Title

Ventral Tegmental Dopamine Neurons Participate in Reward Identity Predictions

Journal

Current Biology, 29(1)

ISSN

0960-9822

Authors

Keiflin, Ronald
Pribut, Heather J
Shah, Nisha B
et al.

Publication Date

2019

DOI

10.1016/j.cub.2018.11.050

HHS Public Access

Author manuscript

Curr Biol. Author manuscript; available in PMC 2020 January 07.

Published in final edited form as:

Curr Biol. 2019 January 07; 29(1): 93–103.e3. doi:10.1016/j.cub.2018.11.050.

Ventral Tegmental Dopamine Neurons Participate in Reward Identity Predictions

Ronald Keiflin[1,4,5,*], **Heather. J. Pribut**[1], **Nisha B. Shah**[1], and **Patricia H. Janak**[1,2,3,*]

[1]Department of Psychological and Brain Sciences, Krieger School of Arts and Sciences, Johns Hopkins University, Baltimore MD 21218 USA

[2]The Solomon H. Snyder Department of Neuroscience, Johns Hopkins School of Medicine, Johns Hopkins University, Baltimore MD 21205 USA

[3]Kavli Neuroscience Discovery Institute, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

[4]Present address: Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara CA 93106 USA

[5]Lead contact

SUMMARY

Dopamine (DA) neurons in the ventral tegmental area (VTA) and substantia nigra (SNc) encode reward prediction errors (RPEs) and are proposed to mediate error-driven learning. However the learning strategy engaged by DA-RPEs remains controversial. RPEs might imbue cue/actions with pure value, independently of representations of their associated outcome. Alternatively, RPEs might promote learning about the sensory features (the identity) of the rewarding outcome. Here we show that although both VTA and SNc DA neuron activation reinforces instrumental responding, only VTA DA neuron activation during consumption of expected sucrose reward restores error-driven learning and promotes formation of a new cue➔sucrose association. Critically, expression of VTA DA-dependent Pavlovian associations is abolished following sucrose devaluation, a signature of identity-based learning. These findings reveal that activation of VTA- or SNc-DA neurons engages largely dissociable learning processes with VTA-DA neurons capable of participating outcome-specific predictive learning, while the role of SNc-DA neurons appears limited to reinforcement of instrumental responses.

[*]**Correspondence:** Ronald Keiflin, Ph.D., Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara CA 93106, rkeiflin@ucsb.edu, Twitter: @RonKeiflin, Patricia H. Janak, Ph.D., Johns Hopkins University, 3400 N. Charles Street, Dunning Hall, room 246, Baltimore, MD 21218, patricia.janak@jhu.edu.

AUTHOR CONTRIBUTIONS

Conceptualization, R.K. and P.H.J.; Methodology, R.K. and P.H.J.; Investigation, R.K., H.J.P, and N.B.S.; Visualization: R.K.; Writing – Original Draft, R.K. and P.H.J.; Writing – Review & Editing, R.K. and P.H.J.; Funding Acquisition, P.H.J.

DECLARATION OF INTERESTS

The authors declare no competing interests.

eTOC BLURB

Keiflin et al. show that phasic activation of dopamine neurons promotes learning about the sensory features (the identity) of upcoming rewards. This challenges the proposal that dopamine signals simply assign value to reward-predictive cues and instead extends the role of dopamine to more complex forms of learning.

Keywords

reward-prediction error; associative learning; blocking; model-free; model-based

INTRODUCTION

Midbrain dopamine (DA) neurons, located in the ventral tegmental area (VTA) and substantia nigra pars compacta (SNc), respond in a characteristic fashion to reward, with increased phasic firing in response to unexpected rewards or reward-predicting cues, little or no response to perfectly predicted rewards, and pauses in firing when predicted rewards fail to materialize [1, 2]. This response pattern largely complies with the concept of a signed reward prediction error (RPE), an error-correcting teaching signal featured in theories of associative learning [3–5]. It has been suggested that the error signal carried by phasic DA responses and broadcast to forebrain regions constitutes a neural implementation of such theoretical teaching signals [2, 4]. In support of this hypothesis, optogenetic studies demonstrated that VTA DA neuron activation or inhibition mimics positive or negative RPEs, respectively, and affects Pavlovian appetitive learning accordingly [6, 7]. Whether phasic activity in SNc DA neurons also contributes to reward prediction learning remains uncertain. Based on their different striatal targets, distinct contributions to learning have been proposed for VTA and SNc DA neurons [8–10]; specifically, that VTA-DA signals contribute to reward predictions while SNc-DA signals contribute to action reinforcement [11, 12].

Another dissociation introduced by formal models of associative learning concerns the nature of reward representation afforded by RPE teaching signals [13]. Reward cues can become associated with the general appetitive value of primary rewards, encoded in some common currency. This form of learning does not allow for a representation of the specific identity of the outcome; therefore, expression of this learning is independent of the desire for that specific outcome at the time of test. Alternatively, reward cues can become associated with sensory features of rewards. As a result, expression of this learning is motivated by internal representations of a specific outcome and inference of its current value. These different learning strategies —value or identity— are broadly captured by model-free or model-based reinforcement algorithms [13–15].

The role of DA teaching signals in value and identity learning remains unclear [16, 17]. Since the original discovery that they track changes in expected value, phasic dopamine signals have predominantly been interpreted as model-free RPEs, promoting pure value assignment. Consistent with this view, direct activation of DA neurons serves as a potent reinforcer of instrumental behavior in self-stimulation procedures [7, 18–22]. More recently,

contributions of phasic DA signals to model-based learning have been suggested, based on evidence that DA neurons have access to higher-order knowledge for RPE computation [23–27]. Moreover, DA neurons were shown to respond to valueless changes in sensory features of expected rewards [28], and DA neuron optogenetic inhibition prevented learning induced by changing either reward identity or value [29]. While these studies reveal model-based influences in DA RPE computation, the exact associative content promoted by these DA signals is uncertain. A recent study intriguingly showed that in absence of a valuable outcome, phasic activation of DA neurons promotes model-based association between two neutral cues [30]. Since the cues were neutral, there was no opportunity for model-free, value-based conditioning. It remains to be determined how DA signals contribute to associative learning when subjects are actively learning about value-laden rewarding outcomes, the canonical situation in which DA signals are robustly observed, and in which both general value and specific identity learning are possible.

Therefore, the purpose of the present study was twofold: 1) assess the contribution of VTA- and SNc-DA neuron activation to Pavlovian reward learning, and 2) when learning was observed as a result of our manipulations, determine the value- or identity-based nature of this learning. To accomplish these goals, rats were trained in a blocking paradigm in which formation of an association between a target cue and a paired reward is prevented, or blocked, if this cue is presented simultaneously with another cue that already signals reward. In this situation the absence of RPEs, presumably reflected in the absence of DA responses, is thought to prevent learning about the target cue. We sought to restore learning by restoring RPEs, either endogenously by increasing the magnitude of reward, or by optogenetically activating VTA- or SNc-DA neurons during reward consumption. When successful, we assessed the associative content of this new learning by determining its sensitivity to post-conditioning outcome devaluation.

RESULTS

Phasic activation of VTA- but not SNc-DA neurons mimics reward prediction errors and promotes Pavlovian learning

Three groups of rats (Reward Upshift, n=24; VTA-DA Stim, n=20; SNc-DA Stim n=16) were trained in a Pavlovian blocking/unblocking task (Figure 1). We refer to our task as an ‘unblocking task’, and to cues as being ‘unblocked’, based on terminology employed in recent papers using a similar design [29, 32]. In the first stage, two visual cues, A and B, were presented individually followed by delivery of a sucrose reward. For the Reward Upshift group, the quantity of sucrose associated with these cues was different: cue A signaled a large reward (3 × 0.1ml, distributed over 30s), while cue B signaled a small reward (0.1ml, at the end of the 30s cue). This was done so that subsequent upshift of sucrose reward magnitude during the compound BY would cause an endogenous RPE and presumably unblock learning about target cue Y. For the other groups (VTA-DA Stim and SNc-DA Stim), cue A and B both signaled a large sucrose delivery, which, in absence of further manipulation should prevent endogenous RPEs during the subsequent compound phase. The purpose of the Reward Upshift group was to demonstrate the appropriateness of

these general training parameters for unblocking per se, as well as to allow a comparison of the magnitude of any optogenetically-induced unblocking with natural unblocking.

Subjects acquired conditioned responding rapidly, as indicated by time spent in the reward port during cue presentation (Figure 2). In the reward upshift group, responding to cue A was greater than cue B (average for last 4 days of individual cue, T=9.703, P<0.001), consistent with the different reward magnitudes associated with these cues. This difference in responding was not observed in VTA and SNc stim groups as in these groups both cues signaled large reward (Ps>.967; average last 4 days of individual cue). In the second stage of the procedure, the individual-cue trials were maintained and two new trial types (compound-cues trials) were introduced consisting of simultaneous presentation of a visual cue (A or B) with an auditory cue (X or Y) to form compounds AX and BY. Both of these compound cues were paired with large sucrose reward. For all subjects, the addition of cue X was redundant: large reward was expected and obtained on the basis of cue A alone. Therefore, in absence of prediction error during AX trials, learning about target cue X should be blocked. In contrast, the introduction of cue Y coincided with prediction errors. For the Reward Upshift group, violation in the expected amount and timing of reward (small and delayed during cue B; large and early during BY) is thought to create endogenous prediction errors that unblock learning about target cue Y. For the other groups, we sought to artificially recreate normally-absent prediction errors by optogenetically activating VTA- or SNc-DA neurons during reward consumption on BY trials. Thus, for each group, this design permits a within-subject test of unblocking by comparison of conditioned responding to X and Y at test. For all groups, the introduction of compound cues in the 2nd phase produced a general increase in conditioned responding (A vs. AX, B vs. BY; Ps<0.001) while responding to the individual cues A and B remained constant (Days 7–10 vs. 11–14: Ps>0.08). This increased responding to the compound cues might reflect the higher salience of auditory cues (X and Y) relative to the visual cues (A and B). This difference in salience might also have contributed to the magnitude of the effects observed in this study.

Finally, to assess the associative strength acquired by each individual cue following reward upshift or DA neuron optogenetic activation, all rats underwent a probe test in which all cues were presented separately in absence of sucrose (Figure 3). A two-way mixed ANOVA (Group x Cue) revealed a main effect of Group ($F_{2,57}$=13.818, P<0.01) and Cue ($F_{3,171}$=17.997, P<0.01) and a significant interaction between these factors ($F_{6,171}$=11.050, P<0.01). Follow-up one-way repeated measures (RM) ANOVAs separately conducted on each group revealed significant effects of cue type on responding (Reward Upshift: $F_{3,69}$=22.078, P<0.001; VTA-DA stimulation: $F_{3,57}$=11.634, P<0.001; SNc-DA stimulation: $F_{3,45}$=7.836, P<0.001). Posthoc comparisons confirmed that responding to the ancillary cues A and B was as expected: subjects in the Reward Upshift group responded more to A than B (T=5.373, P<0.001), and subjects in the other groups responded equally to these cues (VTA-DA stimulation: T=0.904, P=1.000, SNc-DA stimulation: T=0.537, P=1.000), consistent with the magnitude of reward paired with these cues during training. Of primary interest are the responses to target cues X and Y. In the Reward Upshift group, the surprising increase in reward magnitude during the BY compound unblocked learning, resulting in greater conditioned responding to Y than X (T=5.841, P<0.001). Note that both Y and X benefited from equal pairing with sucrose reward during the compound phase, only the presence or

absence of RPE during these cues differed and promoted or blocked learning, respectively. Stimulation of VTA-DA neurons during sucrose consumption in presence of the BY compound also resulted in greater responding to Y than X (T=5.334, P<0.001), indicating that VTA-DA phasic activation mimicked endogenous RPEs and unblocked learning, in agreement with our prior findings [7]. In contrast, activation of SNc-DA neurons did not unblock Pavlovian learning; subjects responded equally to X and Y (T=0.344, P=1) and responding to these cues was low (< 10% of cue time spent in port, on any trial). Analysis of an additional metric of Pavlovian conditioned approach, port entry rate, yielded similar results (Figure S1).

To directly compare consequences of endogenous RPEs and DA neuron activation on Pavlovian learning, we calculated for all individuals an unblocking score defined as the difference in time in port between Y and X (unblocked – blocked)(Figure S2). Comparing this value between groups, we found a general group effect ($F_{2,57}$=8.247, P<0.001), but no difference between Reward Upshift and VTA-DA stimulation groups (T=0.817, P=1) indicating equal unblocking after these manipulations. In contrast unblocking scores of the SNc-DA group were different from all other groups (all Ps≤0.01), confirming the functional dissociation between VTA- and SNc-DA neurons. Because there was a trend towards group differences in response to cue A (P=0.065) —a fully conditioned cue with equal training history across all groups— we then compared the unblocking score between groups while controlling for individual differences in responding to this fully conditioned cue (ANCOVA, with response to A as covariate). This analysis indicated that responding to A had no influence on unblocking scores ($F_{1,56}$=0.464, P=0.499), and confirmed a general group effect ($F_{2,56}$=6.808, P=0.002) with significantly lower score in SNc-DA group compared to all other groups (Ps<0.026).

Cues paired with natural reward or with DA neuron stimulation can elicit behaviors that are not directed towards the reward port, such as orienting to the cue, rearing, and general locomotion/rotations [33, 34]. To determine the role of endogenous- as well as optically induced-RPEs on the acquisition of these behaviors in our procedure, we analyzed animals' behavioral responses to X and Y during the probe test. While the target cues occasionally evoked orienting, rearing, or rotations, these behaviors were equally frequent in response to X and Y (Figure S3), suggesting that, under these experimental parameters, these behaviors are not conditioned responses, but rather reflect unconditioned salient properties of the cues.

After completion of unblocking, we assessed the reinforcing properties of VTA- and SNc-DA neuron activation in an intracranial self-stimulation (ICSS) task in which rats responded on one of two nosepokes to obtain 1-s optical DA neuron stimulation (Figure 4). As shown previously [18, 21, 22, 34], activation of both VTA- and SNc-DA neurons served as a potent reinforcer of ICSS behavior. A 3-way mixed ANOVA (Group x Day x Nosepoke) conducted on responding over two sessions revealed a clear preference for the active nosepoke ($F_{1,34}$=45.522, P<0.001) and a Nosepoke x Day interaction ($F_{1,34}$=54.789, P<0.001) as responding at the active nosepoke increased over time (T=10.712, P<0.001, Bonferroni-corrected *post hoc* tests) while responding at the inactive nosepoke remained virtually absent (T=0.0414, P<0.967). Critically, we found no main effect ($F_{1,34}$=0.876, P=0.356) or interaction with group (Group x Day: $F_{1,34}$=0.244, P=0.625; Group x Nosepoke:

$F_{1,34}$=0.777, P=0.384; Group x Day x Nosepoke: $F_{1,34}$=0.270, P=0.607), indicating that ICSS of VTA- and SNc-DA neurons is equally reinforcing.

Together, these results show that while VTA- and SNc-DA neuron activation are equally potent reinforcers of instrumental behavior, only VTA-DA neurons activation mimics endogenous RPEs in promoting error-correcting Pavlovian learning (unblocking).

Activation of VTA-DA neurons promotes learning about reward identity

Although we demonstrated that endogenous RPEs induced by reward upshift or optogenetic VTA-DA neuron activation results in numerically comparable unblocking, the underlying learning strategies remained unknown. RPEs might imbue predictive cues with a scalar cache value, resulting in conditioned responses largely independent of current outcome value. Alternatively, RPEs might promote association between predictive cues and the sensory features (the identity) of their paired outcome, resulting in conditioned responses motivated by perceptual representations of the outcome and its current value. To determine the learning strategy recruited by endogenous RPEs or VTA-DA neuronal activation, we assessed the effect of devaluing the sucrose outcome on responding to Y, the unblocked cue. New subjects were trained in the blocking/unblocking task and learning about cue Y was unblocked by Reward Upshift (n=24) or by VTA-DA neuron stimulation (n=23) during the BY compound. At the end of compound training, rats in each group were assigned to the “devalued” or “valued” condition. Subjects in the “devalued” condition had sucrose devalued by pairing its consumption with LiCl-induced nausea (conditioned taste aversion). For subjects in the “valued” condition, sucrose consumption and LiCl-induced nausea occurred on alternate days, preserving the value of the sucrose outcome (Figure 5, Figure S4). Two days after the final LiCl injection, rats were tested for conditioned responding to Y (unblocked cue) and A (ancillary cue paired with large reward) in separate probe sessions. A 3-way mixed ANOVA (Group x Devaluation x Cue) conducted on time in port during the cues revealed a main effect of Cue ($F_{1,43}$=6.119, P=0.017) and Devaluation ($F_{1,43}$=10.707, P=0.002) as well as an interaction between these factors ($F_{1,43}$=4.750, P=0.035). This interaction was due to a significant influence of the devaluation procedure on responding to the unblocked cue Y (T=3.563, P<0.001), but not on the ancillary cue A (T=0.514, P=0.609). Reduced responding to Y after sucrose devaluation indicates that this response is normally motivated by the representation of the sucrose outcome and anticipation of its current value (model-based process). Critically, we found no main effect ($F_{1,43}$=0.869, P=0.356) or interaction with Group (Group x Devaluation: $F_{1,43}$=0.005, P=0.943; Group x Cue: $F_{1,43}$=0.000, P=0.993; Group x Devaluation x Cue: $F_{1,43}$=0.339, P=0.564). Planned contrast analyses independently confirmed that, for each group, sucrose devaluation reduced responding to unblocked cue Y (Reward Upshift: T=2.559, P=0.018; VTA-DA Stim.: T= 2.116, P=0.046), but not to A (Reward Upshift: T=1.126, P=0.272; VTA-DA Stim.: T=0.018, P=0.986). Analysis of the port entry rate yielded similar results (Figure S5). Entries and presence in port outside cue presentation (during the ITI) were not affected by sucrose devaluation (Figure 5, Figure S5) indicating that the conditioning chamber context acquired no observable aversive effect on responding. VTA-DA valued and devalued rats later displayed similar ICSS behavior (Figure S4) indicating that reduced responding to Y in devalued subjects cannot be explained by poor efficiency of the optical stimulation. These

results indicate that both endogenous RPEs and VTA-DA neuronal activation during sucrose consumption promoted the formation of sensorily-rich associations and conferred cue Y with the ability to evoke a representation of the sucrose outcome.

DISSCUSION

We have shown that activation of VTA, but not SNc, DA neurons mimics RPEs and promotes the formation of outcome specific cue-reward associations. We used a Pavlovian blocking procedure, in which the formation of a cue-reward association is normally blocked by the absence of RPE (the reward being signaled by other predictive stimuli in the environment). Confirming and extending our previous study [7], we showed that restoring RPEs, either endogenously by manipulating the amount and timing of reward or by optogenetic activation of VTA-DA neurons, unblocks learning and promotes the formation of a cue-reward association. In stark contrast with VTA-DA activation, optogenetic activation of SNc-DA neurons failed to promote Pavlovian learning, i.e., learning remained blocked. This is despite the fact that activation of both VTA- and SNc-DA neurons serves as a potent reinforcer in self-stimulation procedures.

In a separate experiment, we probed the content of the newly formed association by assessing its sensitivity to outcome devaluation. We found that following unblocking by reward upshift or by VTA-DA stimulation, sucrose devaluation almost entirely abolished responding to the unblocked cue. This indicates that responding to the unblocked cue was not automatic but was mediated by an internal representation of the sucrose outcome and was sensitive to the current value of this outcome. This further indicates that both manipulations (reward upshift or VTA-DA stimulation) promote the formation of associations between the predictive cue and some as yet unspecified sensory features of the rewarding outcome. Future experiments that incorporate multiple outcomes differing in physical dimensions (taste, texture, temperature, etc.) will help delineate the nature and precision of perceptual reward expectations afforded by phasic dopamine signals.

Our findings demonstrating DA-enabled reward identity learning are consistent with a recent study by Sharpe and colleagues showing that phasic VTA-DA responses mediate association formation between two neutral stimuli (A➔B), a form of learning that is necessarily strictly identity-based since it involves no value [30]. The status of this association was then assessed by pairing one of the stimuli with food reward (B➔food) and testing conditioned responding to the other stimulus (A); food-seeking responses evoked by the target cue revealed a learned association between the stimuli and inference of upcoming food reward (i.e., if A➔B and B➔food, then A➔food). While Sharpe et al. demonstrated for the first time that VTA-DA signals can promote association between neutral stimuli, this study did not address the nature of *reward* encoding in DA-dependent associations. Indeed, although their study involved natural reward, it was used simply as a necessary means to reveal stimulus-stimulus associations and was not the object of DA manipulations. This distinction is important because unlike stimulus-stimulus associations that by definition involve only the sensory features of the outcome, cue-reward associations can signal the general value *or* the specific identity of the outcome (model-free or model-based association). Therefore, the possibility remains that while capable of promoting model-based learning when only

sensory information is available, VTA DA signals nevertheless engage preferentially model-free learning when (model-free) value can be encoded. In the present study, optogenetic activation of DA neurons was used to promote direct cue-reward associations, a form of learning that presents the opportunity for model-free and model-based algorithms. In these conditions when both learning strategies are equally valid, we showed that VTA-DA signals engage preferentially model-based learning.

Note that our results do not preclude participation of VTA-DA signals in model-free value assignment. Indeed, as shown here (ICSS experiment) and elsewhere [18, 34], the activation of VTA-DA neurons can confer cues and action with incentive/action value in absence of external reward. Ultimately, and consistent with DA's neuromodulatory role, the content of DA-induced learning is likely dependent on the nature of the information encoded and processed in terminal regions when coincident DA surges occur. What we show here is that in the presence of an external reward, the recruitment of a model-based learning strategy is not an exception but rather a central feature of VTA-DA teaching signals. This is consistent with recent studies showing that treatments (pharmacological or dietary restrictions) that globally increase or decrease DA function promote or impair, respectively, model-based processes in humans [35–37].

In this study, we found that phasic activation of VTA-DA neurons reproduces the 'natural' unblocking phenomenon induced by endogenous positive prediction errors -- in this case, violations in expected amounts and precise timing of reward. This result, together with the characteristic encoding of prediction errors by midbrain DA neurons, strongly suggests that VTA-DA neuron activation and endogenous prediction errors engage similar behavioral and neurophysiological processes to promote learning. Future studies aimed at recording and comparing activity of VTA DA neurons in both instances of unblocking (optically- or naturally-induced) are necessary to clarify how VTA DA activation relates to learning. In addition, several other manipulations can unblock learning besides the surprising increase in reward and/or timing violation. Valueless changes in sensory features of rewards can unblock learning, a process that also relies on VTA-DA neurons [29], possibly by engaging model-based learning processes as demonstrated in the present study. In certain conditions, unexpected decreases in reward can also unblock learning and establish cues as predictors of reward. Prior studies showed that unblocking by unexpected reward decreases relies on separate physiological and behavioral processes [modulation of attention by unsigned prediction errors, 38] and might involve SNc-DA neurons [39], although the exact contribution of negative prediction errors encoded by DA neurons to this attention-related process remains largely unknown [40].

An intriguing aspect of our results is the dissociation between the unblocked cue Y and the ancillary cue A in terms of response strategy. Before devaluation, both cue A and Y evoked similar responding and both responses extinguished at the same rate, indicating comparable overall strength of conditioning. However the underlying associative structures driving the response to A and Y appear to differ. Unlike Y, A evoked conditioned responding driven by model-free/value-based associations (unaffected by sucrose devaluation). The reason for this dissociation is unknown but might involve differences in amounts of training of these cues. Compared to Y, A benefited from an extensive training history (224 trials vs. 32 for Y)

which has been shown to promote model-free learning in the context of instrumental conditioning [41, 42], although not in the Pavlovian domain [43, note however that the extended training condition in that last study was only half of the training history of cue A in the present study]. Thus, training amounts and other as yet unknown factors might contribute to the development of model-free Pavlovian approach responses observed here. Alternatively, it is possible that, although consumption of sucrose was at floor, additional pairings between sucrose and illness may have been sufficient to produce a reduction in responding to cue A by further increasing the aversiveness of sucrose, thereby countering the increased appetitive conditioning that cue A received relative to cue Y. Perhaps more interesting are the implications for the role of VTA-DA signals in learning. In the VTA-DA group, the cues A and B are equivalent up to the compound conditioning phase and, based on the lack of effect of devaluation on A, we can assume that responding to both cues is governed by model-free associations. Therefore it appears that activation of VTA-DA neurons promoted formation of model-based associations about Y in subjects that were (presumably) currently engaged in model-free behavior during BY trials. This surprising result suggests that model-based associations could be formed “in the background” independently of the strategy governing behavior at the time these associations are formed or through post-training event replay [44]. Alternatively, activation of VTA-DA neurons could be sufficient to shift response strategy and restore model-based processing [45].

Our results provide strong evidence for a functional dissociation between VTA- and SNc-DA neurons in appetitive learning. While activation of VTA-DA neurons unblocked Pavlovian learning, we found no evidence of unblocking following SNc-DA neurons activation, despite careful analysis of several behavioral responses. This contrasts with recent results from our lab showing that, in absence of a natural reward, activation of VTA- or SNc-DA neurons during cue presentation promotes the development of conditioned cue-evoked locomotion [34]. An important point to consider when comparing these results is the behavior of the animals at the time of stimulation. Although free movement was possible, animals in the present study were relatively immobile during DA stimulation because it occurred as they were consuming sucrose reward. This absence of ambulatory movement during DA stimulation could have prevented the emergence of conditioned locomotion.

In contrast with the selective role of VTA-DA neurons in Pavlovian unblocking, we show here, in agreement with previous studies [21, 34], that instrumental behavior for ICSS is supported by VTA- and SNc-DA neuron stimulation. This partial dissociation between VTA- and SNc-DA neurons in Pavlovian and instrumental learning is reminiscent of the actor-critic reinforcement algorithm. This model is based on the idea of a separation of labor between a prediction module and an action module, with distributed RPEs promoting learning in both modules but with different consequences (updating predictions vs. reinforcing actions). A possible neural implementation of the actor-critic algorithm has been suggested, with ventral striatum/nucleus accumbens and dorsolateral striatum functioning as prediction and action modules, respectively [12]. Consistent with this, we showed that activation of SNc-DA neurons, projecting predominantly to dorsolateral striatum, reinforces prior actions but has no influence on Pavlovian prediction learning, in agreement with the role of RPEs in an action module, while activation of VTA-DA neurons, projecting predominantly to nucleus accumbens, promotes Pavlovian learning, in agreement with the

role of RPEs in a prediction module. Because predictions are updated by RPEs but also influence RPEs computations in return, the actor-critic model predicts that RPEs in the prediction module reinforce Pavlovian cues/states, which can then subsequently evoke back-propagated RPEs, including in the action module. A neural equivalent of this process in which Pavlovian predictions encoded in the nucleus accumbens feed back onto midbrain DA neurons (including SNc-DA neurons) impacting propagation of RPE teaching signals to more dorsal-lateral striatum, could contribute to instrumental reinforcement induced by VTA-DA stimulation. However, a critical difference between our results and the predictions of the actor-critic algorithm is that this algorithm is strictly model-free, while we show here that VTA-DA signals contribute to model-based Pavlovian learning. Therefore, our results suggest a hybrid model incorporating both model-free and model-based processes and in which VTA DA dependent model-based predictions shape SNc-DA signals and train model-free instrumental learning [46]

Finally, these results have important implications for DA-related pathologies. Noisy/deregulated DA signals originating from the VTA, as observed in schizophrenic patients [47], could promote model-based associations between external and/or internal events that are coincident but not causally-related, leading to internal world models out of touch with physical reality and sources of delusional beliefs [48]. In contrast, emergence of cue- or reward-evoked DA signals in the dorsolateral striatum, as reported after repeated drug use [49], could contribute to reinforcement of model-free maladaptive drug-seeking responses that persist despite knowledge of their adverse consequences [50].

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Ronald Keiflin (rkeiflin@ucsb.edu)

EXPERIMENTAL MODEL AND SUBJECT DETAILS Subjects

Th::Cre+ transgenic rats (37 males, 24 females) expressing Cre recombinase under control of the tyrosine hydroxylase promoter and their wild-type littermates (30 males, 16 females; *Th*::*cre*$^{-}$) were used in these studies. Rats were singly housed under a 12 h light/12 h dark cycle with unlimited access to food and water, except during behavioral experiments, when they were food restricted to ~90% free-feeding weight. All experimental procedures were conducted in accordance with UCSF and JHU Institutional Animal Care and Use Committees and the US National Institute of Health guidelines. Males and females were distributed as evenly as possible across groups. No significant effects of sex were found; therefore data for males and females were collapsed.

METHOD DETAILS Surgeries

Th::Cre+ rats (>300g males; >225g females) received unilateral infusions of AAV5-EF1α-DIO-ChR2-eYFP (titer: 1.5–4×10^{12} virus particles/mL) into VTA (AP: −5.4 and −6.2mm from bregma; ML: ± 0.7 from midline; DV: −8.5 and −7.5 from skull) or SNc (AP: −5.0 and −5.8; ML: ± 2.4; DV: – 8.0 and −7.0). This resulted in 4 injection sites for each rat (volume:

1μl per site; 0.1μL/min). Optic fibers aimed at VTA (AP: −5.8; ML: ± 0.7; DV: −7.5) or SNc (AP: −5.4; ML: ± 2.4; DV: −7.2) were also implanted. Behavioral experiments started >2 weeks post-surgery; sessions that included optical stimulation were conducted >4 weeks post-surgery.

Apparatus—Behavioral sessions were conducted in 12 identical sound-attenuated conditioning chambers (Med Associates, St. Albans, VT). A liquid delivery port was in the center of the right wall ~2 cm above the floor and connected to a syringe pump located outside the sound-attenuating cubicle. The left wall had two nosepoke operanda. A houselight was centered on the left wall and a pair of cue lights flanked the liquid delivery port on the right wall. White noise (76dB) and two pure tones (2.9 and 4.5 kHz, both 76dB) could be delivered through 3 wall speakers. The nosepoke operanda were obstructed during the unblocking procedures and accessible only during ICSS sessions. Conversely, the sucrose port was accessible only during unblocking procedures but obstructed during ICSS sessions. Subjects' presence in the port or nosepokes was detected by interruption of infrared beams.

Unblocking by reward upshift—In a brief shaping session, rats were trained to consume sucrose (15%, w/v) delivered in the liquid port (0.1 ml/delivery; 30 deliveries over 45 min). All rats then received 10 daily sessions during which two 30-s visual cues, A and B (flashing of the houselight 1 s on, 2 s off, or steady illumination of the light cues; counterbalanced) were paired with two different quantities of sucrose. Cue A signaled a large sucrose reward: 0.3 ml with 0.1 ml delivered every 9 s of the 30-s cue. Cue B signaled a small sucrose reward: 0.1 ml delivered over the last 3 s of the 30-s cue. These conditioning sessions consisted of 16 presentations of each cue with an average intertrial interval (ITI) of 3 min ± 1.5 (rectangular distribution; average ITI maintained constant throughout the experiment). After this initial phase of individual-cue conditioning, rats were pre-exposed to two uditory stimuli, X and Y (intermittent beeping of the tones 0.1 s on, 0.2 s off, or a steady white noise; counterbalanced) in a single habituation session (six 30-s presentation of each cue, no sucrose delivered). Over the next 4 days, rats received conditioning to the compound cues. Simultaneous presentations of A and X (AX compound), or B and Y (BY compound) were paired with the large sucrose reward. Cues A and B also continued to be presented individually with their respective rewards as in training as a reminder of the individual value of these cues. Each compound conditioning session consisted of 8 presentations of each trial type (AX, BY, A, B). Following compound conditioning, all rats received a probe test consisting of six unrewarded 30-s presentation of A, B, X, Y (in blocks of 3; order counterbalanced).

Unblocking by VTA- or SNC-DA Stimulation—Behavioral procedures were as described for the unblocking by reward upshift, with the following exceptions: 1) during initial conditioning to the individual cues, both cue A and B were paired with a large sucrose reward, which, in absence of further manipulation, should result in the blocking of both cue X and Y; and, 2) during compound conditioning, each delivery of sucrose during compound BY was accompanied by a 3-s train of light pulses (473 nm, 20 Hz, 60 pulses, 5 ms duration) delivered into the VTA or the SNc. The delivery of stimulation required 100 ms of

continuous presence in the baited port in order to coincide with consumption of the sucrose reward. Rats were tethered to optical patch cords for most conditioning sessions with the exception of training day 1, 5, 8, and pre-exposure to X and Y. This was done to habituate rats to perform the task both tethered and untethered. For the final probe test, rats were not tethered to prevent any potential interference on behavior (particularly, on orienting responses).

Outcome devaluation—Rats were initially trained in the unblocking task where learning about target cue Y was unblocked by reward upshift or by photoactivation of VTA-DA neurons. At the end of compound conditioning and before the final probe test, half of the rats in each group had the sucrose outcome devalued by pairing it with lithium chloride (LiCl)-induced nausea (devalued condition). Devaluation took place in the homecage over 4 days. On day 1 and 3, rats in the devalued groups received 10 min free access to sucrose immediately followed by LiCl injection (0.3 M; 6 ml/kg). Rats in the valued condition received similar exposure to sucrose and LiCl-induced illness but on alternate days (LiCl injections on Day 1 and 3; sucrose access on day 2 and 4). To confirm that sucrose devaluation was durable and transferred across contexts, sucrose consumption was measured in the conditioning chambers. Rats were placed in the chambers for 5 min, with 4ml sucrose in the reward cup. After 5 min, rats returned to their homecage and remaining sucrose was measured. This brief sucrose consumption test occurred twice, one day before and one day after cue probe tests. No difference was found between these two consumption tests, therefore these results were collapsed. Cue probe tests consisted of 6 unrewarded presentations of Y (unblocked cue) and A (control cue of comparable high value) on alternate days (order counterbalanced) in order to prevent potential interference between different response strategies (model-free vs. model-based). In these conditions, conditioned responding rapidly extinguished within session, therefore only responding on the first 3 trials was analyzed.

Intra-Cranial Self-Stimulation (ICSS)—Following completion of unblocking procedures, all VTA- and SNc-DA rats were tested for ICSS. During two daily 1-h sessions, rats had access to two nosepoke ports; a response at the active nosepoke (position counterbalanced) resulted in delivery of a 1-s train of light pulses (20 Hz, 5 ms duration). Active nosepoke responses during the 1-s light train were recorded but had no consequence. Inactive nosepoke responses were without consequence.

Video Analysis—A camera located in each conditioning chamber and connected to video acquisition software (Noldus Information Technology, Leesburg, VA) recorded animals' behavior during probe tests. Three types of responses were detected and manually scored: i) *orienting responses*, defined as rapid head movements in the direction of the cue occurring within 3s of cue onset. ii) *rearing responses*, defined as standing on hind legs with front feet off the floor (often against the side walls) and not grooming. iii) *rotation responses*, defined as a full rotation between the onset and termination of the cue.

Histology—Anesthetized animals were perfused with 0.9% saline followed by 4% paraformaldehyde. Brains were extracted, cryoprotected in 25% sucrose for >48 hours, and

sectioned at 50 μm on a freezing microtome. Coronal slices were collected onto glass slides and coverslipped with Vectashield mounting medium with DAPI. Fiber tip position and eYFP-CHR$_2$ virus expression were examined under a fluorescence microscope (Zeiss Microscopy, Thornwood, NY).

QUANTIFICATION AND STATISTICAL ANALYSIS

Counterbalancing procedures were used to form experimental groups balanced in terms of sex, cue identity, and behavioral performance in the sessions preceding the experimental intervention. Conditioned responding was measured by the percentage of time in the port and the rate of port entries during cue presentation, normalized by subtracting behavior during a pre-cue period of equal length. Behavior during pre-cue periods was always extremely low (0.304s ± 0.057 of average presence in the port during the 30s that precede cue presentation, no group difference Ps > 0.752). Statistical analyses were conducted using SPSS Statistics V22, and Systat SigmaPlot 14, and consisted generally of mixed-design repeated measures (RM) ANOVAs with cue and trials as within-subject factors, and group (reward upshift, VTA-DA, or SNc-DA) and devaluation as between-subject factors. On the rare occasions that the sphericity assumption was violated, the Greenhouse-Geisser correction was used to adjust the reported p-value. Post-hoc and planned comparisons were carried with Bonferroni-corrected t-test. Significance was assessed against a type I error rate of 0.05.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health grant R01 DA035943.

REFERENCES

1. Eshel N, Tian J, Bukwich M, and Uchida N (2016). Dopamine neurons share common response function for reward prediction error. Nat Neurosci 19, 479–486. [PubMed: 26854803]
2. Schultz W, Dayan P, and Montague PR (1997). A neural substrate of prediction and reward. Science 275, 1593–1599. [PubMed: 9054347]
3. Waelti P, Dickinson A, and Schultz W (2001). Dopamine responses comply with basic assumptions of formal learning theory. Nature 412, 43–48. [PubMed: 11452299]
4. Glimcher PW (2011). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. Proc Natl Acad Sci U S A 108 Suppl 3, 15647–15654. [PubMed: 21389268]
5. Rescorla RA, and Wagner AR (1972). A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement In Classical conditioning II: current research and theory, Black AH and Prokasy WF, eds. (New York: Appleton-Century-Crofts), pp. 64–99.
6. Chang CY, Esber GR, Marrero-Garcia Y, Yau H-J, Bonci A, and Schoenbaum G (2016). Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. Nat Neurosci 19, 111–116. [PubMed: 26642092]
7. Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, and Janak PH (2013). A causal link between prediction errors, dopamine neurons and learning. Nat Neurosci 16, 966–973. [PubMed: 23708143]

8. Yin HH, and Knowlton BJ (2006). The role of the basal ganglia in habit formation. Nat Rev Neurosci 7, 464–476. [PubMed: 16715055]
9. Haber SN, Fudge JL, and McFarland NR (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. J Neurosci 20, 2369–2382. [PubMed: 10704511]
10. Everitt BJ, and Robbins TW (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. Nat Neurosci 8, 1481–1489. [PubMed: 16251991]
11. Ramayya AG, Misra A, Baltuch GH, and Kahana MJ (2014). Microstimulation of the human substantia nigra alters reinforcement learning. J Neurosci 34, 6887–6895. [PubMed: 24828643]
12. Takahashi Y, Schoenbaum G, and Niv Y (2008). Silencing the critics: understanding the effects of cocaine sensitization on dorsolateral and ventral striatum in the context of an actor/critic model. Front Neurosci 2, 86–99. [PubMed: 18982111]
13. McDannald MA, Takahashi YK, Lopatina N, Pietras BW, Jones JL, and Schoenbaum G (2012). Model-based learning and the contribution of the orbitofrontal cortex to the model-free world. Eur J Neurosci 35, 991–996. [PubMed: 22487030]
14. Daw ND, Niv Y, and Dayan P (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8, 1704–1711. [PubMed: 16286932]
15. Sutton RS, and Barton AG (1998). Reinforcement Learning: an Introduction (Cambridge, Massachusetts: MIT Press).
16. Dayan P, and Berridge KC (2014). Model-based and model-free Pavlovian reward learning: revaluation, revision, and revelation. Cogn Affect Behav Neurosci 14, 473–492. [PubMed: 24647659]
17. Nasser HM, Calu DJ, Schoenbaum G, and Sharpe MJ (2017). The dopamine prediction error: contributions to associative models of reward learning. Front Psychol 8, 244. [PubMed: 28275359]
18. Witten IB, Steinberg EE, Lee SY, Davidson TJ, Zalocusky KA, Brodsky M, Yizhar O, Cho SL, Gong S, Ramakrishnan C, et al. (2011). Recombinase-driver rat lines: tools, techniques, and optogenetic application to dopamine-mediated reinforcement. Neuron 72, 721–733. [PubMed: 22153370]
19. Ilango A, Kesner AJ, Broker CJ, Wang DV, and Ikemoto S (2014). Phasic excitation of ventral tegmental dopamine neurons potentiates the initiation of conditioned approach behavior: parametric and reinforcement-schedule analyses. Front Behav Neurosci 8, 155. [PubMed: 24834037]
20. Pascoli V, Terrier J, Hiver A, and Lüscher C (2015). Sufficiency of mesolimbic dopamine neuron stimulation for the progression to addiction. Neuron 88, 1054–1066. [PubMed: 26586182]
21. Ilango A, Kesner AJ, Keller KL, Stuber GD, Bonci A, and Ikemoto S (2014). Similar roles of substantia nigra and ventral tegmental dopamine neurons in reward and aversion. J Neurosci 34, 817–822. [PubMed: 24431440]
22. Rossi MA, Sukharnikova T, Hayrapetyan VY, Yang L, and Yin HH (2013). Operant self-stimulation of dopamine neurons in the substantia nigra. PLoS ONE 8, e65799. [PubMed: 23755282]
23. Starkweather CK, Babayan BM, Uchida N, and Gershman SJ (2017). Dopamine reward prediction errors reflect hidden-state inference across time. Nat Neurosci 20, 581–589. [PubMed: 28263301]
24. Sadacca BF, Jones JL, and Schoenbaum G (2016). Midbrain dopamine neurons compute inferred and cached value prediction errors in a common framework. elife 5.
25. Bromberg-Martin ES, Matsumoto M, Hong S, and Hikosaka O (2010). A pallidus-habenula-dopamine pathway signals inferred stimulus values. J Neurophysiol 104, 1068–1076. [PubMed: 20538770]
26. Nakahara H, Itoh H, Kawagoe R, Takikawa Y, and Hikosaka O (2004). Dopamine neurons can represent context-dependent prediction error. Neuron 41, 269–280. [PubMed: 14741107]
27. Fonzi KM, Lefner MJ, Phillips PEM, and Wanat MJ (2017). Dopamine Encodes Retrospective Temporal Information in a Context-Independent Manner. Cell Rep 20, 1765–1774. [PubMed: 28834741]

28. Takahashi YK, Batchelor HM, Liu B, Khanna A, Morales M, and Schoenbaum G (2017). Dopamine neurons respond to errors in the prediction of sensory features of expected rewards. Neuron 95, 1395–1405.e3. [PubMed: 28910622]
29. Chang CY, Gardner M, Gonzalez Di Tillio M, and Schoenbaum G (2017). Optogenetic blockade of dopamine transients prevents learning induced by changes in reward features. Curr Biol.
30. Sharpe MJ, Chang CY, Liu MA, Batchelor HM, Mueller LE, Jones JL, Niv Y, and Schoenbaum G (2017). Dopamine transients are sufficient and necessary for acquisition of model-based associations. Nat Neurosci 20, 735–742. [PubMed: 28368385]
31. Stujenske JM, Spellman T, and Gordon JA (2015). Modeling the spatiotemporal dynamics of light and heat propagation for in vivo optogenetics. Cell Rep 12, 525–534. [PubMed: 26166563]
32. McDannald MA, Lucantonio F, Burke KA, Niv Y, and Schoenbaum G (2011). Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. J Neurosci 31, 2700–2705. [PubMed: 21325538]
33. Holland PC (1977). Conditioned stimulus as a determinant of the form of the Pavlovian conditioned response. J Exp Psychol Anim Behav Process 3, 77–104. [PubMed: 845545]
34. Saunders BT, Richard JM, Margolis EB, and Janak PH (2018). Dopamine neurons create Pavlovian conditioned stimuli with circuit-defined motivational properties. Nat Neurosci 21, 1072–1083. [PubMed: 30038277]
35. Sharp ME, Foerde K, Daw ND, and Shohamy D (2016). Dopamine selectively remediates "model-based" reward learning: a computational approach. Brain 139, 355–364. [PubMed: 26685155]
36. de Wit S, Standing HR, Devito EE, Robinson OJ, Ridderinkhof KR, Robbins TW, and Sahakian BJ (2012). Reliance on habits at the expense of goal-directed control following dopamine precursor depletion. Psychopharmacology (Berl) 219, 621–631. [PubMed: 22134475]
37. Wunderlich K, Smittenaar P, and Dolan RJ (2012). Dopamine enhances model-based over model-free choice behavior. Neuron 75, 418–424. [PubMed: 22884326]
38. Holland PC, and Gallagher M (1993). Effects of amygdala central nucleus lesions on blocking and unblocking. Behav Neurosci 107, 235–245. [PubMed: 8484889]
39. Lee HJ, Gallagher M, and Holland PC (2010). The central amygdala projection to the substantia nigra reflects prediction error information in appetitive conditioning. Learn Mem 17, 531–538. [PubMed: 20889725]
40. Esber GR, Roesch MR, Bali S, Trageser J, Bissonette GB, Puche AC, Holland PC, and Schoenbaum G (2012). Attention-related Pearce-Kaye-Hall signals in basolateral amygdala require the midbrain dopaminergic system. Biol Psychiatry 72, 1012–1019. [PubMed: 22763185]
41. Adams CD (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. The Quarterly Journal of Experimental Psychology Section B 34, 77–98.
42. Dickinson A (1985). Actions and habits: the development of behavioural autonomy. Philosophical Transactions of the Royal Society B: Biological Sciences 308, 67–78.
43. Holland PC, Lasseter H, and Agarwal I (2008). Amount of training and cue-evoked taste-reactivity responding in reinforcer devaluation. J Exp Psychol Anim Behav Process 34, 119–132. [PubMed: 18248119]
44. McNamara CG, Tejero-Cantero Á, Trouche S, Campo-Urriza N, and Dupret D (2014). Dopaminergic neurons promote hippocampal reactivation and spatial memory persistence. Nat Neurosci 17, 1658–1660. [PubMed: 25326690]
45. Hitchcott PK, Quinn JJ, and Taylor JR (2007). Bidirectional modulation of goal-directed actions by prefrontal cortical dopamine. Cereb Cortex 17, 2820–2827. [PubMed: 17322558]
46. Russek EM, Momennejad I, Botvinick MM, Gershman SJ, and Daw ND (2017). Predictive representations can link model-based reinforcement learning to model-free mechanisms. PLoS Comput Biol 13, e1005768. [PubMed: 28945743]
47. Ermakova AO, Knolle F, Justicia A, Bullmore ET, Jones PB, Robbins TW, Fletcher PC, and Murray GK (2018). Abnormal reward prediction-error signalling in antipsychotic naive individuals with first-episode psychosis or clinical risk for psychosis. Neuropsychopharmacology 43.
48. Corlett PR, Murray GK, Honey GD, Aitken MRF, Shanks DR, Robbins TW, Bullmore ET, Dickinson A, and Fletcher PC (2007). Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. Brain 130, 2387–2400. [PubMed: 17690132]

49. Willuhn I, Burgeno LM, Everitt BJ, and Phillips PEM (2012). Hierarchical recruitment of phasic dopamine signaling in the striatum during the progression of cocaine use. Proc Natl Acad Sci U S A 109, 20703–20708. [PubMed: 23184975]

50. Keiflin R, and Janak PH (2015). Dopamine prediction errors in reward learning and addiction: from theory to neural circuitry. Neuron 88, 247–263. [PubMed: 26494275]

HIGHLIGHTS

- Phasic activation of VTA dopamine neurons promotes cue➔reward learning
- Expression of this learning is mediated by internal representation of the outcome
- VTA dopamine neurons contribute to perceptual predictions about reward identity
- The role of SNc dopamine neurons appears limited to instrumental reinforcement

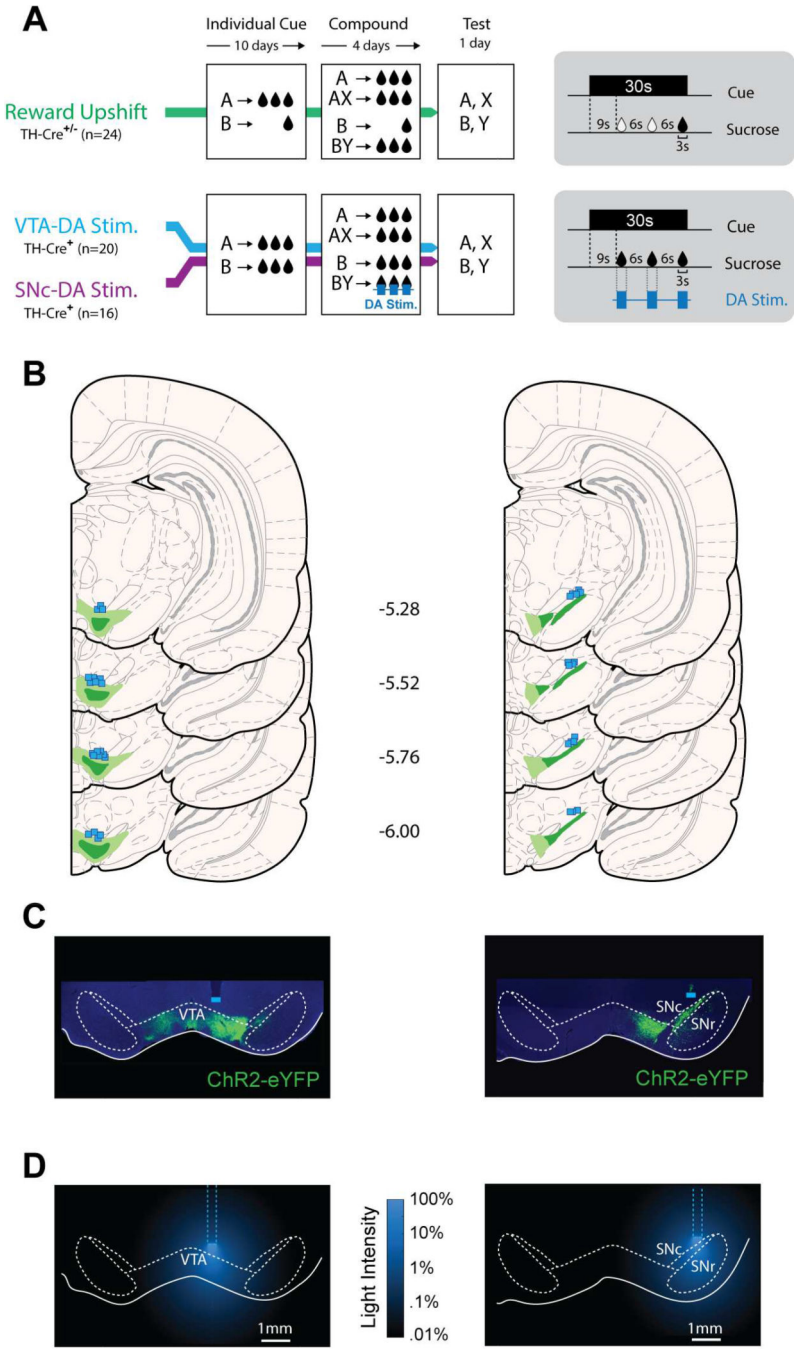

Figure 1. Behavioral task and histology

(A) Three groups of rats were trained in the blocking/unblocking task. During the *Individual Cue* phase, two visual cues (A and B) were paired with sucrose reward. In the *Compound Cue* phase, two new trial types of simultaneous presentation of a visual cue with an auditory cue (X or Y), resulting in two compound stimuli (AX and BY) were introduced. The absence of RPE during compound AX is predicted to block learning about cue X. During compound BY, an RPE was produced by increasing reward magnitude (Reward Upshift group) or by photostimulating DA neurons during sucrose consumption (VTA-DA Stim. and

SNc-DA Stim. groups). A 1-day probe test assessed the associative strength acquired by each individual cue. **(B)** Reconstruction of ChR2-YFP expression and fiber placement in VTA (left) and SNc (right). Light and dark shading indicate maximal and minimal spread of ChR2-YFP, respectively. Square symbols mark ventral extremity of fiber implants. **(C)** Representative ChR2-YFP expression in VTA (left) or SNc (right). **(D)** Laser power from the fiber tip estimated from [31]. Full laser power = 120 mW/mm^2 (corresponds to 34mW at the tip of 300um fibers; http://www.optogenetics.org/calc)

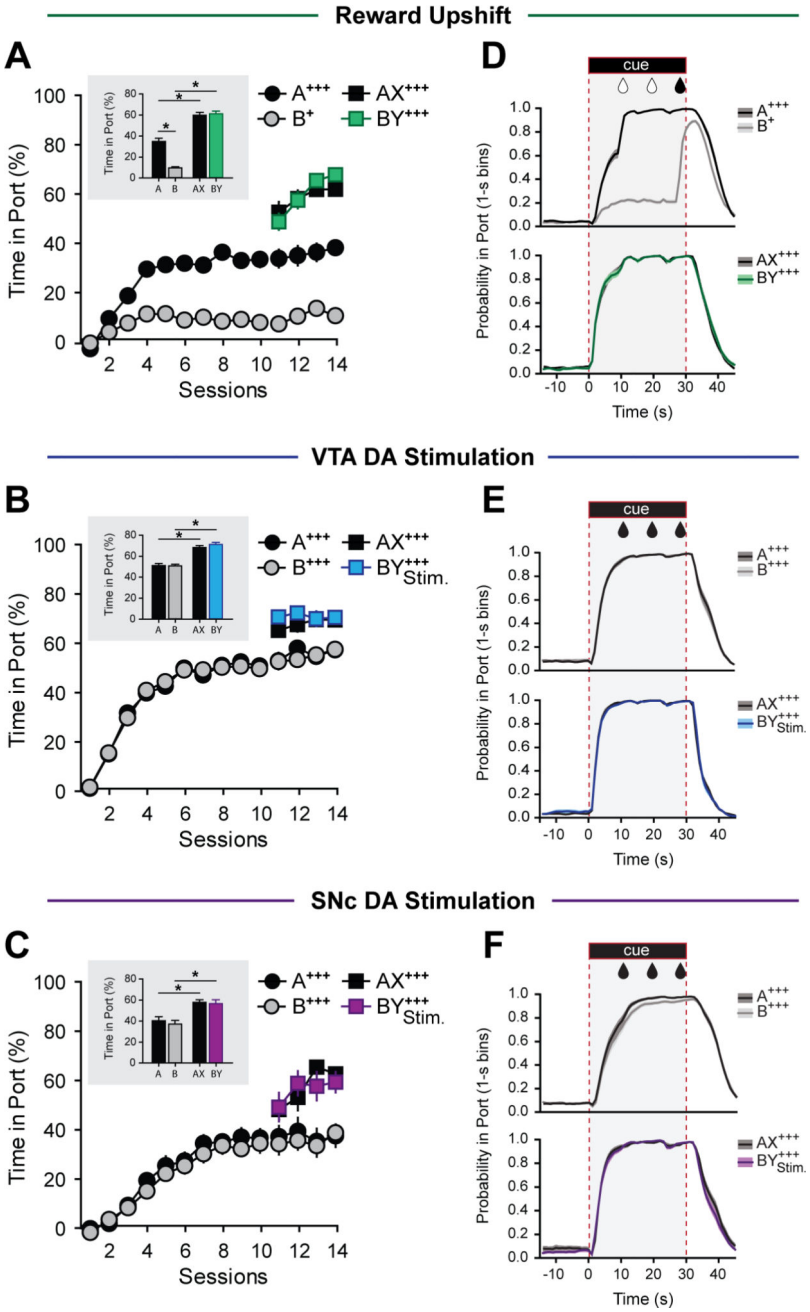

Figure 2. Performance during Individual Cue and Compound Cue training.

(A-C) Time spent in reward port during cue presentation over 10 days of Individual Cue conditioning and 4 days of Compound Cue conditioning for Reward Upshift **(A)** VTA-DA stimulation **(B)** and SNc-DA stimulation **(C)** groups. Values include only the first 9-s after cue onset and prior to sucrose delivery to avoid contamination with the consumption period. Inserts depict average performance over 4 days of Compound Cue conditioning. For all groups, introduction of the auditory stimulus increased performance (A *vs.* AX, and B *vs.*BY, all Ps<0.001, Bonferroni-corrected paired t-tests), but there was no difference in responding between the compound cues (AX *vs.* BY, Ps>0.967, Bonferroni-corrected paired

t-tests). **(D-F)** Probability of presence in port throughout cue presentation during last 4 days of Individual Cue (upper graphs) and 4 days of Compound Cue conditioning (lower graphs), for Reward Upshift **(D)**, VTA-DA stimulation **(E)**, and SNc-DA stimulation **(F)** groups. Note that photostimulation during compound cue BY did not disrupt ongoing behavior. See also Figure S1.

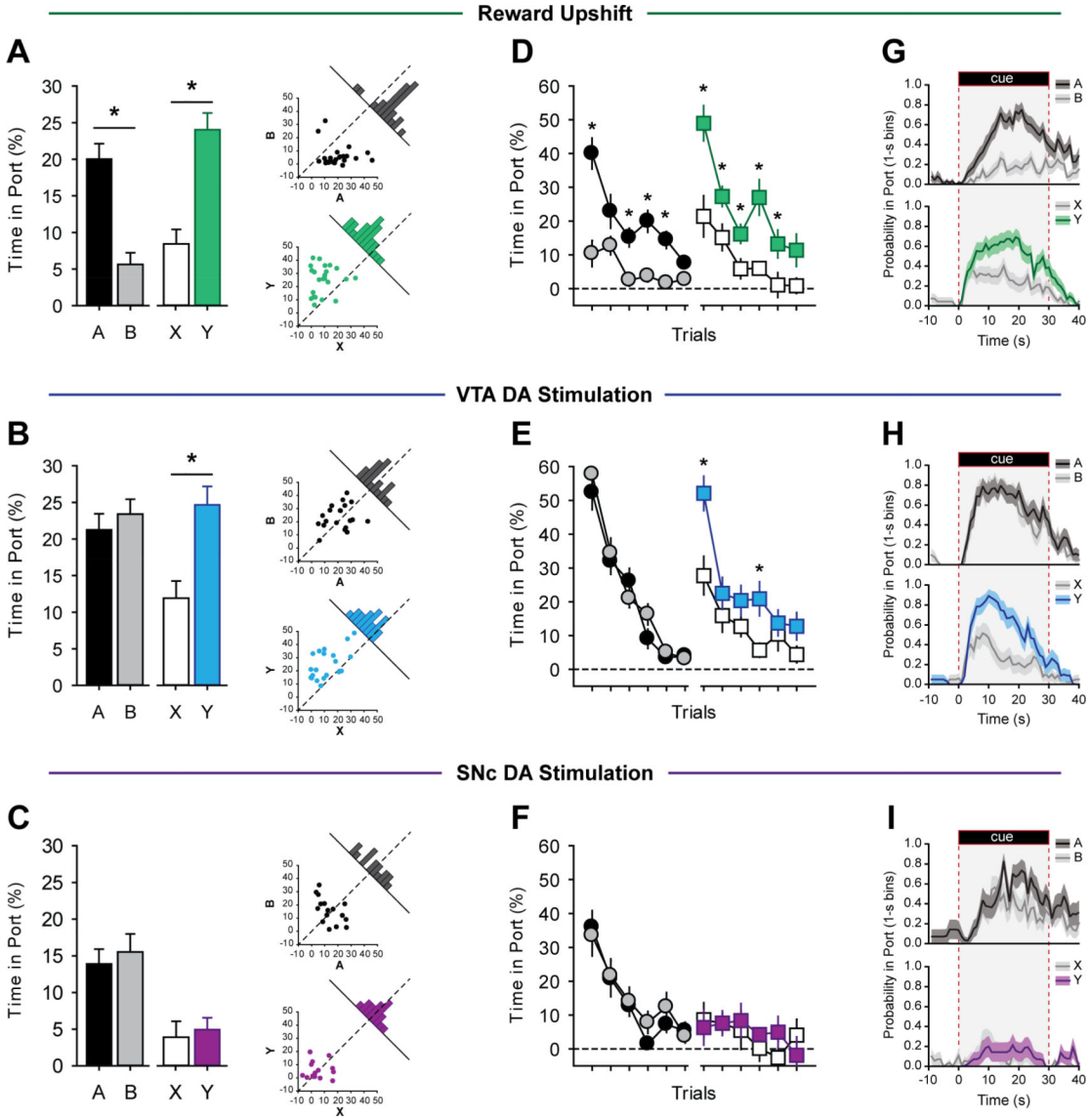

Figure 3. Photoactivation of VTA-DA but not SNc-DA neurons mimics endogenous RPEs and unblocks learning.

Conditioned responding was measured by time spent in the reward port during cue presentation. **(A-C)**: Whole session performance in Reward Upshift **(A)**, VTA-DA stimulation **(B)**, and SNc-DA stimulation **(C)** groups. Scatterplot inserts show individual data distributions for responding to A and B (top inserts) and X and Y (bottom insert). Histograms along the diagonal are frequency distributions (subject counts) for the difference scores (A - B, or X - Y); off-centered distributions reveal higher responding to one of the cues. **(D-F)**. Trial-by-trial test performance after Reward Upshift **(D)**, VTA-DA stimulation **(E)**, and SNc-DA stimulation **(F)**. A 3-way mixed ANOVA (Group x Cue x Trial) analyzed the evolution of responding over the session and found an interaction between all factors ($F_{30,855}$=2.603, P<0.001, after Greenhouse-Geisser correction). **(G-I)** Second-by-second tracking of presence in port during first presentation of each cue (A, B: upper graph; X, Y: lower graph) for Reward Upshift **(G)**, VTA-DA stimulation **(H)**, and SNc-DA stimulation **(I)**

groups. *P<0.05 (A *vs.* B, or X *vs.* Y; Post-hoc Bonferroni-corrected t-test). Error bars = s.e.m. See also Figures S1-S3

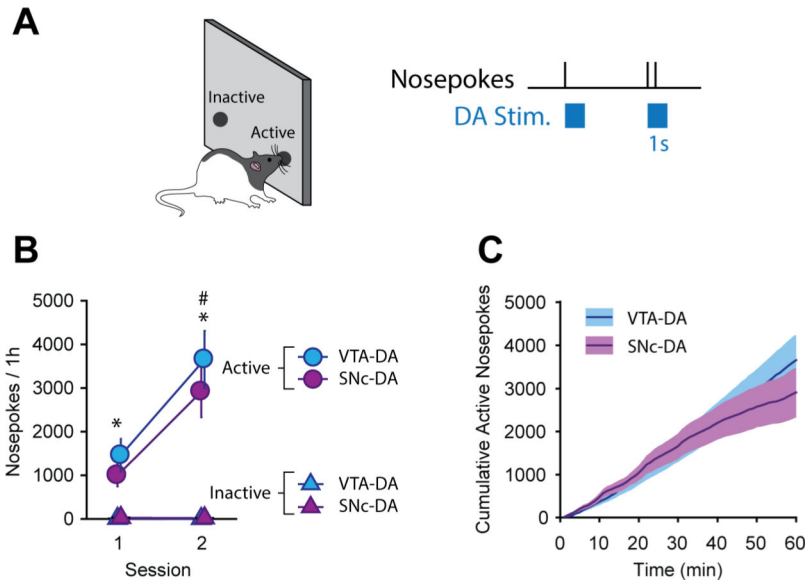

Figure 4. Photoactivation of VTA-DA or SNc-DA neurons serves as an equally potent reinforcer of ICSS behavior.

(A) Rats could respond on one of two nosepokes to obtain optical stimulation of VTA- or SNc-DA neurons. **(B)** Responses at active and inactive nosepokes during daily 1-h sessions. **(C)** Cumulative active nosepoke responses during the last ICSS session. *P<0.05, Active vs. Inactive Nosepoke; #P<0.05, Session 1 vs. Session 2 (active nosepoke). Error bar and error bands = s.e.m.

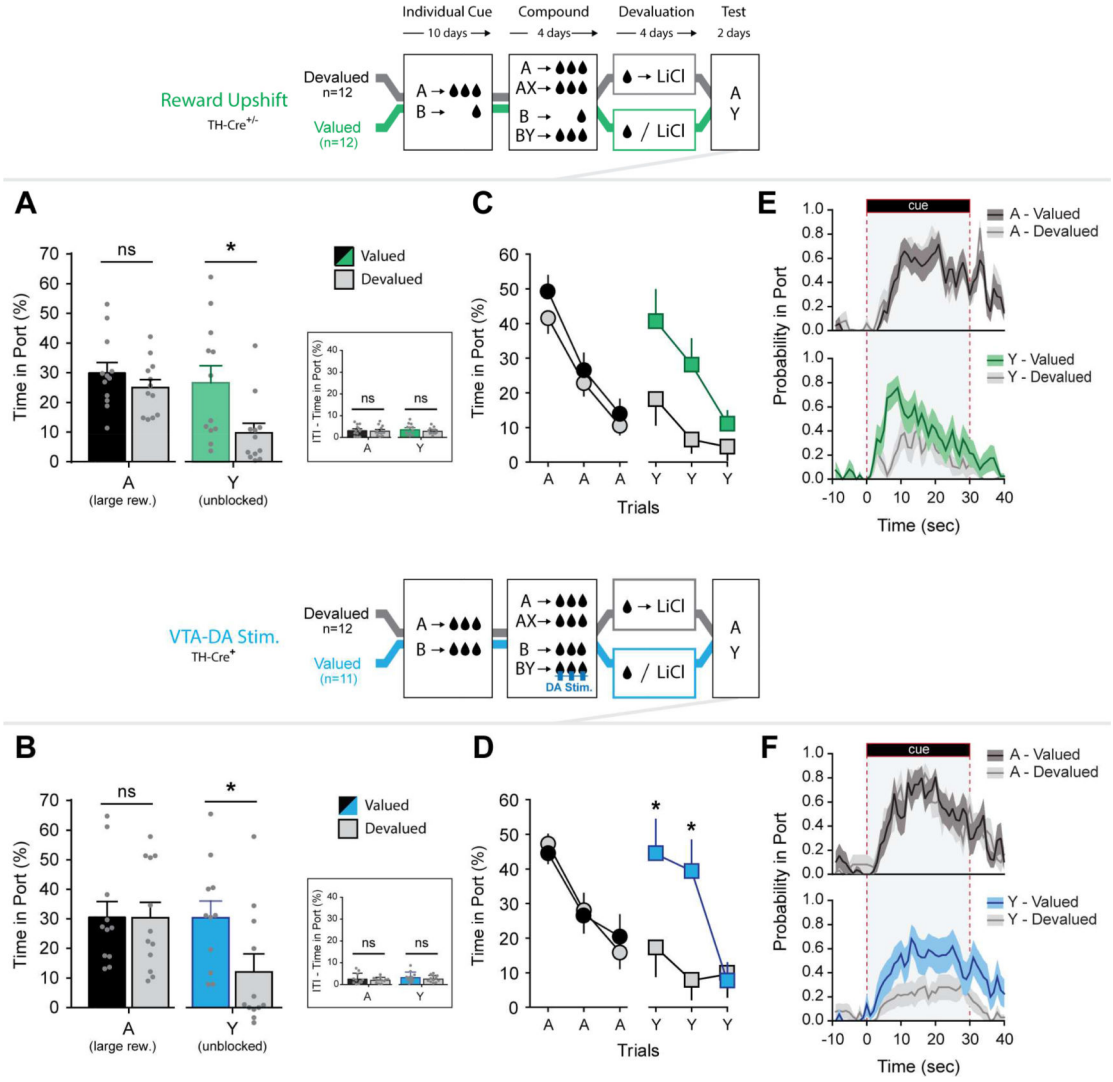

Figure 5. Devaluation of the sucrose outcome abolishes conditioned responding to the unblocked cue Y in Reward Upshift and VTA-DA groups.

Learning about target cue Y was unblocked by reward upshift (top graphs) or activation of VTA-DA neurons (bottom graphs). Following unblocking, sucrose was devalued for half of the subjects in Reward Upshift and VTA-DA groups by pairing sucrose consumption with LiCl (Devalued condition). The remaining subjects were exposed to sucrose or LiCl-induced illness on alternate days, preserving the value of sucrose (Valued condition). Conditioned responding to Y (unblocked cue) and A (cue paired with large reward) was assessed at Test. **(A, B)** Time spent in reward port during cue presentation in Reward Upshift **(A)** and VTA-DA **(B)** groups. Sucrose devaluation reduced responding to Y in both groups. Insets represent inter trial interval (ITI) responding outside cue presentation. **(C, D)** Trial-by-trial performance in Reward Upshift **(C)** and VTA-DA stimulation **(D)** groups. 3-way ANOVAs (Cue x Devaluation x Trial) found an interaction between these factors for VTA-DA ($F_{2,20}$=3.901, P=0.037) but not Reward Upshift ($F_{2,21}$=1.276, P=0.300) subjects. **(E, F)** Second-by-second tracking of presence in port during first presentation of each cue. *P<0.05

(Valued vs. Devalued; Bonferroni-corrected t-test). Error bar and error bands = s.e.m. See also Figures S4-S5.