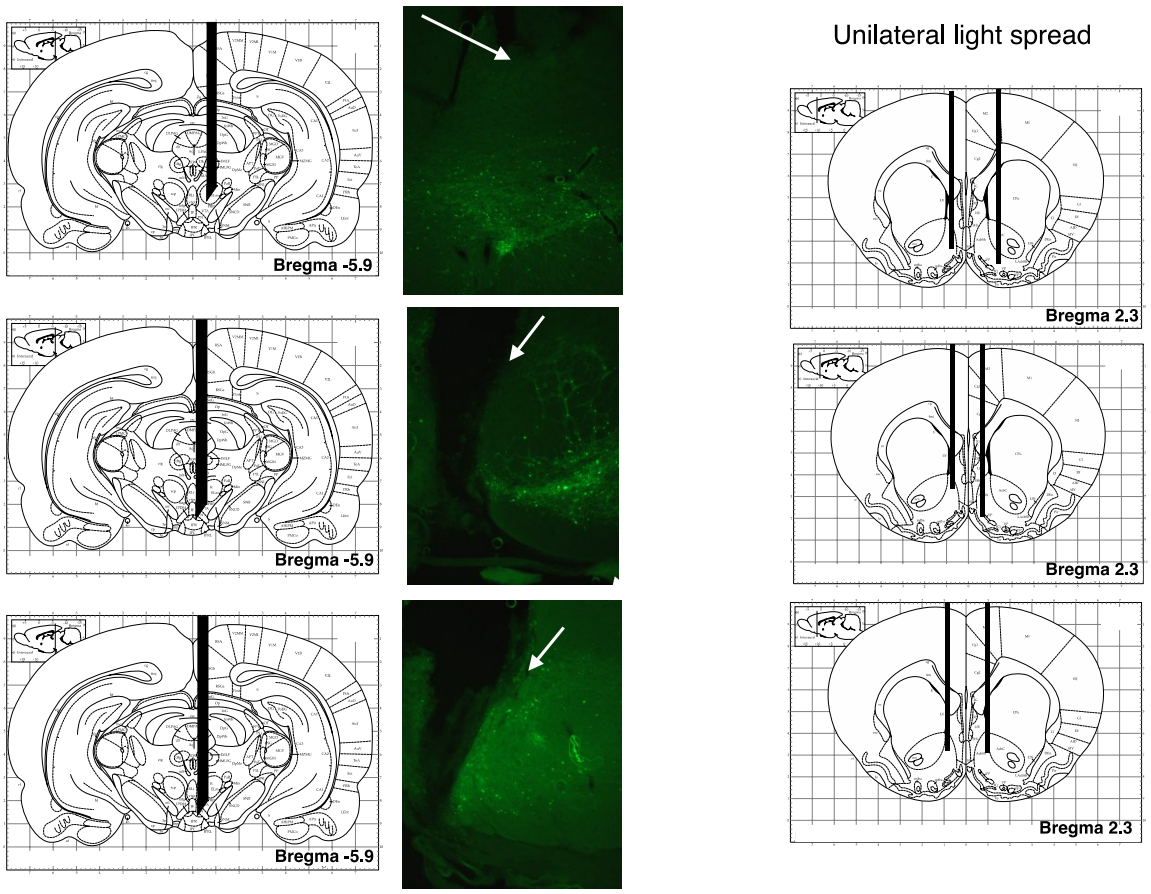


Figure 17. Restricted light distribution: histological localization of optical stimulation and FSCV recording sites

Location of the tips of A) optical implants and B) the FSCV electrodes. The track left by the optical implants is marked with a white arrow and ChR2-EYFP expression is shown in green. Fluorescence of the reported protein is weaker than in Figures 15 and 16 although the amplitude of the stimulation-evoked DA transients in the ipsilateral hemisphere was similar. The weaker fluorescence might be due to longer storage of these brains at 4°C. FSCV, fast scan cyclic voltammetry; ChR2, channelrhodopsin-2; EYFP, yellow fluorescent protein; DA, dopamine.

Bilateral DA transients were recorded in the NAc in response to unilateral optical stimulation when the light was not restricted to one hemisphere and the expression of ChR2 spread to both VTAs (Figure 18).

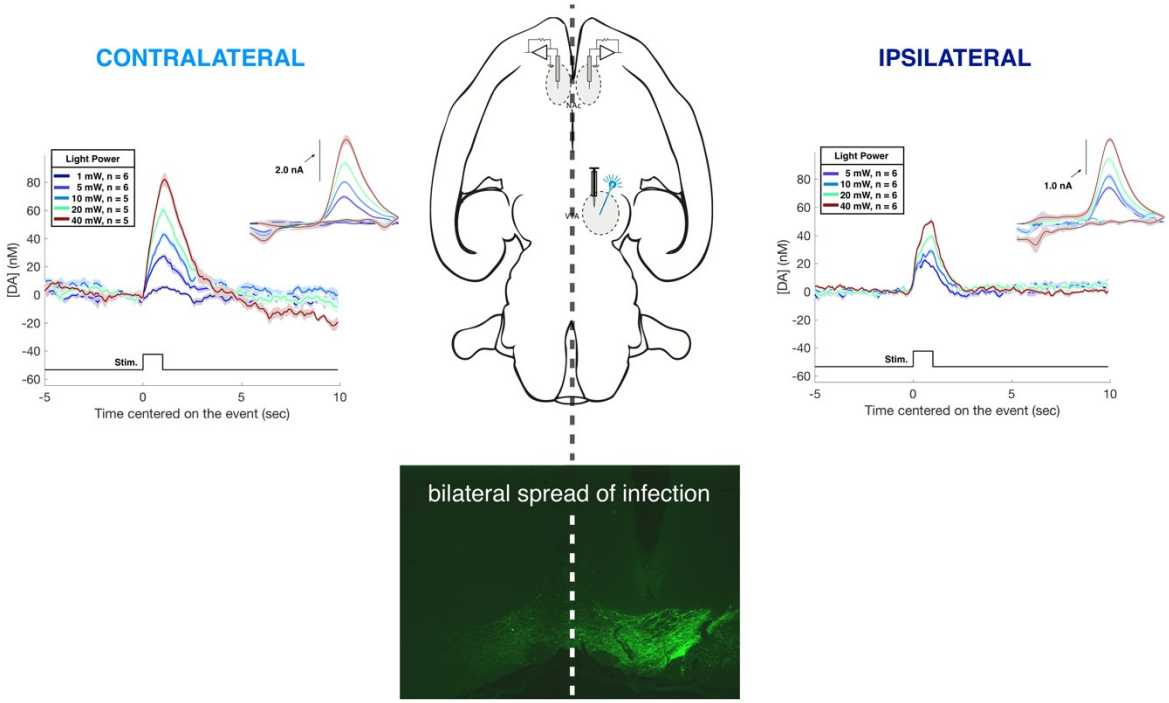


Figure 18. Bilateral response to unrestricted optical stimulation of DA neurons in the VTA

Data from rat nor24. Voltammetric traces and cyclic voltammographs (insets) are shown for DA transients in the contralateral (left panel) and ipsilateral (right panel) hemispheres. The traces are colour-coded according to the optical power used to evoke the transient. The black line below the traces marks the stimulation delivery at time 0 on the x-axis. The centre diagram shows ChR2-EYFP expression. The dotted white line indicates the midline. The damage produced by the optical implant can be seen on the right of the picture. DA, dopamine; VTA, ventral tegmental area; ChR2, channelrhodopsin-2; EYFP, yellow fluorescent protein.

We observed a systematic growth in DA transients as a function of optical power in both hemispheres in all five animals. Raw data and predicted lines in two sample animals are presented in Figure 19.

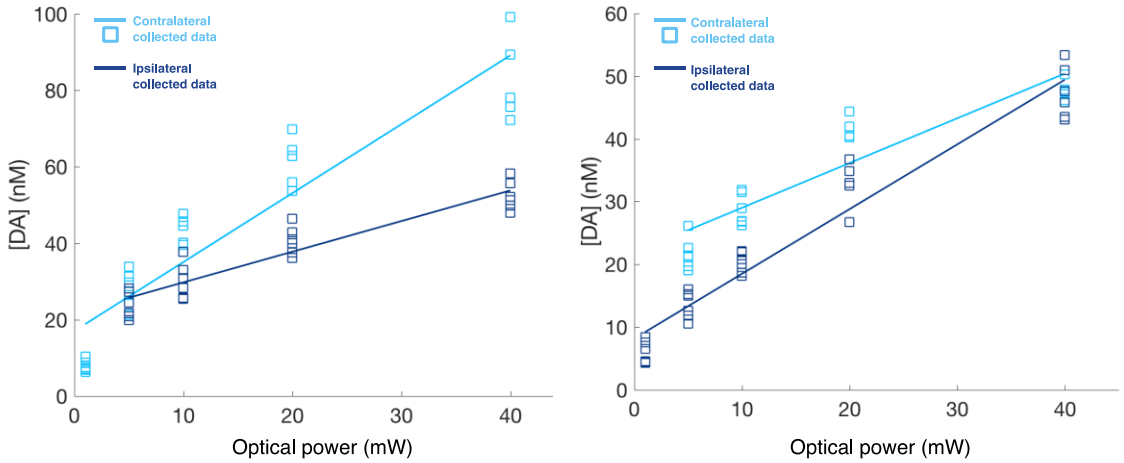


Figure 19. Growth of DA transients as a function of optical power

Raw data with superimposed linear fits for animals nor24 and nor25. Pale and dark blue refer to data acquired in the contralateral and the ipsilateral hemisphere, respectively. DA, dopamine.

Restricting the virus infusion to one hemisphere, however, prevented the detection of contralateral DA transients in all animals (Figure 20A). This result held despite attempts to obtain voltammetric readings at several depths and to place fresh carbon fibers at several locations along the medial-lateral and anterior-posterior axes within the NAc.

Restricting the light distribution, rather than restricting the spread of the virus, prevented the detection of contralateral DA transients in one animal and greatly reduced the amplitude of contralateral DA transients in two other animals (Figure 20B).

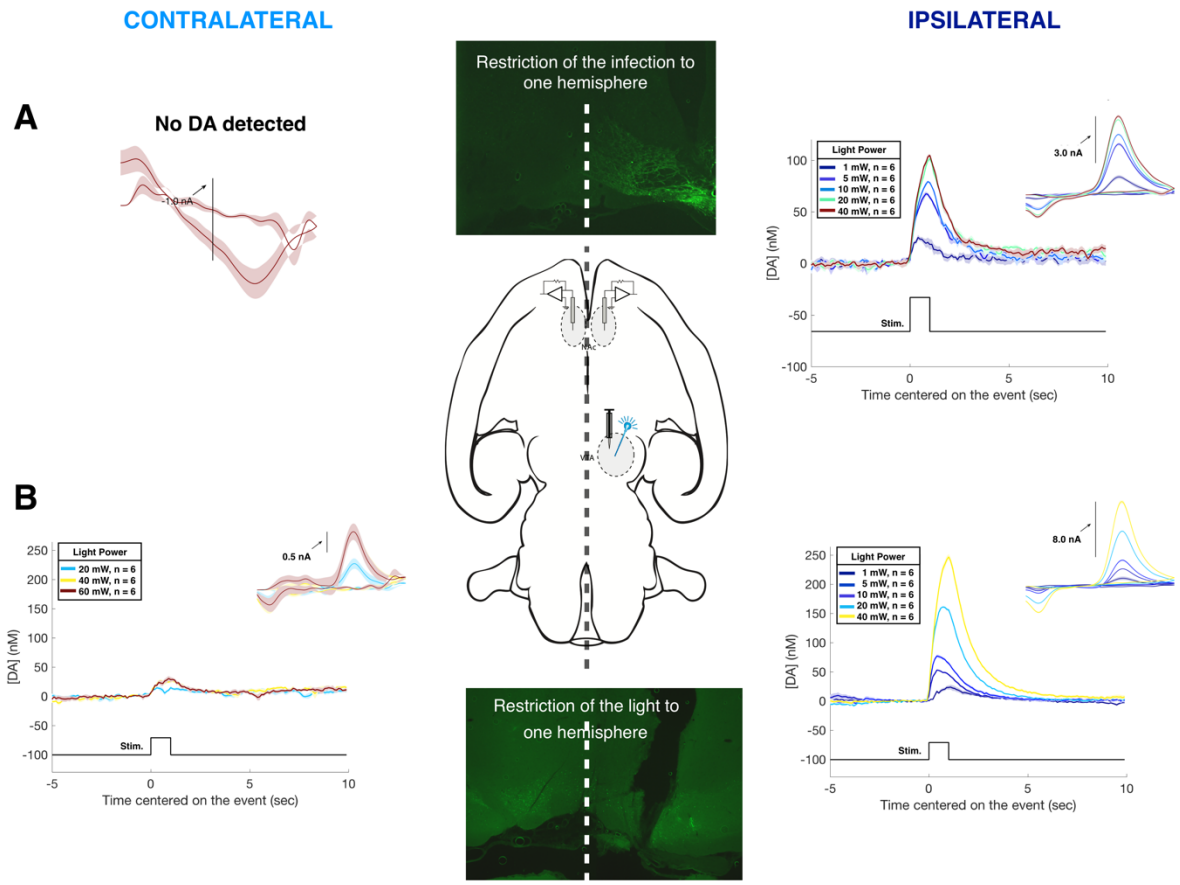


Figure 20. Undetectable or reduced contralateral DA transients following restriction of the spread of virus or light in the VTA

Data from rats Chr39 (laterally displaced viral injection) and Bvta4 (restricted light distribution). Panel A shows FSCV data obtained in the hemisphere contralateral to the optical stimulation sites following lateral displacement of the viral injection site. Panel B shows FSCV data obtained in response to optical stimulation through optical fibers that had been threaded through beveled cannulae so to restrict the spread of light to the contralateral hemisphere. Voltammetric traces and cyclic voltammographs (insets) are shown for DA transients in the contralateral (left column) and ipsilateral (right column) hemispheres. The traces are colour-coded according to the optical power used to evoke the transient. The black line below the traces marks the stimulation delivery at time 0 on the x-axis. The centre diagram shows ChR2-EYFP expression in green. The dotted white line indicates the midline.

The damage produced by the optical implant can be seen on the right of the picture. The expression is weaker in the lower diagram although the amplitude of the stimulation-evoked DA transients in the ipsilateral hemisphere was similar in all optogenetic groups. The weaker fluorescence might be due to significantly longer storage of these brains at 4°C. DA, dopamine; VTA, ventral tegmental area; FSCV, fast scan cyclic voltammetry; ChR2, channelrhodopsin-2; EYFP, yellow fluorescent protein.

The ratio of the amplitudes of DA transients recorded in the ipsilateral and contralateral hemispheres was computed at light powers of 20 mW and 40 mW. These ratios were close to 1 for all five animals in the group in which the light could spread symmetrically from the tip of the optical fiber (Figure 21B), whereas ratios were between 8 and 11 for animals in which the light was introduced by means of a bevelled cannula and thus confined largely to the ipsilateral VTA (Figure 21A). Overall, the contralateral DA transients observed in control animals were small and did not grow with increasing optical power.

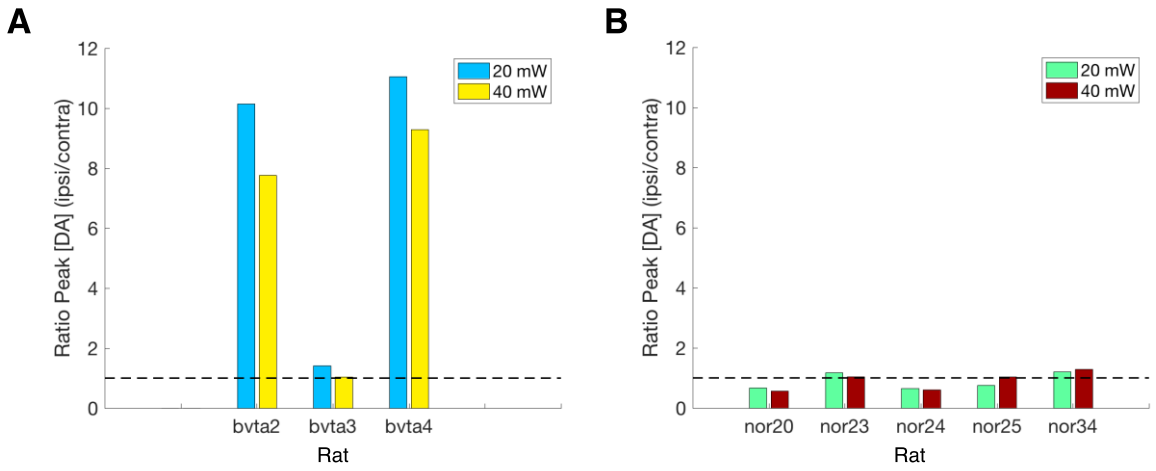


Figure 21. Ratios of the amplitudes of ipsilateral to contralateral transients.

Histograms show the ratios of ipsilateral to contralateral peak DA concentration in response to optical stimulation delivered at 20 and 40 mW for the group with restricted light spread (A) and the group with no restriction of either viral or light spread (B). The colours used correspond to the values assigned to the 20 and 40 mW powers in Figures 18 and 20. DA, dopamine.

Discussion

DA neurotransmission has been heavily implicated in reward seeking based on psychophysical and neurochemical evidence collected using eICSS and MFB stimulation. However, methodological discrepancies between the two types of experiments make it difficult to integrate the findings. Short pulse durations, as used in most psychophysical experiments, predominantly recruit highly excitable neurons whereas longer pulse durations, as used in most neurochemical experiments, recruit a more heterogeneous neural population. Differences in the composition of the recruited populations might lead to differences in the DA transients observed in the terminal fields of midbrain DA neurons. The fact that the electrode is placed most often in the LH for the psychophysical experiments and mostly in the VTA for neurochemical experiments adds variability in the composition of the stimulated substrates. We deemed it necessary to conduct a thorough comparison, using a wide range of stimulation parameters, to determine the degree of similarity in the effects of LH and VTA stimulation on phasic DA activity.

Similarities between DA transients evoked by LH and VTA stimulation

It has been reported that VTA stimulation evokes bilateral DA transients as measured by means of FSCV (Fox et al., 2016). In our experiment we demonstrated that LH stimulation also produces DA transients in both hemispheres. As argued in the introduction, monosynaptic DA activation produced by MFB stimulation is likely to be restricted to the ipsilateral hemisphere (Cossette & Shizgal, SfN poster, 2015; Watabe-Uchida et al., 2012; Menegas et al., 2015) whereas anatomical and functional evidence point to putative trans-synaptic routes of DA activation that could drive bilateral DA transients through the activation of relay brain regions (Cornwall, Cooper & Phillipson, 1990; Semba & Fibiger, 1992; Steininger, Rye & Wainer, 1992; Otake & Nakamura, 1998; Kirouak, 2015; Oakman et al., 1995; Dautan et al., 2014; Forster & Blaha, 2000; Forster & Blaha, 2003; Blaha & Phillips, 1996; Yeomans, Forster & Blaha, 2001; Parsons, Li & Kirouac, 2007). Therefore, polysynaptic activation required to drive contralateral DA transients is not a phenomenon restricted to VTA stimulation but also extends to LH stimulation. At least part of the circuitry involved in producing DA transients relies on polysynaptic connections between MFB and DA neurons for both types of electrical stimulation.

The threshold current for firing a neuron depends on the duration of a stimulation pulse as described by the hyperbolic fits of strength-duration curves (Weiss, 1901; Matthews & Gallistel, 1975; Matthews, 1978). We confirm here, for the first time, that the threshold current required to produce detectable DA transients are higher when short, rather than long, pulse durations are employed. We then demonstrated that threshold values were comparable between hemispheres and stimulation types: LH and VTA stimulation require comparable strengths of stimulation to evoke DA transients detectable by FSCV independently of which hemisphere is being monitored. This information is particularly useful when trying to compare data collected in eICSS experiments using LH stimulation and data collected in FSCV experiments using VTA stimulation. If a stimulation current at a fixed pulse duration is deemed sufficient to produce DA transients when the electrode is aimed at the VTA, it should suffice for LH stimulation. The same logic applies to contralateral DA activation, which is seen in response to stimulation of both the LH and VTA.

Linear growth in the amplitude of DA transients

Another result common to most of the data sets reported here was the linear growth in DA transients as a function of current. This pattern was seen regardless of pulse duration, stimulation site, or the hemisphere in which the FSCV recording was obtained. Monotonic relationships between stimulation strength and the amplitude of DA transients have been reported before (Gratton, Hoffer & Gerhardt, 1988; Blaha & Phillips, 1990; Gonon, 1988, Wightman, Amatore, Engstrom et al., 1988; Kawagoe & Wightman, 1994; Garris & Rebec, 2002; Beyene, Carelli & Wightman, 2010; Cheer, Aragona, Heien et al., 2007). However, this is the first report of a rigorous analysis of the form of the function that maps the stimulation current into the peak amplitude of the stimulation-evoked DA transient. The wide range of stimulation parameters used provided the large data sets required for this purpose. The goodness of fit of both low- and high-ordered polynomials was evaluated. Linear fits described DA growth as a function of current with the best combination of accuracy and parsimony.

To understand how such linearity could emerge let us examine how electrical current is distributed around the tip of an electrode, how the distance between a neuron and the source of current influences the activation threshold, and how this threshold depends on the intrinsic

properties of the neuron. The small, needle-like electrodes used in our experiment can be treated as if they act as a point source of current (Yeomans, Mercouris & Ellard, 1985). If so, and if the impedance of current flow is isotropic, the spread of current will be spherically symmetrical around that point. The current density declines as the radius, and hence the surface area of that sphere, increases (i.e. current flow per unit area declines).

The threshold current density for firing an axon differs between neuronal types. Let us consider the example of a threshold current density for a large, easily excitable axon in the MFB. The threshold is defined in terms of the spatial density of the stimulation current impinging upon it. In the case of spherically symmetrical current flow, the spatial density at a distance, r, from the point source is the ratio of the current (I) divided by the surface area of a sphere of radius r:

$$j_{t\_b} = I(d) / 4*\pi*r_b2 \qquad (3)$$

where j_{t_b} = the threshold current density denoted by t for a large axon denoted by b
r_b is the distance from the point source at which the stimulation current is just sufficient to fire a large axon b
and d is the pulse duration

Note that the threshold current density declines as the inverse square of the distance from the point source. Equation 3 can be rearranged to express stimulation current as a function of the distance between the large axon and the point source:

$$I = j_{t\_b} * 4*\pi * r_b^2$$

Let

$$K_{IS} = 4*\pi*j_t$$

where K_{IS} = the current-distance constant that represents the strength of the current per unit area needed to fire a large axon

$$I = K_{IS} * r_b^2 \qquad (4)$$

Assume that the point source of current lies within a bundle of identical, equally spaced axons. The number of axons recruited will be proportional to the cross-sectional area of the sphere with radius r_b centred on the point source:

$N_{axons} = \pi * r_b^2 *$ the number of axons per unit area

Let K_{pack} = the number of axons per unit area times π (the "packing" constant)
and let us rewrite the relationship as:

$$N_{axons} = K_{pack} * r_b^2 \qquad (5)$$

Thus, the number of stimulated axons increases as a function of the square of the radius.
We can now combine equations 4 & 5 to calculate how the number of stimulated axons grows as a function of the stimulation current, which yields a simple scalar equation:

$$N_{axons} = (K_{pack} / K_{IS}) * I$$

This argument was first advanced by Robert Hawkins (Gallistel, 1988).
Note that the ratio, K_{pack} (number of axons per unit area) / K_{IS} (current per unit area), yields a constant that represents the number of axons recruited per unit of current. If the peak amplitude of an evoked DA transient is proportional to the number of stimulated axons, then the peak amplitude will grow as a scalar function of current under the assumptions specified above and the slope will be determined by the ratio of the packing and current-distance constants.

Steeper growth in DA transients with a longer pulse duration

The slope of DA growth is determined by the ratio of the packing and current-distance constants. The current-distance constant also depends on the threshold current density ($K_{IS} = 4*\pi*j_t$), which itself is a function of pulse duration. The required current to fire a large axon declines as a function of pulse duration, as described by strength-duration curves, is defined as:

$$I(d) = I_{r\_b}*(1+(c_b/d))$$

where I_{r_b} is the required current for an infinitely long pulse duration (the rheobase), c_b is the chronaxie, and d is the pulse duration

and recall

$$j_{t\_b} = I(d) / 4*\pi*r_b2$$

At a longer pulse duration, the required current is smaller, resulting in a lower threshold current density. Consequently, K_{IS} is also smaller, and the ratio of the packing and current-distance constants increases, thus predicting a steeper slope of DA growth at the higher pulse duration. This is exactly what we found, regardless of stimulation site or the hemisphere in which the FSCV recording was obtained.

Linear growth in the amplitude of DA transients in homogeneous and heterogeneous substrates

To this point, a homogeneous population of neurons has been envisaged. What if we were to add a second, less excitable population that also drives DA transients? The current-distance constant (K_{IS}) will be larger, and hence, the radius of excitation for a given current will be smaller (recall that $I = K_{IS} * r^2$). However, as long as the density of these neurons is constant within the sphere defined by this radius, there will again be a scalar relationship between the number of neurons stimulated and the current defined by K_{pack} / K_{IS}. For two neuronal types with identical K_{pack}, a larger K_{IS} results in a smaller number of axons recruited per unit of current. If these axons also sum their outputs in contributing to the peak amplitude of the DA transient, then the addition of this second population merely changes the scale factor. With the help of these additional neurons, peak amplitude grows more steeply as a function of current, but the relationship remains scalar.

Relaxing the simplifying assumptions underlying the derivation of the relationship between the peak amplitude of the DA transient and the current can cause this relationship to deviate from linearity. Indeed, some of the data are fit best by non-linear functions. Nonetheless, the fact that the linear function fits best in most cases implies that these assumptions are not unreasonable under the experimental conditions.

It is important to note that in the current experiment, we did not have sufficient resolution at the lowest stimulation currents to test if DA growth was strictly scalar (i.e. x-intercept of zero). Non-zero, x-intercepts are not implausible, however, one possible cause being the damage around the electrode. The scar tissue creates a dead zone where no action potential can be triggered, thus preventing the lowest currents from firing neurons and producing detectable DA transients. The x-intercept of a linear fit would equal the current that just fails to elicit a detectable DA transient (i.e. at $y = 0$).

The anatomical basis for the growth of DA transients as a function of stimulation current

We demonstrated theoretically that scalar growth in the amplitude of DA transients as a function of current is consistent with proportionality between transient amplitude and the number of stimulated axons. We have yet to explain our ability to detect that scalar growth in DA by means of FSCV. If carbon-fiber electrodes could monitor DA release over the whole terminal field, they would detect the scalar growth. However, the electrodes are only sensitive to DA release in "micro-domains" in the vicinity of the cylinder-shaped, electroactive surface at the tip (radius: 3,5 µm; height: 150 µm). Each increase in stimulation current recruits new neurons farther away from the electrode tip at the fringe of the spherical spread of the threshold current density. For the DA transients to grow as a scalar function of current, the DA neurons activated by the full range of MFB stimulation currents must converge onto the micro-domain.

To illustrate this point, let us consider two anatomical arrangements that could drive convergence of the neural activity triggered by MFB stimulation (Figure 22). The first scenario is the summation of inputs at DA cell bodies (left diagram in Figure 22A). Some MFB axons, shown in yellow, are recruited by the stimulation, whereas others, shown in black, are unstimulated. At the convergence point there is spatial summation between the post-synaptic potentials elicited by the MFB inputs in the DA cell bodies. In our example, the summed effects of the stimulation are

sufficient to activate midbrain neurons as indicated by the lightning symbol. The same integrated signal is outputted by all DA fibers. Note that the integrated signal is proportional to the number of stimulated neurons in the MFB and broadcasted over the terminal field, as shown by the heat map, a schematic representation of DA release. A larger number of neurons stimulated would intensify the colour gradient of the heat map and enlarge its spatial spread at the fringe, proportionally to that number. If a carbon-fiber monitored that terminal field within the boundaries of the heat map, regardless of its location, it would observe a scalar relationship between stimulation current and amplitude of DA transients.

The second scenario places the convergence point in the terminal field (Figure 22B). Once again, some MFB axons, shown in yellow, are stimulated and others, shown in black, are unstimulated. In contrast to the previous example, DA fibers do not all output the same integrated signal: only the DA neurons activated by MFB outputs produce stimulation-evoked DA release. It is the overlapping terminal arbors of the DA fibers that combine the effects of MFB stimulation. Recruiting the black neuron by increasing the current would boost DA release throughout the depicted micro-domain. Once again, the amplitude of the DA transient grows as a function of the number of stimulated MFB neurons.

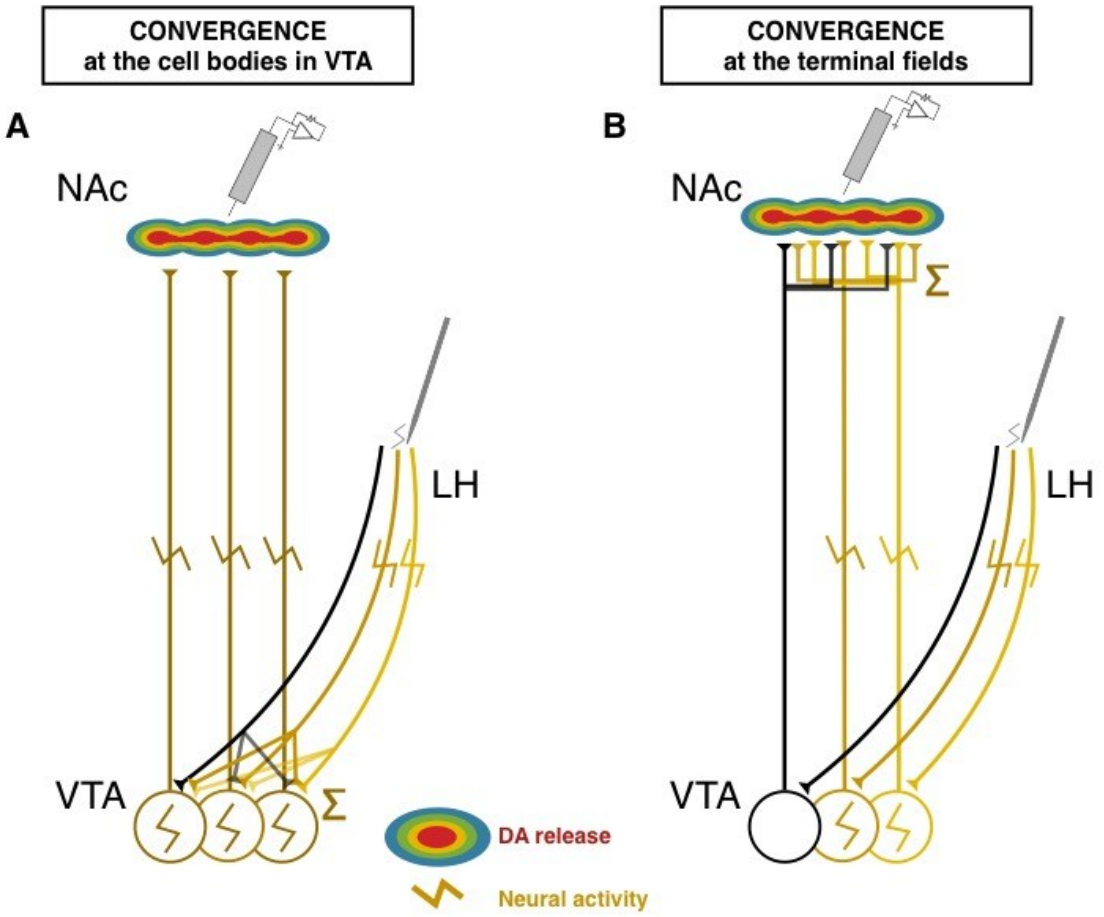


Figure 22. Two projection patterns that could cause the amplitude of DA transients in the NAc to increase as a function of the electrical current delivered to the MFB

Panel A depicts overlapping projections of MFB afferents to DA cell bodies. Some MFB axons, shown in yellow and brown, are recruited by the stimulation, whereas others, shown in black, remain unstimulated. Spatial summation in the DA cell bodies between the post-synaptic potentials elicited by the MFB afferents causes the amplitude of the DA transient to grow as a function of the number of activated MFB axons. Panel B depicts overlapping projections of DA afferents to the NAc terminal field. The overlap between the terminal arbors of the DA fibers causes the amplitude of the DA transient to grow as a function of the number of activated DA neurons. The heat map provides a simplified spatial representation of the extracellular concentration of DA in response to MFB stimulation. Propagating action potentials

are depicted by lightning symbols. DA, dopamine; NAc, nucleus accumbens; MFB, medial forebrain bundle.

The two scenarios (convergence at the level of the DA cell bodies or terminals) are not exclusive. Both types of convergence could contribute to a scalar relationship between the number of stimulated MFB neurons and the amplitude of the recorded DA transient, and empirical evidence indeed suggests that both types of convergence contribute.

Single-cell axon tracing analysis reveals that the telodendria of midbrain DA neurons span large territories in their terminal fields including the ventral striatum (Aransay, Rodrigue-Lopez, Garcia-Amado, 2015; Rodrigue-Lopez, Clasca & Prensa, 2017). There is also evidence for spatial summation between inputs to single DA cell bodies in the VTA (Moisan & Rompré, 1998). Those inputs were driven by stimulation of the posterior mesencephalon. We do not yet know for sure whether this is also the case with stimulation of the MFB, but this can be readily determined.

Inter-hemispheric variability in the rate of growth in DA transients

We have presented linear growth in DA transients as a dominant feature across all conditions tested in our experiment. However, the rate at which the growth occurred varied tremendously within pulse durations, hemispheres, and stimulation sites. The amplitude of contralateral compared to ipsilateral DA transients grew much slower in some data sets, or much faster or similarly in other data sets. As a consequence, the difference in the size of the contralateral DA transients or the slope of DA growth cannot be unambiguously interpreted. For example, a steeper growth in ipsilateral DA transients could be explained by a difference in the neural circuitry involved in producing DA transients in both hemispheres; perhaps the stimulation triggers additional inhibitory inputs to the contralateral DA neurons or triggers additional monosynaptic DA activation that cannot be relayed bilaterally.

A steeper growth could also emerge solely due to a difference in the location of the carbon-fiber electrodes within the spatial distribution of DA release in the terminal field. For example, if the carbon-fiber electrode was at the fringe of the heat map representing the spread of DA release (as

shown in Figure 22), it would detect a shallower increase in evoked-DA release per unit of stimulation current than if it was placed at the centre of the red zone. Without changing the stimulation site or the hemisphere in which the FSCV recording is obtained, the carbon-fiber could be moved and detect a shallow or a steep rate of DA growth.

Differences in sensitivity of the carbon-fiber electrode could also affect the slope of DA growth. Although great care was taken to produce and select sensors that respond similarly to standard concentrations of DA *in vivo* and to reduce damage during the implantation, variations in sensitivity are inevitable. A less sensitive electrode would need a larger change in stimulation current to produce a unit of change in the amplitude of DA transients compared to a more sensitive counterpart.

We demonstrated here that many variables can affect the slope in DA growth, several of which are tied to technical limitations of FSCV. Interpretations based on the amplitudes and the growth of DA transients observed in any FSCV experiments must therefore be formulated with caution. A case in point is one of the conclusions put forward by Fox, Mikhailova, Bass and colleagues in 2016. They reported small amplitudes in contralateral compared to ipsilateral DA transients produced by unilateral, electrical VTA stimulation. They recorded bilateral DA transients using FSCV and found that contralateral DA transients were small compared to ipsilateral transients. They concluded that, since a very small subset of DA projections crosses the midline to reach the contralateral striatum (1% to 3% of DA fibers; reviewed in Molochnikov & Cohen, 2014), the activation of these crossing fibers must produce the small contralateral DA transients.

The researchers only used one set of stimulation parameters: a long pulse duration of 2 ms and a current of 300 μA. In contrast, we used a wide range of stimulation parameters, and we reported several instances in which VTA stimulation produced large contralateral DA transients that were similar or larger to ipsilateral transients. Our findings challenge their interpretation of the parallel they saw between the small subset of crossing DA fibers and the small contralateral DA transients they observed. We predict that if Fox et al. had used near-threshold currents to evoke DA transients, they would have observed similar amplitude distributions for DA transients in the two hemispheres evoked by unilateral electrical stimulation. Such a result might have led to a

different conclusion: that trans-synaptic DA activation plays an important role in producing contralateral DA transients.

DA transients evoked by unilateral optical stimulation of midbrain DA neurons: a minor role at best for DA fibers that cross the midline

Another important finding of the present study is that the effects of direct, optical DA activation restricted to one hemisphere are seen only or primarily in the ipsilateral NAc despite the existence of DA projections that cross the midline. Such stimulation can produce contralateral transients that grow as a function of optical power when steps are not taken to restrict the spread of virus and light. However, only a few small, non-systematic, contralateral transients were detected when control conditions were imposed to ensure that contralateral VTA neurons could not be directly activated by optical stimulation.

Our control conditions likely prevented optical stimulation of the most medially located DA somata. Those specific neurons might be the source of DA fibers that cross the midline. However, these fibers are very few in number, and it seems unlikely that they could be responsible for the large contralateral DA transients often seen in response to optical VTA stimulation in the absence of control measures. Instead, these large contralateral responses could result from a combination of interhemispheric spread of the transgene-containing virus and light spread to the contralateral hemisphere. The histological results shown in Figure 18 show that ChR2 was expressed in some DA neurons contralateral to the injection site. That said, the distance of these neurons from the tip of the optical fiber seems large in terms of what is known about the relationship between optical power and the distance from a fiber tip at which ChR2 neurons can be activated (Aravanis, Wang, Zhang et al., 2007; Adamantidis, Zhang, Aravanis et al., 2007; Bernstein, Han, Henninger et al., 2008; Chow, Han, Dobry et al., 2010; Kahn, Desai, Knoblich et al., 2011; Yizhar, Fenno, Davidson et al., 2011; Yona, Meitav, Kahn et al., 2016). This issue merits further investigation using methods for visualizing optically activated neurons (see below).

Fox et al. also tested whether crossing DA fibers can produce detectable DA transients using unilateral, selective, optical stimulation of DA neurons in the VTA. In addition, they tested the

effect of electrical VTA stimulation. They reported small contralateral DA transients in response to both optical and electrical VTA stimulation. On this basis, they concluded that the effects of electrical VTA stimulation on phasic DA in the contralateral striatum are driven by activation of DA neurons that cross the midline. It is important to note that they used only a single set of stimulation parameters for each of the two types of stimulation, and they did not take steps to restrict the viral expression and light distribution to a single hemisphere. In contrast, we used a wide range of stimulation parameters as well as control conditions to restrict light distribution and viral expression to one hemisphere. Our results challenge the interpretation proposed by Fox et al. and argue against crossing DA fibers as a principal source of the contralateral DA transients.

Interestingly, Fox et al. reported contralateral DA transients in the dorsal striatum (DS) evoked by optical stimulation of the SN. Several millimeters separate the SN in the two hemispheres, making it unlikely that unilateral optical stimulation recruited DA neurons in the contralateral SN. The SN/DS findings of Fox et al. should be replicated and extended using a wide range of stimulation parameters for both the unilateral optical and electrical stimulation of the SN. Will the apparent symmetry of the evoked DA transients hold up under these more demanding conditions?

Neural circuitry that relays electrically evoked MFB signals to midbrain DA neurons

We observed similar amplitude distributions for DA transients in both hemispheres in response to unilateral, electrical MFB stimulation. It seems unlikely that the few DA fibers that cross the midline could be responsible for the contralateral responses. Instead, the results implicate neural circuitry that distributes signals initiated by unilateral, electrical MFB stimulation to DA neurons in both hemispheres. What is known about such circuitry?

Monosynaptic MFB signals to midbrain DA neurons

Monosynaptic DA activation can contribute to DA transients elicited by MFB stimulation. Optical stimulation restricted to VTA-projecting LH neurons can engage midbrain DA neurons, produce changes in phasic DA in the NAc, and support oICSS (Nieh et al., 2015, Nieh et al., 2016, Kempadoo et al., 2013; Stuber & Wise, 2016). However, the effects of monosynaptic DA

activation on phasic DA seem restricted to the ipsilateral terminal fields (Cossette & Shizgal, SfN poster, 2015). Non-DA fibers of passage potentially recruited by the stimulation could also contribute to the monosynaptic DA activation. Anatomical reports on monosynaptic inputs to DA neurons in the VTA identify very few bilateral projections in contrast with the overall large number of inputs. This finding suggests that, if fibers of passage have a role in monosynaptic DA activation, the effects would also be overwhelmingly restricted to the ipsilateral terminal fields (Watabe-Uchida et al., 2012; Menegas et al., 2015).

Polysynaptic relay of MFB signals to midbrain DA neurons

Polysynaptic DA activation can also elicit changes in DA activity. The cases of the mesopontine areas and the PVT are particularly well documented. At least a subset of MFB axons terminates in these two brain regions. In turn, they send bilateral inputs to the midbrain DA neurons (Cornwall, Cooper & Phillipson, 1990; Semba & Fibiger, 1992; Steininger, Rye & Wainer, 1992; Otake & Nakamura, 1998; Kirouak, 2015; Oakman et al., 1995; Dautan et al., 2014). In addition, when electrically stimulated, the mesopontine areas and the PVT produce changes in DA efflux in the NAc, as measured by amperometry (Forster & Blaha, 2000; Forster & Blaha, 2003; Blaha & Phillips, 1996; Yeomans, Forster & Blaha, 2001; Parsons, Li & Kirouac, 2007). Such polysynaptic pathways are the most likely to be involved in relaying the neural activity triggered by MFB stimulation to midbrain DA neurons in both hemispheres.

Future directions

Identifying neural populations that play a role in producing DA transients at the terminal fields

The combination of optogenetic manipulation and electrical stimulation has already proven useful in refining our understanding of the role of direct DA activation in the DA transients evoked by MFB stimulation. We propose to develop a similar approach using our experimental protocol and optogenetic silencing to test the necessity of different non-DA subpopulations recruited by the electrical stimulation in producing DA transients. Previous studies have identified potential components of the neural circuitry underlying polysynaptic DA activation. We would begin our investigation by targeting these regions and asking what the role of the laterodorsal tegmentum (LDTg) and the PVT is in polysynaptic DA activation produced by MFB stimulation.

The LDTg is a major source of bilateral innervation of the VTA and the NAc (Dautan et al., 2014; Oakman et al., 1995), and receives unilateral LH projections (Cornwall, Cooper & Phillipson, 1990; Samba & Fibiger, 1992). One strategy to determine if the LDTg plays a role in DA transients elicited by MFB stimulation is presented in Figure 23A. Stimulating and recording electrodes would be implanted as before. The protocol would also include the injection of a viral construct containing a silencing opsin, such as enhanced halorhodopsin (eNpHR3.0), into the LDTg. An optical implant would then be placed immediately above the injection site for optogenetic silencing. The infected neurons are illustrated in green in the diagram (indicating they are light-sensitive). Following the protocol for unilateral MFB stimulation presented in this chapter, electrical stimulation would be delivered at several currents with the short and long pulse durations. This data set would represent the baseline DA transients produced by electrical stimulation.

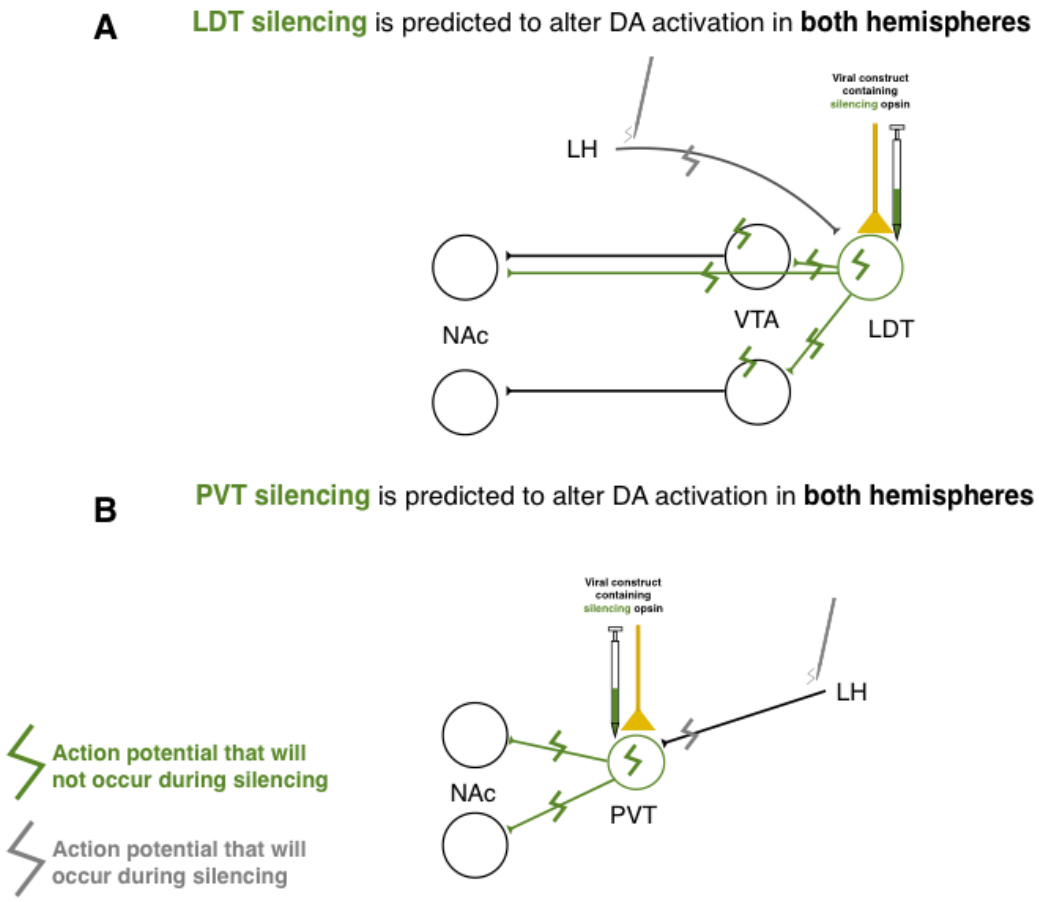


Figure 23. Follow-up experiments to test the role of the LDTg and the PVT in relaying signals induced by MFB stimulation to DA somata or terminals

Two putative polysynaptic routes linking MFB stimulation to DA release in the NAC. In Panel A, LDTg neurons relay afferent signals from stimulated MFB fibers to midbrain DA cell bodies. In Panel B, PVT neurons relay afferent signals from stimulated MFB fibers to DA terminals in the NAc. Expression of a silencing opsin in each of these relay regions is represented by the syringe symbol. The effects of optogenetic silencing during delivery of LH stimulation are shown in green. The green lightning symbols represent the action potentials that will not occur during simultaneous electrical and optical stimulation, whereas the gray symbols represent the action potential that will occur. LDTg, laterodorsal tegmental nucleus; PVT, paraventricular nucleus of the thalamus; MFB, medial forebrain bundle; DA,

dopamine; MFB, medial forebrain bundle; NAc, nucleus Accumbens; LH, lateral hypothalamus.

Another series of electrical stimulations would be delivered, but this time, it would be concurrent with optical silencing of the LDTg (Figure 23A). The light pulse would start a second before and end a second after the electrical train to ensure that LDTg neurons would be hyperpolarized throughout the time when they normally get activated by MFB stimulation. Action potentials produced by electrical stimulation are represented by lightning symbols in Figure 23 to show the hypothesized propagation of the neural activity from the MFB to DA neurons. Gray lightning symbols represent the action potentials that would be generated whereas the green symbols represent the action potential that would fail to occur due to silencing. If the LDTg plays a role in polysynaptic DA activation, silencing should affect DA transients produced by LH stimulation in both hemispheres (see Figure 23A).

Another region that could bilaterally relay the effects of MFB stimulation to DA neurons is the PVT. The LH projects to the PVT (reviewed in Kirouak, 2015) which, in turn, projects to both NAcs (Otake & Nakamura, 1998). The same strategy described above would be followed, substituting PVT silencing for LDTg silencing (see Figure 23B). If the PVT plays a role in polysynaptic DA activation, silencing should affect DA transients produced by LH stimulation in both hemispheres.

Visualizing the effects of optogenetic stimulation

Several variables come into play when evaluating the reach of the optical stimulation in triggering light-gated proteins: the wavelength used, the type of opsin, and the size and numerical aperture of the optical fiber used to deliver the optical train (Aravanis et al., 2007; Adamantidis et al., 2007; Bernstein et al., 2008; Chow et al., 2010; Kahn et al., 2011; Yizhar et al., 2011; Yona et al., 2016). We need empirical data to validate theoretical estimates based on visualization techniques that can map with precision the spatial distribution of optically activated neurons. Immunohistochemistry is routinely used in optogenetic experiments to assess the expression of the viral construct. This strategy provides a census of the neurons that were light-

sensitive and could have been activated, but cannot be used to identify which neurons were activated.

Immunohistochemical labelling of immediate early genes, such as the proto-oncogene, *c-fos*, and its subsequent synthesis of Fos protein has been widely used to identify increases in neural activation in multiple experimental contexts (Morgan & Curran, 1991; Krukoff, 1993; Cruz, Koya, Guez-Barber et al., 2013). Fos-like immunoreactivity has been visualized in the VTA in response to MFB stimulation in several occasions (Arvanitogiannis, Flores, Pfaus et al., 1996; Arvanitogiannis, Flores & Shizgal, 1997), but a double labelling for Fos and tyrosine hydroxylase (TH) revealed than less than 5% of DA cell bodies in the VTA expressed Fos following MFB stimulation (Hunt & McGregor, 1998). We conducted a pilot study at the National Institute of Health in collaboration with Bruce Hope. We used RNAscope, a ribonucleic acid (RNA) *in situ* hybridization technique that tags multiple neuronal types simultaneously, to characterize neural activation following MFB stimulation. In a manner consistent with the study of Hunt & McGregor (1998), we found that most Fos+ cells were glutamatergic neurons located near the midline and GABAergic neurons located throughout the VTA. Although immunohistochemical labelling for Fos is a standard measure of neural activation, it is not a suitable tool to measure activation of VTA DA neurons.

Genetically encoded calcium indicators (GECIs) are used in real-time *in vivo* calcium imaging studies as a means for visualizing neural activity in behaving animals. This technique gives a transient and local readout but can also be extended to provide post-mortem visualization of neurons activated at a specific time point during an experiment. For example, CaMPARI, a fluorescent calcium-sensor, undergoes a permanent photoconversion from green to red fluorescence when calcium levels rise above a certain threshold and violet illumination occurs simultaneously. The photoconversion marks activated neurons during the illumination producing a “snapshot” of neuronal activation at a specific time point (Fosque, Sun, Dana et al., 2015).

Specific expression of CaMPARI in DA neurons, by genetic targeting, is a promising approach to visualizing DA activation. However, validation steps to ensure that photoconversion is activity-dependent, including whole-cell patch clamp, are critical (Zolnik, Sha, Johennings et al.,

2017). Spikes and photoconversion in cortical neurons in an acute slice were found to grow linearly, with the exception of when the stimulation generated fewer than 100 spikes. In that case, changes in red/green fluorescence ratios were difficult to detect using a two-photon system. It remains to be determined if these validation steps would give similar results and be readily applicable to DA activation in the VTA when using a wide range of optical power and pulse frequency.

Refined tools are constantly being engineered to label different metabolic processes, some of which could map activity specific to a subset of neurons. One approach examined in our laboratory has been to label phosphorylated TH. During nerve stimulation, phosphorylation catalyzes TH activity by several kinases including PKA, PKC, CaMKII, MAPKAP-K2, ERK1, ERK2, MSK1, and PRAK to increase the synthesis of catecholamines (Albert, Helmer-Matyjek, Nairn et al., 1984; Daubner, Le & Wang, 2011). Once demand for DA synthesis diminishes TH is quickly dephosphorylated through the activity of phosphatases and is inhibited by DA. Thus, DA activation could be visualized by immunohistochemical labelling of phosphorylated TH. This is one of several molecular approaches that could be used to determine the spread of activating opsins in our experiment. Optogenetic techniques provide unprecedented ways to restrict activation in one type of neuron. We must continue to develop customized tools to map optogenetically-activated neurons that will match the specificity of optogenetic manipulations to ascertain their validity.

Conclusion

Most eICSS experiments have been carried out using short pulse durations and LH stimulation whereas most neurochemical studies have used long pulse durations and VTA stimulation. We questioned the degree of generalization between data acquired in these two different experimental contexts. We concluded that the effects of MFB stimulation on phasic DA were similar regardless of whether the stimulating electrode was placed in the LH or the VTA. A set of stimulation parameters deemed sufficient to produce detectable DA transients for one stimulation site would likely suffice at the other site and in both hemispheres.

The amplitude of DA transients usually grows linearly as a function of stimulation current. A simple explanation of this finding is that the number of stimulated MFB axons is a scalar or linear function of current and that the amplitude of the stimulation-evoked DA transient is a scalar or linear function of the number of activated MFB fibers. Growth of stimulation-evoked DA transients over a broad range of currents implies anatomical convergence, either at the level of the DA cell bodies, the DA terminals, or both.

Unilateral MFB stimulation produces contralateral DA transients without the need for direct DA activation or monosynaptic DA activation. We proposed follow-up experiments to further assess the role that different neuronal subpopulations play in relaying afferent signals from electrically stimulated MFB fibers to DA cell bodies.

Chapter 2

The neural substrates for the rewarding and dopamine-releasing effects of medial forebrain bundle stimulation have partially discrepant frequency responses

General Introduction

Several decades of research employing eICSS have led to the development of powerful tools to study brain reward circuitry (Milner, 1991; Shizgal, 1997). Extensive psychophysical measurements have provided a quantitative portrait of the neurons responsible for eICSS of the MFB and of the function that transforms the firing of these neurons into the rewarding effect (Gallistel, 1978; Simmons & Gallistel, 1994; Gallistel et al., 1981). We took advantage of this trove of quantitive knowledge to test the hypothesis that midbrain DA neurons constitute the final stage of the circuit that computes the rewarding effect. The study is presented next in the form of a published article.

Note: Some of the work was originally presented in my masters thesis. As part of my doctoral work, data acquired in anaesthetized animals was complemented with data acquired in awake animals. In addition, more extensive analysis of the psychometric curves were carried out, and the stability of DA transients across repetitions was also quantified.

ABSTRACT

Midbrain dopamine neurons have long been implicated in the rewarding effect produced by electrical brain stimulation of the medial forebrain bundle (MFB). These neurons are excited trans-synaptically, but their precise role in intracranial self-stimulation (ICSS) has yet to be determined. This study assessed the hypothesis that midbrain dopamine neurons are in series with the directly stimulated substrate for self-stimulation of the MFB and either perform spatio-temporal integration of synaptic input from directly activated MFB fibers or relay the results of such integration to efferent stages of the reward circuitry. Psychometric current-frequency trade-off functions were derived from ICSS performance, and "chemometric" trade-off functions were derived from stimulation-induced dopamine transients in the nucleus accumbens (NAc), measured by means of fast-scan cyclic voltammetry. Whereas the psychometric functions decline monotonically over a broad range of pulse frequencies and level off only at high frequencies, the chemometric functions obtained with the same rats and electrodes are either U-shaped or level off at lower pulse frequencies. This discrepancy was observed when the dopamine transients were recorded in either anesthetized or awake subjects. The lack of correspondence between the psychometric and chemometric functions is inconsistent with the hypothesis that dopamine neurons projecting to the NAc constitute an entire series stage of the neural circuit subserving self-stimulation of the MFB.

KEYWORDS

Brain stimulation reward, medial forebrain bundle, phasic dopamine release, nucleus accumbens, fast-scan cyclic voltammetry, current-frequency trade-off

Introduction

Rats and other vertebrates will work vigorously to trigger electrical stimulation of sites along the medial forebrain bundle (MFB) [1-2]. The rewarding effect of such stimulation can compete with, summate with, and substitute for naturally rewarding commodities, such as sucrose solutions, solid food, and water [3-5]. Midbrain dopamine neurons are activated trans-synaptically by MFB stimulation [6], and pharmacologically induced changes in their synaptic output profoundly alter the proclivity of the subject to reinitiate the electrical stimulation (intracranial self-stimulation: ICSS) [7-9]. Recent work [10-12] argues that the reward signal induced by electrical stimulation of the MFB undergoes spatio-temporal integration before, or as it is relayed to midbrain dopamine neurons.

The present experiment addresses a hypothesis based on a prior proposal by Moisan and Rompré [13] that places midbrain dopamine neurons in series with the directly stimulated substrate for self-stimulation of the MFB. On that view, the dopamine neurons relay the information received from the directly stimulated neurons to downstream elements of the circuitry. The behavioral and electrochemical findings reported here show that the dopamine projection from the ventral tegmental area (VTA) to the nucleus accumbens (NAc) cannot subserve such a role alone. A parallel pathway must shoulder at least some of the transmission load.
Midbrain dopamine neurons have long been implicated in ICSS [7-9]. For example, dopamine receptor antagonists block or reduce ICSS whereas indirect dopamine agonists, such as amphetamine and cocaine, potentiate ICSS [14-15]. However, behavioral studies also implicate neurons with highly excitable, myelinated axons in the rewarding effect of MFB stimulation [16-20, 6], and electrophysiological studies have identified non-dopaminergic neurons, directly driven by MFB stimulation, that have characteristics consistent with those inferred from the behavioral studies [21-23].

Fast-scan cyclic voltammetry (FSCV) confirms that midbrain dopamine neurons are activated trans-synaptically by rewarding MFB stimulation. Sombers, Beyene, Carelli & Wightman [24] demonstrated that, even when the tip of the stimulating electrode is positioned near the cell bodies of VTA dopamine neurons, it activates these neurons, at least in part, via a trans-synaptic

route. They infused the VTA with (±)2-amino,5-phosphopentanoic acid (AP-5), an N-methyl-D-aspartate (NMDA) receptor antagonist, to disrupt afferent glutamate inputs. This reduced the size of electrically-evoked dopamine transients and increased the latency of the rats to press the lever that triggered the electrical stimulation.

Moisan and Rompré [13] compared spatio-temporal integration in the neural circuitry underlying brain stimulation reward (BSR) and in the afferent network responsible for stimulation-induced activation of dopamine neurons. If the dopamine neurons constitute an entire series stage of the neural substrate for BSR, then relationships between input variables that are manifested in the final behavioral output (e.g., lever pressing) must also be evident in the activity of the dopamine neurons [25-26]. The specific relationship they investigated is the one described by the "counter model" of spatio-temporal integration in the neural substrate for BSR [27-28]. According to this model, the magnitude of the rewarding effect produced by a stimulation train of a given duration is a monotonic function of the number of firings induced in the directly-activated neurons. The same number of firings can be produced by using a high current to activate a relatively large number of neurons at a low pulse frequency or by using a lower current to activate fewer neurons at a higher pulse frequency. Moisan and Rompré [13] determined pairs of currents and pulse frequencies that produced equivalent performance for BSR and then assessed the responses of trans-synaptically activated dopamine neurons to these combinations. The neurons showed indistinguishable changes in firing rates to the two pairs of behaviorally equivalent stimulation parameters. On this basis, Moisan and Rompré [13] argued that the dopamine neurons could constitute a trans-synaptically activated stage in the circuitry subserving BSR.

In the present study, we adopted an approach analogous to that employed by Moisan and Rompré [13]. Whereas their stimulating electrodes were aimed at sites in the posterior mesencephalon, ours were aimed at the lateral hypothalamic level of the MFB. Whereas they focused on the range of pulse frequencies over which each stimulation pulse succeeds in firing the directly stimulated neurons, we also included a higher range, over which frequency following breaks down (neurons cease to fire in response to each stimulation pulse). Whereas they used single-unit recording to observe the effect of rewarding stimulation on midbrain dopamine neurons, we used FSCV to measure the synaptic output of these neurons in the NAc. Given that the dose of

amphetamine required to increase the reward effectiveness of electrical brain stimulation is lower when the drug is injected into the shell region of the NAc than into the core [29], we aimed the FSCV probes at the shell, the subregion where sensitivity to dopaminergic modulation appears highest. In this way, we tested the hypothesis that midbrain dopamine neurons projecting to the NAc shell constitute a series stage of the neural circuit underlying the rewarding effect of MFB stimulation.

Methods

Subjects

Subjects were ten Long Evans male rats (Charles-River, St. Constant, QC, Canada), weighing 450–500 g at the time of surgery. They were housed individually with ad libitum access to food and water and maintained on a 12 h light/dark reverse cycle (lights off from 08:00 to 20:00). The experimental procedures were performed in accordance with the principles outlined by the Canadian Council on Animal Care.

First surgery: implantation of stimulating electrode and recording hardware

In preparation for the first surgery, rats were anesthetized with a mixture of Ketamine hydrochloride (87 mg/kg) and xylazine hydrochloride (13 mg/kg) (injected i.p.). An s.c. injection of atropine sulphate (0.05 mg/kg) was given to reduce bronchial secretions during surgery, an s.c. injection of buprenorphine to alleviate post-operative pain (0.05 mg/kg), and an s.c. injection of penicillin procaine to prevent infection (0.3 cc/rat). The animal's snout was placed in an isoflurane nose cone so that volatile anesthetic could be administered continuously throughout the surgery (0.5 to 1.5%). One stimulating electrode, fashioned from 0.25 mm stainless steel insect pin and insulated with Formvar enamel to within 0.5 mm of the tip, was stereotaxically aimed at the right MFB at the level of the LH (AP: -2.8 mm; ML: 1.7 mm; DV: 8.9 mm from the skull; all coordinates are referenced to bregma). An additional electrode was placed in the external plexiform layer of the olfactory bulb (AP: 6.2 mm; ML: -0.1 mm; DV: 5 mm from the skull, all coordinates are referenced to bregma) to serve as the anode of the stimulation circuit. The location of the anode was chosen so as to place the recording site mid-way between the cathode and anode of the stimulation circuit. At that position, the recording site should lie on or near a zero-potential surface, thus reducing interference between the electrical stimulation and the voltammetric measurements. The electrode assembly was secured with dental acrylic and anchored with jeweller's screws.

During the first surgery, five animals were prepared for subsequent acute voltammetric recordings, to be carried out under anaesthesia. The skull above the NAc was temporarily

covered with bone wax, and two stainless-steel tubes were positioned to serve as reference points for the levelling of the head during the second surgery.

Four animals were prepared for subsequent acute voltammetric recordings, to be carried out while the rats were awake and freely moving. One cannula was placed above the NAc (AP: 1.7 mm; ML: 1.0 mm; DV: 4 mm from the skull), and one was placed at a posterior location in the contralateral hemisphere (~ AP: -7; ~ML 4.0 mm; ~DV 2 mm from the skull). Additional acrylic and jeweller's screws secured the cannulae. These cannulae served as guide tubes for the later positioning of a fresh, glass-sealed, carbon fiber (FSCV electrode) and a fresh, sintered, Ag/AgCl reference electrode (In Vivo Metrics, Healdsburg, CA) on the day of voltammetric recordings. Due to the limited space on the skull, the tip of the reference electrode could not be positioned on or near the zero potential surface of the electric field around the stimulation current. Thus, the MFB stimulation produced larger artifacts in the FSCV records obtained from the awake FSCV group than from the anesthetized group. Nevertheless, the artifacts disappear well before peak dopamine oxidation in all recordings.

One animal was implanted with a chronic carbon-fiber microsensor for subsequent acute voltammetric recordings, to be carried out while the rats were awake and freely moving. The chronic carbon-fiber microsensor was fashioned as described by Clark, SandBerg, Wanat et al. [30] and was aimed at the NAc (AP: 1.7 mm; ML: 1.0 mm; DV: 7 mm from the skull). To construct such an electrode, one carbon fiber is pushed through a section of silica tubing while immersed in alcohol; the microsensor is subsequently allowed to dry. Then, the microsensor is sealed with resin epoxy and fixed to a connector with silver epoxy.

An Ag/AgCl reference electrode, fashioned from a silver wire (diameter, AM-systems) chlorinated overnight, was implanted at a posterior location in the contralateral hemisphere (~ AP: -7; ~ML 4.0 mm; spanning ~ 4 mm from the skull). Additional acrylic and jeweller's screws secured the recording electrodes.

Urethane was chosen as the anesthetic for voltammetric recordings in animals under anaesthesia. Given the toxicity of this component, this phase terminates the experiment and thus was

performed at the end of the behavioral phase. Voltammetric recordings in freely moving animals were performed prior to behavioral testing; only upon successful acquisition of dopamine signals did animals continue on to the behavioral phase.

ICSS performance

Each animal was shaped to lever press for a 500 ms train of cathodal, rectangular, constant-current pulses with a pulse duration of 0.1 ms. Shaping took place in a Plexiglas operant chamber (30 cm long × 21 cm wide × 51 cm high) equipped with one retractable lever and a cue light positioned 1.5 cm above the lever. Illumination of the cue light signals that the lever press has been registered by the data acquisition system. A house light flashed during the entire duration of the 10-s intertrial intervals. Each trial consisted of a fixed time during which the stimulation parameters were held constant. After completion of the work requirement, as determined by the price variable (see below), the lever was retracted for 2 s before being made available again for lever pressing.

The behavioral measure was time allocation, the proportion of trial time during which the animal worked for the electrical stimulation of the MFB. Time allocation was corrected as described in Breton, Marcus & Shizgal [31]. The stimulation frequency was held constant during a trial and was decremented systematically from trial to trial, in equal proportional steps. Trial time (the total time that the lever was extended) was 80 s, which allowed the rat to harvest a maximum of 20 rewards at each pulse frequency. The range of frequencies tested was selected so as to track the sigmoidal relationship between pulse frequency and time allocation. Ideally, three data points were positioned along the upper asymptote, three along the lower, and three on the rising portion. Rats were trained on a cumulative handling-time reinforcement schedule which requires the rat to hold down the lever for a fixed, cumulative time, called the price of the stimulation train [31]. Initially rats were trained with the price set at 0.5 s. If the rat displayed signs of aversion or excessive motor effects, it was excluded from the experiment. After the rats performed well for trains offered at the 0.5-s price, the price was increased gradually to 4 s.

Once the psychometric curves relating time allocation to pulse frequency were stable for 3 consecutive days, current sweeps were substituted for the frequency sweeps. Current sweeps

were carried out at each of the following pulse frequencies: 60, 125, 250 and 1000 Hz. This set of pulse frequencies spanned the range tested by Solomon, Trujillo-Pisanty, Conover & Shizgal [32]. The first three values fell within the range over which current and pulse frequency traded off perfectly, or nearly so in the Solomon et al. [32] study; a given increase in frequency necessitated a proportional decrease in the current to maintain a constant level of ICSS performance. The fourth and highest pulse frequency fell within the range over which the current-frequency trade-off broke down. Within that range, the required current was constant, regardless of the pulse frequency. The abrupt transition observed by Solomon et al. [32] provides a well-defined point of comparison for the voltammetric data reported here.

The range of currents tested was selected so as to track the sigmoidal relationship between current and time allocation. Ideally, three data points are positioned along the upper asymptote, three along the lower asymptote, and three on the rising portion. However, each adjustment in stimulation parameters lengthens training substantially and, consequently, we tolerated fewer points on some portions of the curves in some animals. Testing was continued until stable current-sweep data were obtained for six consecutive days. The same procedure was then repeated for each of the 3 remaining pulse frequencies.

During the next phase of the experiment, test sessions consisted of current sweeps carried out at each of the four pulse frequencies; the order of the sweeps was randomized within session. Such test sessions were run repeatedly until stable data were obtained over six consecutive sessions. The mean time allocation and its standard error were computed for each current over these last six sessions.

Second surgery: Fast-scan cyclic voltammetry

Recordings in anesthetized rats

Rats were anaesthetised with urethane (1.5 g/kg, i.p.) and placed in the stereotaxic frame (Kopf Instrument, Tujunga, CA). The bone wax was removed and the head levelled with reference to the two guide tubes. A FSCV electrode, pre-cleaned with a 2-propanol solution containing activated carbon, was secured to a stereotaxic arm, attached to the headstage amplifier, and slowly lowered into the NAc (AP: 1.7 mm; ML: 1.0 mm; DV: 7 mm from the skull). The carbon

fiber was glass-encased, and a seal was produced by heating the glass capillary with a pipet puller (PUL-1, WPI, Sarasota, FL). A wire covered with silver paint was inserted in the capillary to make contact with the carbon fiber and secured with shrink tubing coated with epoxy (150-200 μm exposed tip length, 7 μm diameter). A sintered Ag/AgCl reference electrode (In Vivo Metrics, Healdsburg, CA) was placed in the hemisphere contralateral to the FSCV electrode, on or near the zero-potential surface of the electric field around the stimulation current.
MFB stimulation trains (0.1 ms pulse duration, 500 ms train duration, 400 μA, 60 Hz) were delivered as the carbon fiber was lowered in small steps, and the evoked dopamine transients were measured (see below) by means of FSCV. The position of the fiber was fixed when further lowering failed to increase the amplitude of the evoked dopamine transient. The FSCV electrode was conditioned at 60 Hz for 10 min and allowed to stabilize at 10 Hz for 10 min. Body temperature was monitored constantly and controlled with a thermal pad. An s.c. injection of Ringer's solution (6 ml/kg) was given every hour.

Recordings in awake rats

Animals were anesthetized briefly (~30 mins) with a mixture of Ketamine hydrochloride (87 mg/kg) and xylazine hydrochloride (13 mg/kg) (i.p.). The animal's head was temporary stabilized in the stereotaxic frame. A detachable micromanipulator containing a FSCV electrode (same fabrication as described above for the anesthetized animals; 150-200 μm exposed tip length, 7 μm diameter) was attached to the chronically implanted cannula that had been aimed to terminate above the NAc. The FSCV electrode was slowly lowered to a depth of 7 mm below the skull, and a sintered Ag/AgCl reference electrode was lowered through the second, more posterior, cannula to a depth of 5 mm below the skull.

The depth of the carbon-fiber microsensor was adjusted at the beginning of the recording session until electrically-evoked (0.1 ms pulse duration, 400 μA, 60 Hz) dopamine was detected with a signal to noise ratio (S/N) superior to 3, as measured by the Tarheel CV © software. The animal was placed in the FSCV chamber and allowed to awaken, which took between 5 to 10 minutes. Thereafter, it was connected to the head-mounted voltammetric amplifier via a commutator (Crist Instrument Co., Hagerstown, MD). Further depth adjustments were made to the carbon-

fiber microsensor if the transition from the stereotaxic frame to the FSCV chamber disrupted the quality of electrically-evoked dopamine.

The animal with the chronic carbon-fiber microsensor (150-200 µm exposed tip length, 7 µm diameter) did not require any anaesthesia prior to being placed in the FSCV chamber and was simply connected to the head-mounted voltammetric amplifier.

Fast-scan cyclic voltammetry recordings

All rats received experimenter-delivered electrical stimulation of the MFB for all four pulse frequencies tested in the behavioral phase, i.e. 60, 130, 250 and 1000 Hz. Early voltammetric measurements obtained in this study revealed a non-linear relationship between the amplitude of the dopamine transient and the current, as was also the case in the behavioral data. The range of currents for the FSCV recordings was chosen to be sufficiently broad to capture this non-linearity.

In the anesthetized group, electrically-evoked dopamine transients were obtained at all frequencies and all currents as quickly as feasible, and thus, each combination of parameters was only tested once. The currents were swept in an ascending order for all pulse frequencies to determine, as efficiently as possible, if high currents would lead to asymptotic amplitude in dopamine transients. The testing order for the four pulse frequencies was randomized. Preliminary data in anesthetized animals revealed that an inter-stimulation interval of 60 s ensures stable amplitude of dopamine transients when stimulating the MFB repetitively and thus that inter-stimulation interval was selected for this study. In the awake group, the maximal currents that could be delivered without aversive motor effects was determined first at each pulse frequency and then, animals were exposed three times to each randomly-ordered combination of parameters. Preliminary data in awake animals revealed that an inter-stimulation interval of 180 seconds ensures stable amplitude of dopamine transients when stimulating the MFB repetitively and thus that inter-stimulation interval was selected for this study.

Electrochemistry

Background-subtracted cyclic voltammograms were generated at 10 Hz by applying an 8.5 ms triangular waveform that ramped from -0.4 V to +1.3V and back to -0.4 V at a scan rate of 400 V/s. The potential was held at -0.4 V between each scan to promote cation absorption at the surface of the FSCV electrode. All potentials were measured with respect to the Ag/AgCl reference electrode. The waveform was generated using LabVIEW (National Instruments, Austin, TX) and a multifunction data acquisition board (PCI-6052E, National Instruments, Austin, TX).

A PCI-6711E (National Instruments, Austin, TX) board was used to perform waveform acquisition, and data collection. A synchronization signal from the PCI-6711E board was sent to the external input of a multi-channel pulse generator (Master-8, A.M.P.I, Israel) and used to trigger the electrical stimulation 5 s after the start of each recording. The stimulation was patterned so as to prevent overlap with the voltammetric scans. This was accomplished by confining pulse generation to the 91.5 ms intervals separating the triangle waves. Voltages generated by the Master-8 were converted to constant currents via a stimulus isolation unit (AM-2200, AM-Systems, Carlsborg, WA).

Quantification of dopamine transients

A subset of the carbon-fiber electrodes was pre-calibrated in an electrochemical cell connected to a 6-port pressure-injection valve (Upchurch Scientific, Oak Harbor, WA). FSCV electrodes were cleaned prior to calibration with a 2-propanol solution containing activated carbon. Artificial cerebrospinal fluid (aCSF; 145 mM Na^+ , 2.7 mM K^+ , 1.22 mM Ca^{2+} , 1.0 mM Mg^{2+} , 150mM Cl^-, 0.2mM ascorbate, 2mM Na_2HPO_4, pH 7.4±.05) was passed through the electrochemical cell using a pump (Multi-Phaser, YA-12, Yale Apparatus, Holliston, MA). The loop injector was loaded with dopamine standards at concentrations of 100nM, 200nM, 500nM, and 1000nM. Dopamine was dissolved in ASCF, freshly made at the start, and at every subsequent hour during calibration. Averaged calibration values were used to convert the measured dopamine signal from current to concentration.

Peak detection to quantify dopamine transients was performed using a customized routine in MATLAB (The Mathworks, Natick, MA). The maximal change in peak oxidation currents was measured during the rising phase of the triangle wave. For conversion into molar concentrations, these peak currents were then compared to in vitro post-calibration measurements. If the electrode was damaged upon removal at the end of the experiment, pre-calibration recordings were used instead. For data from the FSCV awake group, a mean and standard error were computed for each current.

Histology

After the completion of the experiment, a lethal dose of sodium pentobarbital (120 mg/rat, i.p.) was administered. For animals in the FSCV awake group, a stimulating electrode was lowered to the sites at which the voltammetric recordings were obtained, and a 1 mA anodal current was applied for 15 s to deposit iron ions at these locations. A 1 mA anodal current was also passed through the MFB stimulating electrode. The animals in the anesthetized FSCV group were tested prior to implementing the lesioning strategy and thus, their brains were not marked for visualization.

The animals were perfused intracardially with 0.9% sodium chloride, followed by a formalin-Prussian Blue solution (10% formalin, 3% potassium ferricyanide, 3% potassium ferrocyanide, and 0.5% trichloroacetic acid) that forms a blue reaction with the iron deposited at the tip of the electrode. Then, the brains were removed and fixed with 10% formalin solution. Sagittal sections, 40 μm thick, were cut with a cryostat (Thermo Scientific) and examined to confirm placements of the FSCV electrodes in the NAc shell and the stimulating electrode in the MFB.

Data analysis

Two current-frequency trade-off functions were derived and compared. As a function of pulse frequency, the psychometric trade-off expresses the current required to sustain half-maximal behavioral performance, whereas the chemometric trade-off expresses the current required to produce half-maximal changes in the amplitude of the stimulation-induced dopamine transient.

Psychometric trade-off functions

Data from the last six sessions were included in the trade-off analysis. The derivation of the psychometric trade-off functions is illustrated in Figure 25. Spline fits were used to describe the data obtained in each current sweep, thus avoiding the assumption of a particular and universal functional form. The maximum and minimum of each fitted function was determined, and the half-maximal time allocation was defined as their midpoint. The current required to sustain this half-maximal time allocation was then derived by interpolation from the fitted function. Next, the mean required current for the current sweeps obtained at each pulse frequency was calculated, and a 95 % confidence interval was constructed using resampling with replacement (bootstrapping with 1000 iteration).

Chemometric trade-off functions

The psychometric functions were generally sigmoidal in shape, or roughly so, but this was true of only a subset of the chemometric curves, which differed not only in shape, but also in their maxima. To derive the chemometric trade-off function, we used the following procedure so as to choose a criterion that intersected all of the curves. The voltammetric data from all 10 rats included at least one non-linear curve. This feature makes it possible to define a maximum. The non-linear chemometric curve with the narrowest range was selected, and the maximal and minimal change in phasic dopamine for that range was determined. A criterial value for half maximal changes in phasic dopamine was derived as the midpoint between the maximal and minimal values. The required current corresponding to half-maximal changes in phasic dopamine was then computed by means of linear interpolation for all pulse frequencies. One dataset included a "problem" curve consisting of only three datapoints with a narrow range. The midpoint of that curve was used as the criterion in that case so as to insure that the criterion intersected all four curves. The derivation of the chemometric trade-off functions is illustrated in Figure 26.

Comparing the psychometric and chemometric trade-off functions

The current experiment asked whether the current required to maintain either a behavioral or neurochemical criterion decreased in similar fashion as pulse frequency was increased. To facilitate visualization and to make the data commensurate, the required currents were

normalized in the same fashion: each value for the behavioral or neurochemical measures was divided by the required current at the 60-Hz pulse frequency. The normalized required currents were then averaged across subjects to obtain two averaged, normalized trade-off functions, which were compared by means of a generalized, linear, mixed-effects model. The results were analyzed using the fitglme function in MATLAB (R2014b), as provided in the Statistical Tool Box. The Wilkinson notation for the completely within-subject design is as follows:

*NrmI ~ 1 + Measurement * LogF + (1 | Subject)*

where:

NrmI = the dependent variable (the normalized required currents),

Measurement = the behavioral or neurochemical phase of the experiment, defined as a categorical variable, and

LogF = the pulse frequency, defined as a continuous variable.

The '1 | Subject' term specifies that all factors are within-subject.

Results

Histology

The tips of the stimulating electrodes were all within the MFB. The electrode tips in the anesthetized FSCV group were somewhat lateral and anterior (~ -2.6 to -2.8 mm relative to bregma; ~1.9 mm lateral to the midline; ~ -8.8 dorsal to the skull) to those in the awake FSCV group (~ -2.8 to -3.2 mm relative to bregma; ~1.5 mm lateral to the midline; ~ -8.8 dorsal to the skull) (see Figure 24A). However, this modest variation in electrode placements did not translate into noticeable group differences in ICSS performances, as discussed below.

The tips of the recording electrodes in the awake FSCV group were all within the medial shell of the nucleus accumbens (see figure 24B). The animals in the anesthetized FSCV group were tested prior to implementing the lesioning strategy now routinely used in the laboratory to identify recording sites, and thus, the recording sites in these animals were not marked for visualization.

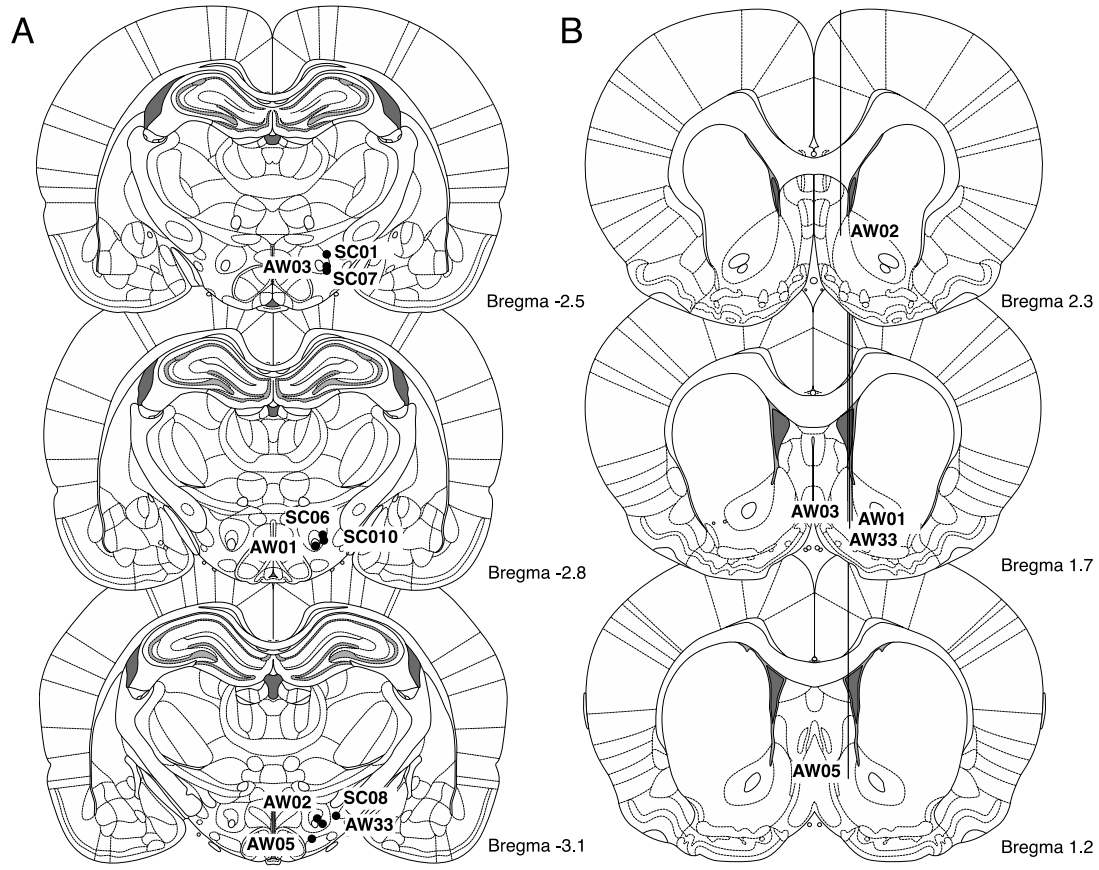


Figure 24. Histological localization of stimulation and recording sites

Location of the tips of A) the stimulating electrodes and B) the carbon fibers (FSCV electrodes). The animals in the anesthetized FSCV group were tested prior to implementing the lesioning strategy to mark the tips of the carbon fibers and thus, do not appear in B. FSCV, fast scan cyclic voltammetry.

ICSS performance

Four sets of psychometric curves relating time allocation to the common logarithm of the current were obtained for each rat, one at each of the tested pulse frequencies. An example is provided in Figure 25A and the left hand panel of Figure 25B. All curves are roughly sigmoidal in form, and they are displaced systematically leftward by increases in the pulse frequency. As predicted by the shape of trade-off functions in previous experiments [28, 32-33], the displacements of the curves are roughly the same as pulse frequency increases from 60 to 125 Hz and from 125 to 250 Hz; a much smaller displacement is seen when the pulse frequency is increased from 250 to 1000 Hz.

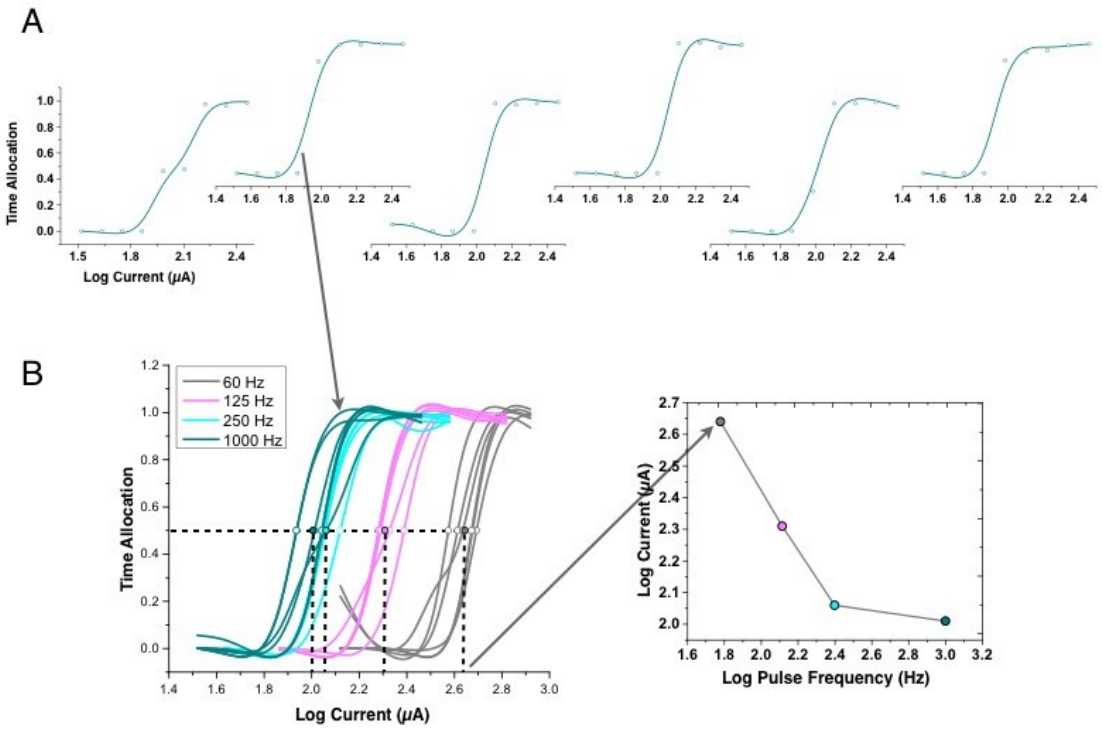


Figure 25. Derivation of the psychometric trade-off functions

A) An example of the raw behavioral data (open circles) and the corresponding spline fits (solid lines) for the last six sessions of training in the 1000 Hz condition. B) The left-hand graph shows the spline fits (solid lines) for the data obtained from one rat at all pulse frequencies. The horizontal dashed line depicts the criterial time allocation derived from the fits. The open circles depict, on each fitted curve, the half-maximal time allocation (y-coordinate) and required current (x-coordinate). The mean of the required-current values is depicted, for each pulse frequency, by a vertical, dashed line segment in the left panel and by a colored circle in both panels. These averaged values constitute the psychometric tradeoff function shown in the right-hand panel.

Phasic dopamine signalling in the nucleus accumbens

Changes in phasic dopamine transients evoked by electrical stimulation of the MFB were measured by means of FSCV in anesthetized (n = 5) and awake animals (n = 5) for all four pulse frequencies tested in the behavioral phase. All dopamine transients were time-locked to the stimulation trains, and all voltammograms had a signature characteristic of dopamine, with a peak oxidation at roughly 0.65 V and a reduction peak at roughly -0.2 V. The quality of the signature was comparable at all pulse frequencies. In most cases, changes in dopamine levels lasted between 2 to 3 s (see Figure 27).

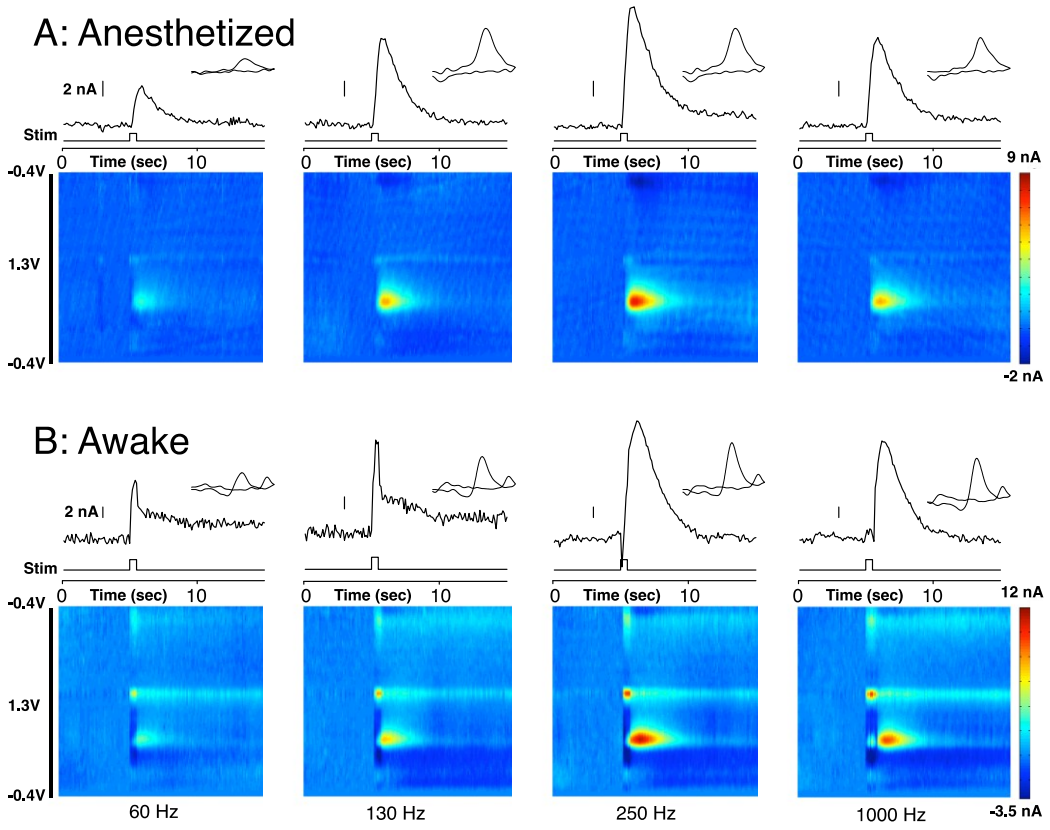


Figure 26. Examples of voltammetric measurements obtained at a stimulation current of 398 µA, at each of the pulse frequencies

Examples of voltammetric measurements obtained at a stimulation current of 398 µA, at each of the pulse frequencies: A) 60 Hz, B) 130 Hz, C) 250 Hz and D) 1000 Hz. Results for one anesthetized subject are shown in row A and for one awake subject in row B. The false-color voltammograms express voltammetric current as a function of time(s) on the x axis and voltage (v) applied at the carbon fiber on the y axis. The voltammetric traces correspond to horizontal cross-sections of the false-color voltammograms at the voltage that elicited peak dopamine oxidation, ~0.65 V. The insets show cyclic voltammetry signatures corresponding to vertical cross-sections of the false-color voltammographs at the time when dopamine oxidation peaks, ~between 5.5 and 6 s. The dashed arrows and lines in the most leftward colorplot of row A show an example of these cross-sections on the false-color

voltammogram. The reference electrode could not be positioned on or near the zero potential surface of the electric field around the stimulation current in the awake FSCV group. Thus, the MFB stimulation produced larger artifacts in the awake FSCV group. Nevertheless, the artifacts do not overlap the oxidation peak or interfere with its measurement. FSCV, fast scan cyclic voltammetry; MFB, medial forebrain bundle.

Four chemometric curves relating the peak amplitude of stimulation-evoked dopamine transients in the NAc to the stimulation current were obtained for each rat. An example is provided in the left-hand panel of Figure 26. Over a substantial portion of the current range in the double logarithmic plot, the dopamine transients triggered by all pulse frequencies grew quasi-linearly as a function of current. The growth of dopamine transients may level off at high currents. Such non-linearity was observed in all ten rats and in 37.5% of the curves overall. Changing the pulse frequency produced some dispersion of the chemometric curves, but the effect was smaller and less systematic than in the case of the psychometric curves. Although the curve obtained at 130 Hz was displaced systematically leftward of the curve obtained at 60 Hz, further leftward displacements were only rarely observed at higher pulse frequencies, and rightward displacements were observed occasionally.

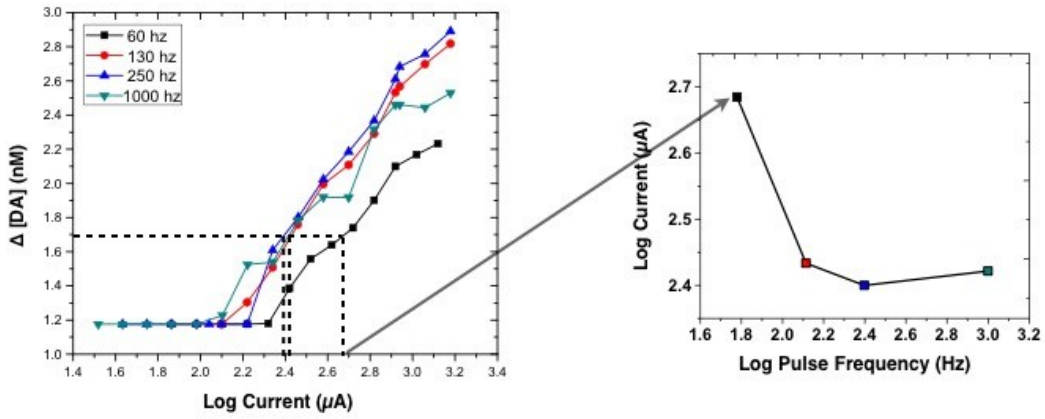


Figure 27. Derivation of the chemometric trade-off functions

The left-hand graph shows the voltammetric data (solid lines and filled symbols depicting the peak amplitude of the dopamine transient) obtained at all pulse frequencies for one rat. The horizontal dashed line is the criterion employed to derive the required currents (dashed vertical lines). These currents constitute the chemometric tradeoff function shown in the right-hand panel.

Trade-off functions

The left-hand panel of Figure 25B shows how the behavioral trade-off function is constructed. The dashed horizontal line lies at the midpoint between the maximal and minimal time allocation over the entire set of four psychometric curves, as determined by a spline fit. This line serves as the criterion for derivation of the trade-off functions expressing the current required to sustain half-maximal time allocation (referred to hereafter as the "required current") as a function of

pulse frequency. Figure 26 shows how a similar strategy is applied to derive the trade-off function expressing the current required to sustain half-maximal changes in phasic dopamine release as a function of pulse frequency.

Comparison between psychometric and chemometric trade-off functions

Figure 28 shows the average, normalized required current as a function of pulse frequency, for both the psychometric and chemometric data. The main effects of measurement type {psychometric, chemometric} ($F(1,56) = 11.22$, p = 0.001) and pulse frequency ($F(1,56) = 22.57$, $p < 0.001$) meet the criterion for statistical significance, as does the interaction between the two factors ($F(1,56) = 15,05$, $p < 0.001$). The interaction effect arises from the divergence of the two trade-off functions beyond 130 Hz. Whereas the normalized psychometric curve (grey, filled circles) continues to decline, the normalized chemometric curve (black, filled squares) levels off.

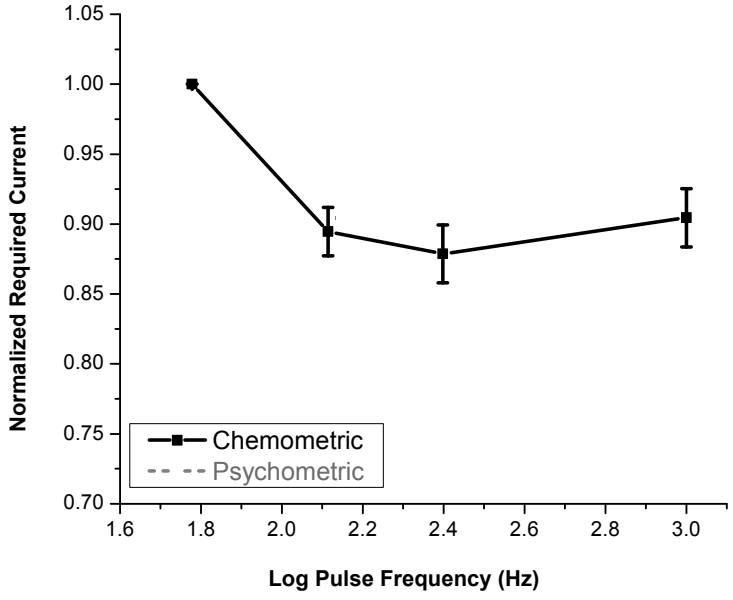


Figure 28. The average normalized change in required current, as a function of pulse frequency

The psychometric data are shown in grey (filled circles), whereas the chemometric data are shown in black (filled squares). The required currents were normalized by dividing each value by the corresponding required current for the 60-Hz condition. Main effect of measurement type {psychometric, chemometric}: $F(1,56) = 11.22$, $p = 0.001$; main effect of pulse frequency: $F(1,56) = 22.57$, $p < 0.001$; interaction effect: $F(1,56) = 15.05$, $p < 0.001$.

The divergence in the shape of the averaged psychometric and chemometric trade-off functions is seen in nine of the ten trade-off functions for the individual subjects in the FSCV anesthetized and awake groups (Figure 29); the distinction between the two trade offs for rat AW03 is more subtle. The slopes for all psychometric trade-off functions decline steeply between 60 and 250 Hz, and all level off between 250 and 1000 Hz except in the case of rat AW03. All chemometric trade-off functions also decline between 60 and 130 Hz. However, the curves rise at some point between 130 and 1000 Hz in half the cases and level off, or nearly so, in four of the five others.

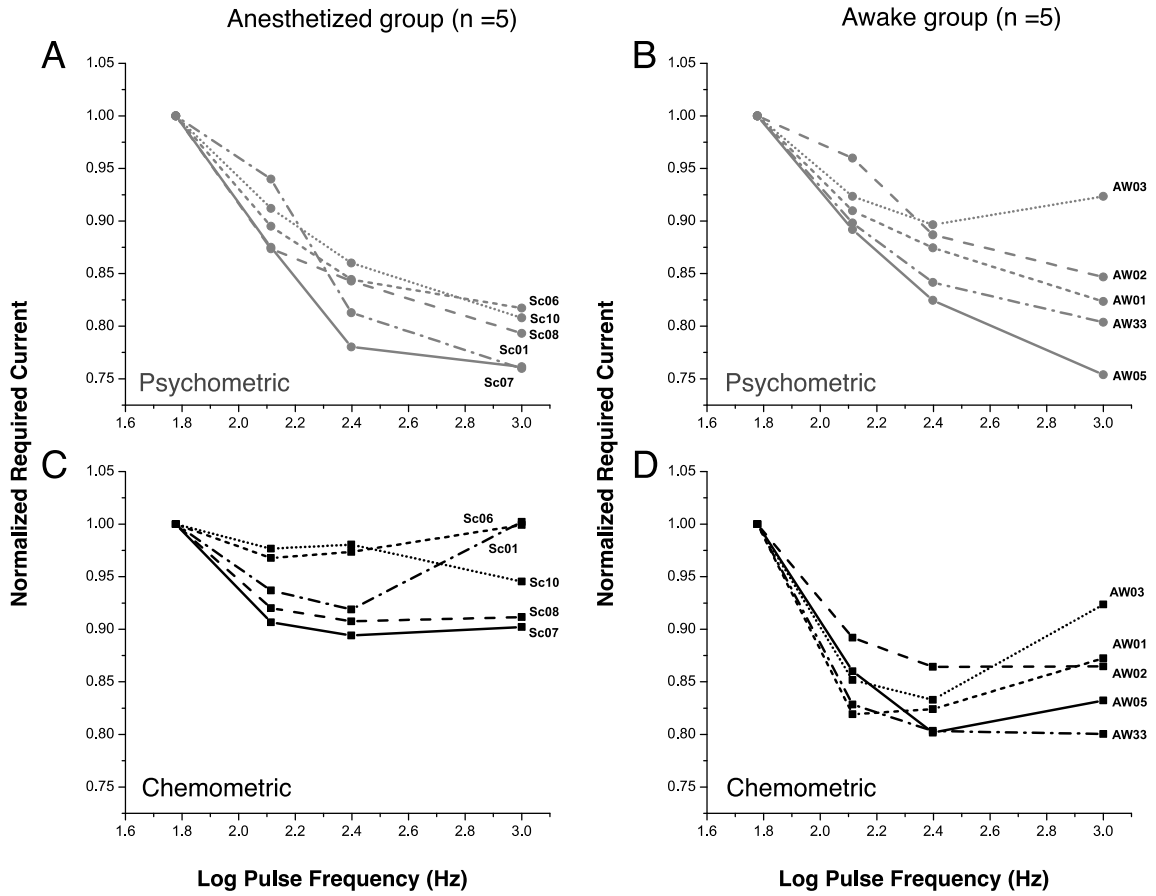


Figure 29. The required current for each rat

Psychometric data for the (A) anesthetized and (B) awake groups are shown in the upper row, whereas chemometric data in the (C) anesthetized and (D) awake groups are shown in the bottom row. The required currents were again normalized by dividing each value by the corresponding required current in the 60-Hz condition.

Discussion

The counter model and the current-frequency trade-off function

The reward effectiveness of MFB stimulation depends on spatio-temporal integration of the signals generated by the volley of action potentials triggered at the tip of the electrode. This integration process has been modelled successfully as a spike counter [27-28]. The counter model holds that when the train duration is held constant, the output of the integrator is determined by the total number of stimulation-elicited firings. If so, the same rewarding effect can be produced by stimulating many fibers at a low frequency or fewer fibers at a higher frequency, provided that the total resulting spike counts are the same. The counter model has fared well in empirical tests [32-33]. As the model predicts, the stimulation current (which determines the number of stimulated fibers) and the pulse frequency (which determines how often they fire) trade off effectively over a broad range of pulse frequencies. This trade-off breaks down abruptly once the pulse frequency exceeds a critical value, usually greater than 300 Hz [32]. As a result, the trade-off function assumes a form resembling a hockey stick. The diagonally oriented "handle" depicts the reciprocity between current and pulse frequency when the latter variable is below its critical value; the flat "blade" depicts the abrupt breakdown of reciprocity when the pulse frequency exceeds the frequency-following capabilities of the stimulated fibers. Although the abruptness of the transition from handle to blade is obscured in the behavioral data from the present study due to the coarse spacing of the pulse-frequency values, the overall form is highly consistent with that reported previously [32-33]. The straightness of the handle is predicted by the counter model, whereas the leveling off along the blade is what is expected as the frequency response of the directly stimulated ICSS substrate approaches asymptote.

Psychometric trade-off functions apply to multiple stages of neural processing

A remarkable feature of behaviorally derived trade-off functions is that when a particular condition has been satisfied, they apply to each and every stage between the final behavioral output and the stage of the neural circuitry that combines the effects of the two input variables [25]. For example, Figure 25A shows monotonic growth of time allocation (the final behavioral output) as a function of current (the initial input). In order for such a monotonic relationship

between the initial input and the final output to arise, the input-output function of each intermediate stage must also be monotonic over the tested range [25]. A failure of monotonicity in any of the intermediate stages would introduce an unresolvable ambiguity: more than one input to that stage would produce the same output. Such ambiguity cannot be removed by subsequent stages in a circuit consisting of concatenated stages because the downstream stages lack information about the initial input other than that relayed to them by the non-monotonic stage.

A trade-off function consists of pairs of input-variable values (e.g., current and pulse frequency) that produce the same output value (e.g., the time-allocation value represented by the horizontal dash line in Figure 25B). If the final stage of the circuitry is monotonic, then its output can be constant only if its input is constant as well. The input to the final stage is the output of the penultimate stage, so the input to the penultimate stage must also be constant across all values along the trade-off function. This argument propagates back to the stage (the integrator) that combines the behaviorally relevant effects of the two input variables (e.g., post-synaptic consequences of the volley of stimulation-induced action potentials).

The logical foundation for the present experiment

The logic described in the previous two paragraphs is the foundation for both Moisan and Rompré's study [13] and the present experiment. They showed that pairs of currents and pulse frequencies that produce equivalent behavioral performance also produce equivalent changes in the firing of midbrain dopamine neurons. We extended their approach to MFB stimulation sites and expanded the number of points along the current-frequency trade-off function so as to carry out a more demanding test. Three of the pulse-frequency values we chose fall along the handle of the trade-off function, whereas the fourth falls along the blade. In order for a neural population to be considered as a viable candidate for a stage between the integrator and the final behavioral output, it must manifest reciprocity between the current and the pulse frequency along the handle and a leveling off along the blade. Thus, if the dopamine projection to the NAc serves as a series stage in the reward pathway downstream from the directly stimulated neurons, then stimulation-induced dopamine transients in the NAc should reflect the way current and pulse frequency trade

off to determine aggregate impulse flow in the directly stimulated stage, and the form of the psychometric and chemometric trade-off functions should be the same.

The contrast between the psychometric and chemometric trade-off functions

We found a lack of correspondence between the psychometric trade-off function and the chemometric trade-off for phasic dopamine release in the NAc. The chemometric trade-off functions leveled off before their psychometric counterparts or inverted in sign at the highest pulse frequencies. This suggests that phasic release of dopamine in the NAc cannot serve as the sole output of a series stage in the neural circuit subserving the rewarding effect of MFB stimulation.

Monotonic relationships between phasic or tonic dopamine release and stimulation parameters that affect reward effectiveness (e.g., pulse frequency, current, pulse duration or train duration) have been described in several instances [34-41]. Results from these experiments have been interpreted as consistent with the notion that dopamine neurotransmission tracks reward effectiveness, as would be expected if midbrain dopamine neurons composed a series stage of the reward circuitry. However, the tests performed in these studies are not particularly demanding because the range of stimulation strengths tested are modest and do not extend into the regions in which behaviorally derived trade-off functions bend.

The present study demonstrates that combinations of stimulation parameters capable of maintaining a criterial level of ICSS performance, can, but do not necessarily, parallel the parameters capable of maintaining a criterial level of change in phasic dopamine release. The findings are in line with those reported by Miliaressis, Emond & Merali [26] who used microdialysis to measure tonic dopamine levels in the NAc in response to MFB stimulation. They obtained psychometric curves relating response rates to currents at different pulse durations and derived trade-off function by selecting combinations of stimulation parameters that sustained a 75% maximal response rate. Tonic level of dopamine varied dramatically in response to a set of four behaviorally equivalent stimulation parameters. As in the present study, there is no simple way to reconcile the results with the hypothesis that dopamine neurons projecting to the NAc constitute an entire series stage of the neural circuit subserving ICSS performance.

Some lack of correspondence has also been reported between measurements of ICSS and of voltammetric measurements of stimulation-evoked changes in dopaminergic neurotransmission. The amplitude of stimulation-evoked dopamine release for a fixed set of parameters declines as the animal learns to lever press for electrical stimulation of the VTA [42-43]. Latency to press the lever stabilized before stimulation-evoked dopamine release. This is consistent with results of an amperometry study conducted by Yavich & Tiihonen [44]. They measured rapid changes in NAc dopamine release in animals lever pressing for MFB stimulation delivered on two fixed-ratio (FR) reinforcement schedules: FR1 and FR8. In the FR1 condition, dopamine transients disappeared within seconds as animals began to work but in the FR8 condition, they persisted throughout the session. Early studies in which the FSCV electrode was in the NAc [45] or dorsal striatum [46] also reported that stimulation-evoked dopamine transients fell quickly below the detection threshold in rats that continued to self-stimulate.

The contrast between the psychometric and chemometric trade-off functions may have been underestimated

The procedures in the present study entailed a bias that may well have minimized the observed discrepancy between the form of the psychometric and chemometric trade-off functions. The bias arises from differences in the time between delivery of successive stimulation trains during collection of the psychometric and chemometric data. This interval was often brief during psychometric data collection: 6 s, on average, (4 s price + 2 s lever retraction) when the stimulation was highly rewarding, and 10 s, on average, (4 s price + 4 s latency + 2 s) when time allocation was 50%. The inter-train interval was much longer during the collection of the FSCV data: 60 s (anesthetized group) or 180 s (awake group). Thus, the possibility that the stimulation depleted dopamine stores or produced other effects of neural fatigue was higher during collection of the psychometric data. Depletion and fatigue should increase as the pulse frequency grows. If so, a bias created by the differences in minimum inter-train intervals would make the stimulation *less* effective at high frequency when measured psychophysically than when measured voltammetrically. The opposite result was obtained: the high pulse frequencies were relatively *more* effective when measured psychophysically than when measured voltammetrically (Figure 28). Thus, an even more dramatic discrepancy between the form of the psychometric and

chemometric trade-off functions could be expected if the two sets of measurements were obtained concurrently in future work, using the same inter-train intervals.

Depletion of the readily-releasable pool of dopamine has been suggested as the reason for the attenuation or disappearance of stimulation-induced dopamine transients at short inter-train intervals [44]. In the present study, three successive measurements of stimulation-induced dopamine release were obtained at each set of stimulation parameters in the awake group. We interrogated the data for signs of dopamine depletion by determining whether the amplitudes of the evoked dopamine transients decreased over successive measurements. As shown in Figure 30, the amplitude of the stimulation-evoked dopamine transients was impressively stable across the three successive measurements, even at the 1000 Hz stimulation frequency. Thus, at the long inter-train intervals employed in the awake group (180 s), there is no evidence of successive dopamine depletion. Given the strong similarity between the results obtained in the awake and anesthetized rats, we infer that dopamine depletion was unlikely to have been substantial in the latter group either, which was tested using a 60 s inter-train interval.

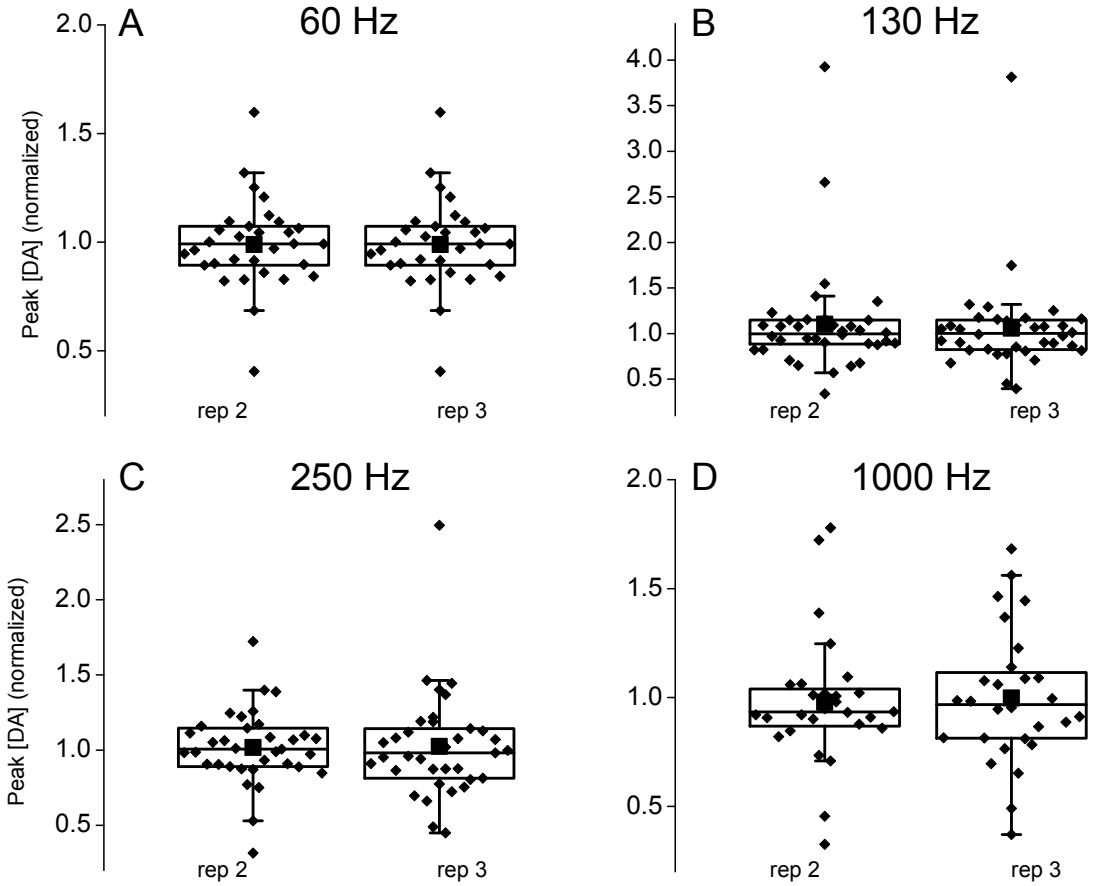


Figure 30. Stability of stimulation-induced dopamine release

Difference between the peak DA concentration recorded on repetitions 2 and 3 and the peak concentration recorded on repetition 1. Data are grouped within pulse frequency (A–D). Each diamond shows the peak DA concentrations recorded at one current in each rat, the filled square represents the mean, the box represents the inter-quartile range, the horizontal line within the box represents the median, and the whiskers represent the 5th and 95th percentiles. DA, dopamine.

Figure 31 shows cumulative time allocation for one rat at all four pulse frequencies during ICSS and from the remaining rats at 1000 Hz. Note that the allocation trajectory is quasi-linear, even when the stimulation was highly rewarding and high frequencies (250 or 1000 Hz) were employed. Thus, the rewarding effect of the stimulation appears to have been stable even when

the inter-train intervals were short, as was the case whenever time allocation was high. The results of prior studies [44-45] predict that substantial dopamine depletion may have occurred when time allocation was high. If so, an even greater discrepancy between the form of the psychometric and chemometric trade-off functions would be expected if the FSCV and ICSS data had been obtained concurrently using the same inter-train intervals. This should be tested in future work.

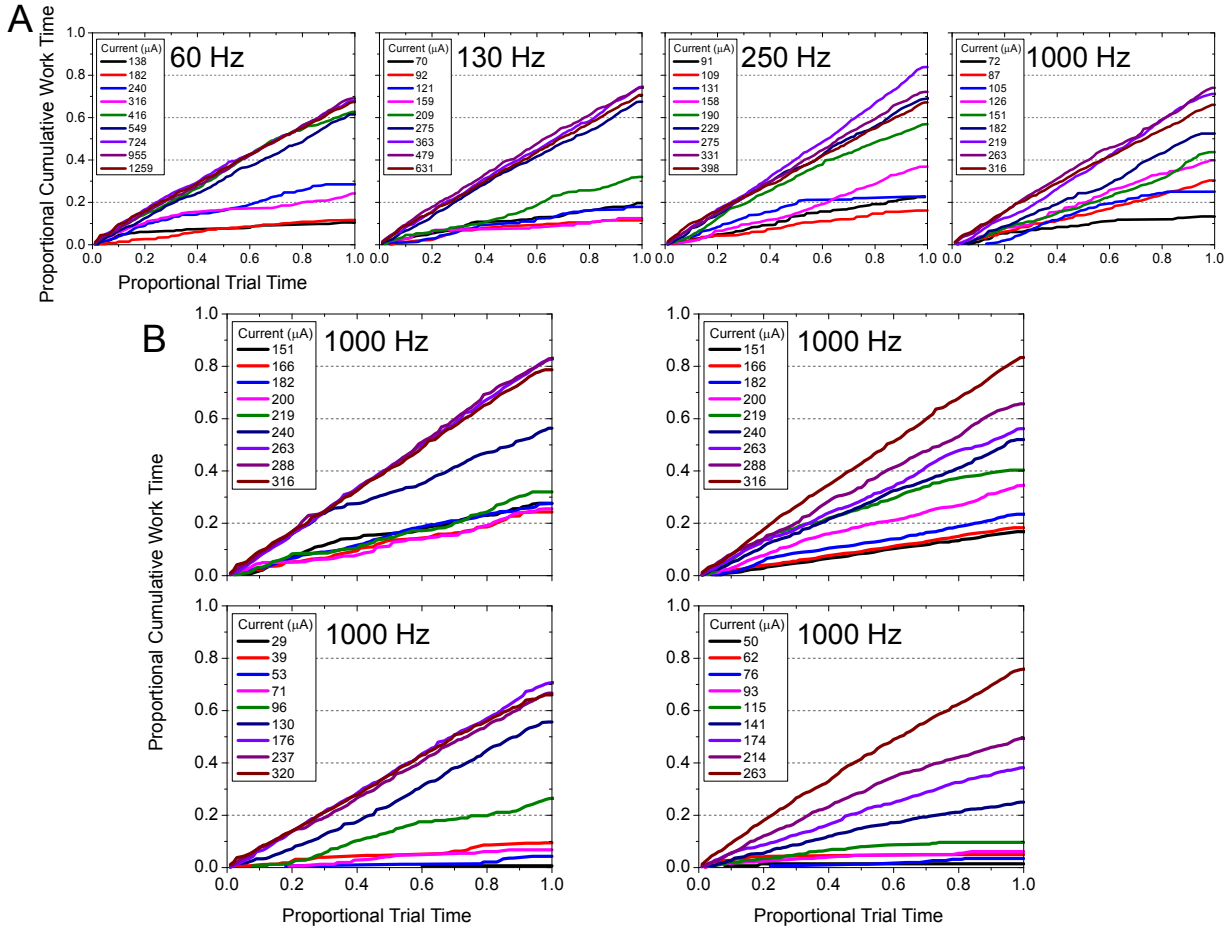


Figure 31. Cumulative records demonstrating stable behavioral allocation during intracranial self-stimulation

The cumulative proportion of trial time spent holding down the lever is plotted as a function of trial time. If the animal held down the lever throughout the session, the time-allocation trajectory would follow the grey diagonal line. Each colored line represents cumulative performance for a different stimulation current. Panel A shows the performance of one rat at all four pulse frequencies, whereas Panel B shows the

performance of the remaining rats at the 1000 Hz pulse frequency. The graphs of cumulative time allocation by these rats for the remaining frequencies resemble those in Panel A. The trajectories followed by all rats are generally straight, particularly when the current is high, even at the highest (1000 Hz) pulse frequency. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

In the present study, the use of long inter-train intervals during collection of the FSCV data appears to have prevented substantial dopamine depletion. Although dopamine depletion may have occurred during the collection of the psychophysical data, any contribution of such an effect was more than offset by the robust response of the neural substrate for the rewarding effect of the stimulation to high pulse frequencies. Thus, another explanation must be sought as to why the higher pulse frequencies proved more effective in supporting ICSS than in driving dopamine release in the present study.

Sharing the load

Combination within the NAc terminal field

The Moisan and Rompré hypothesis pertains to a population property of the dopamine neurons. However, detection of dopamine release by means of FSCV is restricted to the vicinity of the FSCV electrode and thus represents only a small proportion of the dopamine projection to the NAc terminal field. To evaluate our adaptation of the Moisan and Rompré hypothesis, we must examine how local responses could be combined to yield a population response integrated across the terminal field.

Although none of the local dopamine responses measured in the present experiment track the trade-off function derived from the behavioral data, could the entire set of local responses do the job collectively? In principle, different subsets of dopamine neurons could track changes in aggregate impulse flow over different, limited portions of the pulse-frequency range, and, if so, summation of the outputs of these subsets could encode the reward signal over the entire range. However, it is not obvious how the bandpass filtration would be achieved so as to map different portions of the pulse-frequency range onto activation of different “micro-domains” in the ventral

striatum. We know of no data suggesting that such frequency-dependent filtration occurs. Also, none of the recorded transients track the steady decline in the behaviorally derived current-frequency trade-off function beyond 130 Hz, and several of the chemometric trade-off functions have slopes of the wrong sign over at least some portion of the tested range. We see no straightforward way to combine the individual sets of neurochemical responses to produce a trade-off function resembling the one derived from the behavior of the self-stimulating rats.

Combination across terminal fields

The FSCV recordings reported here were obtained exclusively in the NAc terminal field. We cannot rule out the possibility that the information missing from the chemometric trade-off functions is represented by dopamine release in other terminal fields that receive stimulation-induced input. Additional recordings would be required to determine the plausibility of such a scheme. Prior work suggests that the NAc field is the principal behaviorally relevant target of the stimulation-induced dopamine release [47-49].

Combined effects of co-released neurotransmitters

Midbrain dopamine neurons projecting to the NAc co-release glutamate [50-52]. Therefore, we need to distinguish the hypothesis that midbrain dopamine neurons projecting to the NAc constitute a series stage of the neural circuitry subserving ICSS from the more restricted hypothesis that they accomplish this solely via their release of dopamine. Could release of dopamine and glutamate in the NAc conjointly encode the rewarding effect of MFB stimulation? Evidence that the dependence of release on firing frequency differs in the case of these two neurotransmitters [51] is interesting in this regard. To evaluate this hypothesis, it will be necessary to monitor the release of the two neurotransmitters concurrently under the conditions of an experiment such as the present one. Progress toward achieving such dual monitoring has been reported [53-54].

Conclusion

The behavioral and electrochemical findings reported here show that phasic release of dopamine in the NAc cannot serve as the sole output of a series stage in the neural circuit subserving the rewarding effect of MFB stimulation. Electrochemical measurements of dopamine release in additional terminal fields and of co-released glutamate [55] are required before this statement can be generalized to the dopamine neurons that are affected by the rewarding stimulation. Such measurements could readily be based on the same logical foundation as the one on which the present experiment rests. Should such measurements fail to rescue our adaptation of the Moisan and Rompré hypothesis [13], this would lend further credence to alternative accounts of the profound effects on ICSS performance produced by manipulation of dopaminergic neurotransmission. Among these are rescaling of the reward-intensity signal (gain adjustment) and alteration of the subjective effort cost of the reward-procuring behavior [10-12].

General Discussion of chapter 2

The relevance of a psychophysical framework to understand eICSS at other stimulation sites and oICSS

Brain reward circuitry is often conceptualized as a unitary system that receives converging inputs from rewarding stimuli and computes a common currency to enable choices between a wide variety of rewards such as warmth, food and money (Conover & Shizgal, 1994 ; Conover, Woodside & Shizgal, 1994; Shizgal, 1997). Data in chapter two demonstrate that the phasic DA release in the NAc does not suffice to compute or relay the reward signal underlying eICSS of the MFB to the behavioural final common path. This result is inconsistent with series-circuit hypothesis: At least part of the reward signal appears to reach the behavioural final-common path by another route. This may well be the case for eICSS of other brain sites and for the effects of natural rewards as well.

Several studies report a correlation between accumbal DA concentration and the reward strength of natural or artificial reinforcers (Beyene, Carelli & Wightman, 2010; Gan, Walton & Phillips, 2010; Howe, Tierney, Sandberg et al., 2013; Hernandez, Cossette, Shizgal et al., 2016; Sackett, Saddoris & Carelli, 2017). These findings support the role of DA signaling in relaying the rewarding effects of electrical MFB stimulation but do not provide definitive evidence. Our objective was to developed a more demanding test, using a psychophysical approach. Whereas a correlation can arise from many different relationships, we made predictions about both the functional form (the shape) *and* the parameters (how the shape must scale) that relate pulse frequency and stimulation current to the amplitude of stimulation-induced DA transients. The psychometric function (the trade-off between pulse frequency and the stimulation current that elicits criterial behaviour (i.e. half-maximal time allocation)) would have allowed us to predict the amplitude of stimulation-induced DA transients if phasic DA signaling consisted of an intermediate stage of reward computation that relays the reward signal underlying eICSS of the MFB to the behavioural final-common path. Any intermediate series stage must reflect this relationship (Gallistel, Shizgal& Yeomans, 1981).

The psychometric trade-off functions described in this chapter represent the relationship between pulse frequencies and currents that jointly determine behavioural output, measured here as time allocation. These two stimulation parameters determine the aggregate rate of firing in the population of directly stimulated neurons that give rise to the rewarding effect. According to the counter model (Gallistel, Stellar & Rubis, 1974; Gallistel, 1978; Simmons & Gallistel, 1994), an early stage of the reward pathway generates a signal proportional to the product of the number of stimulated neurons and the number of times they are fired by a pulse train of a given duration. This signal is the input to the neural stages that compute the intensity of the rewarding effect and determine the proclivity of the rat to seek out further stimulation.

In our experiment, all of the psychometric current-frequency trade-off functions decline between 130 and 250 Hz; 9 of the 10 psychometric trade-off functions decline further between 250 and 1000 Hz. The decline in required current between 130 and 1000 Hz implies that the induced firing rate grows over this range because the higher the firing rate of each individual neuron, the lower the number of neurons required to produce a given aggregate spike rate. If the DA projection from the VTA to the NAc were to constitute a stage of processing between the output of the spike counter and the input to the behavioural allocation function, stimulation-induced DA release in the NAc would have to grow between 130 and 1000 Hz in a manner similar to the growth of the firing rate implied by the psychometric trade-off functions. It did not: the drop in chemometric trade-off functions decelerates between 130 and 1000 Hz and turns into a rise in 7 of 10 cases. The inflection of the chemometric trade-off functions would be manifested in the final behavioural output as a TA-versus-Log(current) curve for a higher pulse frequency (e.g., 1000 Hz) that lies to the right of the TA-versus-Log(current) curve for a lower pulse frequency (e.g., 250 Hz). Instead, the TA-versus-Log(current) curves for eICSS march steadily leftwards as the pulse frequency is increased to 250 Hz and, in 9 of 10 cases, move at least somewhat further to the left between 250 and 1000 Hz. Such a result could not be generated by a pathway containing an intermediate stage with the same profile as the chemometric trade-off functions. Therefore, the DA projection from the VTA to the NAc cannot serve as the sole constituent of an intermediate stage, in contradiction to the series-circuit hypothesis.

It is the condition that intermediate stages in series between the directly stimulated neurons and the final stage of reward computation must reflect the shape of psychometric functions that makes our test of the series-circuit hypothesis a stringent one. This approach has been laid out by pioneering work done decades ago using psychophysical methods (Gallistel, 1978; Simmons & Gallistel, 1994; Gallistel et al., 1981) and remains an invaluable tool to challenge and refined our theories on the role of DA neurotransmission in reward seeking.

Technological advances, including optogenetic perturbations of neural activity, are providing exciting new opportunities to probe the reward circuitry (Deisseroth, 2011; Bernstein & Boyden, 2011; Tischer & Weiner, 2014). A modified version of the approach presented here, substituting oICSS for eICSS, could further leverage the application of psychophysical methods. Selective optical stimulation of midbrain DA neurons elicits oICSS and evokes DA transients in the NAc. If changes in DA concentration become the only input to the reward circuitry, the trade-off between pulse frequency and optical power required to elicit a criterion level of oICSS performance should correspond to the trade-off to produce a criterion amplitude of optically-induced DA transients. Combining direct DA stimulation with other selective manipulations of the reward circuitry (e.g. selective optical stimulation of other neurotransmissions) could help examining the role of phasic DA release and its convergence with other reward-related signals in ICSS.

It was shown in this chapter that even if an increase in activity in a given subset of neurons, here DAergic, can modulate reward seeking and be sufficient in producing self-stimulation, activity in those same neurons do not necessarily compute the reward signal. Precise optogenetic perturbations and precise real-time monitoring of brain activity (fiber photometry, biosensors, genetically-encoded calcium sensors) are now routinely performed in numerous neuroscience laboratory (Akerboom et al., 2013; Boyden, 2015; Chernov et al., 2017). It is important to extend this level of sophistication to our behavioural and neurochemical protocols, thus providing demanding tests for our theories.

Chapter 3

Use of optical and electrical intracranial self-stimulation to test the hypothesis that dopamine-mediated reward-prediction errors shape reward predictions and the selection of reward-seeking actions

Introduction

Reinforcement learning (RL) is defined by the learning problem it addresses: how to maximize the harvest of rewards (Dayan & Niv, 2008; Sutton & Barto, 2018). RL researchers build learning algorithms, or agents, with the goal of assigning optimal credit to reward-seeking actions through successive approximations. These methods allow learning to proceed without complete knowledge of the environment and without providing prior instructions on how to proceed. The essential components of RL models consist of incremental and iterative search and selection principles.

RL emerged from the convergence of research on dynamic programming, trial-and-error learning and temporal difference learning. Dynamic programing is widely used to solve planning problems in complex environments that change over time. A seminal approach to dynamic programming was developed by Bellman in the mid-1950s and gave rise to what is now known as the Bellman equation (Sutton & Barto, 2018). The value function expressed by the Bellman equation is a key feature of reinforcement-learning. One major difference, however, is that dynamic programming is designed to find solutions in systems for which we have complete knowledge and does not necessary involve learning, which is the focus of the trial-and-error mechanisms at the heart of RL.

The roots of trial-and-error learning are in experimental psychology. One of the founding principles, the "Law of Effect," describes how actions that lead to rewarding outcomes are more likely to be repeated, while actions that lead to aversive consequences are more likely eschewed (Thorndike, 1911). Bush and Mosteller (1953) were the first to use an iterative error-based rule to mathematically formalize such an account of trial-and-error learning. The algorithm compares each encounter with an unconditioned stimulus to the probability of occurrence of that stimulus formed from all past encounters. If a large difference in prediction is detected, represented by the

error term, a large update in prediction is implemented. This model was extended further by Rescorla and Wagner (1972) to explain changes in associative strength between predictive cues and a reward delivery. The addition of a common error-term takes into account the modulation of the associative strength of a stimulus by other stimuli concurrently present. This is a significant departure from the Bush and Mosteller's linear model (1953). Such insights into the computational basis of learning inspired scientists in the field of machine learning to build robots that could learn by trial-and-error, including the maze-learning machine developed by Thomas Ross in 1933 and the mechanical tortoise developed by W. Grey Walter in 1950 (as cited in Sutton & Barto, 2018).

A third tributary of modern RL consists of temporal-difference algorithms that compute the difference between temporally successive estimates of a quantity to guide learning. An early example of such a method is a checkers-playing program developed by Samuel in 1959 (as cited in Sutton & Barto, 2018). Richard Sutton, in 1978, developed a seminal temporal-difference algorithm which he and Andrew Barto, in 1981, applied in an account of Pavlovian conditioning (Sutton & Barto, 2018).

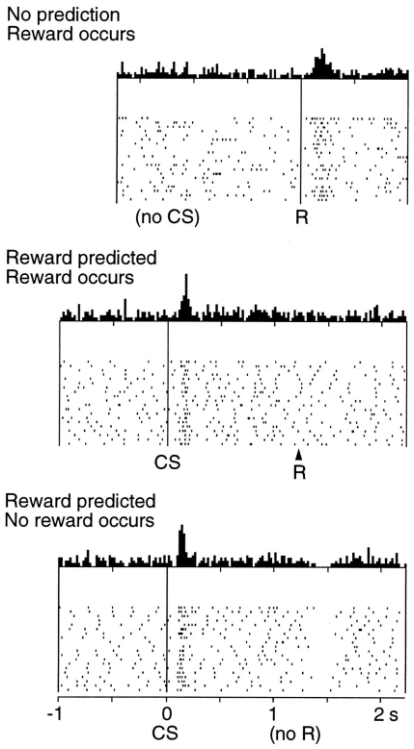
The parallel advances in dynamic programming, trial-and-error learning, and temporal difference learning eventually intertwined. An early paper by Minsky (1961) discusses many of the core issues, such as how to distribute credit appropriately among the multiple decisions that led up to a success or failure. Sutton and Barto (2018) credit Harry Klopf for integrative ideas that were eventually developed into founding principles of RL. They modified and extended Klopf's ideas in their work on temporal-difference learning. Temporal-difference algorithms were later combined with the fruits of research on dynamic programming to build agents that can optimize their reward-seeking actions on the basis of their experience (as cited in Sutton & Barto, 2018; Watkins & Dayan, 1992).

During the 1990s, there emerged growing recognition that RL was highly pertinent to neuroscientific analyses of the brain mechanisms underlying learning and action selection (Sutton & Barto, 2018). Particularly impactful was the realization by Montague, Dayan & Sejnowski (1996) that electrophysiological recordings obtained by Wolfram Schultz's group in

monkeys undergoing behavioural conditioning could be well explained in terms of temporal-difference reinforcement-learning.

The DA-RPE hypothesis

A series of experiments entailing electrophysiological recordings during conditioning procedures characterized the role of midbrain DA neurons in learning. Monkeys learned the contingency between visual cues and drops of juice delivered orally (Shultz & Romo, 1990; Schultz, Apicella, Ljungberg, 1993; reviewed in Schultz, 1998). In the operant paradigm, animals learned a conditioned reaching response to earn the juice. In the Pavlovian paradigm, anticipatory licking developed as the animals learned to predict juice delivery. A crucial feature of these experiments was the random duration of the inter-trial interval (ITI), which made the visual cue at the beginning of each trial the earliest reliable predictor of reward availability. Key findings from these studies were: 1) Before animals learn to anticipate rewards, the DA firing rate increases immediately after the unexpected juice delivery; 2) Once animals learn to predict the timing and magnitude of the reward, the DA firing rate increases at cue onset but not at reward delivery, and 3) if the reward is omitted, the DA firing rate decreases at the time when the juice was expected (Figure 32).



Reproduced from: Schultz, W., Dayan, P & Montague, P. R., (1997). A neural substrate of prediction and reward. *Science*. *275 (5306)*, 1593–1599.

Figure 32. Midbrain DA firing rate encodes an error in prediction of reward during Pavlovian learning

DA firing rate during delay conditioning. Monkeys learned the contingency between a visual cue and a drop of juice delivered orally at cue offset. The duration of the ITI varied unpredictably. Thus, the onset of the visual cue was the earliest reliable predictor of reward availability. Each panel comprises a raster plot, with a summary histogram above it. Each horizontal row of the raster plot represents the times of occurrence of recorded action potentials (dots) within each trial. The number of impulses within each 10-ms time bin is summed over the trials and represented as a bar in the histogram. CS refers to the conditioned, reward-predicting stimulus and R refers to the juice reward. The top panel shows an increase in the DA firing rate immediately after an unexpected juice delivery. The middle panel shows an increase in the DA firing rate at the onset of the reward-predictive cue but not at the delivery

of the expected reward. The bottom panel shows an increase in the DA firing rate at the onset of a reward-predictive cue but a decrease in the DA firing rate when an expected reward was omitted. Data are reproduced from Schultz, W., Dayan, P & Montague, P. R., (1997). A neural substrate of prediction and reward. *Science*. *275 (5306)*, 1593–1599. DA, dopamine; CS, conditioned stimulus. R, juice reward.

RL researchers noticed a resemblance between the changes in midbrain DA firing rate described above and the error signal used in machine-learning algorithms (Montague, Dayan & Sejnowski, 1996).

Reward-Prediction Errors

A RPE is the difference between the rewards expected and received at each point in time, *t*. The magnitude of the RPE reflects the accuracy of reward prediction: the smaller the RPE, the more accurate the prediction. A highly influential reinforcement-learning algorithm, called the "temporal-difference reinforcement-learning" (TDRL) algorithm, employs RPEs to drive learning. This algorithm is composed of four main components. A value function represents the predictive weights of reward-related stimuli. A temporal difference (TD) signal tracks changes in the temporally discounted sum of future rewards through a derivative that compares the current prediction, at time *t*, with the prediction made one moment ago, at time *t-1*. A TD error (RPE) combines the change in the expected reward value estimated by the TD function and the magnitude of the current reward. A weight-changing rule adjusts predictions, in proportion to the size of the TD error, by changing weights used to compute the TD signal and select reward-seeking actions.

Figure 33 illustrates how DA-mediated RPEs drive learning. It shows the evolution of the value function, the TD signal, the RPE expected under Schultz's experimental design (delay conditioning), and the DA firing observed while monkeys learn and perform the task. The key experimental times are the onset of the cue, depicted by vertical grey shaded areas, and the simultaneous offset of the cue and time of the reward delivery, depicted by red and green shaded areas, respectively. We will first focus on the time points depicted in red and green during the first trial before learning occurs (left column). The visual representations of the different

mathematical expressions, depicted by the grey horizontal lines in Figure 33, can assist the reader with the upcoming description of the TD model.

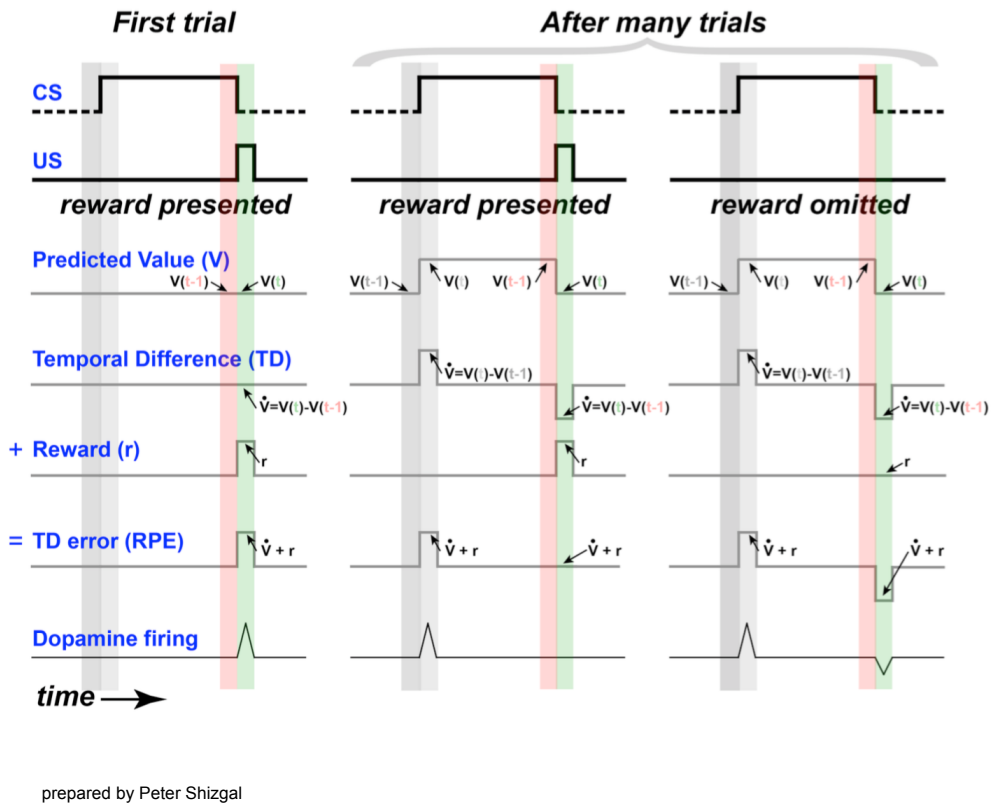


Figure 33. The TDRL algorithm and the DA-RPE hypothesis

This diagram shows how RPEs are generated by the TDRL algorithm and mapped into phasic firing in DA neurons. Three conditions are illustrated: the first trial before learning occurs (left column), after many training trials when performance has become asymptotic (middle column), and a probe trial on which the reward has been omitted after performance has become asymptotic (right column). Trial time runs from left to right as indicated by the black arrow at the bottom. The key trial phases are the onset of the cue, depicted by vertical grey shaded areas, and the simultaneous offset of the cue and time of the reward delivery, depicted by red and green shaded areas, respectively. The two first rows depict the presentation of the CS and US. The third row shows the value function, *V(t),* which assumes the same form as the CS as a function of learning. The fourth row

shows the TD function, which is the discrete derivative of the value function, *V(t) - V(t-1)*. The fifth row depicts the delivery of the reward, *r(t),* whereas the sixth row shows the TD error, the result of summing the TD (fourth row) and the reward magnitude (fifth row). The last row is a stylized representation of the observed firing rate of midbrain DA neurons, which closely resembles the TD error (RPE). According to the DA-RPE hypothesis, it is the phasic firing of midbrain DA neurons (last row) that drives the learning of reward predictions (third row) and reward-seeking actions (not shown). TDRL, temporal-difference reinforcement-learning; DA, dopamine; RPE, reward prediction error; CS, conditioned stimulus; US, unconditioned stimulus; TD, temporal difference.

The expected reward is represented by a predicted value function, *V(t)*. This function is the product of the state of all reward-related stimuli (x) and the reward-predictive weight (w):

$$V(t) = x(t)*w(t)$$

where x(t) represents the state of the reward-related stimulus, internal or external, as a function of time. w(t) represents the reward-predictive weight, or predicted value, of all these stimuli as a function of time.

Note that due to the random component of the ITI, the time of delivery of only one future reward, the one to be delivered on the current trial, can be predicted. Thus V comes to assume the value of the reward to be delivered on the current trial, rather than the discounted sum of the rewards to be delivered on all trials during the session.

Before the formation of an association between the cue and the juice reward, on the first trial (left column), the cue has no reward-predictive properties. The value function, *V(t),* has a value of zero at all time points, including time *t,* when the reward is delivered.

In the TDRL algorithm, the evolution of the reward prediction over time is tracked by comparing the current prediction, at time t, with the prediction made one moment ago, at time *t-1*. The resulting TD function is written as:

$V_{TD} = V(t)-V(t-1)$

where *V(t)* is the expected value of the reward predicted by the reward-related stimulus.

The TD function is an approximation of what is known about changes in the world rather than an exact representation of all past and future experiences. On the first trial (left column), since the value function, *V(t),* equals zero at all time points, the derivative, *V(t)-V(t-1,)* is also zero.

The TD error (RPE) sums the step-wise change in the reward prediction, as estimated by the TD function, with the juice reward. It is written as:

TD error = V(t)-V(t-1)+r(t)

where the value of the juice reward is described by a reward function, *r(t).*

At the time of reward delivery on the first trial (green shaded area, left column), the reward function, *r(t)*, has a positive value. A large TD error is produced, proportional to the magnitude of the reward. This is the only time point prior to learning at which the resulting TD error is positive.

Let us now examine changes in the value function, the TD signal, and the TD error after the task has been fully learned (middle column). Recall that the presentation of the reward-predicting cue (the CS) is unexpected due to the random component of the ITI. At the onset of the CS that as a result of learning now predicts the upcoming delivery of a juice reward (grey shaded areas), the value function, *V(t),* has a positive value that persists until the offset of the CS (red shaded area). The TD signal, *V(t)-V(t-1),* has a positive value as the value function transitions between predicting no reward one moment ago, *V(t-1),* and now predicting a reward at time *t*, *V(t)* (lighter gray shaded area). Conversely, the TD signal has a negative value at the time of reward delivery, as the value function transitions between predicting a reward one moment ago, *V(t-1),* and now predicting no reward (during the ITI), *V(t)* (green shaded area). Because the reward prediction

has been fully learned, the magnitude of the TD signal at the time of reward delivery is exactly equal to the magnitude of the reward. The TD error at the time of reward delivery thus adds two signals of equal magnitude but opposing sign and is thus equal to zero. The error is said to have been predicted away.

The right column in Figure 33 shows the value function, the TD signal and the TD error after the task has been fully learned but the reward has been omitted. At the onset of the cue, the three expressions take the values described in the last paragraph, and a positive TD error occurs at the onset of the CS. However, after the offset of the CS (green shaded area), a different result is obtained than in the training trials, when the reward was always delivered. As before, the TD signal, $V(t)-V(t-1)$, has a negative value at the moment when reward delivery is expected, but the reward function now has a value of zero due to the omission of the reward. Thus, the resulting TD error has a negative value.

TD errors drive the weight changes underlying learning

Learning is implemented through weight changes that adjust the value function and the likelihood of repeating the action that procures a reward (Sutton, 1988). These weights can be conceptualized as representations of synaptic strength within the reward circuitry (Montague, Dayan & Sejnowski, 1996). The TD error optimizes the weights. The size of the TD error indicates how far the weight deviates from its optimal value, and thus the magnitude of the required adjustment is proportional to the error. The abruptness of the weight changes also depends on a scalar called the "learning rate." If the learning-rate scalar is too low, predictions will lag far behind changes in the environment, whereas if the value of the learning-rate scalar is too high, predictions will fail to smooth out external noise (Doya, 2008):

$$w(t-1)_{new} = w(t-1)_{old} + \eta * \text{TD error}$$

where η is the learning rate, $w(t-1)_{old}$ is the current weight associated with the reward-related stimulus at time t-1, and $w(t-1)_{new}$ is the updated weight. Note that a scalar relationship between the magnitude of the TD error and the learning rate is assumed for simplicity sake, although it implies that, in theory, weights can grow to infinity. That

possibility is not biologically realistic, but since the TD error cancels itself away during learning, it is seldom perceived as problematic (see the discussion section for more details on concerns specific to our findings).

The weight changes driven by the error at time *t* are applied to the state of all reward-related stimuli at time *t-1, **x(t-1)* w(t-1)***. Thus, the weight changes propagate backward over successive trials to the earliest reliable predictor of reward delivery. In an associate conditioning paradigm implemented with a randomly varying ITI, the earliest reliable predictor is the onset of the conditioned stimulus. Thus, over the course of learning, the value function, *V(t)*, becomes non-zero from the time of reward delivery to the onset of the predictive cue, producing a long pulse-like value function as in seen in the middle and right columns in Figure 33. The same rule would apply to update weights associated with actions that previously generated rewards. Thus, this mechanism reinforces specific actions to maximize reward harvest.

Changes in DA firing rate correlate with TD errors (RPEs)

Montague, Dayan & Sejnowski (1996) were the first to formally apply a TDRL framework to explain the results of Schultz's experiments. The two last lines of Figure 33 show that RPEs correlate with the changes in DA burst firing observed in those experiments. The delivery of an unexpected juice reward before learning occurs (green shaded area in the left column) produces a large TD error and a corresponding burst of DA-neuron firing. Once the animal has been trained, the onset of the CS (gray shaded area in the middle column) produces a large TD error and a corresponding burst of DA-neuron firing. However, at the time of reward delivery (green shaded area in the middle column), the TD signal is nulled by the reward, and the DA firing rate remains at baseline. Omitting the reward after learning generates a negative TD error at the time when reward delivery is expected and a corresponding decrease in DA firing.

The striking resemblance between the predictions of the TDRL model and the firing of midbrain DA neurons led to the hypothesis that phasic DA activity reflects TD errors (mostly called RPEs in the neuroscientific literature) and adjusts synaptic weights to update reward predictions.

RPEs modulate actor-critic architecture

McClure, Daw & Montague (2003) highlighted that TDRL models implement RPEs not only to update prediction about the world but also to bias action selection, as was formally proposed by Witten in 1977; Barto, Sutton and Anderson in 1983; and Sutton in 1984; and later used to explain eICSS by Montague, Dayan & Sejnowski in 1996. They argue for a dual effect of RPEs that employs DA burst firing to incentivize the initiation of actions and to drive learning by means of reward predictions.

Additional variants of TD models to explain the role of DA-mediated RPEs in behavioural tasks continued to emerge (Barto, 1995; Houk, Adams & Barto, 1995; Egelman, Person & Montague, 1998; Doya, 1999; Niv, Daw, Joel et al., 2007) but the core concepts remained the same: RPEs drive learning within an actor-critic architecture. The error updates the value of predictive cues (the expected reward value) through the critic module *and* the proclivity to select the best reward-seeking action through the actor module. The idea that phasic DA firing is responsible for *both* types of weight adjustments is common in the literature and rarely presented as a point for debate (Steinberg et al., 2013; Colombo, 2014; Schultz, 2015; Keiflin & Janak, 2015; Chang et al., 2016; Chang et al., 2017; Chang et al., 2018; Parker et al., 2016; Watabe-Uchida, 2017; Nasser et al., 2017; Sharpe et al., 2017; Keiflin et al., 2019).

Limited evidence in support of a causal role of phasic DA activity in RPE-driven learning

There is a large and growing body of evidence demonstrating that phasic DA activity has the properties of RPEs. This interpretation is becoming a well-accepted fact, now rarely disputed in the field of neuroscience. There is less evidence that DA signalling causes learning, although this assumption is commonly made. Most results interpreted as confirmation that DA-mediated RPEs cause learning (Keiflin & Janak, 2015; Chang, Gardner, Gonzalez Di Tillio et al., 2018) were obtained in a series of recent blocking experiments. In such a paradigm, a cue, A, that fully predicts a reward is compounded with a neutral cue, B. The resulting cue, AB, is then repeatedly paired with the same reward. Cue B fails to acquire associative strength with the reward because cue A already suffices to predict the reward (Kamin, 1969). Steinberg et al. (2013) selectively stimulated midbrain DA neurons during the compound presentation and thereby unblocked associative learning between cue B and the reward. Although this experiment supports the DA-

RPE theory, it does not rule out the possibility that stimulating DA neurons could produce a perceptible difference in the properties of the delivered reward, and thus allow B to acquire predictive power. In the same study, sucrose alone or sucrose combined with the stimulation were evaluated in a preference test. Both options were equipreferred. Nevertheless, the lack of preference does not necessarily confirm that the two rewards, sucrose alone or sucrose with stimulation, were comparable on all dimensions. Based on Kamin's work (1969), changes in reward identity are as efficient in unblocking learning as are changes in intensity. Thus, it is possible that stimulation of midbrain DA neurons can enable learning without acting as an RPE; the stimulation simply needs to make the reward perceptibly different.

A stringent test of the DA-RPE hypothesis

The limited evidence suggesting a causal link between phasic DA signalling and learning is restricted to one main type of behavioural paradigm. In the current experiment, we applied a new approach using ICSS. More specifically, we tested the premise that the calibration of ICSS performance to the strength of the stimulation is due to the action of DA transients as RPEs. We used selective optical stimulation of midbrain DA neurons, careful manipulation of reward strength, quantitative behavioural measurement, and concurrent FSCV to test the predictions of the TDRL model.

Measurement of DA transients during oICSS

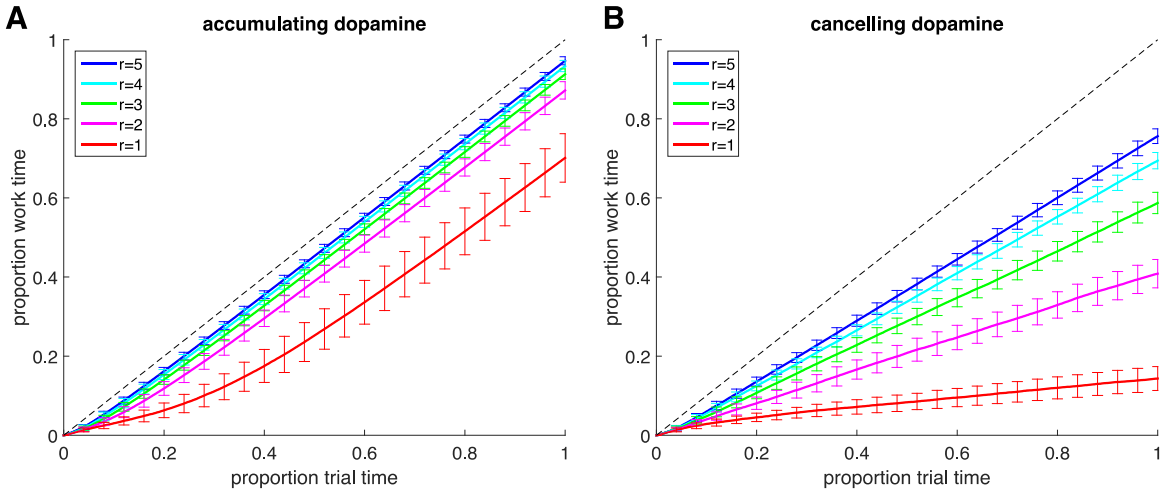
A stimulation-induced DA transient is expected at receipt of every optical train (Witten et al., 2011). Nonetheless, FSCV recordings have yet to be performed to monitor the persistence and amplitude of DA transients during lengthy sessions of oICSS such as the long oICSS sessions required to obtained detailed, quantitative behavioural measurements adequate to test the TDRL hypothesis. The first objective of our study was to monitor the persistence and amplitude of DA transients in this experimental context.

Persistent stimulation-induced DA transients is predicted to accelerate oICSS

A second objective of the study was to determine whether stable, non-maximal rates of oICSS can be obtained while persistent stimulation-induced DA transients are observed. According to the TDRL algorithm, weights must eventually saturate when repeated optical stimulation trains

persistently and reliably drive DA transients. Thus, *V(i,t)* and the resulting bias on action selection should attain asymptotic values. This implies that if the amplitude of the optically induced DA burst remains constant and non-zero over time, oICSS cannot be stable. Instead, the proclivity to work should grow larger with every weight change, and thus oICSS should accelerate throughout trials.

A key feature of the TDRL algorithm is that under stable circumstances, RPEs "predict themselves away." As the predictions become increasingly accurate, the errors diminish and eventually disappear (see the second trace from the bottom, middle panel in Figure 33). Figure 34A shows simulations done by Peter Dayan that contrast this standard scenario ("cancelling dopamine," right panel) with an alternative in which optically induced DA transients cannot be nulled by improving reward predictions ("accumulating dopamine," left panel). The proportion of work, normalized to the maximal work, is graphed as a function of trial time.



simulations made by Peter Dayan

Figure 34. Simulated ICSS for persistent and self-cancelling stimulation-induced RPEs

Simulated ICSS for persistent stimulation-induced RPEs (A) and for RPEs that predict themselves away (B). We graphed a cumulative ICSS trajectory for each trial, showing the cumulative proportion of work time performed as a function trial time (also expressed as a proportion). On the y-axis, the amount of work done from time 0 to time *t* is divided by the total work time for that specific trial. On the x-axis, the time *t* is the amount of time elapsed in the trial up to that point divided by the total trial time. The different curves are simulations obtained using different reward magnitudes (*r(t)*), ranging from 1 (smallest) to 5 (largest), in arbitrary reward units. The dashed gray line represents the trajectory produced by continuous work for the entire duration of the trial. The error bars are standard deviations. In A, the acceleration in work rate over the early portion of the trial is most apparent in the case of the smallest reward (r = 1, red), which produces the smallest weight changes. In B, cumulative work time grows linearly as a function of trial time, at a rate proportional to the reward magnitude. ICSS, intracranial self-stimulation; RPE, reward prediction error.

In the accumulating-dopamine scenario, even the reward produced by a low pulse frequency (red curve), arbitrarily set to R = 1, would result in small weight changes and accelerating oICSS as predicted above. In contrast, higher pulse frequencies would result in large weight changes and quickly elicit maximal oICSS. All curves terminate in straight lines with the same slope, which represents the maximum achievable work rate. The time required to attain this terminal slope decreases as the pulse frequency increases, and hence the reward intensity increases. Stable sub-maximal slopes are incompatible with the accumulating-dopamine scenario because these would imply that action weights could remain constant at non-saturating values in the face of persistent RPEs.

Figure 34A implies that weights are not saturated over the portion of the trial during which the rate of work is non-maximal. Therefore, the most stringent test using oICSS consists of

manipulating light pulse frequencies to obtain groups of trials in which animals work at different rates (maximal, non-maximal, and minimal) and test the prediction that persistent stimulation-induced DA transients, with stable amplitude, will produce accelerating rates of work. We demonstrate that this approach can detect violation of the DA-RPE hypothesis.

eICSS within the framework of the RPE-DA hypothesis

There is a fundamental difference between oICSS and eICSS: electrical stimulation produces DA transients predominantly through trans-synaptic activation (Yeomans, Kofman & McFarlane, 1985; Yeomans, 1989; Murray & Shizgal, 1994; Shizgal et al., 1980; Wise, 1980; Maeda & Mogenson, 1981) whereas optical stimulation directly activates DA neurons. In contrast with optical stimulation, the effects of electrical stimulation can be nulled by other inputs also converging onto DA cell bodies. This possibility has received some support from voltammetric recordings in behaving animals. A rapid decline in the amplitude of stimulation-induced DA transients was reported during short eICSS sessions leading to the disappearance of DA transients after 30 minutes (Garris, Kilpatrick, Bunin et al., 1999; unpublished data from our laboratory) or to small but persisting DA transients after earning roughly 50 rewards (Owesson-White, Cheer, Beyene et al., 2008).

The DA-RPE hypothesis makes a different prediction in the case of eICSS compared with oICSS. The nulling of DA-mediated RPEs during eICSS, but not oICSS, should result in the gradual disappearance of stimulation-induced DA transients and stable rates of work. Figure 34B shows a second simulation done by Peter Dayan assuming that DA-RPE nulls itself (i.e., the "cancelling-dopamine" scenario). No matter how high the pulse frequency of the stimulation, arbitrarily set to R = {1,2,3,4,5}, eICSS grows linearly, but at a rate proportional to the reward magnitude.

TD algorithms are built to implement learning through trial and error, a type of learning also referred to as model-free. However, a large body of literature demonstrates that animals can create mental maps ("world models") to guide their navigation through the environment and to maximize the harvest of rewards (Dayan & Niv, 2008; Doll, Simon & Daw, 2012). This type of learning is called "model-based." Doll, Simon & Daw (2012) have proposed that there is a

continuum of learning forms ranging from model-free at one extreme to model-based at the other. We constructed ICSS sessions using a mix of trial types configured to determine whether the adjustment of performance as a function of changes in stimulation strength reflect model-free learning or a more complex form of learning on the continuum toward model-based learning. Due to the limited number of ICSS sessions with simultaneous FSCV measurements, however, training was discontinued before it could be determined whether a form of learning other than the model-free variant had emerged. Therefore, the approach presented in this chapter focuses exclusively on the DA-RPE hypothesis in the context of model-free learning.

Methods

Subjects

Long Evans TH::Cre male rats (n=2; rat Elop7 trained for eICSS; rat Elop1 trained for oICSS) were bred in-house and maintained on a 12-hour reverse light/dark cycle (lights off from 08:00 to 20:00) with ad libitum access to food and water. After surgery, the animals were housed individually. The experimental procedures were performed in accordance with the principles outlined by the Canadian Council on Animal Care.

First surgery: virus infection and implantation of optical implants and stimulation electrode

Viral transfection

Rats were anesthetized with a mixture of ketamine hydrochloride (87 mg/kg) and xylazine hydrochloride (13 mg/kg) (injected i.p.). They received a s.c. injection of atropine sulphate (0.05 mg/kg) to reduce bronchial secretions during surgery, a s.c. injection of buprenorphine to alleviate post-operative pain (0.05 mg/kg), and a s.c. injection of penicillin procaine to prevent infection (0.3 cc/rat). Burr holes were made to provide access for lowering the injector in the VTA.

We transfected VTA DA neurons in TH::Cre males with a serotype-5 adeno-associated virus (AAV5) containing an EF1a-DIO-hChR2(H134R)-EYFP-WPRE transcript suspended in phosphate buffered saline. Animals received six injections of 0.5 µL in each hemisphere (AP: 5.4, ML: 1.0, DV: 8.3, 7.7, 7.2; AP: 6.2, ML: 1.0, DV: 8.3, 7.7, 7.2). We injected at 0.1 µL per min via a Hamilton syringe connected to a Harvard pump, waiting 10 mins between injections.

Implantation of optical implants

An optical implant was lowered into each hemisphere at a 10-degree angle along the medial-lateral axis and aimed at the following coordinates: AP 5.8, ML 2.2 and DV 8.12. The optical implants were fashioned from a 300-micron core optical fiber, stripped of its cladding. One end was inserted in a stainless alloy ferrule, glued in place and polished. The optical implants were secured with dental acrylic and anchored with jeweller screws.

Implantation of stimulation electrode

A stimulating electrode was fashioned from a 0.25 mm stainless steel insect pin and insulated with Formvar enamel to within 0.5 mm of the tip. The electrode was aimed at the right MFB at the level of the LH (AP: −2.8 mm; ML: 1.7 mm; DV: 8.9 mm from the skull; all coordinates are referenced to bregma). Once tacked into place with dental acrylic, the stimulating-electrode connector was secured to the head cap with a small strip of removable adhesive. The anode of the stimulation circuit was lowered during a subsequent surgery. Delaying the implantation of the anode prevented damage to the skull where the chronic carbon-fiber microsensors would ultimately be implanted.

The animal's temperature was controlled throughout the surgical procedures with a heating pad. An s.c. injection of Ringer's solution (6 mL/kg) was given every hour.

Second surgery: implantation of the stimulation anode and chronic carbon-fiber microsensors

Implantation of carbon-fiber microsensors

Rats were anesthetized and prepared following the same procedure as in the first surgery. A burr hole was made in the skull to provide access for lowering the anode into the brain. The anode was fashioned in the same way as the stimulating electrode with the exception that the stainless-steel insect pin was insulated with Formvar enamel to within 3 mm instead of 0.5 mm of the tip. The electrode was placed in the external plexiform layer of the olfactory bulb (AP: 6.2 mm; ML: −0.1 mm; DV: 5.0 mm from the skull; all coordinates are referenced to bregma). This location was chosen to ensure the recording site was roughly mid-way between the cathode and anode of the stimulation circuit. At the mid-way position, the recording site should lie on or near a zero-potential surface, thus reducing interference between the electrical stimulation and the voltammetric measurements. The electrode assembly was secured with dental acrylic and anchored with jeweller screws.

Implantation of chronic carbon-fiber microsensors and reference electrode

Burr holes were made bilaterally above the NAc and the carbon-fiber microsensors were lowered, one at the time. The microsensors were fashioned as described by Clark, SandBerg, Wanat et al. (2010). One carbon fiber was pushed through a section of silica tubing while immersed in alcohol and was subsequently allowed to dry. Then, the microsensor was sealed with resin epoxy and fixed to a connector with silver epoxy. The sensors were pre-cleaned with a solution of 2-propanol containing activated carbon, and slowly positioned in the NAc shell (AP: 1.7 mm; ML: 1.0 mm; DV: 7 mm from the skull). The sensor was advanced slowly, by 0.2 mm every 30 s, to minimize damage.

An Ag/AgCl reference electrode was fashioned from a silver wire secured with silver epoxy to a gold-plated connector and soaked in a 10% solution of sodium hypochlorite for 60 mins. Its final conductive surface was approximately 5 mm. The reference electrode was placed in the left hemisphere in the remaining space posterior to the head cap built during the first surgery. The location was roughly AP: −9 mm; ML: 4 mm; DV 5 mm from bregma.

Behavioural training

Behavioural paradigm to test the DA-RPE hypothesis

To test the DA-RPE hypothesis in the most stringent way using ICSS, we ensured that our behavioural paradigm would produce non-maximal rates of work. In this way, work rates were free to increase, decrease, or remain stable. Recall that the DA-RPE hypothesis predicts accelerating (i.e., non-stable) work rates in the oICSS paradigm but is compatible with stable, non-maximal work rates in the eICSS paradigm.

Animals were presented with a lever that delivered one of three stimulation trains that differed in pulse frequency. The high frequency produced maximal ICSS, the low frequency produced near non-zero rates, and the medium frequency produced intermediate, non-maximal rates. Note that only the pulse frequency varied, all other stimulation parameters remained constant.

Trials

Each trial consisted of a fixed time during which the stimulation parameters were held constant. Animals had to hold the lever down for a total of 2 secs (referred to as price) by means of a single hold or multiple shorter holds interrupted by breaks. The stimulation was delivered when the cumulative hold time equaled the price. After completion of that work requirement, the lever was retracted for 8 secs, a variable referred to as the black-out delay. All trials lasted 240 secs, excluding the black-out delays. A maximum of 120 stimulation trains could be earned per trial.

A computer-controlled digital pulse generator was used to set the pulse frequency of the stimulation earned after each completion of that work requirement. A Master-8 pulse generator (A.M.P.I., Jerusalem, Israel) was used to generate biphasic electrical stimulation pulses and to control their amplitude and duration. The voltage output of this device was routed through an analog stimulus isolator unit (A-M Systems, Sequim, Washington), which was connected to the stimulation electrode and current return via an electrical rotary joint (Crist Instrument Co., Hagerstown, MD). To generate optical pulse trains, the Master-8 pulse generator was connected to a 473 nm DPSS laser (Laserglow, Toronto, Ontario); the output of the laser was routed to the chronically implanted optical probe via an optical rotary joint (Doric lenses Inc, Quebec, Quebec) and a custom-built optical patch cord (Trujillo-Pisanty, Sanio, Chaudhri et al. 2015). A custom-written computer program ("PREF", Steve Cabilio, Concordia University, Montreal, QC, Canada) controlled the experiments and logged the data.

Random-world lite

Trials were arranged in triads. The first and last trials of a triad are called leading and trailing trials. The high pulse frequency was always on offer during the leading trial, whereas the low pulse frequency was always on offer during the trailing trial. These trials flanked each test trial on which either a high (HI), medium (MED), or low (LOW) pulse frequency was on offer (Figure 35). Each of these three pulse frequencies were presented three times during a session, generating nine triads and a total of 27 trials. The triads followed a new randomized sequence in each session. The lever for the leading and trailing trials extended from one wall of the cage and the lever for the test trials extended from the opposite wall.

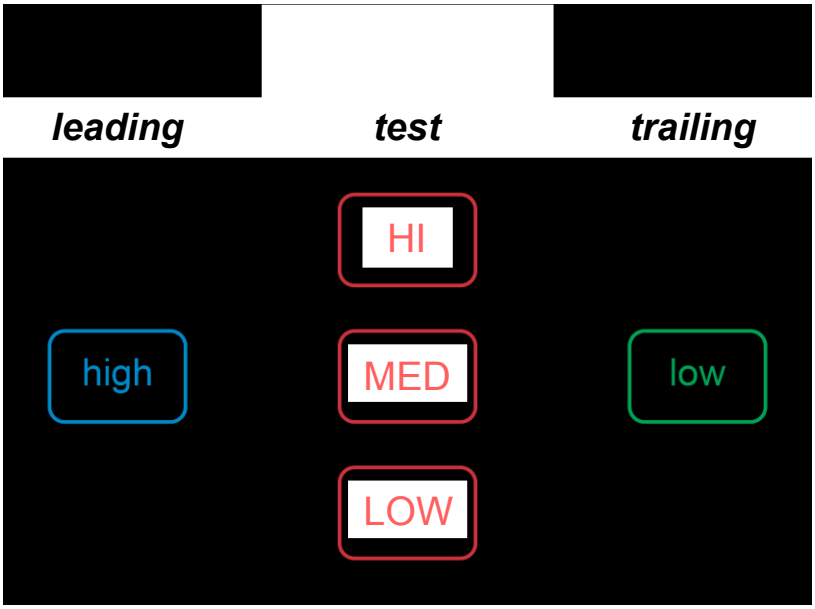


Figure 35. Trial structure in "random-world lite"

Trials were arranged in triads. The leading and trailing trials are shown in blue and green, respectively. The high (HI) pulse frequency was always on offer during the leading trial and the low (LOW) pulse frequency was always on offer during the trailing trial. In contrast to these fixed conditions, the pulse frequency offered on the test trials (shown in red) was drawn pseudorandomly (indicated on the diagram as p = 0.33) from a set of three values containing the HI and LOW pulse frequencies used on the leading and trailing trials as well as an intermediate, medium (MED) frequency. The lower arrow denotes the ITI of variable duration with a random component. Note that the trial structure was designed to examine the effects of persistent stimulation-induced DA transients on different forms of learning arrayed along the continuum from model-free to model-based. However, experimental

constraints restricted the focus to model-free learning (see Chapter 3 Methods section for more details). ITI, inter-trial interval; DA, dopamine.

TD algorithms are built to implement model-free learning. However, complex environments can also promote model-based learning: the construction of mental maps can guide behaviour and maximize the harvest of reward (Dayan & Niv, 2008; Doll, Simon & Daw, 2012). The trial structure was designed to determine whether the behaviour that emerged during extensive training was model-free or whether it was positioned along the continuum that extends between model-free and model-based forms.

Data acquired previously from a different group of subjects demonstrated that model-free learning was not used exclusively in the acquisition of our task. Animals anticipated the beginning of the trials on which high pulse frequencies were on offer (HI leading trials, see below), orienting toward the location where the lever would extend. At the beginning of the trials on which low pulse frequencies were on offer, they often ignored the lever (LOW trailing trials, see below). These behaviours were manifested after a trial transition prior to receipt of reward. However, due to the limited number of ICSS sessions with simultaneous FSCV measurements, training was discontinued before rats Elop1 and Elop7 showed evidence of having learned the triadic trial structure. They sampled the lever with essentially the same latency for all trials. Consequently, this chapter focuses on model-free learning.

Test chamber and shaping

Rat Elop7 was shaped to lever press for a 500-ms train of cathodal, rectangular, constant-current pulses with a pulse duration of 0.1 ms. Rat Elop1 was shaped to lever press for a 1000-ms optical train of rectangular pulses with a pulse duration of 5 ms.

Shaping took place in standard modular operant chamber (Med Associates Inc.), 30 cm long × 25-cm wide × 19-cm high, equipped with one retractable lever and a cue light positioned 1.5 cm above the lever. Illumination of the cue light signals that the lever press has been registered. The duration of the ITI was drawn from a lagged exponential distribution. During the ITI, the house light flashed continuously.

The electrical and optical rotary joints were both positioned above the chamber, with the optical fiber passing through a central lumen of the electrical rotary joint. The electrical rotary joint connected the head-mounted voltammetric amplifier to the data acquisition system.

Training

Rats were trained on a cumulative handling-time reinforcement schedule (Breton, Marcus & Shizgal, 2009). On such a schedule, the stimulation is delivered after the lever has been depressed for a predetermined, cumulative time, called the "price" of the stimulation train. Animals can invest time in holding the lever down intermittently or satisfy the price requirement with a single lever press.

Initially rats were trained with a price set to 0.5 secs and a black-out delay set to 2 secs. These values were increased incrementally to reach a price of 2 secs and a black-out delay of 8 secs. Previous experiments in our laboratory typically used a black-out delay of 2 secs. We extended that duration to an average value of 8 secs to prevent the superimposition of DA transients evoked by stimulation trains in close temporal proximity, and we made the duration of the ITI unpredictable to render the lever extension the earliest reliable predictor of reward availability. During this training phase, a combination of high pulse frequency and high power or current was used (optical stimulation: 56 Hz, 60 mW; electrical stimulation: 130 Hz, 450μA). We monitored the rats' behaviour for signs of aversion, but no such sign were observed. Once steady ICSS was obtained over three consecutive days using the longest duration price and black-out delay, animals were introduced to a trial structure that we termed "random-world lite."

On the first day of "random-world lite", animals were shaped to alternate between levers on leading, test and trailing trials and to continue sampling the lever while transitioning out of LOW trailing trials. Acquisition of voltammetric data began immediately because FSCV electrodes are known to degrade over time. After three consecutive days of early training, the stimulation parameters were adjusted, if needed, to ensure that the medium pulse frequency produced non-maximal rates of work. We tried to collect as many sessions with simultaneous FSCV and ICSS observations as possible.

FSCV recordings

Fabrication of FSCV microsensors

FSCV electrodes were fabricated for chronic implantation following the procedure described by Clark, SandBerg, Wanat et al. (2010). A carbon-fiber electrode for acute voltammetric recording is sealed by inserting it in a glass capillary tube, heating until the glass becomes viscous, and applying a longitudinal tensile force until the glass coats the fiber. In contrast, the procedure introduced by Clark et al. for fabricating electrodes for chronic voltammetric recording entails sealing the carbon fiber within the silica tube by means of an epoxy dome. The latter strategy results in a more uniform fabrication process. To ensure the seals on all our FSCV microsensors were consistent, we took considerable care to inspect them visually by means of a stereomicroscope. The length of exposed carbon fiber at the tip was also carefully assessed and adjusted. Consequently, the variation in currents obtained as a function of holding potentials spanned by each scan—the Faraday current—varied little across microsensors.

Electrochemistry

Cyclic voltammograms were generated at 10 Hz by applying an 8.5 ms triangular waveform that ramped from -0.4 V to +1.3V and back to -0.4 V at a scan rate of 400 V/s. The potential was held at -0.4 V between scans to promote cation absorption at the surface of the FSCV electrode. All potentials were measured with respect to the Ag/AgCl reference electrode. The waveform was generated using LabVIEW (National Instruments, Austin, TX) and a multifunction data acquisition board (PCI-6052E, National Instruments, Austin, TX).

A PCI-6711E (National Instruments, Austin, TX) board was used to perform waveform acquisition and data collection. A synchronization signal from the PCI-6711E board was sent to the external input of the Master-8 pulse generator and used to trigger the electrical stimulation 5 secs after the start of each recording. The stimulation was patterned to prevent overlap with the voltammetric scans. This was accomplished by confining pulse generation to the 91.5 ms intervals separating the triangular waves.

Principal component regression (PCR) and principal component analysis (PCA)

A background subtraction is typically performed as the first step before visualizing voltammograms or quantifying DA concentrations in FSCV recordings. This data transformation is intended to isolate the current produced by the oxidation and reduction of DA and DA-O-quinone, as well as to remove the slow drift of the Faraday current over time.

Background subtraction can be problematic because of unavoidable arbitrariness in determining what constitutes the background. A recent paper by Kishida, Saez, Lorenhz et al. (2015), demonstrated that the pertinent information necessary to extract DA concentrations from FSCV recordings can be obtained without resorting to background subtraction. They also found that even if maximal DA oxidation occurs at roughly 0.6 V and maximal reduction at -0.2 V, voltages recorded at additional time points during the scan have predictive power for detecting changes in phasic DA. We originally intended to follow the method proposed by Kishida et al., but given the structure of our experiment, their *in vitro* approach to acquiring calibration data was not applicable. Whereas the Faraday currents registered by freshly implanted FSCV microsensors in an intact brain are similar to those observed *in vitro* (as was the case in the experiment by Kishida et al.), the Faraday currents registered by chronically implanted microsensors are more complex than those observed *in vitro,* and they evolve over time as the FSCV electrode interacts with the brain environment and is etched by continuous usage (Keithley, Takmakov, Bucher et al., 2011; Rodeberg, Johnson, Cameron et al., 2015). Consequently, we obtained our calibration data from measurements acquired in awake animals presented with unexpected stimulation trains (see the section on training sets for details).

To perform an elastic-net regression to model changes in DA concentration according to the method of Kishida et al., all 1000 voltages swept during each scan are entered as predictors. *In vitro* measurements can yield thousands of usable cyclic voltammograms in a few minutes, thus providing sufficient data for such a regression (see Kishida et al., 2015 for more detailed information). It is not feasible to obtain an equivalent dataset by delivering unexpected stimulation trains to behaving rats. The temporal density of the stimulation trains would have to be much lower, and hence this would take an inordinate amount of time. Moreover, conditions would most certainly evolve over the course of such a long acquisition period, rendering

observations obtained at different time points incomparable. Therefore, we retained the PCR-PCA methods used in most voltammetry experiments (Rodeberg et al., 2015) but omitted background subtraction, thus retaining all informative parts of the voltammograms. This allowed us to benefit from the increased predictive power of the approach developed by Kishida et al., while circumventing the distortions that can arise from background subtraction and minimizing the number of unexpected stimulation trains delivered to acquire the training set.

Training sets

The ideal way to obtain voltammetric data associated with known concentration of DA and pH is to introduce FSCV microsensors in a flow cell and expose them to standard solutions made from synthetic DA, with the pH adjusted using NaOH and HCl. However, whereas the Faraday currents of freshly implanted FSCV microsensors in an intact brain are similar to the Faraday currents observed *in vitro,* chronically implanted microsensors register much more complex Faraday currents that cannot be fully replicated *in vitro.* To confirm the training sets were adequate, we determined whether the *in vitro* training sets could predict the amplitude of DA transients in recordings acquired in awake animals that received unexpected optical stimulations. The results were analysed manually using the TarHeel software supplied with the FSCV system. The best model did predict small simulation-induced DA transients, but the sensitivity was below acceptable levels; only changes produced with 40 and 56 Hz were detectable, even if manual inspection of voltammograms clearly showed detectable DA transients for pulse frequencies as low as 10 Hz. Thus, an alternate strategy was required.

Instead of employing *in-vitro* training sets, we used cyclic voltammograms from data acquired in awake animals. Six animals were presented with unexpected, experimenter-triggered optical pulse trains delivered at different pulse frequencies (light power: 40 mW; pulse frequencies: 5, 10, 20, 40, and 56 Hz). The selection of cyclic voltammograms was based on their shapes; known features of the cyclic voltammograms known to occur when FSCV microsensors are exposed to rapid changes in DA or pH (Keithley & Wightman, 2011). Due to the high degree of similarity across microsensors, data from all six animals were combined in the same training set.

As a result, we constructed one matrix with cyclic voltammograms (Matrix **A**: n x 1000; number of samples by number of voltages tested for each scan) and one matrix with the corresponding change in concentration of DA and pH (Matrix **C**: 2 x n; number of analytes by number of samples).

Multivariate regression

To build a linear model to predict unknown DA concentrations and unknown pH from the cyclic voltammograms associated with known DA concentrations and known pH, we solved for the inverse of the Beer-Lambert law:

$$\mathbf{C} = \mathbf{PA}$$

where **A** is the matrix containing the voltammograms, **C** is the matrix containing the known concentration associated with these voltammograms for both analytes, and P is the coefficient matrix

Before conducting the regression, we used PCA to reduce the dimensionality of matrix A. This step lowers the number of samples of known concentration required to obtain a statistically reliable model. Because each voltammogram consists of 1000 data points, solving **C** = **PA** would require several thousand scans with known concentrations. Such numbers can be reached with *in vitro* calibration data, but obtaining that large a number in awake animals is not feasible. Principal components were calculated using singular value decomposition in MATLAB. A scree plot was used to determine the number of components that would account for at least 90% of the variance in matrix **A**. Fifteen components were retained. We expressed the voltammograms as a function of the 15 dimensions determined by PCA to obtain $\mathbf{A_{proj}}$.

We solved for **P** using the least squares method:

$$\mathbf{P} = \mathbf{C}\mathbf{A_{proj}}^{t}[\mathbf{A_{proj}}\mathbf{A_{proj}}^{t}]$$

where A_{proj}^{t} is the transposition of matrix A_{proj}

The coefficient matrix **P** could then be used to predict the unknown concentrations associated with the voltammograms acquired during ICSS.

$$\mathbf{C}_{\mathbf{unknown}} = \mathbf{P}^{*}\mathbf{D}_{\mathbf{proj}}$$

where $\mathbf{D}_{\mathbf{proj}}$ is the data expressed as a function of the 15 dimensions determined by PCA and $\mathbf{C}_{\mathbf{unknown}}$ is a matrix predicting DA concentration and pH during ICSS

The number of components retained is much more than the usual two or three recommended when data are background-subtracted (Heien, Johnson & Wightman, 2004; Rodeberg et al., 2015). However, it is consistent with the observation of Kishida et al., that voltages obtained at multiple time points during the scan have predictive power to detect and quantify DA transients.

Results

Testing the DA-RPE hypothesis predictions with oICSS and eICSS

We tested two predictions of the DA-RPE hypothesis: 1) Non-maximal work rates imply that weights on action selection are not saturated. Thus, persistent simulation-induced DA transients must lead to accelerating **oICSS**. 2) Stable work rates during **eICSS** imply that the amplitude of evoked DA transients will be zero. We compared DA transients and ICSS elicited by stimulation trains by means of simultaneous accumbal voltammetry and behavioural measures in two animals: Elop1 worked for optical stimulation of midbrain DA neurons and Elop7 worked for electrical stimulation of the MFB at the level of the LH. We examined the persistence and stability in the amplitude of DA transients as animals earned multiple optical and electrical stimulations.

FSCV measurements were stable over several long recording sessions

The tip of the optical implant in rat Elop1 was positioned just above the VTA, whereas the tip of the stimulating electrode in rat Elop7 was within the MFB at the level of the LH. The tip of the FSCV electrodes in both rats was within the NAc shell (see Figure 36).

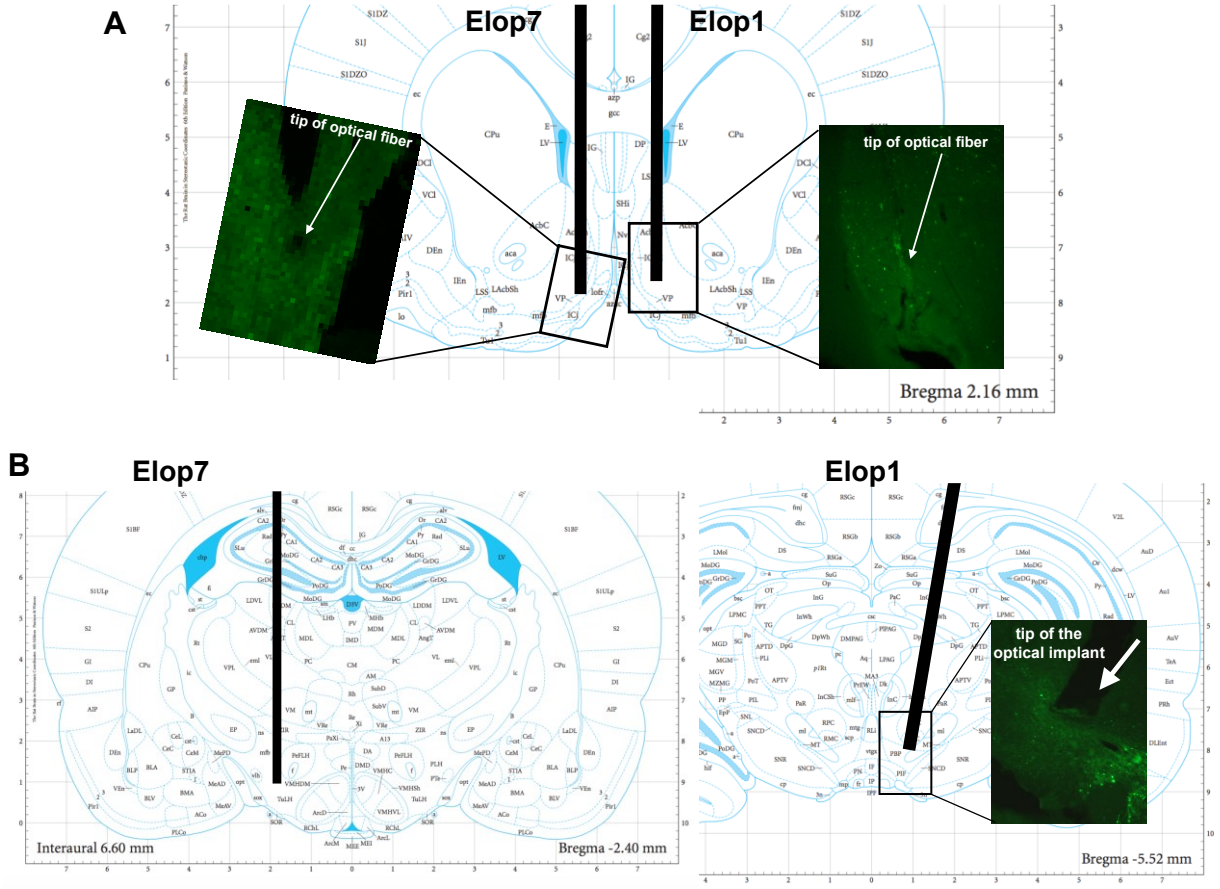


Figure 36. Histological localization of FSCV recording sites and electrical or optical stimulation

Location of A) the FSCV electrodes aimed at the NAc and B) the stimulation electrode aimed at the LH (Elop7: black line in the left diagram) or the optical implant aimed at the VTA (Elop1: black line in the right diagram). ChR2-EYFP expression is shown in green in the three insets. The tip of the track left by the FSCV electrodes and the optical implant are indicated with white arrows. Note that both animals were infected with a virus bearing a ChR2 transcript and implanted with both a stimulating electrode and an optical implant. However, this chapter reports only eICSS data from rat Elop7 and oICSS data from rat Elop1. FSCV, fast scan cyclic voltammetry; LH, lateral hypothalamus; VTA, ventral tegmental area; ChR2, channelrhodopsin-2; EYFP, yellow fluorescent protein; eICSS, electrical intracranial self-stimulation; oICSS, optical intracranial self-stimulation.

We carefully tracked the sensitivity and performance of the FSCV electrodes throughout the experiment. This was important because the surface of the FSCV electrode is etched with every scan acquired (Keithley et al., 2011), limiting the number of reliable recording sessions. Although chronically implanted FSCV electrodes have been used successfully to detect simulation-induced DA transients over days, and even months (Clark et al., 2010), the etching problem is particularly serious in the present experiment due to the very long duration of the testing sessions.

In each session, we delivered an unexpected, high pulse-frequency stimulation train to assess the sensitivity of the FSCV electrode. The peak amplitude of the resulting DA transient was used to normalize data within each session (Figure 37). The amplitudes of the DA transients induced by the unsignalled high pulse-frequency trains ranged from 1 to 10 nA and 3 to 10 nA for the optical and electrical stimulation, respectively. Multiple usable data sets were acquired for both stimulation types.

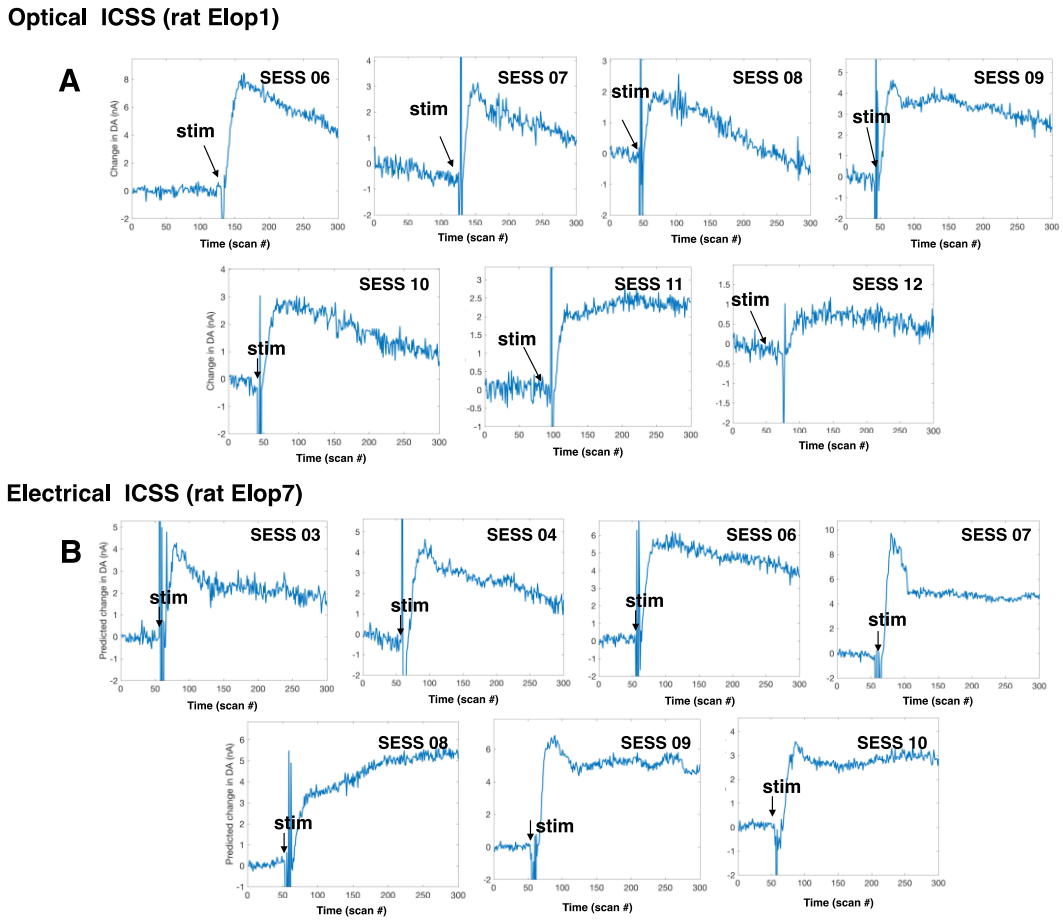


Figure 37. Reliability of FSCV signals over several ICSS sessions

The amplitude of DA transients in response to unexpected delivery of electrical stimulation was used to assess changes in the sensitivity of FSCV electrodes across sessions. The responses show the DA concentration achieved during a stimulation-induced DA transient as a function of time (A: Data from rat Elop1; B: Data from Elop7). Prior to normalization, this concentration was estimated from the DA oxidation current recorded at 0.6 V on the ascending limb of the triangular wave. The responses to unexpected delivery of a stimulation train were recorded prior to the start of an ICSS session. Black arrows mark stimulation delivery. The stimulation frequency was 130 Hz at a current of 400 µA, a pulse width of 0.1 ms and a train duration of 500 ms for the electrical stimulation delivered to Elop7. Note the amplitude of the traces was normalized by expressing it as a proportion of the average

of the 10 time points prior to stimulation delivery. FSCV, fast scan cyclic voltammetry; ICSS, intracranial self-stimulation; DA, dopamine.

Despite inter-session variability in the shape of the DA transients following the unexpected stimulation, a typical sharp rise in DA release was observed in all voltammograms. Noteworthy was the systematic absence of return to pre-stimulation baseline in the voltammetric traces, even after several secs. A quick return to baseline was also absent from records obtained following the first stimulation earned in HI leading, HI test, and MED test trials, as described in the next section. Overall, FSCV measurements were stable over several long recording sessions.

Persistent DA transients following optical and electrical stimulations

ChR2-expressing midbrain DA neurons have been shown to respond to low-frequency optical pulse trains with high spike fidelity (Tsai, Zhang, Adamantidis, 2009). Not surprisingly, DA transients have been detected reliably by means of FSCV in terminal regions in response to such stimulation. Nonetheless, in no previous study were DA transients measured over the course of oICSS sessions lasting as long as those reported here. Thus, it was not known in advance whether reliable stimulation-induced DA release would continue throughout the test session.

In the present study, persistent simulation-induced DA transients of roughly stable amplitude were indeed observed throughout HI leading, HI test, and MED test trials for both stimulation types, following an early decline. This pattern held true when all voltammograms were pooled by trial type and the median changes were presented in 3-D plots (Figures 38A and 39A), or when single trials were examined with raster plots, even in trials extracted at the end of the session after several hours of ICSS (Figures 38B and 39B).

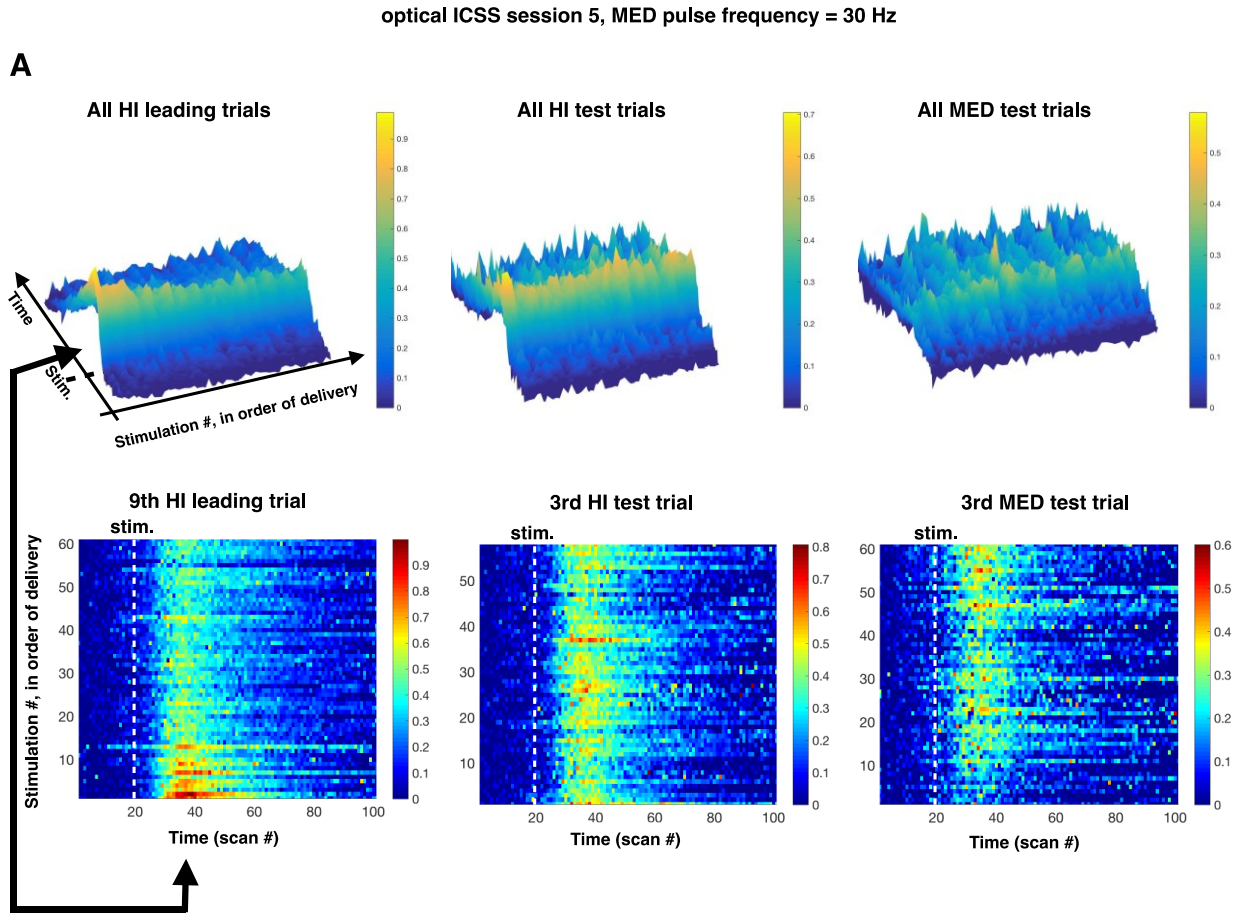


Figure 38. Persistent stimulation-induced DA transients for HI leading, HI test, and MED test trials during oICSS (rat Elop1)

A) Stimulation-induced DA transients are presented for HI (high) leading trials (n = 9) in the left column, HI test trials (n = 3) in the middle column, and MED (medium) test trials (n = 3) in the right column. The 3-D colour plots show the median values (within trial types) of successive stimulation-induced DA transients. The estimated DA concentration is represented in false colour along the z-axis, normalized first to the response to unexpected delivery of a stimulation train (Figure 37) and then to the 10 time points preceding delivery of each response-contingent stimulation train. Time runs along the y-axis. The DA transients recorded in response to each successive stimulation train (z vs. y plot) are stacked in the order they were earned, from first to last, along the x-axis (stimulation number). The time of stimulation delivery is indicated by the short, black dashed line intersecting the y axis ("Stim.").

The smallest number of stimulations earned on a given trial within trial types was used as an exclusion cutoff; any stimulation earned beyond that value was excluded to base the medians on equal numbers of observations (HI leading: n = 9; HI test: n = 3; MED test: n = 3). The DA concentration increases gradually at the start of the trial, as documented in Figures 40 and 41. This gradual increase is omitted here by excluding the responses to the first five stimulation trains. B) The bottom colour plots are 2-D representations of the 3-D plots in A. Stacked voltammetric traces are aligned along the x-axis and ordered by stimulation number along the y-axis. This reorientation is signified by the heavy line terminating in double arrow heads on the left of the figure. The white dotted lines on the 3-D colour plots mark stimulation delivery. The pulse frequency was 56 Hz on the HI leading or test trials and 30 Hz on the MED test trials. The optical power was 50 mW, the pulse duration was 5 ms, and the train duration was 1000 ms. DA, dopamine; oICSS, optical intracranial self-stimulation.

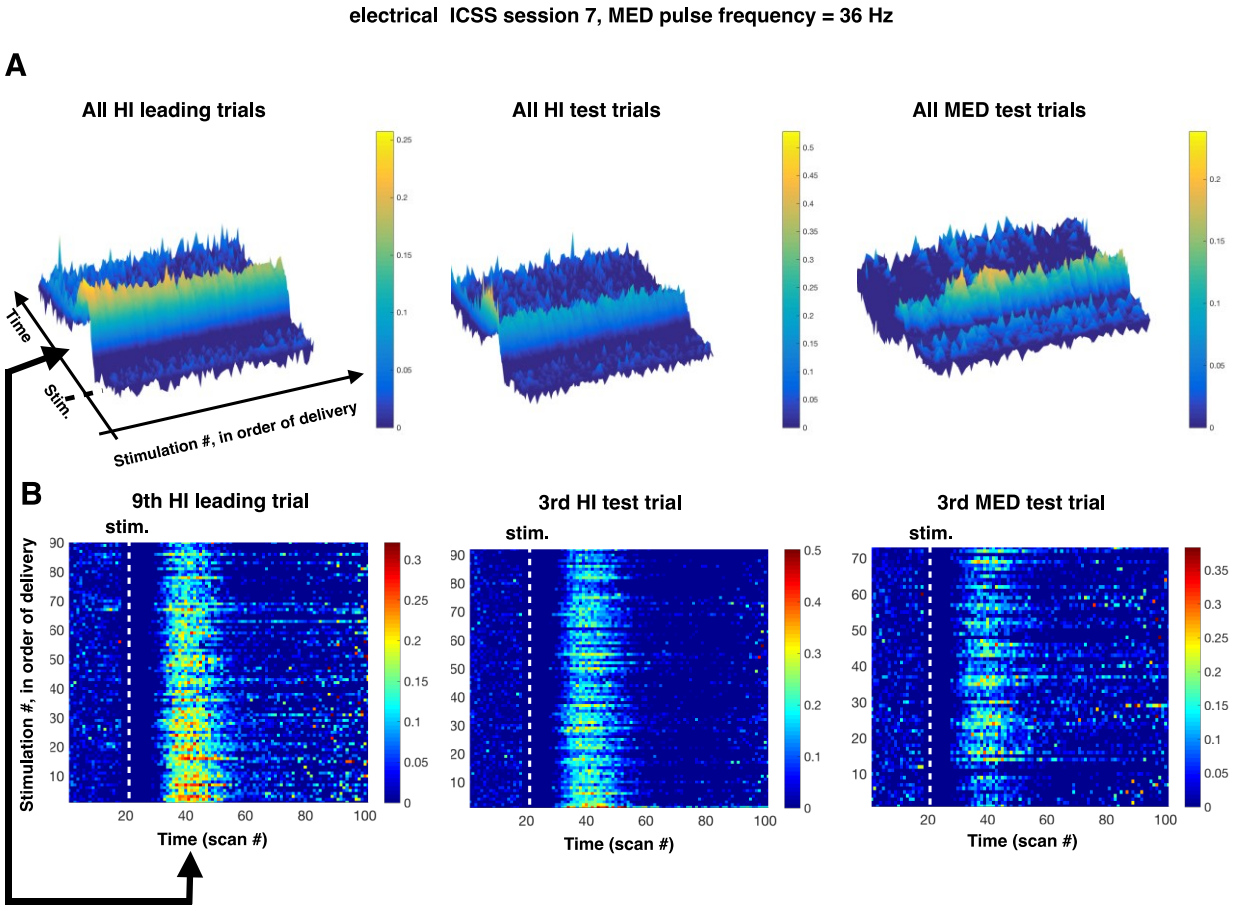


Figure 39. Persistent stimulation-induced DA transients for HI leading, HI test, and MED test trials during eICSS (rat Elop7)

This figure shows FSCV data acquired during eICSS sessions conducted with rat Elop7 and shares the same format as Figure 38. The pulse frequency was 135 Hz on the HI (high) leading or test trials and 36 Hz on the MED (medium) test trials. The stimulation current was 400 µA, the pulse duration was 0.1 ms, and the train duration was 500 ms. DA, dopamine; eICSS, electrical intracranial self-stimulation; FSCV, fast scan cyclic voltammetry.

oICSS failed to produce systematic simulation-induced DA transients after session 8. In sessions acquired past this point, persistent DA transients disappeared even if unexpected stimulation delivery produced a detectable DA transient (confirming adequate sensitivity of the FSCV electrode) (Figure 40A). In contrast, eICSS produced detectable simulation-induced DA

transients in all sessions in which delivery of an unexpected train evoked a detectable DA transients (Figure 39B).

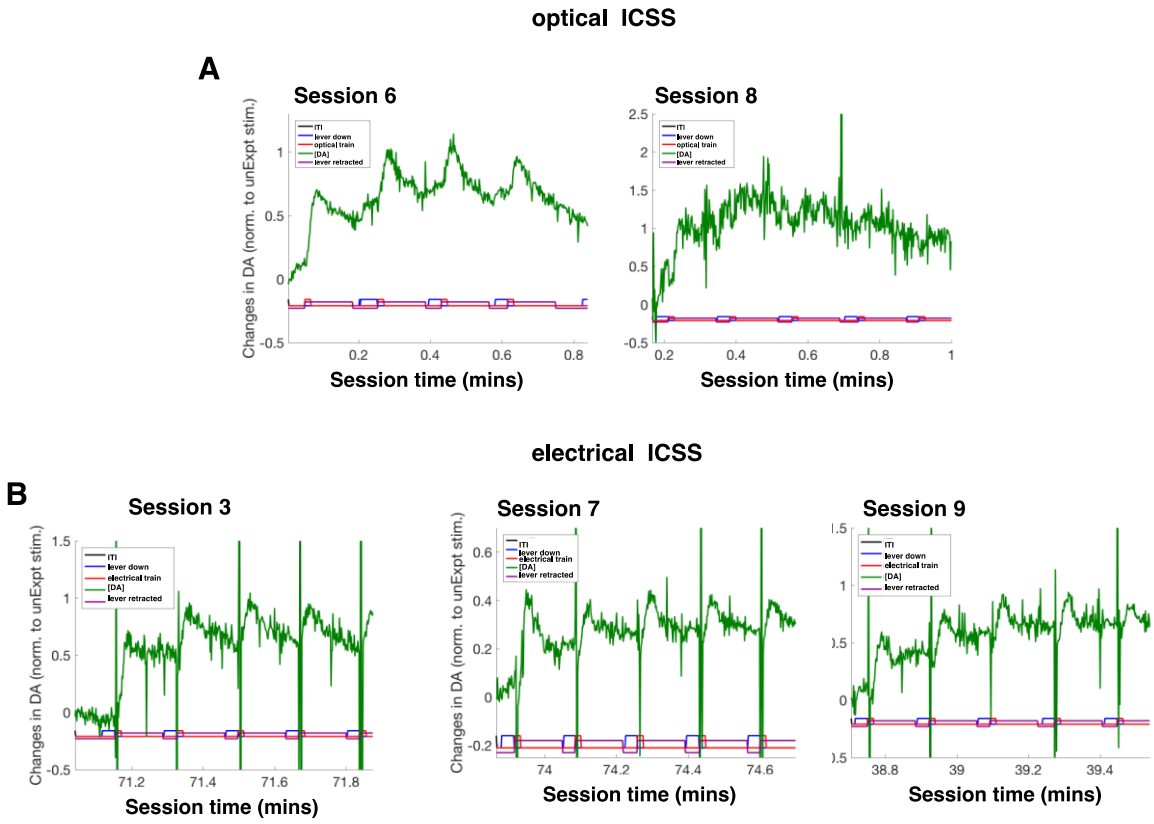


Figure 40. Examples of DA transients evoked by stimulation trains

DA concentration (green trace) estimated from the DA oxidation current recorded at 0.6 V on the ascending limb of the triangle wave. The concentration estimate was normalized first to the peak value obtained in response to an unexpected stimulation train (see Figure 37) and then to the first 10 time points on the x-axis. Panel A shows oICSS data from sessions 6 and 8, whereas panel B shows eICSS data from sessions 3, 7, and 9. Each graph shows the time course of the estimated DA concentration over the initial portion of a HI (high) leading trial. The lines at the bottom of the graphs designate lever presses (blue), stimulation delivery (red), and blackout periods when the lever was retracted (purple). In A, the pulse frequency was 56 Hz, the optical power was 50 mW, the pulse duration was 5 ms, and the train duration was

1000 ms. In B, the pulse frequency was 135 Hz, the stimulation current was 400 μA, the pulse duration was 0.1 ms, and the train duration was 500 ms. DA, dopamine; oICSS, optical intracranial self-stimulation; eICSS, electrical intracranial self-stimulation.

The amplitude of DA transients during eICSS was previously reported to decline rapidly and disappear within a few minutes (Garris et al., 1999; Kilpatrick, Rooney, Michael et al., 2000) or persist but decline over several minutes when a cued, variable time-out, fixed-ratio 1 reinforcement schedule was employed (Owesson-White et al., 2008). In contrast, during our HI leading and HI test trials, rat Elop7 worked at a maximal rate to earn high pulse frequency stimulation for roughly 15 mins and the amplitude of the stimulation-evoked DA transients was essentially stable (Figures 38 and 39).

Careful examination of the voltammetric traces captured during the first 50 secs of each trial (examples in Figure 40) revealed a gradual increase in DA concentration that began in response to the first stimulation and leveled off after several seconds. The DA response to the stimulation trains are best described as a superimposition of sharp, phasic changes onto an increasing baseline that sometimes built gradually (example) and in some cases increased in roughly a step-wise manner.

To ensure that the gradual increase in DA was caused by the delivery of the first train, we extracted trials with long initial response latencies. The voltammetric traces for these scarce trials were flat or slightly declining until the animal completed the work requirement and was granted the first stimulation train (Figure 41). Note in the lower right panel that each of the three DA transients reached a similar peak value. This fact would have been obscured had the DA concentration immediately prior to train onset been taken as a baseline and subtracted from the subsequent record, as has been commonly done in previous studies (Hamid, Pettibone, Mabrouk et al., 2016).

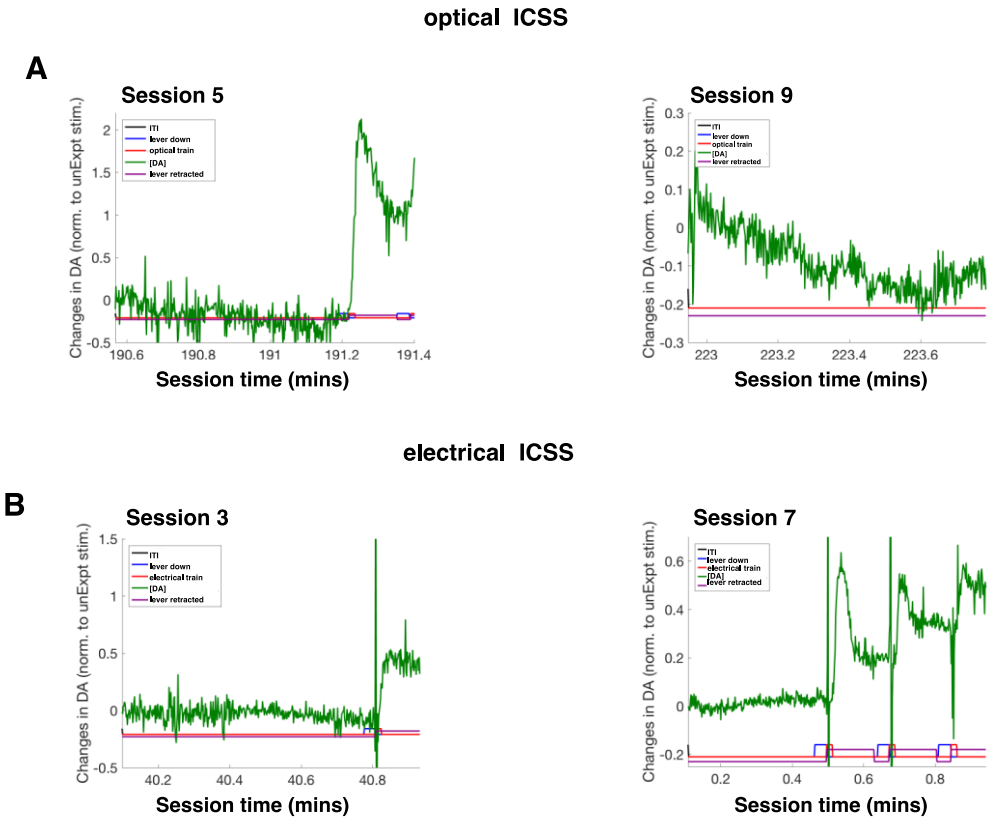


Figure 41. Gradual or step-wise increase in DA concentration time-locked to the delivery of the first stimulation earned

Estimated DA concentration as a function of time during selected segments from the early portion of HI (high) leading trials when the latency to press was long. The DA concentration was estimated from the DA oxidation current recorded at 0.6 V on the ascending limb of the triangle wave and then normalized, first to the peak estimate obtained in response to an unexpected stimulation train and then to the first 10 time points on the x-axis. Examples for oICSS are shown in panel A during session 5 when optical stimulation evoked DA transients (left graph) and during session 9 when optical stimulation no longer evoked detectable DA transients (right graph). Examples for eICSS are shown in panel B during session 3 (left graph) and session 7 (right graph). The leftmost value on the time axis corresponds to the end of the inter-trial period and the extension of the lever. The lines at the bottom of the graphs

designate lever presses (blue), stimulation delivery (red), and blackout periods when the lever was retracted (purple). In panel A, the pulse frequency was 56 Hz, the optical power was 50 mW, the pulse duration was 5 ms, and the train duration was 1000 ms. In panel B, the pulse frequency was 135 Hz, the stimulation current of was 400 μA, the pulse duration was 0.1 ms, and the train duration was 500 ms. DA, dopamine; oICSS, optical intracranial self-stimulation; eICSS, electrical intracranial self-stimulation.

Non-maximal rates of oICSS are stable despite persistent stimulation-induced DA transients

We observed persistent stimulation-induced DA transients in all trial types associated with stimulation trains of sufficiently high pulse frequency to support oICSS. According to the DA-RPE hypothesis, persistent DA transients lead to accelerating oICSS when the rate of work is non-maximal because weights on action selection are not saturated. Due to the limited number of sessions with simultaneous FSCV measurements, training was incomplete and the attempts of the experimenter to adjust the medium pulse frequency did not always succeed in producing non-maximal bouts of oICSS or eICSS in MED test trials. When these attempts did succeed, the resulting subset of MED test trials allowed us to carry out a stringent test of the DA-RPE hypothesis.

To determine the stability of ICSS and DA transients within trials, we graphed their cumulative trajectories (Figure 42). In the case of ICSS, the cumulative trajectory shows, on the y-axis, the total amount of work (holding the lever down to earn stimulation trains) done from time 0 to time *t* divided by the total time spent working during the entire trial. Time *t* on the x-axis is the proportion of trial duration that has elapsed (see the simulated examples in Figure 34). In the case of the DA transients, the cumulative trajectory expresses the sum of the peak amplitude of all stimulation-induced DA transients recorded during a given portion of the trial (e.g., the first 20%) as a proportion of the corresponding sum over the entire trial.

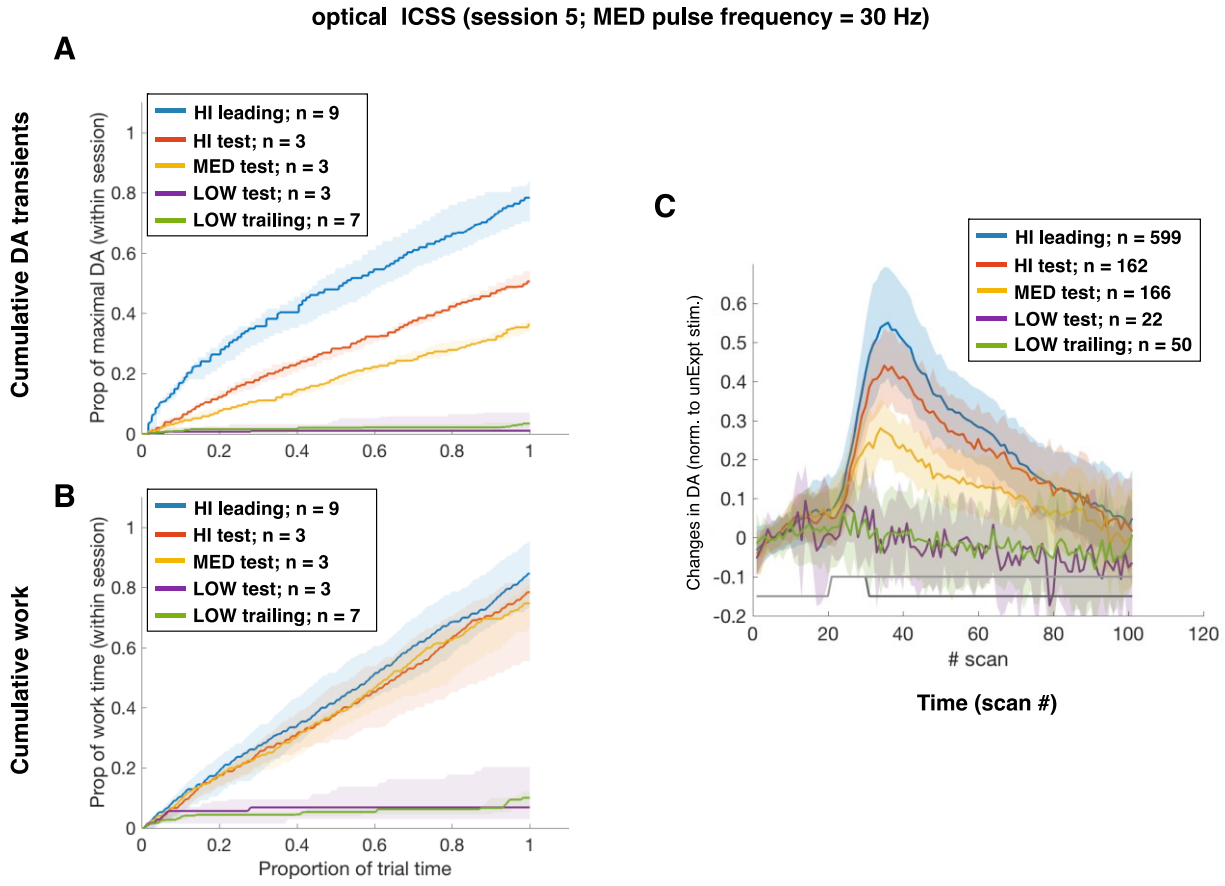


Figure 42. Work rates and amplitudes of stimulation-induced DA transients during the 5th session of oICSS (rat Elop1)

The cumulative trajectories in A express the sum of the peak amplitude of all stimulation-induced DA transients recorded from time 0 to time *t* in a given trial as a proportion of the maximal sum over an entire trial within a session. The x-axis shows elapsed trial time as a proportion of trial duration. The DA concentration was estimated from the DA oxidation current recorded at 0.6 V on the ascending limb of the triangle wave. The cumulative trajectories in B show, on the y-axis, the total amount of work (holding the lever down to earn stimulation trains) from time 0 to time *t* in a given trial divided by the maximal time spent working during an entire trial within a session. The x-axis shows elapsed trial time as a proportion of trial duration. Each line corresponds to the median trajectories obtained for each trial type (HI (high) leading: blue; HI test: orange; MED (medium) test: yellow; LOW (low)

test: purple; and LOW trailing: green; n = {9, 3, 3, 3, 7}). The surrounding shaded area represents a 95% bootstrapped confidence interval. The average DA concentration during stimulation-induced DA transients is plotted as a function of time in C (HI leading: blue; HI test: orange; MED test: yellow; LOW test: purple; and LOW trailing: green; n = {569, 162, 166, 22, 50}). The lighter gray line demarcates the blackout period during which the lever was retracted, and the darker gray demarcates the stimulation delivery. The pulse frequencies were 56, 30, and 4 Hz. The optical power was held constant at 50 mW, the pulse duration was 5 ms, and the train duration was 1000 ms. DA, dopamine; oICSS, optical intracranial self-stimulation.

Session 5 was acquired before adjustment of the pulse frequency for the MED test trials. The 30-Hz pulse frequency used for oICSS in that session produced DA transients that were intermediate in amplitude (Figures 42A and C), but the median cumulative work was comparable for HI leading, HI test, and MED test trials (Figure 42B). In contrast, individual MED test curves plotted against the median curve for the HI leading trials in Figure 43 (top graph) reveals that the animal worked at non-maximal rates of work during the first MED test trial (black line). Plotting the cumulative amplitude of DA transients (red lines) and cumulative work (gray to black lines) for each individual MED test trial in the bottom graphs of Figure 43 shows that the progression in the two variables is strikingly similar regardless of whether the rates of work were maximal or sub-maximal. The black and red curves are essentially superimposed when values were normalized to the maximal amount of cumulative work or peak DA within a trial. According to the DA-RPE theory, persistent simulation-induced DA transients obtained when oICSS is non-maximal lead to an accelerating rate of work. If so, the cumulative amplitude of DA transients should increase at a steady, linear rate whereas the cumulative work time should increase exponentially and deviate substantially from the progression of DA transients. The correspondence between the two variables reported here does not support the DA-RPE hypothesis.

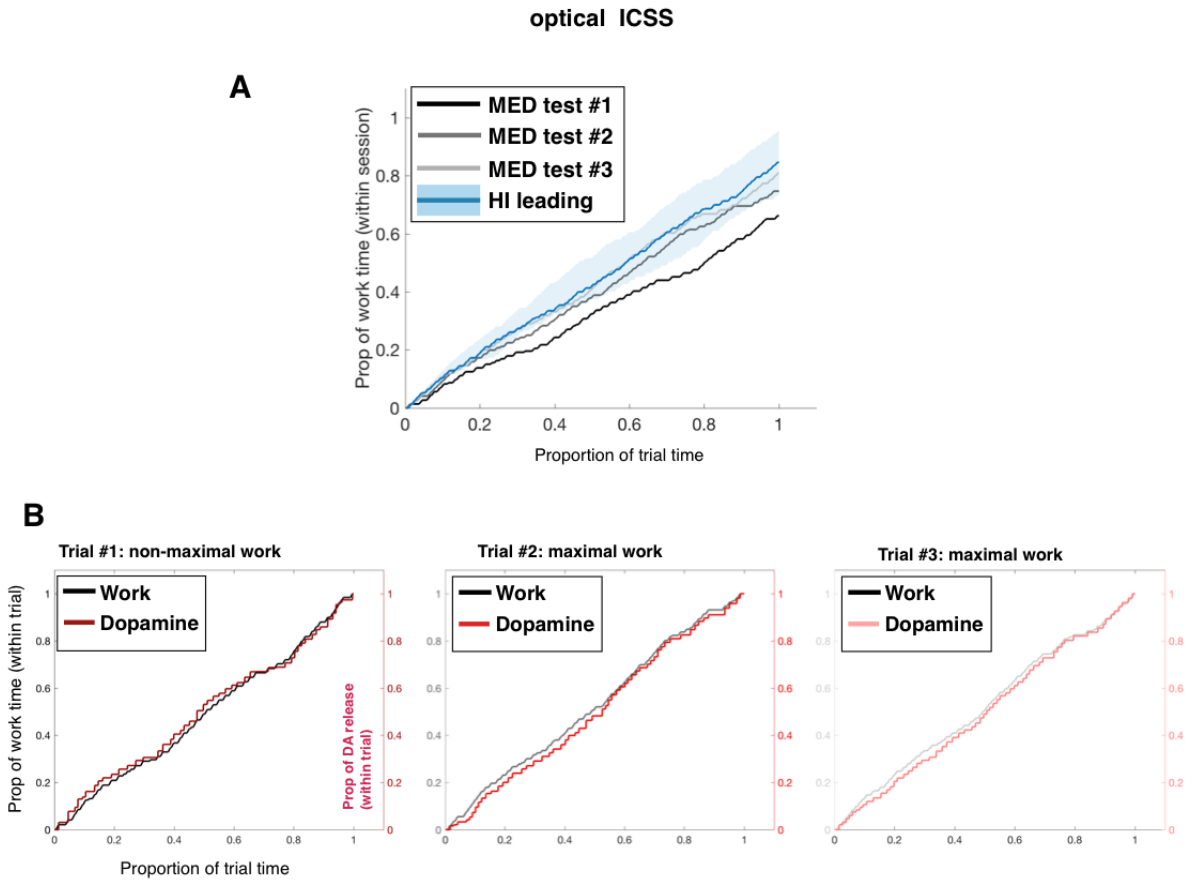


Figure 43. Cumulative work and cumulative peak amplitude of DA transients elicited during the MED (medium) test trials of the 5th session of oICSS (rat Elop1)

A) Cumulative trajectories showing, on the y-axis, the total amount of work (holding the lever down to earn stimulation trains) done from time 0 to time *t* in a given trial divided by the maximal time spent working during an entire trial within a session. The x-axis shows elapsed trial time as a proportion of trial duration. The blue line is the median cumulative work time during the HI (high) leading trials shown in Figure 42B, and the shaded area is a 95% bootstrapped confidence interval. The gray lines represent cumulative work time in each MED test trial. The slopes of the blue and gray lines were compared, and if the slope on a MED trial was lower than the slope of the median trajectory on HI leading trials, work on that MED trial was categorized as non-maximal. The results of this classification are reported in the titles of the graphs in B, which compare cumulative work time (gray) to the cumulative peak

amplitude (red) of the corresponding stimulation-induced DA transients within MED trials (trial 1 on the right, trial 2 in the middle, and trial 3 on the left). The pulse frequencies were 56 Hz and 30 Hz on the HI-leading and MED test trials, respectively. The optical power was held constant at 50 mW, the pulse duration was 5 ms, and the train duration was 1000 ms. DA, dopamine; oICSS, optical intracranial self-stimulation.

Lowering the pulse frequency to 20 Hz in oICSS sessions 6 and 7 reduced the amplitude of the induced DA transient substantially (Figure 44A and C) and produced a median cumulative work rate that was non-maximal for the MED test trials (Figure 44B).

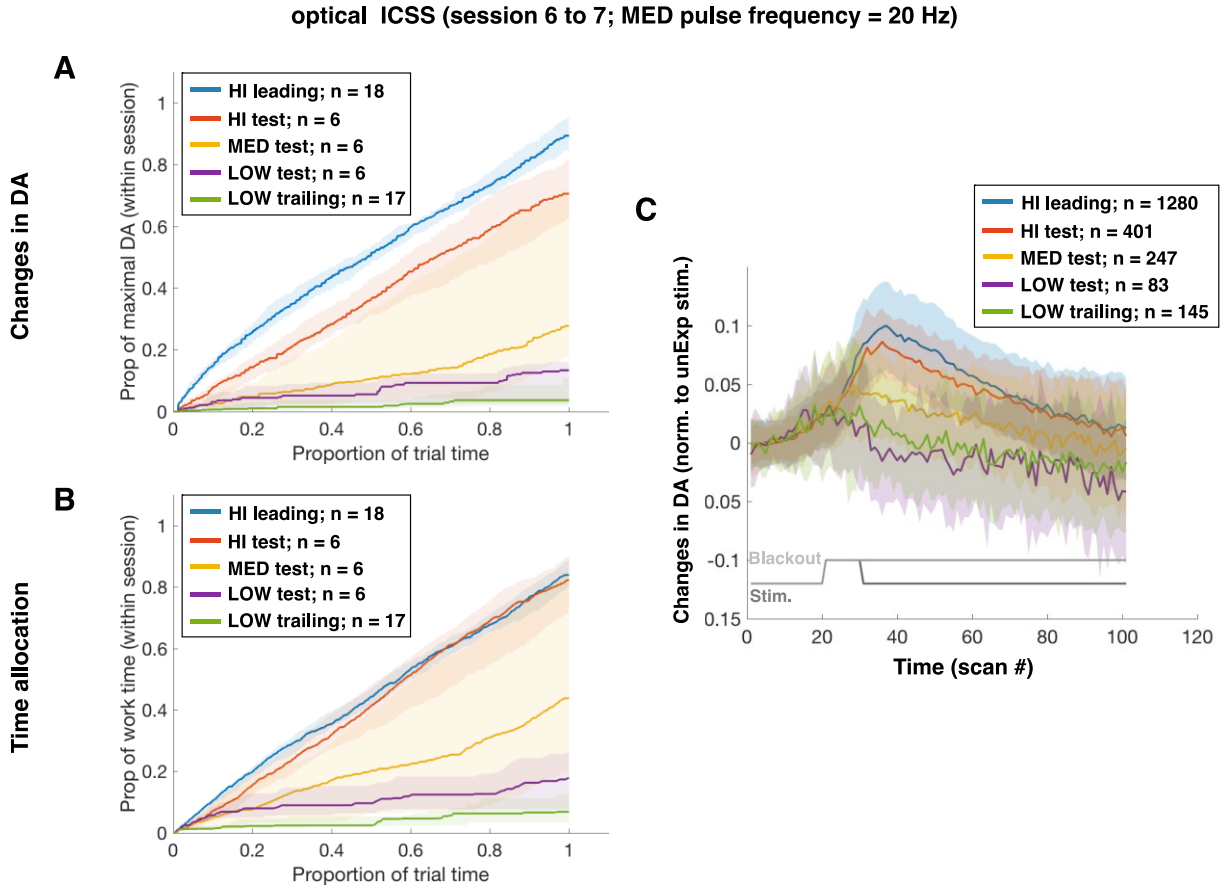


Figure 44. Work rates and amplitudes of stimulation-induced DA transients during the 6th and 7th sessions of oICSS (rat Elop1)

The cumulative trajectories in A express the sum of the peak amplitude of all stimulation-induced DA transients recorded from time 0 to time *t* in a given trial as a proportion of the maximal sum over an entire trial within a session. The x-axis shows elapsed trial time as a proportion of trial duration. The DA concentration was estimated from the DA oxidation current recorded at 0.6 V on the ascending limb of the triangle wave. The cumulative trajectories in B show, on the y-axis, the total amount of work (holding the lever down to earn stimulation trains) done from time 0 to time *t* in a given trial divided by the maximal time spent working during an entire trial within a session. The x-axis shows elapsed trial time as a proportion of trial duration. Each line corresponds to the median trajectories obtained for each trial type (HI (high) leading: blue; HI test: orange; MED (medium) test: yellow; LOW (low) test: purple; and LOW trailing: green; n = {16, 6, 6, 6, 17}). The surrounding shaded area represents a 95% bootstrapped confidence interval. The average DA concentration during stimulation-induced DA transients is plotted as a function of time in C (HI leading: blue; HI test: orange; MED test: yellow; LOW test: purple; and LOW trailing: green; n = {1280, 401, 247, 83, 145}). The lighter gray line demarcates the blackout period during which the lever was retracted, and the darker gray demarcates the stimulation delivery. The pulse frequencies were 56, 30, and 4 Hz. The optical power was held constant at 50 mW, the pulse duration was 5 ms, and the train duration was 1000 ms. DA, dopamine; oICSS, optical intracranial self-stimulation.

Individual MED test curves were plotted against the median curve for the HI leading trials (Figure 45, middle graphs, session 6 on the left and session 7 on the right). Non-maximal rates of work were identified in five out of six MED test trials; the rates of work were lower than in the median curve for the HI leading trials. In some cases, rates were systematically lower throughout the trials whereas, in other cases, lower rates were observed over circumscribed time windows within trials. Comparisons between cumulative amplitude of DA transients and cumulative work for each individual MED test show that the progression in the two variables is strikingly similar no matter if rates of work were maximal or sub-maximal (Figure 45, top graphs for session 6 and bottom graphs for session 7). Of particular interest is the third MED test trial in session 6. The

rate of work began to rise sharply for roughly ten percent of the trial but then transitioned to a slower rate for the remaining of the trial. The slower transition violates the RPE hypothesis, which predicts an acceleration in the rate of ICSS with each optical stimulation.

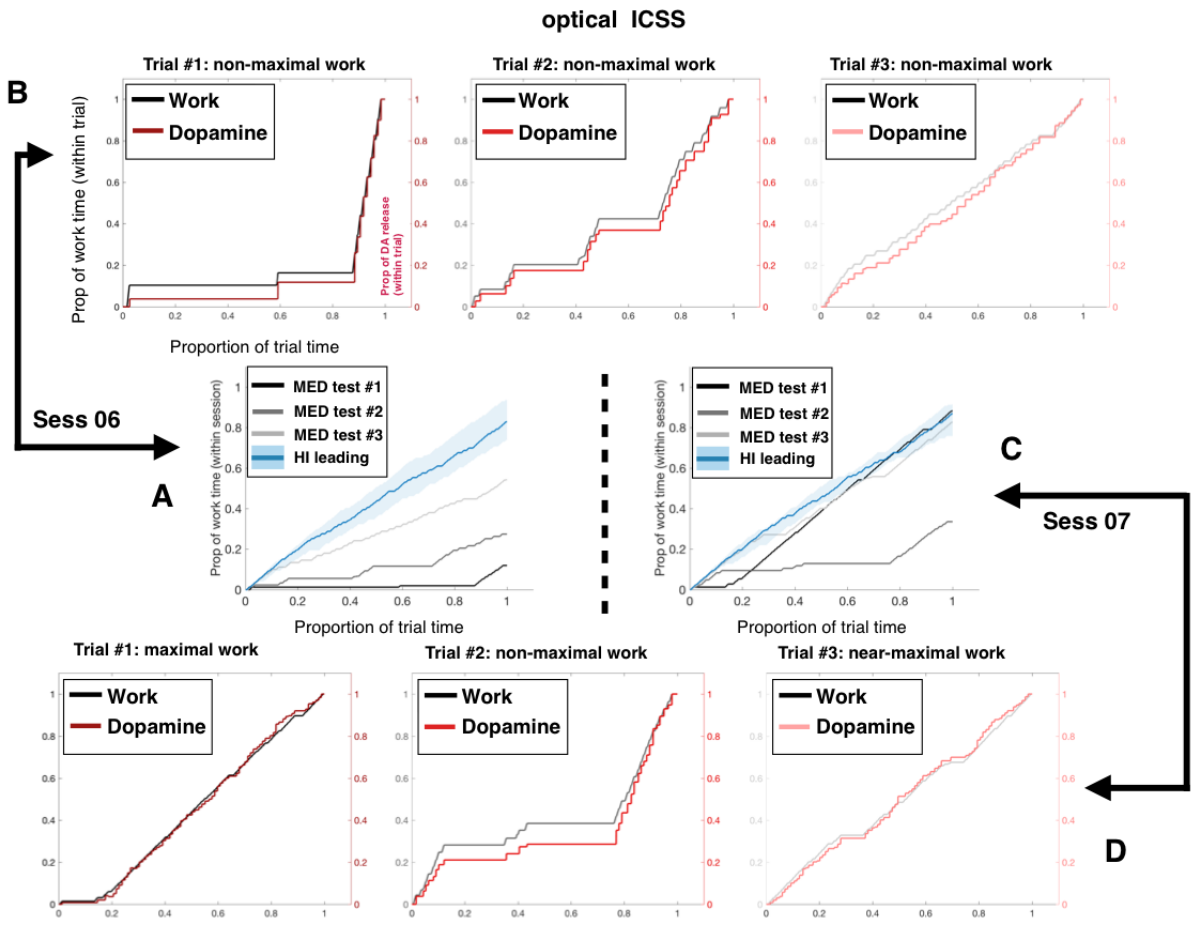


Figure 45. Cumulative work and cumulative peak amplitude of DA transients elicited during the MED (medium) test trials of the 6th and 7th sessions of oICSS (rat Elop1)

A) Cumulative trajectories showing, on the y-axis, the total amount of work (holding the lever down to earn stimulation trains) done from time 0 to time *t* in a given trial divided by the maximal time spent working during an entire trial within a session. The x-axis shows elapsed trial time as a proportion of trial duration. The blue line is the median cumulative work time during HI (high) leading trials (also shown in Figure 44B), and the shaded area is a 95% bootstrapped confidence interval. The gray lines represent cumulative work time in each MED test trial. C) Analogous data from session 7. The slopes of the blue and gray lines were compared (within graphs

A and C), and if the slope on a MED trial was lower than the slope of the median trajectory on HI leading trials, work on that MED trial was categorized as non-maximal. The results of this classification are reported in the titles of the graphs in B (session 6) and D (session 7), which compare cumulative work time (gray) to the cumulative peak amplitude (red) of the corresponding stimulation-induced DA transients within MED trials (trial 1 on the right, trial 2 in the middle, and trial 3 on the left). The thick lines terminating in double arrowheads link the graphs that contain behavioural data taken from the same MED test trials. The pulse frequencies for the HI leading and MED trials were 56 Hz and 20 Hz, respectively. The optical power was held constant at 50 mW, the pulse duration was 5 ms, and the train duration was 1000 ms. DA, dopamine; oICSS, optical intracranial self-stimulation.

Several cumulative graphs in sessions 6 and 7 were best described as work bouts punctuated by pauses (Figure 45). For example, during the second MED test trial in session 7, the animal engaged firmly in ICSS for about 80 secs, took a pause of about the same duration, then earned a few more stimulations, took an even longer pause, and then resumed work at a near-maximal rate. Initial pauses, or pauses following one or two stimulations, could stem from distraction and may not necessary pose a challenge to the DA-RPE hypothesis. However, 80 secs of oICSS supported by persistent RPEs of stable amplitudes should not lead to long pauses if those RPEs signal that the expected reward value is better than expected and should accelerate the proclivity to lever press. Together, the robust similarity in the progression of cumulative DA release and work in all trials, combined with long pauses after vigorous behavioural bouts, challenge the hypothesis that DA transients can act as an RPE signal and dictate oICSS.

Non-maximal rates of eICSS are also stable despite persistent stimulation-induced DA transients

In the animal undergoing eICSS training, the 35-Hz medium pulse frequency produced DA transients of intermediate amplitude and a stable, non-maximal rate of ICSS in the four sessions that gave usable FSCV readings (Figure 46).

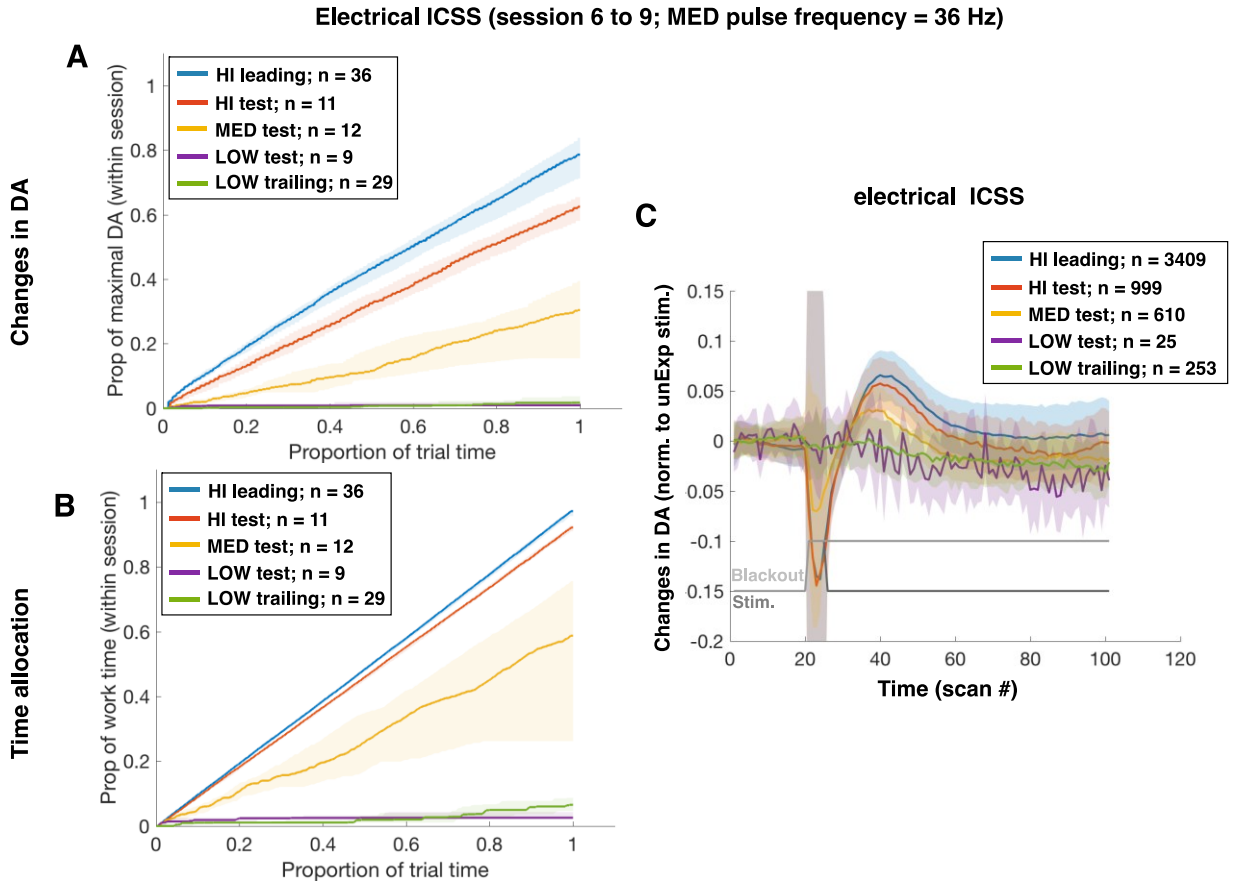


Figure 46. Work rates and amplitudes of stimulation-induced DA transients during the 6th to the 9th session of eICSS (rat Elop7)

The cumulative trajectories in A express the sum of the peak amplitude of all stimulation-induced DA transients recorded from time 0 to time *t* in a given trial as a proportion of the maximal sum over an entire trial within a session. The x-axis shows elapsed trial time as a proportion of trial duration. The DA concentration was estimated from the DA oxidation current recorded at 0.6 V on the ascending limb of the triangle wave. The cumulative trajectories in B show, on the y-axis, the total amount of work (holding the lever down to earn stimulation trains) done from time 0 to time *t* in a given trial divided by the maximal time spent working during an entire trial within a session. The x-axis shows elapsed trial time as a proportion of trial duration. Each line corresponds to the median trajectories obtained for each trial type (HI (high) leading: blue; HI test: orange; MED (medium) test: yellow; LOW (low)

test: purple; and LOW trailing: green; n = {36, 11, 12, 9, 29}). The surrounding shaded area represents a 95% bootstrapped confidence interval. The average DA concentration during stimulation-induced DA transients is plotted as a function of time in C (HI leading: blue; HI test: orange; MED test: yellow; LOW test: purple; and LOW trailing: green; n = {3409, 999, 610, 25, 253}). The lighter gray line demarcates the blackout period during which the lever was retracted, and the darker gray demarcates the stimulation delivery. The stimulation pulse frequencies were 135, 36, and 4 Hz. The stimulation current was held constant at 400 μA, the pulse duration was 0.1 ms, and the train duration was 500 ms. DA, dopamine; eICSS, electrical intracranial self-stimulation.

Electrical stimulation produces DA transients predominantly through trans-synaptic activation (Yeomans, Kofman & McFarlane, 1985; Yeomans, 1989; Murray & Shizgal, 1994; Shizgal et al., 1980; Wise, 1980; Maeda & Mogenson, 1981), raising the possibility that the effects of electrical stimulation can be progressively nulled by other inputs also converging onto DA cell bodies. However, if such nulling had occurred, the transients would have declined progressively over the course of the MED test trials depicted in Figure 46. According to the DA-RPE hypothesis, this would reflect the action of RPEs as they "predicted themselves away." Instead, the cumulative DA trace rose at a steady rate with every electrical stimulation earned. According to the DA-RPE hypothesis, the rate of eICSS should have accelerated as a result of the persistent simulation-induced DA transients. The fact that it was stable contradicts the hypothesis.

Individual cumulative work-time curves for the MED test trial were plotted against the median curve for the HI leading trials (Figures 47 and 48, middle graphs). These graphs reveal that the animal worked at non-maximal rates in most of the MED test trials. Comparisons between cumulative DA release and cumulative work show that the progression in the two variables are strikingly similar no matter if the rates of work were maximal or sub maximal (Figures 47 and 48, top and bottom graphs). The behavioural patterns reported during oICSS were also seen during eICSS, including long pauses after intense work bouts. According to the DA-RPE theory, persistent simulation-induced DA transients obtained when eICSS is non-maximal lead to

accelerating rates of work, not steady rates of work. Once again, these findings are inconsistent with the hypothesis that DA transients can act as an RPE signal and dictate eICSS.

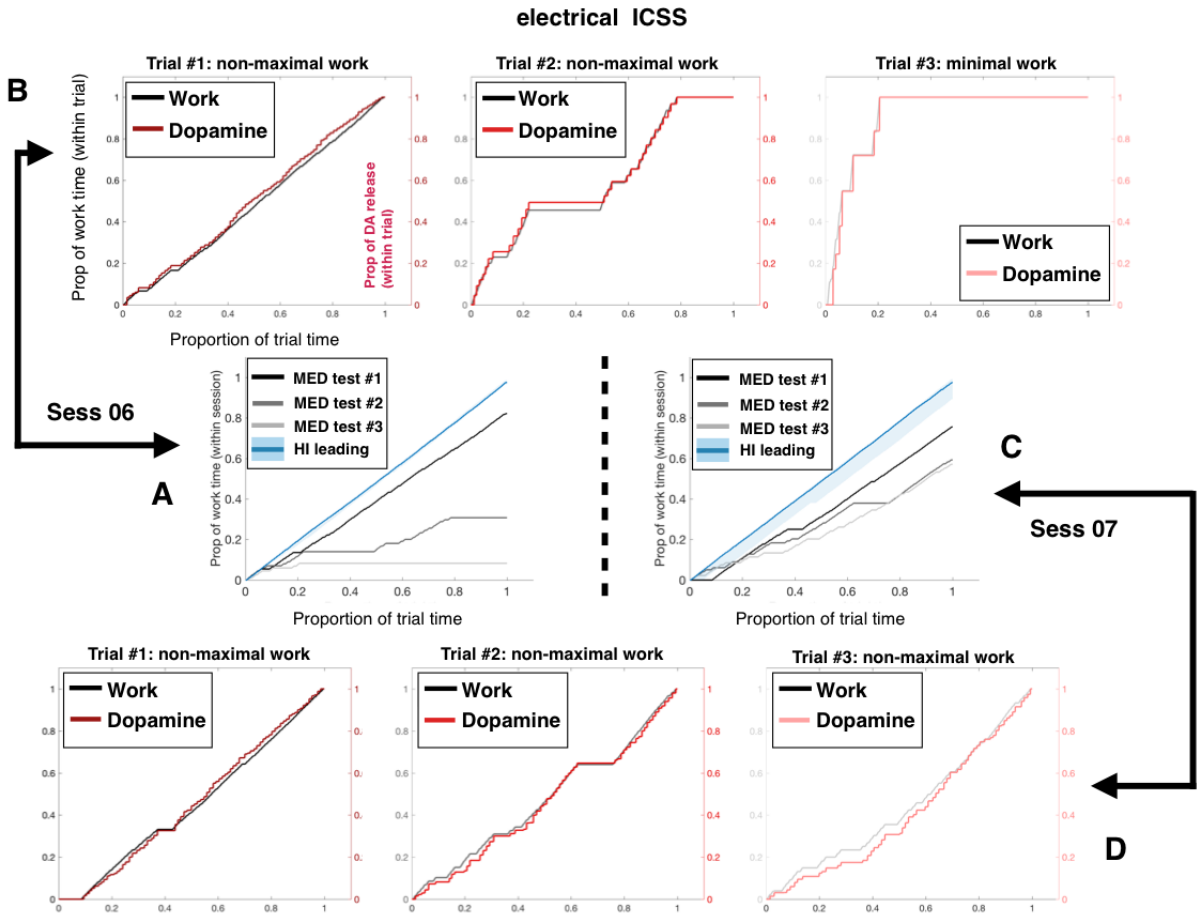


Figure 47. Cumulative work and cumulative peak amplitude of DA transients elicited during the MED (medium) test trials of the 6th and 7th sessions of eICSS (rat Elop7)

A) Cumulative trajectories showing, on the y-axis, the total amount of work (holding the lever down to earn stimulation trains) done from time 0 to time *t* in a given trial divided by the maximal time spent working during an entire trial within a session. The x-axis shows elapsed trial time as a proportion of trial duration. The blue line is the median cumulative work time during HI (high) leading trials (also shown in Figure 46B), and the shaded area is a 95% bootstrapped confidence interval. The gray lines represent cumulative work time in each MED test trial. C) Analogous data from session 7. The slopes of the blue and gray lines were compared (within graphs A and C), and if the slope on a MED trial was lower than the slope of the median

trajectory on HI leading trials, work on that MED trial was categorized as non-maximal. The results of this classification are reported in the titles of the graphs in B (session 6) and D (session 7), which compare cumulative work time (gray) to the cumulative peak amplitude (red) of the corresponding stimulation-induced DA transients within MED trials (trial 1 on the right, trial 2 in the middle, and trial 3 on the left). The thick lines terminating in double arrowheads link the graphs that contain behavioural data taken from the same MED test trials. The pulse frequency for the HI leading trial was 135 Hz and 36 Hz for the MED test trials. The stimulation current was held constant at 400 µA, the pulse duration was 0.1 ms, and the train duration was 500 ms. DA, dopamine; eICSS, electrical intracranial self-stimulation.

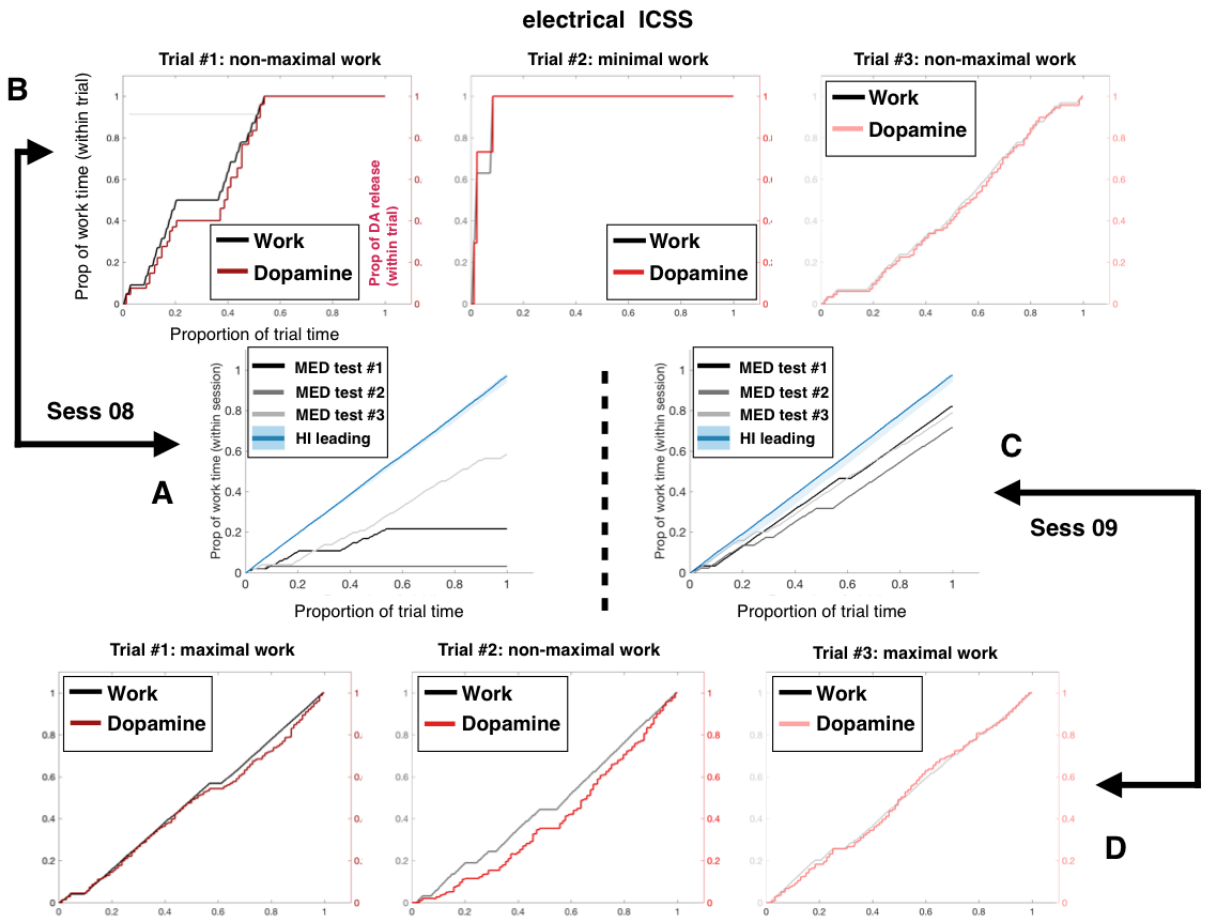


Figure 48. Cumulative work and cumulative peak amplitude of DA transients elicited during the MED (medium) test trials of the 8th and 9th sessions of eICSS (rat Elop7)

A) Cumulative trajectories showing, on the y-axis, the total amount of work (holding the lever down to earn stimulation trains) done from time 0 to time *t* in a given trial divided by the maximal time spent working during an entire trial within a session. The x-axis shows elapsed trial time as a proportion of trial duration. The blue line is the median cumulative work time during HI (high) leading trials (also shown in Figure 46B), and the shaded area is a 95% bootstrapped confidence interval. The gray lines represent cumulative work time in each MED test trial. C) Analogous data from session 8. The slopes of the blue and gray lines were compared (within graphs A and C), and if the slope on a MED trial was lower than the slope of the median trajectory on HI leading trials, work on that MED trial was categorized as non-maximal. The results of this classification are reported in the titles of the graphs in B (session 8) and D (session 9), which compare cumulative work time (gray) to the cumulative peak amplitude (red) of the corresponding stimulation-induced DA transients within MED trials (trial 1 on the right, trial 2 in the middle, and trial 3 on the left). The thick lines terminating in double arrowheads link the graphs that contain behavioural data taken from the same MED test trials. The pulse frequency for the HI leading trial was 135 Hz and 36 Hz for the MED test trials. The stimulation current was held constant at 400 μA, the pulse duration was 0.1 ms, and the train duration was 500 ms. DA, dopamine; eICSS, electrical intracranial self-stimulation.

Absence of DA transients evoked by the earliest cue predictive of stimulation availability

If we assume that the effects of the optical stimulation on phasic DA activity act as a fictive RPE, DA transients should back propagate to the earliest predictor of stimulation availability for HI leading and test trials. Due to the random component of the ITI duration, the first lever extension was the earliest predictor of reward availability at the onset of each trial. No DA transients were detected for any of the trials at this critical moment (Figure 49). This finding presents another challenge to the hypothesis that DA transients act as RPEs.

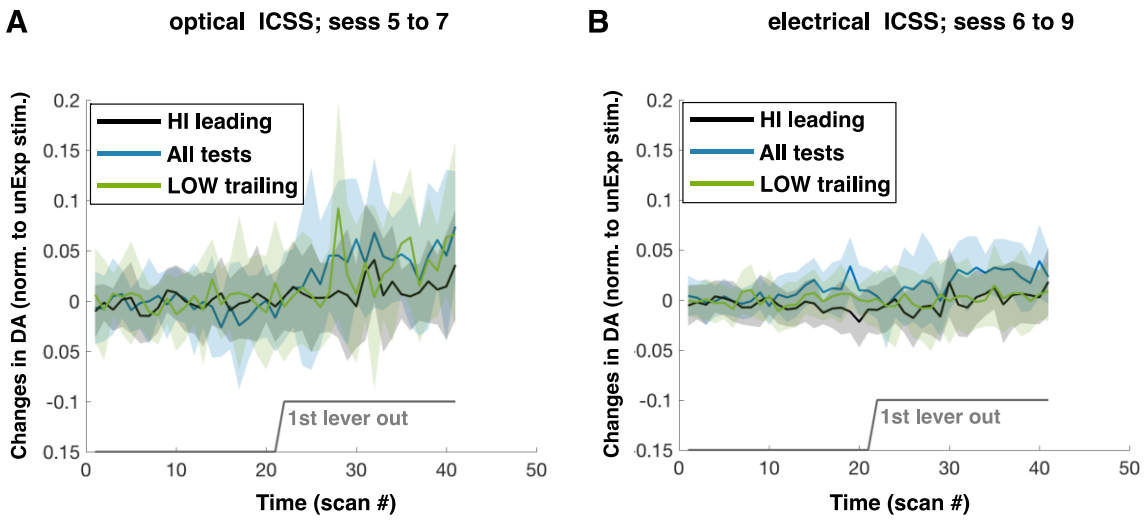


Figure 49. No detectable DA transient at the earliest cue predicting the availability of optical or electrical trains

Due to the random component of the ITI, lever extension at the beginning of trials was the earliest predictor of stimulation availability. Nonetheless, this event did not produce a detectable DA transient with optical (A) or electrical (B) stimulation. The DA concentration (y-axis) was estimated from the DA oxidation current recorded at 0.6 V on the ascending limb of the triangle wave and then normalized, first to the peak estimate obtained in response to an unexpected stimulation train and then to the first 10 time points on the x-axis. Each line represents the median DA concentration for a given trial type (HI (high) leading: blue; HI, MED (medium), or LOW (low) test: black; and LOW trailing: green; n = {27, 9, 9, 9, 27}), and each shaded area represents the interquartile range. The gray line marks the first lever extension. DA, dopamine; ITI, inter-trial interval.

It is noteworthy that the onset of the ITI was accompanied by a gradual increase in DA release only during the transition between the LOW trailing and HI leading trials (Figure 50). Although the gradual increase in DA release occurred during both oICSS and eICSS, the associated neurochemical profiles differed between stimulation methods. During oICSS, the increase in DA release began precisely at the onset of the ITI and continued to rise until the next lever extension. In contrast, during eICSS, the increase in DA release was much smaller and began with a 2-sec delay regardless of the randomized duration of the ITI, and plateaued quickly instead of ramping up.

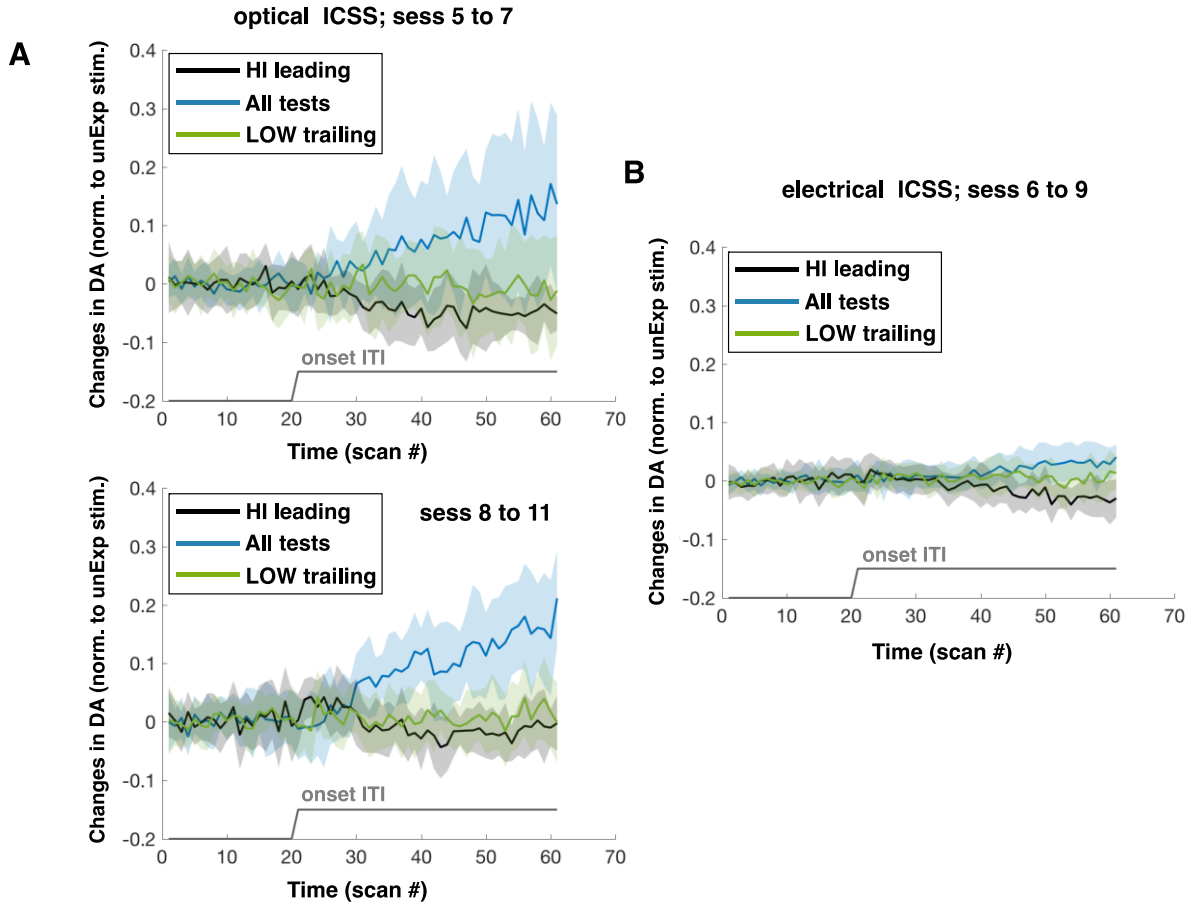


Figure 50. A gradual increase in DA release at the onset of the inter-trial period between LOW (low) trailing and HI (high) leading trials during oICSS sessions

A) The effect of lever retraction and ITI onset on DA concentration when transitioning out of LOW trailing trials during oICSS sessions. B) A corresponding change in DA concentration is not seen at the analogous point during eICSS sessions. The DA concentration (y-axis) was estimated from the DA oxidation current recorded at 0.6 V on the ascending limb of the triangle wave and then normalized, first to the peak estimate obtained in response to an unexpected stimulation train and then to the first 10 time points on the x-axis. Each line represents the median DA concentration for a given trial type (HI leading: blue; HI, MED, or LOW test: black; and LOW trailing: green; in A: n = {27, 9, 9, 9, 27} for sessions 5 to 7 and n = {36, 12, 12, 12, 36} for sessions 8 to 10; in B: n = {27, 9, 9, 9, 27}), and each shaded area represents the interquartile range. The gray line designates lever retraction and ITI onset. DA, dopamine; oICSS, optical intracranial self-stimulation; eICSS, electrical intracranial self-stimulation; ITI, inter-trial interval.

Discussion

Learning can be conceptualized along a continuum of model-free and model-based learning mechanisms (Dayan & Niv, 2008; Doll, Simon & Daw, 2012). On this view, reward predictions can be learned through trial-and-error; by constructing complex world models, or by means of intermediate, composite mechanisms. We designed a trial structure to examine the effects of persistent simulation-induced DA transients on ICSS in a context that makes it possible to determine whether learning is model-free or whether it assumes a more complex form. Although learning more complex than the model-free variant has been manifested by subjects in a prior experiment (Trujillo-Pisanty et al., SfN poster, 2016), rats Elop1 and Elop7 completed only a limited number of ICSS sessions with simultaneous FSCV measurements and did not yet show evidence of more complex forms of learning. Therefore, we focused on model-free TDRL, the form of RL originally tied to ICSS (Montague, Dayan & Sejnowski, 1996). We tested the premise that DA transients act as RPEs to update reward value *and* bias action selection. Several of the findings appear inconsistent with the DA-RPE hypothesis.

Principal-component regression without background subtraction

One of the principal innovations introduced here is the method developed to isolate the DA transients. Background subtraction is typically performed as the first step before visualizing voltammograms or quantifying DA concentrations in FSCV recordings. This data transformation is intended to isolate the currents produced by the oxidation of DA and reduction of DA-O-quinone, as well as to remove the slow drift of the Faraday current over time.

Implementation of background subtraction can be problematic because of unavoidable arbitrariness in determining what constitutes the background. In addition, a recent paper by Kishida et al. (2015) demonstrated that the pertinent information necessary to extract DA concentrations from FSCV recordings can be obtained without resorting to background subtraction. They obtained FSCV recordings *in vitro* from solutions that varied in DA concentration and pH. Elastic-net regression was then used to estimate the DA concentration and pH from the FSCV data. Given that no background subtraction was carried out, the cyclic

voltammograms included the background Faraday current. They found DA concentrations were accurately predicted by this method in a manner independent of pH levels. The elastic-net regression also revealed that data at many time-points within the scan contributed significantly to the model. They concluded that, although maximal DA oxidation occurs at roughly 0.6 V and maximal reduction at -0.2 V, voltages sampled at other points during the scan have predictive power to detect changes in phasic DA activity.

We had initially hoped to follow the method explained by Kishida et al., but we could not successfully acquire the necessary calibration data *in vitro*. However, whereas the Faraday currents registered by freshly implanted FSCV microsensors in an intact brain are similar to those observed *in vitro* (as was the case in the experiment by Kishida et al.), the Faraday currents registered by chronically implanted microsensors are more complex than those observed *in vitro,* and they evolve over time as the FSCV electrode interacts with the brain environment and is etched by continuous usage (Keithley et al., 2011; Rodeberg et al., 2015). Consequently, we had to use, as calibration data, cyclic voltammograms acquired in awake animals presented with unexpected stimulation trains (see the section on training sets for details).

We took a novel approach, inspired by the findings of Kishida et al. We employed the PCR-PCA methods used in most voltammetry experiments (Rodeberg et al., 2015) but omitted the background subtraction, thus retaining all informative parts of the voltammograms. Our approach scales the size of the required training set in proportion to the number of principal components retained, conserves the enhancement in predictive power provided by the approach of Kishida et al., and circumvents the potential distortion caused by background subtraction. We confirmed that, in doing so, we could resolve the stimulation-induced DA transients. Interestingly, we also detected fluctuations in DA activity over a longer time scale. We found that the baseline DA concentration rose in response to the very first stimulation received at the onset of trials in which the animal worked at more than minimal rates and then levelled off after several seconds. This time course had yet to be documented in prior work (see below for more details).

Previous studies have characterized the amplitude of stimulation-induced DA transients during short eICSS sessions (Garris et al., 1999; Owesson-White et al., 2008). In both cases, multiple backgrounds were used to obtain background-subtracted data and quantify the amplitude of DA transients. Garris et al., selected a different background for each 30-sec recording shown in their paper. Owesson-White et al. averaged ten cyclic voltammograms prior to stimulation and subtracted this average from cyclic voltammograms after the stimulation. If we apply this rule to a subset of our data, at the onset of the first HI leading trial, in session 7 for Elop 7 (working for electrical stimulation), we can see clearly how the background-subtraction approach obscures gradual changes in DA concentration. This is illustrated in Figure 51. The changes in DA concentration, normalized to the median of the first 10 scans on the x-axis, are plotted in 51A, whereas the effect of multiple background subtractions is plotted in 51B. Ten time points were selected (indicated by yellow dots below the green line) when Owesson-White et al. would have performed background subtractions, and the average DA concentration at these time points was subtracted from subsequent DA concentrations (indicated by arrows below the green line). In so doing, the gradual increase in DA concentration becomes undetectable and the amplitude of the stimulation-induced DA transients appears to decline as a function of the number of stimulations earned, as reported in Garris et al. (1999) and Owesson-White et al. (2008). In contrast, the data in Figure 51A show that the DA transients are superimposed on a more gradual increase in DA concentration. Whereas Figure 51B makes it appear as if the effect of the stimulation on DA concentration falls as repeated trains are delivered, Figure 51A shows that the net effect actually grows.

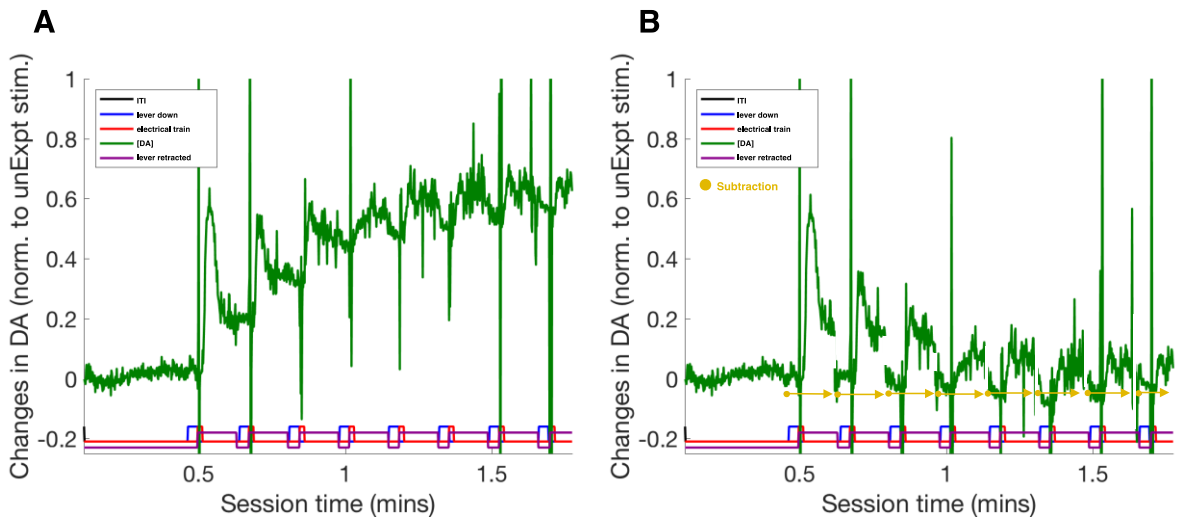


Figure 51. Multiple background subtractions obscure gradual changes in DA release

A) When our method for estimating the DA concentration was employed, a gradual increase in DA release was observed when rat Elop7 first started earning electrical stimulation during session 7. B) In contrast, the gradual increase in DA release was obscured when the conventional method, which entailed multiple background subtractions, was employed to estimate the DA concentration. The DA concentration (y-axis) was estimated from the DA oxidation current recorded at 0.6 V on the ascending limb of the triangle wave and then normalized, first to the peak estimate obtained in response to an unexpected stimulation train. In A, data were further normalized to the average of the first 10 time points on the x-axis. In B, a separate baseline was used for the response to each stimulation train (indicated by yellow dots below the green line) as per the method employed by Owesson-White, Cheer, Beyene et al. (2008). The resulting baseline values were subtracted from subsequent

DA concentrations over the intervals designated by the yellow arrows below the green line. The lines at the bottom of the graphs demarcate lever presses (blue), stimulation delivery (red), and blackout periods when the lever was retracted (purple). The pulse frequency was 135 Hz, the stimulation current was 400 μA, the pulse duration was 0.1 ms, and the train duration was 500 ms. DA, dopamine.

Combining oICSS and chronic voltammetry can test the DA-RPE hypothesis

We established that every stimulation earned by rat Elop 1 elicited persistent DA transients with stable amplitudes. This is the first report, to our knowledge, that optical stimulation *can* produce persistent simulation-induced DA transients during long oICSS sessions lasting four to five hours. Such long sessions are needed to obtain sufficient, quantitative behavioural measurements under the variety of conditions required to test the DA-RPE hypothesis of ICSS.

Under the assumption that DA transients act as RPEs, a fictive RPE would change weights in the value function, *V(t)*, and weights on action selection, leading to increases in the proclivity to lever press. The DA-RPE hypothesis predicts that repeated increases in weights would lead to an accelerating rate of work during non-maximal oICSS. Nonetheless, cumulative DA release closely paralleled lever pressing in all trials, even when the animal worked at non-maximal rates. The lack of an accelerating rate of work deviates dramatically from the prediction of the DA-RPE hypothesis. In addition, there was no evidence that weight changes caused by optically induced DA transients propagated to the earliest cue predictive of stimulation availability: no DA transient was registered when the lever was re-introduced into the test cage at the onset of trials (Figure 49). Nonetheless, the rats often re-initiated responding shortly after the lever was extended, indicating the rats were attending to the lever and regarded its extension as behaviourally salient. The absence of a DA transient at this critical moment also challenges the DA-RPE hypothesis.

Can the DA-RPE hypothesis fend off these challenges? One possible line of defence might be based on the fact that not all DA neurons are activated by midbrain DA optical stimulation, as suggested by our collaborator, Peter Dayan (personal communication).

The compensation hypothesis

Could unstimulated DA neurons produce negative RPEs to counter the effects of positive RPEs in stimulated DA neurons?

We targeted the VTA, a large brain region that extends roughly 2 mm along the anterior-posterior axis, 1.5 mm along the medial-lateral axis and 1 mm along the dorsal-ventral axis (Ikemoto, 2010). The optical implant was 300 µm in diameter and the nominal half-angle of divergence of the cone of emitted light was 0.29 rad (Optical fiber: FT300EMT, Thorlabs, NA: 0.39). Taking into consideration models of light intensity distribution once transmitted out of the fiber (Aravanis et al., 2007; Adamantidis et al., 2007; Bernstein et al., 2008; Chow et al., 2010; Kahn et al., 2011; Yizhar et al., 2011; Yona et al., 2016) and histological data on ChR2 expression confirming not all DA neurons are successfully transfected with this technique (Witten et al., 2011), a substantial number of DA neurons in our preparation were likely unstimulated.

Wolfram Schultz (1998) has argued that most midbrain DA cell bodies respond to unexpected rewards and reward-predicting stimuli. If so, the unstimulated DA neurons should register negative RPEs during oICSS because they receive a TD signal in the absence of a primary reward signal, *r(t)* (also called the reward function in the RL literature). This rationale is presented through highly simplified diagrams of the population of VTA DA neurons in Figure 52. The three diagrams in this figure represent DA-mediated RPEs at the following time points: receipt of the first, second, and third optical trains. Green neurons represent DA neurons expressing ChR2, and superimposed red action potentials differentiate the neurons successfully activated by the optical stimulation from ChR2-expressing neurons positioned too far from the light source to be stimulated. Black neurons do not express ChR2 and also remain unstimulated during optical stimulation. Each neuron computes an RPE, and the value is shown at its output. The optical stimulation is assumed to override the inhibitory input from the reward prediction (see below), and thus the optically activated DA neurons produce an RPE of constant amplitude. As the prediction is learned, unstimulated neurons signal a negative RPE at their output by reducing their firing. To broadcast a net RPE, as suggested by Schultz (1998), DAergic outputs must be integrated downstream. A simple way to perform such integration is to average the RPEs

over the entire population of DA neurons. The resulting average is then used to adjust the weights on the value function, *V(t)*, represented in purple in the figure.

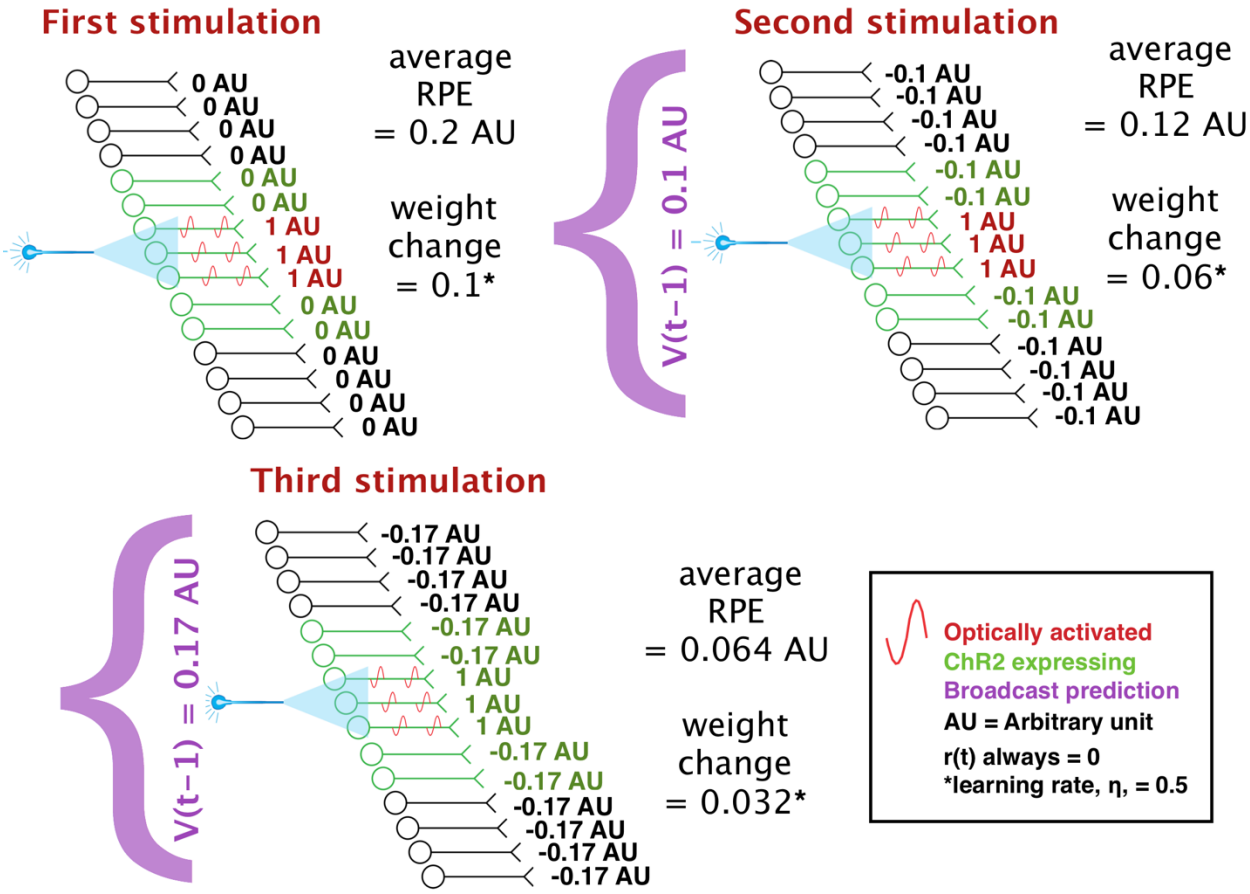


figure adapted from a presentation prepared by Peter Shizgal

Figure 52. The compensation hypothesis: Negative stimulation-induced RPEs signalled by unstimulated DA neurons balance persistent RPEs signalled by stimulated neurons

The three diagrams in this figure represent DA-mediated RPEs at the following time points: receipt of the first, second, and the third optical trains. Green neurons represent DA neurons expressing ChR2 and superimposed red action potentials differentiate the neurons successfully activated by the optical stimulation (blue laser icon) from ChR2-expressing neurons positioned too far from the light source to be stimulated. Black neurons do not express ChR2 and also remain unstimulated during optical stimulation. The reward prediction is provided to all the neurons by their afferents (not shown). Each neuron computes an RPE, and the value is shown at its output. An average RPE is shown on the left of the diagrams to represent the

integrated signal of all DAergic outputs. The weight change caused by the average RPE is also shown on the left ($w(t-1)_{new} = w(t-1)_{old} + \eta$* net TD error; $\eta = 0.5$ in this example). The corresponding adjustment to the weights on the value function, *V(t)*, is shown in purple. RPE, reward prediction error; DA, dopamine; ChR2, channelrhodopsin-2.

The RPE consists of a discrete derivative of the value function, the TD signal, *V(t) - V(t-1)*, compared with the actual reward represented by the reward function*, r(t)*. Note that if the optically induced DA transients act as RPEs rather than contributing to the primary reward signal, the value of the reward function, *r(t)*, is zero at all times, which means that the TD signal, *V(t) - V(t-1)*, uniquely influences RPEs in rats engaging in oICSS. When the first stimulation is earned, unstimulated DA neurons output an RPE of zero because the value function at time *t, V(t)*, and at time *t-1*, *V(t-1)*, is still zero. Stimulated DA neurons, in contrast, output positive RPEs, set arbitrarily to 1 on the diagram, due to the changes in phasic DA produced by the optical stimulation. The overall RPE is positive (0.2 in our hypothetical example), which increases the weights in the value function at time *t-1*, *V(t-1),* to produce a nonzero prediction. That increase in weight is proportional to the size of the net RPE and the learning rate, η, arbitrarily set to 0.5. (The increase in weight is 0.1 in our hypothetical example, and the weight increase equals RPE* η).

Recall that a scalar relationship between the magnitude of the TD error and the resulting weight change is assumed for the sake of simplicity, although this implies that weights can grow to infinity, which is not biologically possible. Since the TD error in conventional conditioning tasks cancels itself away during learning, the weights stabilize at values that reflect the magnitude of the reward; growth towards infinity cannot occur. However, this issue must be revisited in the case of direct optical stimulation, which, as we have shown (Figure 52), produces persistent DA transients. If these act as TD errors, then the problem of growth towards infinity must be addressed.

Upon delivery of the second stimulation, unstimulated neurons compute a TD signal equal to 0-*V(t-1)*, which results in a negative RPE equal to *V(t-1)* (-0.1 in our hypothetical example). Even

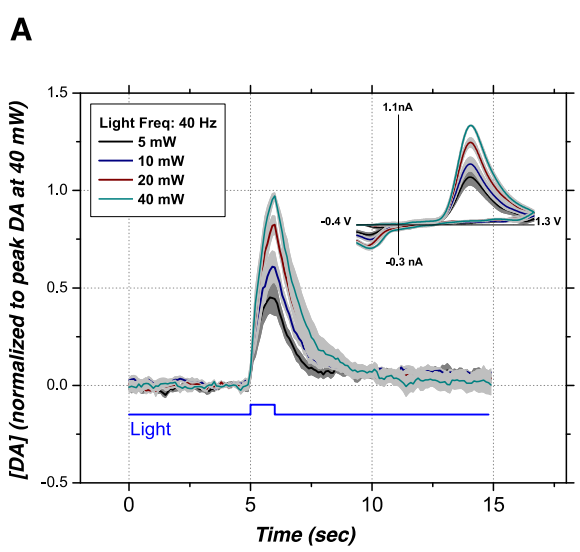
if the stimulated DA neurons continue to output the same positive RPE, the net RPE begins to decrease compared with the first stimulation (to 0.12 in our hypothetical example). The resulting small, but positive, net RPE increases the weights in the value function, proportionally to the value of the smaller RPE (by 0.06 in our hypothetical example). The updating of the value function makes *V(t-1)* larger (0.17 in our hypothetical example). When the animal earns a third stimulation, the TD signal equals 0-*V(t-1)* and results in a larger negative RPE in unstimulated neurons (-0.17 in our hypothetical example). The net RPE is still positive, but smaller than the previous net RPE (0.064 in our hypothetical example). Given this pattern, if a net RPE is broadcasted to all DA neurons, the value function at time *t*-1, *V(t-1)*, in unstimulated neurons increases until the net RPE nulls itself, in accordance to the DA-RPE hypothesis. Stimulating a larger proportion of DA neurons would result in the same nulling of the net RPE, but at a faster pace. We will call this rationale, which was proposed by Peter Dayan, the "compensation hypothesis." This hypothesis eliminates the requirement for an accelerating work trajectory, even on trials associated with non-maximal work rates. Thus, on face value, the compensation hypothesis appears to provide a way to reconcile the DA-RPE hypothesis with our evidence of stable, intermediate work rates in the face of repeated DA transients.

Are negative DA transients observed in unstimulated neurons?

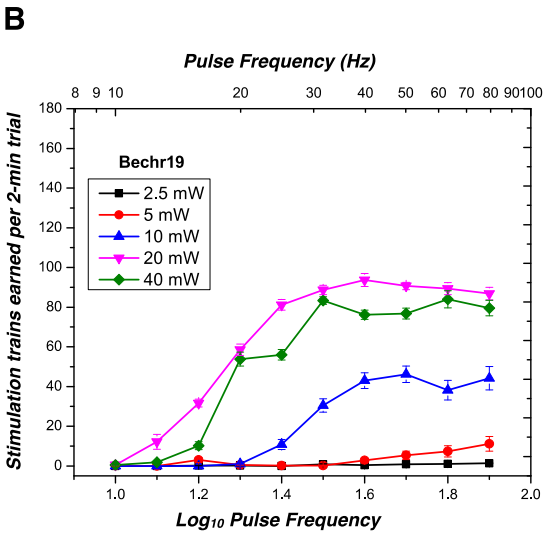
The negative DA transients predicted by the compensation hypothesis might be observed by recording simultaneously from two FSCV electrodes, one aligned with the stimulated DA neurons and the other with unstimulated neurons. For example, the virus infection and the optical implant could be modified to restrict stimulated DA neurons to one hemisphere (see procedures presented in the first chapter for unilateral transfection of the VTA and the insertion of the optical implant in a beveled cannula to block light delivery to the contralateral VTA); one FSCV electrode could be positioned in each hemisphere. Prior to training animals to engage in oICSS, FSCV recordings during the unexpected delivery of food would determine if both FSCV electrodes can detect RPE-like signals and are thus positioned in DA terminal fields. The unexpected delivery of optical stimulation would confirm that only one of the two FSCV electrodes is aligned with DA neurons activated by stimulation trains. According to the compensation hypothesis, the latter electrode would register positive simulation-induced DA transients whereas the former would register negative ones.

Negative DA-RPEs in unstimulated neurons are likely to be small

Any negative DA transients observed in unstimulated neurons are likely to be small. DA neurons have a baseline firing rate around 3 to 5 Hz, and a low firing rate can only decrease by a small amount until it is no longer firing. In contrast, optical stimulation can produce large DA transients with increases in amplitude over a wide range of pulse frequency, from 1 to 56 Hz, at several optical powers, including 40 mW, the optical power used in our experiment (Figure 53). This asymmetry works against the compensation hypothesis, especially in HI leading and HI test trials, in which individual DA neurons would be maximally activated by the high pulse frequency. If only small decreases in DA firing rate can occur in neurons not recruited by the optical stimulation, a large number of unstimulated neurons would be needed to quickly null the net positive RPE obtained after the first stimulation at trial onset.



A) Reproduced from Cossette, Trujillo-Pisanty & Shizgal, Colloquium poster for the 30th Anniversary of the Center for Studies in Behavioural Neurobiology at Concordia University, 2013)



B) Reproduced from Trujillo-Pisanty, I. (2016) Dissecting the role of dopamine in brain stimulation reward: neuroeconomic, pharmacological, and optogenetic studies (Unpublished doctoral dissertation; p.114) Concordia University, Montreal, Canada.

Figure 53. The neurochemical and rewarding effects of optical stimulation of midbrain DA neurons as a function of optical power

Optically-induced DA transients in rat Avchr2 (from a distinct pilot study) in A. The effect of the stimulation is shown as a set of superimposed voltammetric traces, representing the oxidation current as a function of time for the potential at which DA oxidation current was maximal. Corresponding cyclic voltammograms are shown in the insets. These represent the background-subtracted current as a function of the potential applied to the carbon fiber at the time when the DA oxidation was maximal. Each line corresponds to the averaged response to one optical power (n = 6), normalized to the peak DA obtained at the highest power tested; shaded areas represent standard errors of the mean. The blue line demarcates stimulation delivery. The rewarding effects of the optical stimulation, inferred from the behaviour of rat Bechr19 (trained as part of the doctoral thesis work of Ivan Trujillo-Pisanty, 2016) are shown in B. The number of stimulation trains earned during 2-min trials is plotted on the y-axis as a function of pulse frequency and optical power (coded by the symbols and colours). The plotted values are averages of five values obtained at each pulse frequency, and the error bars are standard errors of the mean. The animal was working on a fixed-ratio, FR1, schedule of reinforcement. DA, dopamine.

A test of the compensation hypothesis

An effective test of the compensation hypothesis could entail three changes to our method. First, multiple fibers would be used to deliver light to a larger area. Second, a red-shifted opsin such as Chrimson would be substituted for ChR2 to stimulate midbrain DA neurons in a larger area below the fiber tip. Third, the proportion of activated DA neurons would be estimated by a method analogous to immediate-early-gene expression. If unstable, accelerating time allocation to pursuit of intermediate-strength rewards were observed (Figure 34A) following activation of a large proportion of the population of midbrain DA neurons, then the compensation model and the DA-RPE hypothesis underlying it would be confirmed. However, if data similar to those reported here were again obtained (Figure 34B), then the challenge posed by the compensation hypothesis would be parried, and the DA-RPE hypothesis would be more convincingly contradicted. Our approach to extract DA concentrations is particularly well-suited to detect the

small drops in voltammetric recordings predicted by the compensation hypothesis, especially since data are not background subtracted.

Combining eICSS and chronic voltammetry reveals persistent simulation-induced DA transients produced by electrical stimulation

The DA-RPE hypothesis makes distinct predictions when applied to eICSS depending on how the effect of the electrical stimulation is portrayed within the TDRL model. The diagrams in Figure 54 illustrates two different accounts of how the eICSS electrode activates DA neurons, within the framework on the TDRL model. The top row shows the computation of the TD, the middle row shows the computation of the TD error (RPE), and the bottom row shows the consequent changes in the weights of reward predictions and reward-seeking actions. When Montague, Dayan & Sejnowski (1996) first presented the TDRL algorithm to show how midbrain DA neurons could compute an RPE, they proposed that electrical stimulation of the MFB stimulates midbrain DA neurons directly, thus mimicking an RPE and boosting reward seeking. This account places the signal evoked by the electrode at the stage that relays the RPE to the weight-changing mechanism, as shown by the electrode icon in Figure 54A. Direct DA activation overrides incoming inputs to midbrain DA neurons because it acts after the computation of the RPE. The resulting, persistent RPE updates the value function at time, *t-1,* but the weight adjustments fail to affect the output of the midbrain DA neurons because of the artificial RPE introduced by the electrical stimulation. Under this view, optical stimulation (depicted by the blue laser on the left diagram) and electrical stimulation both increase the weight assigned to the reward-seeking action with each stimulation train and accelerate the rate of work until a maximum is reached.

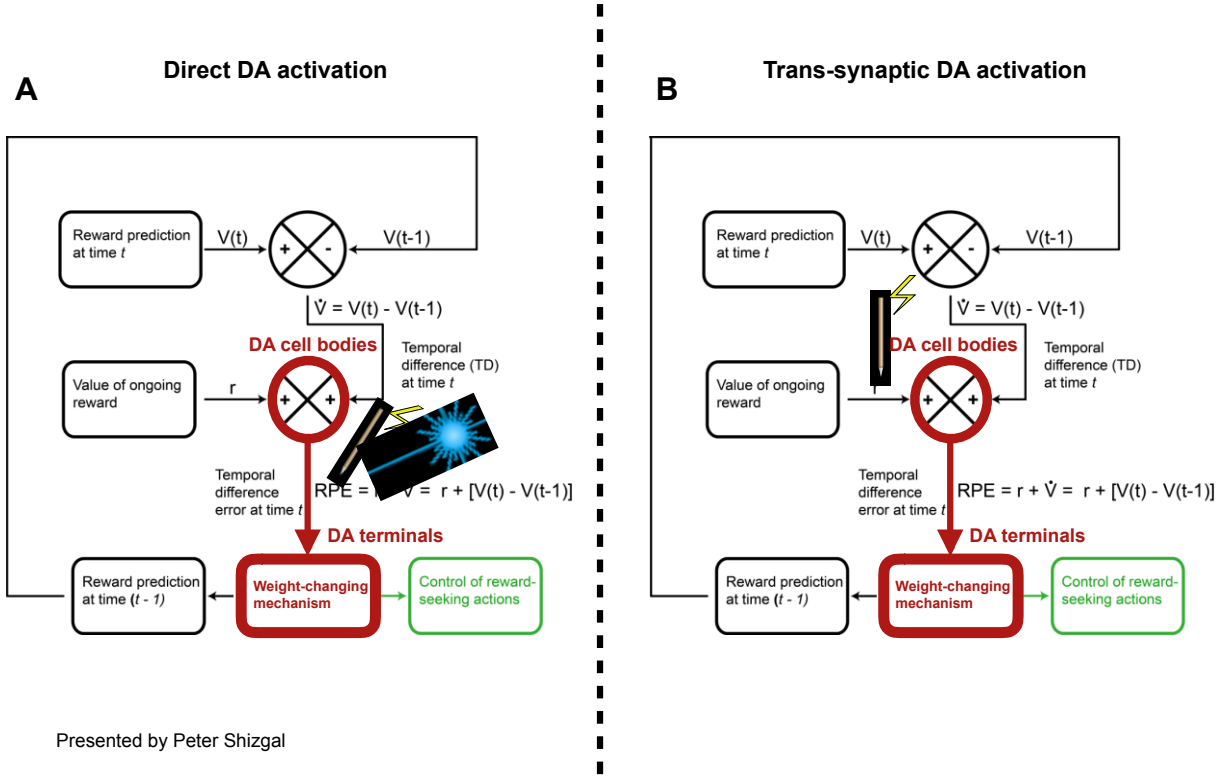


Figure 54. Direct and indirect DA activation in the context of the DA-RPE hypothesis

Schemata depicting the generation of stimulation-induced RPEs in eICSS when the electrical stimulation of the MFB stimulates midbrain DA neurons directly (A) or via trans-synaptic input (B). The top row shows the computation of the TD, the middle row shows the computation of the TD error (RPE), and the bottom row shows the consequent changes in the weights of reward predictions and reward-seeking actions. The electrode icon is positioned over the signal it is assumed to mimic: an RPE in A, and a primary reward signal in B. Panel A depicts the events assumed by Montague, Dayan & Sejnowski (1996). Direct activation of DA axons bypasses incoming inputs to midbrain DA neurons and acts beyond the point at which an RPE would be computed in response to delivery of a natural reward. Given the assumptions discussed in the main text, the same prediction would be made for optical stimulation

of midbrain DA neurons, as indicated by the icon of a blue laser. An alternative scenario is shown in B, based on numerous studies implying that the rewarding effect produced by electrical stimulation of the MFB is due mainly or exclusively to indirect DA activation (Yeomans, Kofman & McFarlane, 1985; Yeomans, 1989; Murray & Shizgal, 1994; Shizgal et al., 1980; Wise, 1980; Maeda & Mogenson, 1981) and on an hypothesis by Shizgal and colleagues (Conover & Shizgal, 1994; Conover, Woodside & Shizgal, 1994; Shizgal, 1997) that the signal produced by the neurons activated by rewarding MFB stimulation constitutes a primary reward, *r(t)*. In panel B, the electrically induced, primary reward signal can be nulled by inhibitory inputs representing the reward prediction. No such nulling occurs in the schema depicted in panel A. DA, dopamine; DA-RPE, dopamine-reward-prediction-error; eICSS, electrical intracranial self-stimulation; TD, temporal difference; MFB, medial forebrain bundle.

The depiction in Figure 54A does not correspond to what is known about the neural substrate for electrical ICSS of the MFB. It is firmly established (and was already well known at the time the DA-RPE hypothesis was proposed) that the rewarding effect produced by electrical stimulation of the MFB is due mainly or exclusively to the activation of non-DA neurons at the level of the electrode tip. The stimulation activates midbrain DA neurons mainly or exclusively via trans-synaptic inputs (Yeomans, Kofman & McFarlane, 1985; Yeomans, 1989; Murray & Shizgal, 1994; Shizgal et al., 1980; Wise, 1980; Maeda & Mogenson, 1981). Shizgal and colleagues hypothesized (Conover & Shizgal, 1994 ; Conover, Woodside & Shizgal, 1994; Shizgal, 1997) that the signal produced by the neurons activated by rewarding MFB stimulation constitutes a primary reward, *r(t)* (called the output of the reward function in the RL literature). That places the reward signal triggered by the electrode prior to the computation of the RPE, where it can be countered by inhibitory inputs representing the reward prediction. This is shown by the electrode icon in Figure 54B. If initially the value function is zero, which implies that the TD signal, *V(t) - V(t-1)*, is also zero, then the RPE will be the same size as the output of the reward function, *r(t)*, as is the case with receipt of an unexpected reward. At time *t-1*,RPEs will update the value function, ***V(t-1)***, and the prediction will back-propagate to the earliest reliable predictor. As the reward function, *r(t)*, produced by the electrical stimulation becomes fully predicted, the TD

signal, *V(t) - V(t-1)*, takes a negative value of the same amplitude as the reward function, *r(t)*. Thus, the inhibitory TD signal will counter the effect of the electrical stimulation, and the amplitude of the resulting RPEs will be zero. According to the DA-RPE hypothesis, simulation-induced DA transients in the schema shown in Figure 54B should gradually cancel themselves away during eICSS. This prediction was not confirmed by our observations: the electrical stimulation elicited persistent DA transients with stable amplitude.

In Figure 54B, the indirect DA activation evoked by electrical stimulation produces DA transients by introducing a reward signal, *r(t)*. Alternative explanations can easily be conceived. For example, the electrical stimulation could activate neurons carrying the TD signal, *V(t) - V(t-1)*. In other words, the stimulation would act as a fictive prediction rather than as a fictive reward. If the activated neurons output a positive TD signal, in the absence of a reward signal, *r(t)*, it would result in a positive RPE at every stimulation delivery earned by the animal. Under this hypothesis, the fictive TD signal produces a fictive RPE that updates weights on the value function, *V(t)*, and on action selection. The electrical stimulation overrides changes in the value function, *V(t)*, and continues to produce a positive RPE throughout the trials. The prediction becomes the same as the one originally proposed for oICSS; that is, persistent simulation-induced DA transients should be observed, and the resulting weight changes are expected to yield accelerating work rates. Thus, this explanation can account for our neurochemical finding of persistent stimulation-induced DA transients, but not for our behavioural finding of stable, non-maximal work rates.

Could phasic DA signalling act as an RPE only some of time?

One hypothesis offered by Hamid et al. (2016) suggests that changes in phasic DA release can both contribute to learning and play a motivational role in reward seeking. They view the time course of the DA concentration as a continuously varying signal that can be demultiplexed into functionally distinct components. The instantaneous DA concentration represents how much a rewarding outcome is worth pursuing, a quantity they call the value of work. In contrast, abrupt changes in the DA concentration are used as learning signals (RPEs) that adjust reward predictions and action proclivities.

A recent perspective paper by Berke (2018) reformulated the proposition made by Hamid et al. with a particular emphasis on a window of opportunity for new learning. In this view, local microcircuit mechanisms at DA terminals modulate synaptic plasticity to determine if, based on learning opportunities, changes in phasic DA activity can act as an RPE. Applying this logic to our experiment, persistent simulation-induced DA transients could initially bias animals to lever press by acting as an RPE and strengthen weights on action selection but then lose its ability to update weights as trials progress. Given this scenario, oICSS could be sustained through the remaining trials due to a high value of work signalled by increase in phasic DA release and no longer be influenced by RPEs.

Berke attempts to reconcile data implicating DA activity in processes such as incentive salience, facilitation of flexible approach patterns, and investment of effort in procuring rewards (Berridge & Robinson, 1998; Ikemoto & Panksepp 1999; Salmon & Correa, 2002) with the DA-RPE hypothesis. However, we can offer an anecdotal piece of evidence that casts some doubt on the idea that the window for new learning is restricted to trial onset. When we first began training animals on oICSS, lasers failed or patch cords detached from the headgear in the middle of some sessions. Well-trained animals, after a few lever presses that failed to deliver an optical stimulation, quickly stopped responding and chose instead to rest. Such abrupt adjustment of behaviour to the change in stimulation strength cannot be easily explained, unless learning mechanisms are continuously available to update the value function and the proclivity to engage in oICSS. Therefore, the dual function proposed for phasic DA release by Hamid and colleagues (2016) and Berke (2018) may not succeed in rescuing the DA-RPE hypothesis.

Functional specialization of midbrain DA neurons

Whereas Berke proposes that phasic DA release can carry multiple DA signals, other researchers focus on the heterogeneity of subpopulations of midbrain DA neurons, their anatomical differences, and the distinct roles some of these subpopulations might play in reward-related contexts (Ikemoto, 2007; Lammel, Hackel, Jones et al., 2008; Lammel, Ion, Roeper et al., 2011, Lammel, Lim, Ran et al., 2012). Moreover, phasic DA signalling, as measured by means of FSCV, differs throughout the striatal terminal fields. For example, in untrained rats, unpredicted food delivery reliably evokes DA transients in the NAc core but not in the NAc shell, the DMS,

or dorsolateral striatum (Brown, McCutcheon, Cone et al., 2011; Stalnaker, Calhoon, Owaga et al., 2012). In trained animals, unexpected food delivery and food-predictive cues evoked DA transients in the NAc core and in the DMS exclusively (Brown et al., 2011). FSCV recordings obtained throughout the striatum would determine to which extent simulation-induced DA transients resemble an RPE in the different subpopulations of midbrain DA neurons during oICSS and eICSS. Monitoring the post-synaptic responses in the DA terminal fields, by means of fiber photometry or molecular engineering to alter their fluorescence as a function of DA concentration, should also provide crucial information about the behavioral importance of DA signalling. We cannot rule out the possibility that DA responses in terminal fields other than the ones sampled in the present study might better match the requirements of the DA-RPE hypothesis.

Optical stimulation produces both transient and sustained increases in phasic DA release

Optogenetic stimulation of midbrain DA neurons is particularly well-suited to generate rapid increases in phasic DA release, one of the main reasons why it is routinely used to study the DA-RPE hypothesis. DA stimulation is also said to slowly increase DA neurotransmission on a timescale of minutes, as measured by microdialysis (Martig, Underhill, Defrancesco et al., SFN poster, 2013); however, it is unclear to what degree this reflects the limitations of the measurement method and to what degree it reflects the dynamics of neural signalling. Gradual changes in phasic DA release on a timescale of seconds have also been observed during reward seeking (Fiorillo, Tobler & Schultz, 2003; Howe et al., 2013; Fonzi et al., 2017, Hamid et al., 2016). To our knowledge, we are the first to report a gradual increase in phasic DA release in response to optical stimulation of midbrain DA neurons.

Slow and rapid increases in phasic DA release and everything in between

A gradual increase in phasic DA release on the timescale of seconds does not fit well with the traditional dichotomized slow tonic and rapid phasic patterns of DA neurotransmission. The latter is believed to correspond to DA burst firing, time-locked to specific external or internal stimuli (Gonon, 1988; Wightman et al., 1988; Kawagoe & Wightman, 1994; Kilpatrick et al., 2000) whereas the former has been attributed to changes in the proportion of DA neurons in the "ON" state (i.e., neurons that are not silenced by profound hyperpolarization) (Grace & Bunney,

1984; Floresco, West, Ash et al., 2003). However, Berke (2018) questions the tonic-phasic dichotomy. Slow changes in DA release correlate with locomotion and other indices of motivation, but there is no evidence, according to him, that variation in tonic firing rates account for these correlations, and there is no reason to assume an inherently slow DA signal.

In our experiment, the initial stimulation-induced DA transients are best described as either a superimposition of sharp, phasic changes onto a distinct, more gradual change in DA release or a set of step-wise changes. A more advanced analysis of the signal must be performed to better describe these two types of changes. Nevertheless, this finding suggests the existence of two separate DA signals. Berke (2018) makes a strong case for the lack of evidence supporting the categorization of DA neurotransmission into either slow or fast modes. Nonetheless, our data caution that DA neurotransmission can consist of multiple distinct DA signals. It is interesting to consider that the DA transients and the gradual increase in DA release produced by optical stimulation could play distinct roles in reward seeking but conjointly reinforce lever pressing during oICSS. That proposition remains to be examined.

The superimposition of sharp, phasic changes onto a distinct, more gradual change in DA release or a set of step-wise changes was observed during both oICSS and eICSS. Previous studies have reported a decline in the amplitude of stimulation-induced DA transients at the beginning of short eICSS sessions (Garris et al., 1999; Owesson-White et al., 2008). If our hypothesis is correct and the initial DA transients are a combination of sharp and gradual increases in DA release, it is possible that the sharp component approaches a constant peak amplitude and that the transition of the peak oxidation current from just prior to stimulation delivery to peak amplitude is largest initially because the gradual component has not yet grown. It seems likely that compensatory mechanisms, such as enhanced reuptake, limit the maximum DA concentration achieved. Thus, the greater the growth of the gradual component, the more difficult it is for an individual train to push the DA concentration transiently higher.

Our findings raises the possibility that the traditional method for extracting DA concentration from FSCV recordings might be misleading: It fails to distinguish and depict both the sharp and

gradual components of stimulation-induced changes in DA concentration. In contrast, both component are evident when the FSCV signal is analyzed by means of our method.

It is likely that behaviours other than ICSS might produce gradual or step-wise changes in DA concentration. If these behaviours coincide with DA transients evoked by reward-predictive cues or rewards, our ability to understand the link between DA signalling and reward seeking will require new FSCV analyses, such as the one we proposed above, that dissociate sharp from gradual changes in DA activity.

Conclusion

The advent of optogenetic techniques have given researchers tools to precisely control phasic DA activity on a timescale of milliseconds and to examine the viability of the DA-RPE hypothesis (Steinberg et al., 2013; Keiflin & Janak, 2015; Takahashi, Langdon, Niv et al., 2016; Nasser et al., 2017). There is a tendency in the field to assume sufficient evidence has been gathered to confirm the hypothesis. In the current experiment, we applied a novel approach using optical and electrical ICSS with simultaneous voltammetric recordings and quantitative behavioural measurements to impose a stringent test of the dominant perception that changes in phasic DA signalling act as RPEs to adjust reward prediction *and* action selection. We provided proof-of-principle evidence that violations of the DA-RPE hypothesis can be detected using our method, and we have proposed follow-up experiments to more definitively establish the role of phasic DA signalling in learning.

General Discussion

My first encounter with a self-stimulating rat provoked a mixture of wonder and curiosity—an animal was lever pressing to earn bursts of electricity aimed at the MFB—Just that. But why was the animal working so hard? What was the electricity mimicking in the brain to entice the animal into doing so. Since this first exposure, my fascination has extended to the mechanisms that might cause organisms to approach or forgo rewards and the challenges entailed in unraveling them. ICSS is a powerful experimental paradigm particularly suitable to test, with empirical precision existing theories about the fundamental principles of reward seeking. We focused on the role of phasic DA activity during ICSS.

Chapter one

Chapter one examine the circuitry linking midbrain DA neurons to the non-DA neurons recruited at the tip of the MFB self-stimulation electrode. We observed the output of DA neurons under controlled conditions and found that unilateral, electrical, MFB stimulation evoked DA transients in both hemispheres. We extended the range of stimulation parameters typically used to evoke DA transients and detected contralateral DA transients over the entire range of parameters that also induced ipsilateral transients; the amplitudes of these transients in the two hemispheres were similarly distributed. Given what is known about the physiology and anatomy of the MFB, these findings suggest that action potentials triggered in non-DA neurons are relayed to midbrain DA cell bodies, largely through polysynaptic circuitry. We have proposed candidate neuronal components of this circuitry as well as optogenetic experiments to assess the contributions of these candidate neurons to the effects of MFB stimulation on phasic DA activity.

Our findings raise the question: why might the input circuitry be configured such that a MFB signal in one hemisphere is relayed to midbrain DA neurons in both hemispheres? Fox et al. (2016) observed that most spontaneously-generated DA transients are synchronized between hemispheres. They proposed that such a synchronization might optimize sensitivity to the precise timing of synaptic inputs and contribute to synchronous neural patterns; that is, the inter-hemispheric connections might play a key role in synchronizing DA transients. Further investigation of inter-hemispheric synchrony in DA firing is merited.

Our work bridges two approaches to the study of eICSS that have previously been largely separate: experiments carried out applying psychophysical or neurochemical methods at two different levels of the MFB. We closed the empirical gaps, thus ensuring that behavioural findings obtained using LH stimulation could better parallel neurochemical data obtained using VTA stimulation. We also provided neurochemical data for the whole range of stimulation parameters used in the two experimental contexts. We can now more confidently integrate data acquired using MFB stimulation.

Lastly, we used optical stimulation to selectively activate DA neurons to better understand the mechanisms linking contralateral DA transients to MFB stimulation. In this vein, procedures for limiting the spread of light or viral transfection were introduced. Nonetheless, we encountered several limitations. One in particular was the difficulty confirming that activated DA neurons are restricted to one hemisphere. Several variables can affect the reach of the optical stimulation in triggering light-gated proteins. To tease apart the contributions of these variables, we need better methods for visualizing optogenetically-activated neurons.

Chapter two

The idea that non-DA neurons provide inputs to midbrain DA neurons critical for reward seeking is the basis of the series-circuit hypothesis, the focus of the second chapter. On that view, the signal representing the intensity of the stimulation-induced rewarding effect must obligatorily pass through midbrain DA neurons, even if it arises in their direct or indirect afferents. For the series-circuit hypothesis to be true, relationships between input variables that are manifested in the final behavioural output (e.g., lever pressing) must also be evident in the activity of the DA neurons. Using a psychophysical approach, we characterized the frequency response of the directly stimulated MFB neurons subserving the rewarding effect of electrical stimulation. Using a neurochemical approach, we measured the frequency response of the directly stimulated MFB neurons that drive DA transients in the NAc. According to the series-circuit hypothesis, these two frequency responses should be the same. We found, however, that they are not the same. This discrepancy suggests that there is a route from the MFB stimulation site to the behavioural final-common path that bypasses midbrain DA neurons.

The directly stimulated neurons responsible for the rewarding effects of MFB stimulation have yet to be identified. Finding these neurons remains a high priority for future research. A promising avenue to this goal entails a combination of psychophysical inference, optogenetic perturbation of neural signalling, and direct measurement of neural activity. The crucial psychophysical component has heretofore been largely neglected in optogenetic research on brain reward circuitry. That this omission is unfortunate was presaged by the insightful observation of J.C.A. Read (2015): "Physiology is fascinating in its own right, but acquires its full meaning and significance when related to perceptual experience by psychophysics."

Previous attempts to link DA signalling to the computation and relay of the reward signal were predicated on the assumption that an increase in the amplitude of DA transients should correspond to an increase in eICSS. Instead, we asked whether the psychophysically inferred characteristics of the MFB neurons subserving eICSS are reflected in the activity of DA neurons. This more demanding test revealed that DA transients do not fully replicate the psychophysically inferred frequency response of the directly stimulated substrate. This discrepancy has important implications for the structure of brain reward circuitry and suggests novel experimental directions to pursue. The pairing of psychophysics with direct neural measurements might also be applied profitably in the search for the neural circuitry responsible for the rewarding effects of natural goal objects.

Chapter three

The last chapter addresses the dominant theory linking phasic DA signalling to learning. A highly influential machine-learning algorithm, the "temporal-difference reinforcement-learning" (TDRL) algorithm, employs reward-prediction error signals (RPEs) to drive learning. The firing rate of midbrain DA neurons appears to encode such RPEs, a finding that led to the hypothesis that DA-mediated RPEs drive learning (DA-RPE hypothesis) (Montague, Dayan & Sejnowski, 1996; Schultz, 2015). On that view, DA transients signal the discrepancy between expected and experienced rewards, adjusting synaptic weights to update reward predictions *and* bias action selection.

We tested this premise using selective optical stimulation of midbrain DA neurons to evoke DA transients and elicit oICSS. More precisely, we asked if the learning observed during oICSS could be caused by stimulation-evoked DA transients. We first determined that optical stimulation of midbrain DA neurons produces persistent DA transients throughout long oICSS sessions. Then, we used optical stimulation parameters, presented in an unpredictable order, that elicited different work rates (maximal, non-maximal and minimal). The randomized component forced the animal to update its seeking behaviour according to the strength of the optical stimulation (different pulse frequencies produced DA transients of different amplitudes). Trials with non-maximal work rates were examined to compare the persistence of DA transients with the steadiness of oICSS. The DA-RPE hypothesis predicts that persistent DA transients will act as RPEs and bias lever pressing, accelerating the rate of work until maximal oICSS is reached. We did not observed such acceleration. We also determined that electrical stimulation of MFB at the level of the LH produces persistent DA transients throughout long eICSS sessions. The DA-RPE hypothesis predicts that the DA transients, and the errors they encode, will "predict themselves away" over the course of repeated reward encounters under constant conditions. We saw no evidence of this. Moreover, in the face of repeated DA transients of stable amplitude, the engagement in reward seeking should have accelerated. Again, no evidence of this was seen.

There is a tendency for researchers who study the link between DA activity and reward seeking to assume that there is already sufficient evidence to confirm the DA-RPE hypothesis (DA transients can update reward value and bias action selection) (Steinberg et al., 2013; Colombo, 2014; Schultz, 2015; Keiflin & Janak, 2015; Chang et al., 2016; Chang et al., 2017; Chang et al., 2018; Parker et al., 2016; Watabe-Uchida, 2017; Nasser et al., 2017; Sharpe et al., 2017; Keiflin et al., 2019). Researchers have used optogenetic techniques to precisely control phasic DA activity to examine the viability of the DA-RPE hypothesis (Steinberg et al., 2013; Keiflin & Janak, 2015; Takahashi et al., 2009; Nasser et al., 2017). Most experiments used a blocking design in which optical stimulation of DA neurons was delivered to mimic an RPE, during the learning of the contingency between cues and a rewarding outcome. Reward seeking was enhanced in groups that received the optical stimulation, and that outcome was deemed as a confirmation that DA-mediated RPEs drive learning. We used a different approach. We trained an animal to perform oICSS using a specific trial structure and set of stimulation parameters

designed to test if accelerating work rates would be obtained, as predicted by the DA-RPE hypothesis. This stringent test revealed a discrepancy between DA signalling and predictions under the DA-RPE hypothesis. Less restrictive approaches have brought valuable evidence to the field of reward learning. However, a switch toward more stringent tests is becoming increasingly important as the DA-RPE theory has gained such wide acceptance and is used as the foundation for new experiments.

Another important finding in chapter three, although not central to the research question, is that fluctuations in phasic DA activity can be measured without resorting to background subtraction. Implementation of background subtraction can be problematic because of unavoidable arbitrariness in determining what constitutes a background. The process can also filter out or distort valuable variations in DA activity. Omitting this transformation resolved DA transients time-locked to optical stimulation of DA neurons, as expected. In addition, it revealed a gradual increase in phasic DA activity that began in response to the first stimulation received at the onset of trials in which the animal worked at more than minimal rates and then levelled off after several seconds, a pattern of DA release that has yet to be documented. We will continue to examine gradual increases in phasic DA activity concomitant with the stimulation-evoked DA transients and hope to develop a procedure to relate such changes to behavioural components of reward seeking.

Concluding remarks

We are often like the hypothetical observer described in the general introduction. We examine a small subset of behaviours and we try to determine the causal mechanisms that drive these behaviours. We have come a long way, mapping the brain circuitry implicated in reward seeking with increasing precision, monitoring fluctuations in neurotransmission in real-time while animals work to harvest rewards, and precisely mimicking patterns of activation in specific subtype of neurons to test learning theories. Despite the unprecedented proliferation of new ways to probe our existing theories, we still need to carefully design experiments to implement demanding tests of working hypotheses. We have proposed ways to do so. Being the best observers that will help us reach the most meaningful conclusions.

While reflecting on whether we will ever understand why a self-stimulating rat works so hard, I knew I was also asking why we all work so hard. I was aware of the conundrum that we all face; we need our brains to understand brain processes. I now realize how much we also need to find research rewarding to glimpse at reward mechanisms, a thought that inspired me throughout my doctoral work.

REFERENCES

Adamantidis, A. R., Zhang, F., Aravanis, A. M., Deisseroth, K. & De Lecea, L. (2007). Neural substrates of awakening probed with optogenetic control of hypocretin neurons. *Nature, 450*, 420–424.

Akerboom, J., Carreras Calderón, N., Tian, L., Wabnig, S., Prigge, M., Tolö, J., Gordus, A., Orger, M. B., Severi, K. E., Macklin, J. J., Patel, R., Pulver, S. R., Wardill, T. J., Fischer, E., Schüler, C., Chen, T.-W., Sarkisyan, K, S., Marvin, J. S., Bargmann, C. I., KimD. S., Kügler, S., Lagnado, L., Hegemann, P., Gottschalk, A., Schreiter, E. R. & Looger, L. L. (2013). Genetically encoded calcium indicators for multi-color neural activity imaging and combination with optogenetics. *Frontiers in Molecular Neuroscience*, *6*, 1–29.

Albert, K. A., Helmer-Matyjek, E., Nairn, A. C., Müller, T. H., Haycock, J. W., Greene, L. A., Goldstein, M. & Greengard, P. (1984). Calcium/phospholipid-dependent protein kinase (protein kinase C) phosphorylates and activates tyrosine hydroxylase. *Proceedings of the National Academy of Sciences of the United States of America*, *81*(24), 7713–7717.

Aransay, A., Rodriguez-Lopez, C., Garcia-Amado, M., Clasca, F. & Prensa, L. (2015). Long-range projection neurons of the mouse ventral tegmental area: a single-cell axon tracing analysis. *Frontiers in Neuroanatomy*, *9*, 1–24.

Aravanis, A. M., Wang, L.-P., Zhang, F., Meltzer, L. A., Mogri, M. Z., Schneider, M. B. & Deisseroth, K. (2007). An optical neural interface: in vivo control of rodent motor cortex with integrated fiberoptic and optogenetic technology. *Journal of Neural Engineering, 4*(3), S143-156.

Arvanitogiannis, A., Flores, C., Pfaus, J. G. & Shizgal, P. (1996). Increased ipsilateral expression of Fos following lateral hypothalamic self-stimulation. *Brain research*, *720(1-2)*, 148-54.

Arvanitogiannis, A., Flores, C. & Shizgal, P. (1997). Fos-like immunoreactivity in the caudal

diencephalon and brainstem following lateral hypothalamic self-stimulation. *Behavioural Brain Research, 88*(2), 275-279.

Barto, A. G. (1995). Adaptive critics and the basal ganglia. In Houk, J. C., Davis J. L. & Beiser, D. G. (Eds.), Models of Information Processing in the Basal Ganglia, pp. 215–232. MIT Press, Cambridge, MA.

Barto, A. G., Sutton, R. S. & Anderson, C. W. (1983). Neuron-like elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics*, *13*(5):835–846.

Bement, S. L. & Ranck, J. B. (1969). Quantitative Study Central of Electrical Stimulation Fibers of Myelinated. *Experimental neurology, 24*, 147–170.

Berke, J. D. (2018). What does dopamine mean? *Nature Neuroscience*, *21*(6), 787–793.

Bernstein, J. G. & Boyden, E., S. (2011). Optogenetic tools for analyzing the neural circuits of behavior. *Trends in cognitive Science, 15*(12), 592-600.

Bernstein, J. G., Han, X., Henninger, M. A., Ko, E. Y., Qian, X., Franzesi, G. T., McConnell, J. S., Stern, P., Desimone, R. & Boyden, E. (2008). Prosthetic systems for therapeutic optical activation and silencing of genetically targeted neurons. *Proceedings of Society of Photo-Optical Instrumentation Engineers*, *6854,* 68540H.

Berridge, K. C. & Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain research. Brain research reviews*, 28(3), 309-69.

[40] Beyene, M., Carelli, R. M. & Wightman, R. M. (2010). Cue-evoked dopamine release in the nucleus accumbens shell tracks reinforcer magnitude during intracranial self-stimulation. *Neuroscience*, *169(4)*, 1682-8.

[19] Bielajew, C. & Shizgal, P. (1982). Behaviorally derived measures of conduction velocity in the substrate for rewarding medial forebrain bundle stimulation. *Brain research*, *237*, 107-119.

[20] Bielajew, C. & Shizgal, P. (1986). Evidence implicating descending fibers in self-stimulation of the medial forebrain bundle. *The Journal of neuroscience, 6(4)*, 919-929.

[35] Blaha, C. D. & Phillips, A. G. (1990). Application of in vivo electrochemistry to the measurement of changes in dopamine release during intracranial self-stimulation. *Journal of neuroscience Methods, 34,* 125-133.

Blaha, C. D. & Phillips, A. G. (1996). A critical assessment of electrochemical procedures applied to the measurement of dopamine and its metabolites during drug-induced and species-typical behaviours. *Behavioural pharmacology, 7,* 675-708.

Boyden, E. S. (2015). Optogenetics and the future of neuroscience. *Nature Neuroscience, 18,* 1200-1201.

[30] Breton, Y.-A., Marcus, J. C. & Shizgal, P. (2009). Rattus Psychologicus: construction of preferences by self-stimulating rats. *Behavioural brain research*, *202(1)*, 77-91.

Brown, H. D., McCutcheon, J. E., Cone, J. J., Ragozzino, M. E. & Roitman, M. F (2011). Primary food reward and reward-predictive stimuli evoke different patterns of phasic dopamine signaling throughout the striatum. *European journal of neuroscience, 34*(12), 1997-2006.

Burnham, K. P. & Anderson, D. R. (2002). Model Selection and Inference: A Practical Information-Theoretic Approach. 2nd Edition, New York: Springer-Verlag.

Bush, R. R. & Mosteller, F. (1953). A stochastic model with applications to learning. *Annals of Mathematical Statistics*, *24*, 559-585.

Chang, C. Y., Esber, G. R., Marrero-garcia, Y., Yau, H., Bonci, A. & Schoenbaum, G. (2016). Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. *Nature Neuroscience, 19*(1), 111–116.

Chang, C. Y., Gardner, M., Di Tillio, M. G. & Schoenbaum, G. (2017). Optogenetic Blockade of Dopamine Transients Prevents Learning Induced by Changes in Reward Features. Current Biology, 27(22), 3480–3486.

Chang, C. Y., Whitaker, L. R., Conroy, J. C., Gardner, M. P. H. & Schoenbaum, G. (2018). Brief, But Not Prolonged, Pauses in the Firing of Midbrain Dopamine Neurons Are Sufficient to Produce a Conditioned Inhibitor. The Journal of Neuroscience, 38(41), 8822–8830.

[41] Cheer, J. F., Aragona, B. J., Heien, M. L., Seipel, A. T., Carelli, R. M. & Wightman, R. M. (2007). Coordinated accumbal dopamine release and neural activity drive goal-directed behavior. *Neuron, 54*(2), 237-44.

Chernov, K. G., Redchuk, T. A., Omelina, E. S. & Verkhusha (2017). Near-Infrared Fluorescent Proteins, Biosensors, and Optogenetic Tools Engineered from Phytochromes. *Chemical Reviews, 117*(9), 6423-6446.

Chow, B. Y., Han, X., Dobry, A. S., Qian, X., Chuong, A. S., Li, M., Henninger, M. A., Belfort, G. M., Lin, Y, Monahan, P. E. & Boyden, E. S. (2010). High-performance genetically targetable optical neural silencing by light-driven proton pumps. *Nature, 463*, 98-102.

[49] Chuhma, N., Zhang, H., Massion, J., Zhuang, X., Sulzer, D. & Rayport, S. (2004) Dopamine neurons mediate a fast excitatory signal via their glutamatergic synapses. *Journal of Neuroscience, 24*(4), 972-81.

[50] Chuhma, N., Choi, W. Y., Mingote, S. & Rayport, S. (2009). Dopamine neuron glutamate cotransmission: frequency-dependent modulation in the mesoventromedial projection.

Neuroscience, 164, 1068-83.

[29] Clark, J. J., Sandberg, S. G., Wanat, M. J., Gan, J. O., Hart, A. S., Akers, C. A., Parker, J. G., Willuhn, I., Martinez. V., Evans, S. B., Stella, N. & Phillips, P. E. (2010). Chronic microsensors for longitudinal, subsecond dopamine detection in behaving animals. *Nature Methods, 7(2)*, 126-129.

Colombo, M. (2014). Deep and beautiful. The reward prediction error hypothesis of dopamine. *Studies in History and Philosophy of Science Part C : Studies in History and Philosophy of Biological and Biomedical Sciences*, *45*(1), 57–67.

[4] Conover, K. L. & Shizgal, P. (1994). Competition and summation between rewarding effects of sucrose and lateral hypothalamic stimulation in the rat. *Behavioral neuroscience*, *108(3)*, 537-48.

[5] Conover, K. L., Woodside, B. & Shizgal, P. (1994). Effects of sodium depletion on competition and summation between rewarding effects of salt and lateral hypothalamic stimulation in the rat. *Behavioral neuroscience*, *108(3)*, 549-58.

Cornwall, J., Cooper, J. D. & Phillipson, O. T. (1990). Afferent and efferent connections of the laterodorsal tegmental nucleus in the rat. *Brain Research Bulleting, 25*, 271-284.

Cossette, M.-P. & Shizgal, P. (2015). Contrasting effects of rewarding electrical and optical stimulation (Poster). Chicago: The annual meeting for the Society for Neuroscience.

[7] Crow, T. J. (1970). Enhancement by cocaine of intra-cranial self-stimulation in the rat. *Life* science*, 9,* 375-381.

Cruz, F. C., Koya, E., Guez-barber, D. H., Bossert, J. M., Lupica, C. R., Shaham, Y. & Hope, B. T. (2013). New technologies for examining the role of neuronal ensembles in drug addiction and fear. *Nature Reviews. Neuroscience, 14*(11), 743–754.

Daubner, S. C., Le, T. & Wang, S. (2011). Tyrosine Hydroxylase and Regulation of Dopamine Synthesis. *Archives of Biochemistry and Biophysics*, *508*(1), 1–12.

Dautan, D., Huerta-Ocampo, I., Witten, I. B., Deisseroth, K., Bolam, J. P., Gerdjikov, T. & Mena-Segovia, J. (2014). A Major External Source of Cholinergic Innervation of the Striatum and Nucleus Accumbens Originates in the Brainstem. *Journal of Neuroscience*, *34*(13), 4509–4518.

Dayan, P. & Niv, Y. (2008). Reinforcement learning: The Good, The Bad and The Ugly. *Current Opinion in Neurobiology*, *18*(2), 185–196.

Deisseroth, K. (2011). Optogenetics. *Nature Methods, 8,* 26-29.

Doll, B. B., Simon, D. A. & Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Current opinion in neurobiology*, *22*(6):1075-81

Doya, K. (1999). What are the Computations of the Cerebellum, the Basal Ganglia, and the Cerebral Cortex? *Neural Networks*, *12*, 961–974.

Doya, K. (2008). Modulators of decision making. *Nature neuroscience, 11*(4), 410-416.

Egelman, D. M., Person, C. & Montague, P. R. (1998). A Computational Role for Dopamine Delivery in Human Decision-Making. *Journal of Cognitive Neuroscience*, *10*(5), 623–630.

Fiorillo, C. D., Tobler, P. N. & Schultz, W. (2003). Discrete Coding of Reward Dopamine Neurons. *Science, 299*, 1898–1902.

[33] Fiorino, D. F., Coury, A., Fibiger, H. C. & Phillips, A. G. (1993). Electrical stimulation of reward sites in the ventral tegmental area increases dopamine transmission in the nucleus accumbens of the rat. *Behavioural brain research*, *55(2)*, 131-41.

Floresco, S. B., West, A. R., Ash, B., Moore, H. & Grace, A. A. (2003). Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission. *Nature neuroscience*, *6(9)*, 968-73.

Fonzi, K. M., Lefner, M. J., Phillips, P. E. M. & Wanat, M. J. (2017). Dopamine Encodes Retrospective Temporal Information in a Context-Independent Manner. *Cell reports, 20*(8), 1765-1774.

Forster, G. L. & Blaha, C. D. (2000). Laterodorsal tegmental stimulation elicits dopamine efflux in the rat nucleus accumbens by activation of acetylcholine and glutamate receptors in the ventral tegmental area. *The European journal of neuroscience*, *12(10)*, 3596-604.

Forster, G. L. & Blaha, C. D. (2003). Pedunculopontine tegmental stimulation evokes striatal dopamine efflux by activation of acetylcholine and glutamate receptors in the midbrain and pons of the rat. *European journal of neuroscience, 17*(4), 751-62.

Fosque, B. F., Sun, Y., Dana, H., Yang, C. T., Ohyama, T., Tadross, M. R., Patel R., Zlatic M., Kim D. S., Ahrens M. B., Jayaraman V., Looger L. L. & Schreiter, E. R. (2015). Labeling of active neural circuits in vivo with designed calcium integrators. *Science*, *347*(6223), 755–760.

Fox, M. E., Mikhailova, M. A., Bass, C. E., Takmakov, P., Gainetdinov, R. R., Budygin, E. A. & Wightman, R. M. (2016). Cross-hemispheric dopamine projections have functional significance. *Proceedings of the National Academy of Sciences*, *113*(25), 6985–6990.

[27] Gallistel, C. R. (1974). Note on temporal summation in the reward system. *Journal of comparative and physiological psychology, 87*(5), 870-875.

[28] Gallistel, C. R. (1978). Self-stimulation in the rat: quantitative characteristics of the reward pathway. *Journal of comparative and physiological psychology, 92(6),* 977-998.

Gallistel, C. R. (1988). Determining the quantitative characteristics of a reward pathway. In M. L. Commons, R. M. Church, J. R. Stellar, & A. R. Wagner (Eds.), Quantitative analyses of behavior, Vol. 7. Biological determinants of reinforcement (pp. 1-30). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

[14] Gallistel, C. R. & Karras, D. (1984). Pimozide and amphetamine have opposing effects on the reward summation function. *Pharmacology, biochemistry, and behavior*, *20*(1), 73-7.

[25] Gallistel, C. R., Shizgal, P. & Yeomans, J. S. (1981). A portrait of the substrate for self-stimulation. *Psychological review*, *88(3)*, 228-73.

Gallistel, C. R., Stellar, J. R. & Bubis, E. (1974). Parametric analysis of brain stimulation reward in the rat: I. The transient process and the memory-containing process. *Journal of comparative and physiological psychology*, *87*(5), 848-859.

Gan, J. O., Walton, M. E. & Phillips, P. E. M. (2010). Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. *Nature Neuroscience*, *13*(1), 25–27.

Garrigues, A. M. & Cazala, P. (1983). Central catecholamine metabolism and hypothalamic self-stimulation behaviour in two inbred strains of mice. *Brain research*, *265(2)*, 265-71.

[44] Garris, P. A., Kilpatrick, M., Bunin, M. A., Michael, D., Walker, Q. D. & Wightman, R. M. (1999). Dissociation of dopamine release in the nucleus accumbens from intracranial self-stimulation. *Nature*, *398*(6722), 67–69.

[39] Garris, P. A. & Rebec, G. V. (2002). Modeling fast dopamine neurotransmission in the nucleus accumbens during behavior. *Behavioural brain research, 137,* 47-63.

Geisler, S. & Zahm, D. S. (2005). Afferents of the ventral tegmental area in the rat-anatomical substratum for integrative functions. *Journal of Comparative Neurology*, *490*(3), 270–294.

[36] Gonon, F. G. (1988). Nonlinear relationship between impulse flow and dopamine released by rat midbrain dopaminergic neurons as studied by in vivo electrochemistry. *Neuroscience*, *24*(1), 19-28.

Grace, A. & Bunney, S. (1984). The control of firing neurons in nigral dopamine neurons: single spike firing. *Journal of Neuroscience*, *4*(11), 2866-2876.

[34] Gratton, A., Hoffer B. J. & Gerhardt, G. A. (1988). Effects of electrical stimulation of brain reward sites on release of dopamine in rat: an in vivo electrochemical study. *Brain Research Bulletin*, *21*(2), 319-324.

[3] Green, L. & Rachlin, H. (1991). Economic substitutability of electrical brain stimu- lation, food, and water. *Journal of the experimental analysis of behavior, 55*(2), 133– 143.

Hamid, A. A., Pettibone, J. R., Mabrouk, O. S., Hetrick, V. L., Schmidt, R., Vander Weele, C. M., Kennedy, R.T., Aragona, B.J. & Berke, J. D. (2015). Mesolimbic dopamine signals the value of work. *Nature Neuroscience*, *19*(1), 117–126.

Heien, M. L., Johnson, M. A. & Wightman, R. M. (2004). Resolving Neurotransmitters Detected by Fast-Scan Cyclic Voltammetry. *Analytical chemistry, 76*(19), 5697–5704.

Herberg, L. J. & Rose, I. C. (1990). Excitatory amino acid pathways in brain-stimulation reward. *Behavioural brain research*, *39(3)*, 230-9.

[10] Hernandez, G., Breton, Y.-A., Conover, K. & Shizgal, P. (2010). At what stage of neural processing does cocaine act to boost pursuit of rewards? *PloS one*, *5*(11), e15081-e15081.

Hernandez, G., Cossette, M.-P., Shizgal, P. & Rompré, P.-P. (2016). Ventral Midbrain NMDA Receptor Blockade: From Enhanced Reward and Dopamine Inactivation. *Frontiers in Behavioral Neuroscience*, *10*, 1–11.

[11] Hernandez, G., Trujillo-Pisanty, I., Cossette, M.-P., Conover, K. & Shizgal, P. (2012). The Role of Dopamine Tone in the Pursuit of Brain Stimulation Reward. *Journal of Neuroscience*, *32*(32), 11032–11041.

Houk, J. C., Adams, J. L. & Barto, A. G. (1995). A model of how the basal ganglia generates and uses neural signals that predict reinforcement. In J. C. Houk, J. L. Davis, and D. G. Beiser (Eds.), Models of Information Processing in the Basal Ganglia, pp. 249–270. MIT Press, Cambridge, MA.

Howe, M. W., Tierney, P. L., Sandberg, S. G., Phillips, P. E. M. & Graybiel, A. M. (2013). Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature*, *500*(7464), 575–579.

Hunt, G. E. & McGregor, I. S. (1998). Rewarding brain stimulation induces only sparse Fos-like immunoreactivity in dopaminergic neurons. *Neuroscience*, *83*(2), 501-15.

Ikemoto, S. (2007). Dopamine reward circuitry: two projection systems from the ventral midbrain to the nucleus accumbens-olfactory tubercle complex. *Brain research reviews, 56(1),* 27-78.

Ikemoto, S. (2010). Brain reward circuitry beyond the mesolimbic dopamine system: a neurobiological theory. *Neuroscience & biobehavioral Review, 35*(2)*,* 129-150.

Ikemoto S. & Panksepp J. (1999). The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. Brain research. *Brain research reviews, 31*(1), 6-41.

Kahn, I., Desai, M., Knoblich, U., Bernstein, J., Henninger, M., Graybiel, A. M., Boyden, E. S., Buckner, R. L. & Moore, C. I. (2011). Characterization of the Functional MRI Response Temporal Linearity via Optical Control of Neocortical Pyramidal Neurons. *Journal of Neuroscience*, *31*(42), 15086–15091.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell and R. M. Church (Eds.), Punishment and Aversive Behavior. Appleton-Century-Crofts, New York.

[38] Kawagoe, K. T. & Wightman, R. M. (1994). Characterization of amperometry for in vivo measurement of dopamine dynamics in the rat brain. *Talanta*, *41(6)*, 865-74.

Keiflin, R. & Janak, P. H. (2015). Dopamine Prediction Errors in Reward Learning and Addiction: From Theory to Neural Circuitry. *Neuron*, *88*(2), 247–263.

Keiflin, R., Pribut, H. J., Shah, N. B. & Janak, P. H. (2019). Ventral Tegmental Dopamine Neurons Participate in Reward Identity Predictions. Current Biology, 29(1), 93–103.

Keithley, R. B., Takmakov, P., Bucher, E. S, Belle, A. M., Owesson-White, C. A., Park, J. & Wightman, R. M. (2011). Higher sensitivity dopamine measurements with faster-scan cyclic voltammetry. *Analytical Chemistry, 83*(9):3563-71.

Keithley, R. B. & Wightman, R. M. (2011). Assessing principal component regression prediction of neurochemicals detected with fast-scan cyclic voltammetry. *ACS Chemical. Neuroscience*, *2*(9), 514–525.

Kempadoo, K. A., Tourino, C., Cho, S. L., Magnani, F., Leinninger, G.-M., Stuber, G. D., Zhang, F., Myers, M. G., Deisseroth, K., De Lecea, L. & Bonci, A. (2013). Hypothalamic Neurotensin Projections Promote Reward by Enhancing Glutamate Transmission in the VTA. *Journal of Neuroscience*, *33*(18), 7618–7626.

[45] Kilpatrick, M. R., Rooney, M. B., Michael, D. J. & Wightman, R M. (2000). Extracellular dopamine dynamics in rat caudate-putamen during experimenter-delivered and intracranial self-stimulation. *Neuroscience*, *96(4)*, 697-706.

Kirouac, G. J. (2015). Placing the paraventricular nucleus of the thalamus within the brain circuits that control behavior. *Neuroscience and Biobehavioral Reviews*, *56*, 315–329.

Kishida, K. T., Saez, I., Lohrenz, T., Witcher, M. R., Laxton, A. W., Tatter, S. B., White, J. P., Ellis, T. L., Phillips, P.E. & Montague, P. R. (2016). Subsecond dopamine fluctuations in human striatum encode superposed error signals about actual and counterfactual reward. *Proceedings of the National Academy of Sciences*, *113*(1), 200–205.

[47] Kolodziej, A., Lippert, M., Angenstein, F., Neubert, J., Pethe, A., Grosser, O. S., Amthauer, H., Schroeder, U. H., Reymann, K. G., Scheich, H., Ohl, F. W. & Goldschmidt, J. (2014). SPECT-imaging of activity-dependent changes in regional cerebral blood flow induced by electrical and optogenetic self-stimulation in mice. *Neuroimage, 103*, 171-180.

Krukoff, T.L. (1993). Expression of c-fos in studies of central autonomic and sensory systems. *Molecular Neurobiology,7*(3-4), 247-263.

Lammel, S., Hetzel, A., Häckel, O., Jones, I., Liss, B. & Roeper J. (2008). Unique properties of mesoprefrontal neurons within a dual mesocorticolimbic dopamine system. *Neuron, 57*(5), 760-73.

Lammel, S., Ion, D. I., Roeper, J. & Malenka, R. C. (2011). Projection-specific modulation of dopamine neuron synapses by aversive and rewarding stimuli. *Neuron, 70*(5), 855–862.

Lammel, S., Lim, B. K., Ran, C., Huang, K. W., Betley, M. J., Tye, K. M., Deisseroth, K. & Malenka, R. C. (2012). Input-specific control of reward and aversion in the ventral tegmental area. *Nature, 491*(7423), 212–217.

[48] Lee, T., Cai, L. X., Lelyveld, V. S., Hai, A. & Jasanoff, A. (2014). Molecular-Level Functional Magnetic Resonance Imaging of Dopaminergic Signaling. *Science, 344*, 533-535.

Maeda, H. & Mogenson, G. J. (1981). A comparison of the effects of electrical stimulation of the lateral and ventromedial hypothalamus on the activity of neurons in ventral tegmental area and substantia nigra. *Brain Research Bulletin*, *7,* 283-291.

Martig, A.,K., Underhill, S., DeFrancesco, A., Tock, J., Rinaman, L., Witten, I., Deisseroth, K., Amara, S. & Moghaddam, B. (2013). Transient activation of VTA neurons causes a sustained post-activation increase of terminal extrasynaptic dopamine (Poster). San Diego: The annual meeting for the Society for Neuroscience.

Matthews, G. (1978). Strength-duration properties of single units driven by electrical stimulation of the lateral hypothalamus in rats. *Brain Research Bulletin*, *3*(2), 171–174.

Matthews, G. & Gallistel, C. R. (1975). Bilateral interactions in single units driven by MFB self-stimulation electrodes. *Physiology & Behavior, 15*(5), 543-549.

McClure, S. M., Daw, N. D. & Read Montague, P. (2003). A computational substrate for incentive salience. *Trends in Neurosciences, 26*(8), 423–428.

Menegas, W., Bergan, J. F., Ogawa, S. K., Isogai, Y, Venkataraju, K. U., Osten, P., Uchida, N. & Watabe-Uchida, M. (2015). Dopamine neurons projecting to the posterior striatum form an anatomically distinct subclass. *elife, 4:*e, 100032.

[26] Miliaressis, E., Emond, C. & Merali, Z. (1991). Re-evaluation of the role of dopamine in intracranial self-stimulation using in vivo microdialysis. *Behavioural brain research*, *46(1)*, 43-8.

Milner, P. M. (1991). Brain-stimulation reward: A review. *Canadian Journal of Psychology/Revue canadienne de psychologie, 45*(1), 1-36.

Minsky, M. L. (1961). Steps toward artificial intelligence. *Proceedings of the Institute of Radio Engineers*, *49*, 8–30.

[13] Moisan, J. & Rompré, P. P. (1998). Electrophysiological evidence that a subset of midbrain dopamine neurons integrate the reward signal induced by electrical stimulation of the posterior mesencephalon. *Brain research*, *786(1-2)*, 143-52.

Molochnikov, I. & Cohen, D. (2014). Hemispheric differences in the mesostriatal dopaminergic system. *Frontiers in Systems Neuroscience*, *8*, 1–14.

Montague, P., Dayan, P. & Sejnowski, T. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *The Journal of Neuroscience*, *16*(5), 1936–1947.

Morgan, J. & Curran, T. (1991). Stimulus-Transcription Coupling In The Nervous System: Involvement Of The Inducible Proto-oncogenes fos And jun. *Annual Review of Neuroscience*, *14*, 421–451.

[23] Murray, B. & Shizgal, P. (1996). Behavioral measures of conduction velocity and refractory period for reward-relevant axons in the anterior LH and VTA. *Physiology & behavior*, *59*(4-5), 643-52.

Murray, B. & Shizgal, P. (1994). Evidence implicating both slow- and fast-conducting fibers in the rewarding effect of medial forebrain bundle stimulation. *Behavioural Brain Research*, *63*, 47-60.

Nakahara, D., Ishida, Y., Nakamura, M., Kuwahara, I., Todaka, K. & Nishimori, T. (1999). Regional differences in desensitization of c-Fos expression following repeated self-stimulation of the medial forebrain bundle in the rat. *Neuroscience*, *90*(3), 1013-20.

Nakahara, D., Ozaki, N., Kapoor, V. & Nagatsu, T. (1989). The effect of uptake inhibition on dopamine release from the nucleus accumbens of rats during self- or forced stimulation of the medial forebrain bundle: a microdialysis study. *Neuroscience letters*, *104*(1-2), 136-40.

Nasser, H. M., Calu, D. J., Schoenbaum, G. & Sharpe, M. J. (2017). The dopamine prediction error: Contributions to associative models of reward learning. *Frontiers in Psychology*, *8*, 1–17.

Nieh, E. H., Matthews, G. A., Allsop, S. A., Presbrey, K. N., Leppla, C. A., Wichmann, R., Neve1, R., Wildes, C. P. & Tye, K. M. (2015). Decoding Neural Circuits that Control Compulsive Sucrose-Seeking, *Cell, 160*(3), 528–541.

Nieh, E. H., Vander Weele, C. M., Matthews, G. A., Presbrey, K. N., Wichmann, R., Leppla, C. A., Izadmehr, E. M. & Tye, K. M. (2016). Inhibitory Input from the Lateral Hypothalamus to the Ventral Tegmental Area Disinhibits Dopamine Neurons and Promotes Behavioral Activation. *Neuron, 19,* 1286-1298.

Nieuwenhuys, R., Geeraedts, L. M. & Veening, J. G. (1982). The medial forebrain bundle of the rat. I. General introduction. *Journal of comparative neurology, 206*(1)*,* 49-81.

Niv, Y., Daw, N. D., Joel, D. & Dayan, P. (2007). Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology*, *191*(3), 507–520.

Oakman, S. A, Faris, P. L., Kerr, P. E., Cozzari, C. & Hartman, B. K. (1995). Distribution of pontomesencephalic cholinergic neurons projecting to substantia nigra differs significantly from those projecting to ventral tegmental area. *Journal of Neuroscience*, *15*(9), 5859-69.

[1] Olds, J. (1962). Hypothalamic substrates of Reward. *Physiological reviews, 42*, 554-604.

[2] Olds, M. E. & Olds, J. (1963). Approach-avoidance analysis of rat diencephalon. *The Journal of Comparative Neurology*, *120*(2), 259-295.

Otake, K & Nakamura, Y. (1998). Single midline thalamic neurons projecting to both the ventral striatum and the prefrontal cortex in the rat. *Neuroscience, 86*(2), 635-649.

[42] Owesson-White, C. A., Cheer, J. F., Beyene, M., Carelli, R. M. & Wightman, R. M. (2008). Dynamic changes in accumbens dopamine correlate with learning during intracranial self-stimulation. *Proceedings of the National Academy of Sciences, 105(33)*, 11957–11962.

Parker, N. F., Cameron, C. M., Taliaferro, J. P., Lee, J., Choi, J. Y., Davidson, T. J., Daw, N. D. & Witten, I. B. (2016). Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nature Neuroscience*, *19*(6), 845–854.

Parsons, M. P., Li, S. & Kirouac, G. S. (2007). Functional and anatomical connection between the paraventricular nucleus of the thalamus and dopamine fibers of the nucleus accumbens. *Journal of comparative neurology, 500*(6), 1050-1063.

Phillips, A. G., Jakubovic, A. & Fibiger, H. C. (1987). Increased in vivo tyrosine hydroxylase activity in rat telencephalon produced by self-stimulation of the ventral tegmental area. *Brain research*, *402*(1), 109-16.

Ranck, J. B. (1975). Which elements are excited in electrical stimulation of mammalian central nervous system: a review, *Brain Research, 98*, 417–440.

Read, J. C. A. (2015). The place of human psychophysics in modern neuroscience. *Neuroscience, 296*, 116–129.

Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black and W. F. Prokasy (Eds.), Classical Conditioning II, pp. 64–99. Appleton-Century-Crofts, New York.

Rodeberg, N. T., Sandberg, S. G., Johnson, J. A., Phillips, P. E. M. & Wightman, R. M. (2017). Hitchhiker's Guide to Voltammetry: Acute and Chronic Electrodes for in Vivo Fast-Scan Cyclic Voltammetry. *ACS Chemical Neuroscience*, *8*(2), 221–234.

Rodriguez, E. A., Tran, G. N., Gross, L. A., Crisp, J. L., Shu, X., Lin, J. Y. & Tsien, R. Y. (2016). A far-red fluorescent protein evolved from a cyanobacterial phycobiliprotein. *Nature Methods*, *13*(9), 763–769.

[21] Rompré, P. P. & Shizgal, P. (1986). Electrophysiological characteristics of neurons in forebrain regions implicated in self-stimulation of the medial forebrain bundle in the rat. *Brain research*, *364*(2), 338-49.

[9] Rompré, P-P. & Wise, R. A. (1989). Behavioral evidence for midbrain dopamine depolarization inactivation. *Brain research, 477*, 152-156.

Sackett, D. A., Saddoris, M. P. & Carelli, R. M. (2017). Nucleus Accumbens Shell Dopamine Preferentially Tracks Information Related to Outcome Value of Reward. *Eneuro*, *4*(3), ENEURO.0058-17.

Salamone, J.D. & Correa, M. (2002). Motivational views of reinforcement: implications for understanding the behavioral functions of nucleus accumbens dopamine. *Brain research. Brain research reviews, 137*(1-2), 3-25.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology, 80*(1), 1–27.

Schultz, W. (2015). Neuronal Reward and Decision Signals: From Theories to Data. *Physiological Reviews*, *95*(3), 853–951.

Schultz, W., Apicella, P. & Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *The Journal of Neuroscience, 13* (3), 900–913.

Schultz, W. & Romo, R. (1990). Dopamine neurons of the monkey midbrain: contingencies of responses to stimuli eliciting immediate behavioral reactions. *Journal of Neurophysiology,*

63(3), 607–624.

Semba, K. & Fibiger, H. C. (1992). Afferent connections of the laterodorsal and pedunculopontine tegmental nuclei in the rat: a retro- and antero-grade transport and immunohistochemical study. *Journal of comparative neurology, 323,* 387-410.

[6] Shizgal, P. (1997). Neural basis of utility estimation. *Current opinion in neurobiology*, *7(2)*, 198-208.

[18] Shizgal, P., Bielajew, C., Corbett, D., Skelton, R. & Yeomans, J. (1980). Behavioral methods for inferring anatomical linkage between rewarding brain stimulation sites. *Journal of comparative and physiological psychology*, *94*(2), 227-37.

[22] Shizgal, P., Schindler, D. & Rompré, P. P. (1989). Forebrain neurons driven by rewarding stimulation of the medial forebrain bundle in the rat: comparison of psychophysical and electrophysiological estimates of refractory periods. *Brain research*, *499*(2), 234-48.

[32] Simmons, J. M. & Gallistel, C. R. (1994). Saturation of subjective reward magnitude as a function of current and pulse frequency. *Behavioral neuroscience*, *108*(1), 151-60.

[31] Solomon, R. B., Trujillo-Pisanty, I., Conover, K. & Shizgal, P. (2015). Psychophysical inference of frequency-following fidelity in the neural substrate for brain stimulation reward. *Behavioural brain research, 292*, 327-341.

[24] Sombers, L., A., Beyene, M., Carelli, R. M & Wightman, R. M. (2009). Synaptic overflow of dopamine in the nucleus accumbens arises from neuronal activity in the ventral tegmental area. *Journal of neuroscience, 29*(6)*,* 1735-1742.

Stalnaker, T. A., Calhoon, G. G., Ogawa, M., Roesch, M. R. & Schoenbaum, G. (2012). Reward Prediction Error Signaling in Posterior Dorsomedial Striatum Is Action Specific. *Journal of Neuroscience, 32*(30), 10296–10305.

Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K. & Janak, P. H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nature Neuroscience*, *16*(7), 966–973.

Steininger, T. L., Rye, D. B. & Wainer, B., H. (1992). Afferent projections to the cholinergic pedunculopontine tegmental nucleus and adjacent midbrain extrapyramidal area in the albino rat. I. Retrograde tracing studies. *The journal of comparative neurology, 321,* 515-543.

[46] Stellar, J. R. & Corbett, D. (1989). Regional neuroleptic microinjections indicate a role for nucleus accumbens in lateral hypothalamus self-stimulation reward. *Brain research, 477*, 126-143.

[51] Stuber, G. D., Hnasko, T. S., Britt, J. P., Edwards, R. H. & Bonci, A. (2010) Dopaminergic terminals in the nucleus accumbens but not the dorsal striatum corelease glutamate. *Journal of neuroscience, 30*(24), 8229-33.

Stuber, G. D. & Wise, R. A. (2016). Lateral hypothalamic circuits for feeding and reward. *Nature Neuroscience*, *19*(2), 198–205.

Stujenske, J. M., Spellman T. & Gordon J. A. (2015). Modeling the Spatiotemporal Dynamics of Light and Heat Propagation for In Vivo Optogenetics. *Cell reports, 12*(3), 525-534.

Sutton, R. S. (1984). Temporal Credit Assignment in Reinforcement Learning. Ph.D. thesis, University of Massachusetts, Amherst.

Sutton, R. S. (1988). Learning to predict by the method of temporal di↵erences. *Machine Learning, 3*(1), 9–44

Sutton, R. S. & Barto, A. G. (2018). Reinforcement Learning: An Introduction. 2nd Edition.

Cambridge, MA: MIT Press.

Takahashi, Y. K., Langdon, A. J., Niv, Y. & Schoenbaum, G. (2016). Temporal Specificity of Reward Prediction Errors Signaled by Putative Dopamine Neurons in Rat VTA Depends on Ventral Striatum. *Neuron, 91*(1), 182–193.

Thorndike, E. L. (1911). Animal intelligence: Experimental studies. New York: Macmillan.

Threlfell, S., Lalic, T., Platt, N. J., Jennings, K. A., Deisseroth, K. & Cragg, S. J. (2012). Striatal dopamine release is triggered by synchronized activity in cholinergic interneurons. *Neuron*, *75*(1), 58–64.

Tischer, D. & Weiner, O. D. (2014). Illuminating cell signalling with optogenetic tools. *Nature Reviews Molecular Cell Biology, 15*, 551–558.

[12] Trujillo-Pisanty I., Conover K. & Shizgal P. (2014). A new view of the effect of dopamine receptor antagonism on operant performance for rewarding brain stimulation in the rat. *Psychopharmacology, 231*, 1351-64.

Trujillo-Pisanty, I., Hernandez, G., Moreau-Debord, I., Cossette, M.-P., Conover, K., Cheer, J. F. & Shizgal, P. (2011). Cannabinoid Receptor Blockade Reduces the Opportunity Cost at Which Rats Maintain Operant Performance for Rewarding Brain Stimulation. *Journal of Neuroscience*, *31*(14), 5426–5435.

Trujillo-Pisanty, I., Sanio, C., Chaudhri, N. & Shizgal, P. (2015). Robust optical fiber patch-cords for in vivo optogenetic experiments in rats. *MethodsX*, *2*, 263–271.

Trujillo-Pisanty, I., Solis, P., Conover, K., Dayan, P. & Shizgal, P. (2016). On the forms of learning supported by rewarding optical stimulation of dopamine neurons (Poster). San Diego: The annual meeting for the Society for Neuroscience.

Tsai, H.C., Zhang, F., Adamantidis, A., Stuber, G. D., Bonci, A., De Lecea, L. & Deisseroth K. (2009). Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science, 324*(5930), 1080-4.

Veening, J. A., Swanson, Larry W, Cowan, W. M., Nieuwenhuys, R. & Geeraedts L. E., (1982). The medial forebrain bundle of the rat. II. An autoradiographic study of the topography of the major descending and ascending components. *Journal of comparative neurology, 206*(1), 82-108.

[52] Wassum, K. M., Tolosa, V. M., Walker, E., Monbouquette, H. G. & Maidment, N. T. (2008). Silicon Wafer-Based Platinum Microelectrode Array Biosensor for Near Real-Time Measurement of Glutamate in Vivo. *Sensors, 8*(8), 5023-36.

Watabe-Uchida, M., Eshel, N. & Uchida, N. (2017). Neural Circuitry of Reward Prediction Error. *Annual Review of Neuroscience*, *40*(1), 373–394.

Watabe-Uchida, M., Zhu, L., Ogawa, S. K., Vamanrao, A. & Uchida, N. (2012). Whole-Brain Mapping of Direct Inputs to Midbrain Dopamine Neurons. *Neuron*, *74*(5), 858–873.

Watkins, C. & Dayan, P. (1992). Q-learning. *Machine Learning, 8*(3-4), 279–292.

Weiss, G. (1901). Sur la possibilité de rendre comparables entre eux les appareils servant a l'excitation électrique. *Archives italiennes de biologie, 35,* 413-446.

West, D. C. & Wolstencroft, J. H. (1983). Strength-duration characteristics of myelinated and non-myelinated bulbospinal axons in the cat spinal cord. *The Journal of Physiology*, *337*(1), 37–50.

[37] Wightman, R. M., Amatore, C., Engstrom, R. C., Hale, P. D., Kristensen, E. W., Kuhr, W. G. & May, L. J. (1988). Real-time charaterization of dopamine oveflow and uptake in the rat striatum. *Neuroscience, 25*(2)*,* 513-523.

Wise, R. A. (1980). Action of drugs of abuse on brain reward systems. *Pharmacology Biochemistry & Behavior, 13*(S1)*,* 213-223.

[8] Wise, R. A. (1978). Catecholamine theories of reward: a critical review. *Brain Research, 152,* 215-247.

[15] Wise, R. A. (2004). Dopamine, learning and motivation. *Nature reviews. Neuroscience*, *5*(6), 483-94.

Witten, I. H. (1977). An adaptive optimal controller for discrete-time Markov environments. *Information and Control, 34*(4), 286–295.

Witten, I. B., Steinberg, E. E., Lee, S. Y., Davidson, T. J., Zalocusky, K. A, Brodsky, M., Yizhar, O., Cho, S. L., Gong, S., Ramakrishnan, C., Stuber, G. D., Tye, K. M., Janak, P. H. & Deisseroth, K. (2011). Recombinase-driver rat lines: tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron*, *72*(5), 721–733.

[43] Yavich, L. & Tiihonen, J. (2002). Patterns of dopamine overflow in mouse nucleus accumbens during intracranial self-stimulation. *Neuroscience Letters, 293*, 41-44.

[16] Yeomans, J. S. (1975). Quantitative measurement of neural post-stimulation excitability with behavioral methods. *Physiology and behavior, 15*, 593-602.

[17] Yeomans J. S. (1979). The absolute refractory periods of self-stimulation neurons. *Physiology and behavior, 22*, 911-9.

Yeomans, J. S. (1989). Two substrates for medial forebrain bundle self-stimulation: myelinated axons and dopamine axons. *Neuroscience and biobehavioral reviews*, *13*(2-3), 91-8.

Yeomans, J., Forster, G. & Blaha, C. (2001). M5 muscarinic receptors are needed for slow activation of dopamine neurons and for rewarding brain stimulation. *Life sciences*, *68*(22-

23), 2449-56.

Yeomans, J. S., Kofman, O. & McFarlane, V. (1985). Cholinergic involvement in lateral hypothalamic rewarding brain stimulation. *Brain research, 329,* 19-26.

Yeomans, J. S., Maidment, N. T. & Bunney, B. S. (1988). Excitability properties of medial forebrain bundle axons of A9 and A10 dopamine cells. *Brain Research*, *450*(1–2), 86–93.

Yeomans, J. S., Mercouris, N. & Ellard, C. (1985). Behaviorally measured refractory periods are lengthened by reducing electrode tip exposure or raising current. *Behavioral Neuroscience, 99*(5), 913-928.

Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M. & Deisseroth, K. (2011). Optogenetics in neural systems. *Neuron, 71*(1), 9-34.

Yona, G., Meitav, N., Kahn, I. & Shoham, S. (2016). Realistic Numerical and Analytical Modeling of Light Scattering in Brain Tissue for Optogenetic Applications. *ENeuro*, *3*(1), 1–9.

Zolnik, T. A., Sha, F., Johenning, F. W., Schreiter, E. R., Looger, L. L., Larkum, M. E. & Sachdev, R. N. S. (2017). All-optical functional synaptic connectivity mapping in acute brain slices using the calcium integrator CaMPARI. *Journal of Physiology*, *595*(5), 1465–1477.