Evaluation of the hypothesis that phasic dopamine constitutes a cached-value signal

Melissa J. Sharpe[a,b,c,*], Geoffrey Schoenbaum[a,d,e,*]

[a] National Institute on Drug Abuse, Baltimore, MD, USA
[b] Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA
[c] School of Psychology, UNSW Australia
[d] Department of Anatomy and Neurobiology, University of Maryland School of Medicine, Baltimore, MD 21201, USA
[e] Solomon H. Snyder Department of Neuroscience, The John Hopkins University, Baltimore, MD, USA

ABSTRACT

The phasic dopamine error signal is currently argued to be synonymous with the prediction error in Sutton and Barto (1987, 1998) model-free reinforcement learning algorithm (Schultz et al., 1997). This theory argues that phasic dopamine reflects a cached-value signal that endows reward-predictive cues with the scalar value inherent in reward. Such an interpretation does not envision a role for dopamine in more complex cognitive representations between events which underlie many forms of associative learning, restricting the role dopamine can play in learning. The cached-value hypothesis of dopamine makes three concrete predictions about when a phasic dopamine response should be seen and what types of learning this signal should be able to promote. We discuss these predictions in light of recent evidence which we believe provide particularly strong tests of their validity. In doing so, we find that while the phasic dopamine signal conforms to a cached-value account in some circumstances, other evidence demonstrate that this signal is not restricted to a model-free cached-value reinforcement learning signal. In light of this evidence, we argue that the phasic dopamine signal functions more generally to signal violations of expectancies to drive real-world associations between events.

1. Introduction

The finding that dopamine neurons signal errors in reward prediction has ushered in a revolution in behavioral neuroscience. For decades before this signal was discovered in the brain, errors in reward prediction- referred to as 'surprise' signals- have been the lynchpin in associative learning models, in which they are proposed to be the critical force driving the acquisition of associations between events (Rescorla & Wagner, 1972; Wagner & Rescorla, 1972). Take, for example, the typical blocking experiment (Kamin, 1969). Here, a light leads to presentation of food. Then, the light and a novel tone are presented simultaneously and followed by the same food reward. Humans and other animals learn to use the light to predict delivery of reward (Corlett et al., 2004; Hinchy, Lovibond, & Ter-Horst, 1995; Kamin, 1969). However, they typically do not appear to learn that the novel tone predicts reward delivery, despite it being repeatedly paired with the very same reward (Corlett et al., 2004; Hinchy et al., 1995; Kamin, 1969). This simple experiment demonstrates the importance of prediction errors for driving the learning of relationships between events; we will only learn to relate them insofar as they tell us something new about the associative structure of our world. Thus the discovery that dopamine neurons were broadcasting such a signal throughout key associative learning circuits in the brain was a milestone.

However when the phasic dopamine signal was discovered in the midbrain, it was quickly interpreted as reflecting a cached-value signal described in model-free temporal-difference reinforcement learning algorithms (TDRL; Schultz, Dayan, & Montague, 1997; Sutton & Barto, 1981, 1987, 1998). Specifically, the finding that dopamine neurons exhibit a phasic increase to unexpected reward which gradually transfers to the beginning of the reward-predictive cue was argued to constitute the transfer of cached-value from the reward to the cue, proposed to occur in the TDRL model (Schultz et al., 1997). Importantly, in the TDRL models, this transfer or learning endows the cue with a scalar value representing a knowledge of how good the reward was, however it does not allow the formation of any associative relationship between the neural or psychological representations of the cue and the actual reward (Sutton & Barto, 1981, 1987). Hence the term "cached value" to

describe what is learned. A host of studies have shown that the dopamine prediction error *correlates* with putative measures of such cached value. Dopamine neurons show a phasic increase in activity when unexpected rewards are delivered or are better than expected, and the magnitude of the response correlates with the size of the unexpected reward (Fiorillo, Tobler, & Schultz, 2003; Lak, Stauffer, & Schultz, 2014; Schultz, 1986; Schultz, Apicella, & Ljungberg, 1993; Stauffer, Lak, & Schultz, 2014). Further, the signal which accrues to the reward-predictive cue also reflects with impressive accuracy the subjective value of the upcoming future sum of expected rewards (Lak et al., 2014; Schultz et al., 1993; Stauffer et al., 2014).

While the close correspondence in these studies between dopaminergic correlates and cached-value errors is impressive, it is problematic because we know that humans and other animals form detailed representations of the associative relationship between events in a manner that transcends value (Balleine & Dickinson, 1991; Blundell, Hall, & Killcross, 2001, 2003; Colwill & Rescorla, 1985; Dickinson & Balleine, 1994; Holland & Rescorla, 1975; Rescorla, 1973). For example, if we were to pair our light and tone together prior to any experience with reward and then subsequently pair the light with reward by itself, subjects will learn that both the light and tone will lead to reward, despite only one of them being directly paired with reward (Brogden, 1939). Termed sensory preconditioning, this procedure demonstrates, in a simple way, the formation of a rich associative structure of the world that allows us to make novel inferences about rewards even when we have not directly experienced these associative relationships. Even such a simple phenomenon as sensory preconditioning cannot be explained by a cached-value model of learning, which argues that value transfers back to the cue from the reward, as the tone in this case has never been directly paired with reward. Thus, interpreting phasic dopamine as reflecting a cached-value signal dramatically limits the role of dopamine in the development of the more complex associative relationships that truly characterize cognitive behavior.

So, is phasic dopamine activity restricted to signalling cached-value errors? The hypothesis that phasic dopamine acts as a cached-value signal makes three notable predictions about when changes in phasic dopamine activity should be observed and what sorts of learning this phasic activity can support. Firstly, this theory predicts that stimulation or inhibition of dopamine neurons should act as a value signal to produce increments of decrements in responding to reward-paired cues. Secondly, such manipulations should not produce learning about the relationships between events outside of a scalar expectation of value. Finally, phasic activity in dopamine neurons should not be evident in response to valueless changes in reward or to cues which have come to predict a particular reward indirectly. We will now discuss these predictions in light of several recent studies that we believe provide particularly strong tests of their validity.

2. Prediction one: Phasic stimulation or inhibition of dopamine neurons should substitute as a cached-value prediction error to drive learning

The first prediction of the hypothesis that phasic dopamine constitutes a cached-value signal is that stimulation or inhibition of dopamine neurons should serve to increase or decrease the value attributed to the antecedent reward-paired cue. The advent of optogenetics affords us the cell-type and temporal specificity to causally assess this hypothesis (Deisseroth, 2011; Deisseroth et al., 2006). Indeed, Steinberg et al. (2013) recently demonstrated that phasic stimulation of putative dopamine neurons in the ventral tegmental area (VTA) during the blocking procedure could drive an increase in learning. Here, rats first learnt that cue X leads to a food reward (X → US). Subsequently, novel cue A and X are presented as a simultaneous compound with the same reward (AX → US). Under normal circumstances, rats will show little learning about cue A as it has been blocked by prior training with

cue X and reward. However, phasic stimulation of VTA dopamine neurons during reward after presentations of compound cue AX restored learning about cue A, as indexed by greater levels of entry into the food port during presentation of cue A under extinction (Steinberg et al., 2013). These results are compatible with an interpretation that phasic dopamine reflects a cached-value signal. Specifically, introduction of a phasic dopamine signal could function as an error signal that allows excess value to accrue to cue A despite the predictability of the reward. This would permit cue A to become associated with food-port entry being made during presentation of the AX compound and lead to the enhanced responding in the presence of cue A.

However, the finding that phasic stimulation of dopamine increases learning about cue A could also be construed as increasing the salience of cue A (Berridge & Robinson, 1998; Ungless, 2004). If dopamine does in fact function as a salience signal which determines the rate of learning, then inhibiting dopamine should result in less learning. If, on the other hand, phasic dopamine acts as a cached-value error signal, then phasic inhibition of dopamine should cause a reduction in the value attributed to a cue and produce extinction. In order to dissociate between these hypotheses Chang et al. (2016) briefly inhibited dopamine neurons in rats to introduce a negative error during an overexpectation task. Overexpectation usually involves first pairing two cues individually with reward (e.g. A → US; X → US). Then, these two cues are paired together with the same magnitude of reward (AX → US). Here, rats would usually extinguish learning about cue X as the reward is now "over-expected" by the summed expectations elicited by cue A and X (i.e. 2US; (Rescorla, 1970). However, in a modified version of the task, Chang et al. (2016) presented the compound AX with the expected reward during the second phase of learning (AX → 2US). This functioned to maintain learning to cue X. In half the rats, VTA dopamine was briefly inhibited during the reward delivery in the second phase after presentation of compound AX. Chang et al. (2016) found that inhibition of dopamine during this phase restored normal extinction learning to cue X. That is, inhibition of dopamine neurons resulted in greater amounts of learning- in the form of extinction learning- rather than less learning as would be predicted if turning down dopamine resulted in a decrease of salience. These results cannot be explained by the proposal that phasic dopamine functions as a salience signal since in that case as less dopamine should result in less learning (and a failure to show extinction learning). Rather, these results are again consistent with the cached-value hypothesis of dopamine, where dopamine functions as a bidirectional error signal to increase or decrease value attributed to a reward-predictive cue.

3. Prediction two: What is stamped in by manipulating phasic dopamine activity should be related to cached value

Experiments showing that optogenetic stimulation or inhibition can drive increases or decreases in responding to reward-predictive cues are consistent with the idea that phasic dopamine constitutes a scalar value which increases or decreases the value attributed to a reward-paired cue. However, in the studies described above (Chang et al., 2016; Steinberg et al., 2013) as well as many others (Tsai et al., 2009; Adamantidis et al., 2011; Witten et al., 2011) the learning induced by manipulating the firing of the dopamine neurons is not probed to determine what information is actually being acquired. The simple behaviors that were assessed in these studies could be easily supported by cached-value learning. However they could equally well reflect the formation of a more detailed associations between the cue and reward in the case of unblocking and the cue and reward omission in the case of extinction. The former would constitute a learning mechanism consistent with that described in the model-free reinforcement algorithm postulated by Sutton and Barto (1987, 1998), whereas the latter would reflect more complex associations between events that transcend the backpropagation of value to the reward-predictive cue. The experimental designs described above confound these two possibilities.

To avoid this confound, we assessed whether manipulating dopamine neurons would alter learning in the sensory preconditioning procedure. As mentioned above, sensory preconditioning usually involves pairing two neutral cues together in close succession such that a relationship forms between them (e.g. A → X). Then, cue X is paired directly with reward. Subsequently, both cues A and X will elicit an appetitive response to enter the food port. As cue A has never been directly paired with reward, it can only enter into a relationship with reward through its association with cue X. This is supported by our recent findings showing that preconditioned cues do not support conditioned reinforcement (Sharpe, Batchelor, & Schoenbaum, 2017). Here, we trained rats on a standard preconditioning procedure (A → X; X → US). Following this training, we tested whether rats will press a lever to receive presentation of cue A or cue X. We found that rats would readily press a lever to receive presentations of cue X. However, they would not press to receive presentations of cue A. These data suggest that the preconditioned cue A did not have any cached value outside of its model-based association with X and reward. These features of the sensory preconditioning procedure make it an ideal procedure to test whether dopamine is involved in the development of more complex association, independent of reward.

Using a modified version of the sensory preconditioning procedure, we investigated whether dopamine transients are sufficient for this more complex form of associative learning (Sharpe, Chang, et al., 2017; Fig. 1). To do this, we reduced the likelihood that rats would form an association between the two neutral cues during preconditioning. Specifically, we first paired cues A and X together in close succession (A → X), in line with the standard design. However, prior to pairing X with reward, we introduced a blocking phase. Here, we presented an additional cue C in compound with cue A, followed by presentation of cue X (AC → X). Then we paired X with reward. In normal rats, we found that learning about cue C was blocked. This demonstrated that learning the relationship between neutral cues in the sensory preconditioning procedure is subject to an error mechanism. However, brief stimulation of dopamine neurons at the beginning of cue X when it was preceded by compound AC restored learning about the C → X association. Importantly there was no reward present during this preconditioning phase when dopamine neurons were being stimulated nor were the rats engaged in food cup responding. Thus dopamine did not directly drive the acquisition of the response. Further there was no change in learning to X, the cue present when dopamine was triggered, in the subsequent conditioning phase. This suggests that dopamine did not directly alter the value, salience or associability of the cues present when it was delivered, since if it had then learning for X should have been facilitated. Instead, the simplest and by far most likely explanation of this effect is if dopamine acted to endow rats with the knowledge of the associative relationship between cue C and X, which then allowed cue C to predict reward after X was paired with food. Consistent with this assertion, the responding in the probe test was sensitive to devaluation of the food reward.

4. Prediction three: Phasic changes in dopamine should only reflect associations acquired through direct experience

While the findings from Sharpe et al. (2017) suggest that dopamine can support the acquisition of complex associations between events (rewarding or otherwise), this does not require that the content of information contained in the prediction error itself go beyond errors in cached value. That is, stimulation or inhibition of dopamine could be allowing other neural structures to form more complex associations about the relationship between events, yet phasic activity in dopaminergic neurons may be ignorant of these associations under normal circumstances, changing only in response to cached-value errors. If this is the case, then phasic activity in dopamine neurons should not reflect associations that have been inferred from prior associative relationships (as is the case in sensory preconditioning) or a change in the current

state of the environment. This is because a cached-value error only receives predictions based on value that back propagates from the reward to the cue after the cue and reward have been paired together in close succession. This cannot happen if a contingency has not been directly experienced.

Assessing whether the dopamine prediction error has access to information about the relationship between events requires examining how dopamine neurons or dopamine release changes in response to errors that reflect such associative information. There are now a growing number of studies that do this (Aitken, Greenfield, & Wassum, 2016; Bromberg-Martin & Hikosaka, 2009; Nakahara, Itoh, Kawagoe, Takikawa, & Hikosaka, 2004; Papageorgiou, Baudonnat, Cucca, & Walton, 2016; Sadacca, Jones, & Schoenbaum, 2016; Takahashi et al., 2011). For example, dopamine activity to reward-paired cues changes depending on the physiological state of the subject (Aitken et al., 2016; Papageorgiou et al., 2016). In one study, Papageorgiou et al. (2016) monitored dopamine release using fast scan voltammetry in the nucleus accumbens (NaCC) as rats were performing an instrumental learning task. Here, rats had a choice of pressing one of two levers for one of two rewards (R1 → O1 or R2 → O2). On some of the trials, rats were presented with one lever option (forced trials; R1 or R2) while on others they could make a choice between pressing either one of the two levers (choice trials; R1 and R2). Prior to test sessions, rats were given free access to one of the rewards (e.g. devaluing O1). Subsequently, rats exhibited a preference for the lever associated with the non-devalued reward they had not had access to prior to the session (R2 → O2). Papageorgiou et al. (2016) found that dopamine release to the reward-paired cues (i.e. the insertion of the lever into the behavioral chamber) was modulated by outcome devaluation prior to the rats experiencing the lever producing the now devalued outcome. That is, the dopamine response to lever presentation on forced trials reflected the new value of the devalued reward before it had been experienced with the lever-press response. Further, the dopaminergic response to presentation of the other lever was increased, showing an increased preference for the non-devalued option. This demonstrates that dopamine responses to reward-paired cues can update in response to the current physiological state of the subject without the subject directly experiencing the association between the cue and now devalued reward. These data are at odds with an interpretation of the dopamine signal as the model-free reinforcement learning algorithm described by Sutton and Barto (1981, 1998), since the cue and the devalued reward have never been paired, and so the new value of the reward cannot be attributed to the cue which precedes its occurrence.

The data from Papageorgiou et al. (2016) beg the question of whether the phasic dopamine signal might also reflect information about an entirely new association developed in the absence of experience. In line with this possibility, Sadacca et al. (2016) showed that phasic activity of dopamine neurons can reflect associations between cues and rewards that have been inferred from prior knowledge of associative relationships in the experimental context. Specifically, Sadacca et al. (2016) recorded the activity of putative dopamine neurons in the VTA during sensory preconditioning. In this study, rats were first presented with two neutral cues in close temporal succession (A → X). Following this training, one of these cues was paired with reward (X → US). During conditioning, putative dopamine neurons exhibited the expected reward prediction-error correlates, firing to reward early in conditioning and transferring this response back to the cue later in learning. After conditioning, in the probe test in which both cues A and X were presented in the absence of reward, putative dopamine neurons continued to exhibit increased firing to X, the cue paired with reward, while also now firing to A, the cue paired with X in the preconditioning phase. Further, dopamine neuron firing to A and X was correlated, suggesting that the information signalled in response to A was the same as what was signalled in response to X. The simplest interpretation of these data is that dopamine neurons in the VTA signal reward prediction errors similarly whether they are based on directly experienced
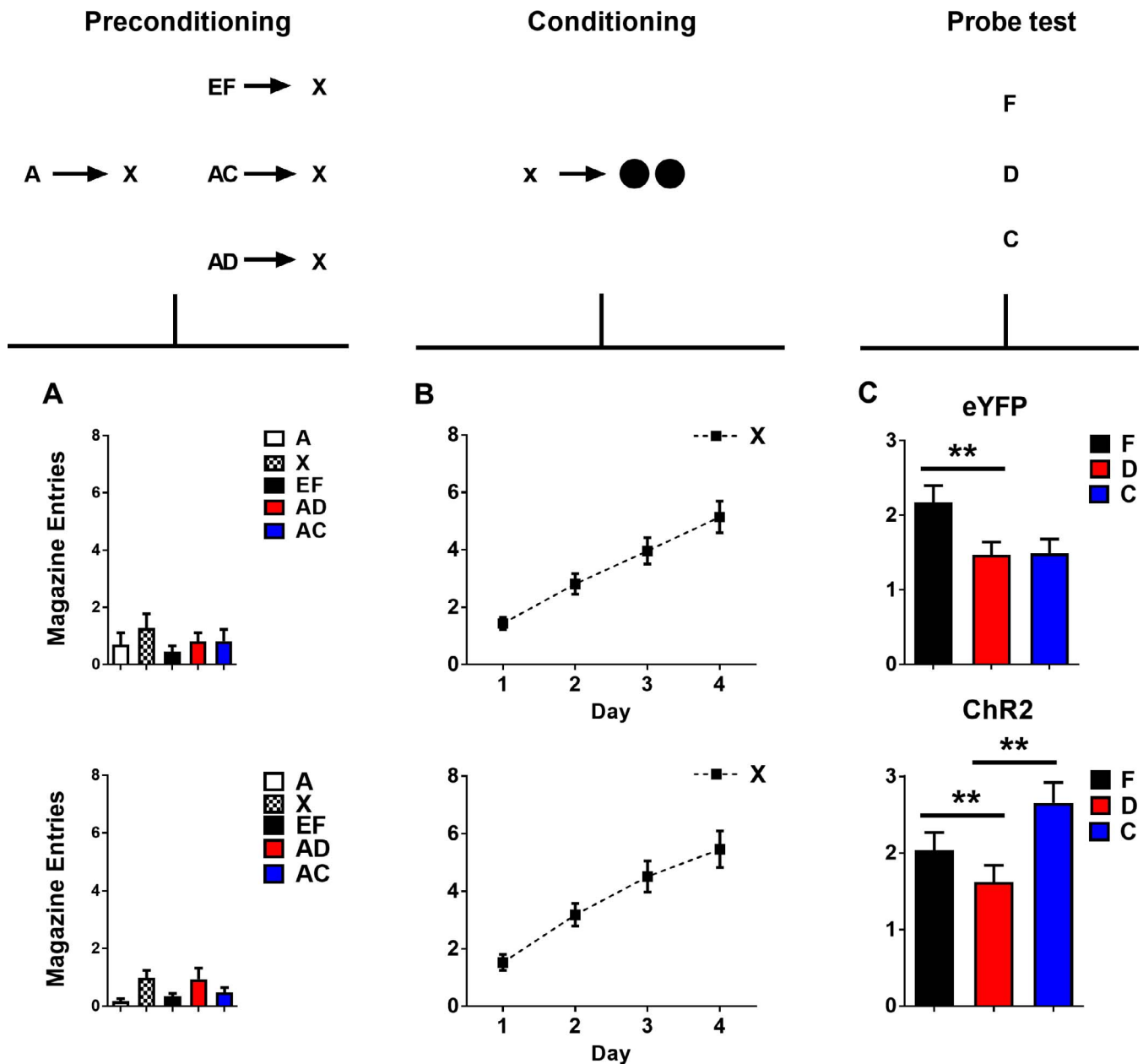
Fig. 1. Brief optogenetic activation of VTA dopamine neurons strengthens associations between cues (adapted from Sharpe, Chang, et al., 2017). Plots show number of food cup entries occurring during cue presentation across all phases of the blocking of sensory preconditioning task for the eYFP control group (top) and the ChR2 experimental group (bottom): (A) preconditioning, (B) conditioning, and (C) the probe test. Brief stimulation of dopamine neurons in the ChR2 group during the presentation of X when it was preceded by compound AC unblocked learning of the C–X association. This allowed C to enter into an association with sucrose-pellet reward and promote conditioned responding directed towards the food port. ** indicates significance at the $p < .05$ level for either a main effect (F vs D) or simple main effect following a significant interaction (D vs C).

associations or whether they require inference. Again this is not accomodated by a theory which argues that the dopamine signal reflects value which has back propagated from the reward to a cue from their pairing (a notion reinforced by data showing a preconditioned cue does not acquire general value during the preconditioning procedure; Sharpe, Batchelor, et al., 2017). Rather, these data suggest that dopamine neurons may make more general predictions about the nature of upcoming rewards, garnered from associative model of the world and based on past experience.

5. Where to now?

Here we have discussed recent data that provide strong tests of key predictions of the hypothesis that phasic changes in dopamine are restricted to signalling the cached-value errors to support cached value

learning, as described in model free reinforcement learning algorithms (Sutton & Barto, 1981, 1987). Consistent with this proposal, optogenetic stimulation of dopamine neurons acts to increase learning about reward-paired cues (Steinberg et al., 2013). However such manipulations appear to produce complex associations between sensory information, which allow rats to make inferences about associative relationships they have not directly experienced (Sharpe et al., 2017). Such learning cannot be easily explained as reflecting cached value. Further, phasic activity in dopamine neurons also reflects the value derived from these complex associative models of the world, including sensory preconditioning (Sadacca et al., 2016) and also in response to changes in physiological state (Aitken et al., 2016; Papageorgiou et al., 2016). On the whole, these data challenge the conception that transient changes in dopamine are restricted to carrying the cached-value prediction error described in the models currently applied to interpret

dopamine function, since in these models, value cannot transfer back to a cue which has not been paired with something valuable and a value signal cannot facilitate the acquisition of associations between neutral stimuli.

So how do we accommodate these data into a framework which describes dopamine function? Two similar models have recently been put forward that attempt to reconcile such findings with existing models of dopamine function. Specifically, Nakahara (2014) and Gershman (2017) argue that the dopaminergic error system can be influenced by more than the expectation elicited by the cue which is currently present. Rather, the prediction error has access to associative models of the world which are distributed across the brain. This allows the cached-value prediction error to take into account prior associative relationships garnered from past experience when making predictions about the scalar value of upcoming rewards in novel circumstances and, in turn, update knowledge of these associative networks. However, what is critical about the theories posited by Nakahara (2014) and Gershman (2017) is that in each of case the error exhibited by dopamine neurons remains a cached-value error. That is, while the error has access to a knowledge of associative relationships that transcend computations of value, the error which is elicited is explicitly value based. In essence phasic activity in dopamine neurons still reflects the future expected sum of rewards in scalar form, despite it's ability to make predictions using knowledge which transcends this information. Thus, according to these theories the dopamine error is still a cached-value signal that should not facilitate the acquisition associative relationships between the neural or psychological representations of events in the environment.

These models expand the sorts of learning we might expect phasic dopamine activity to support and when might expect to see changes in phasic activity. For example, both the Nakahara (2014) and Gershman (2017) models can explain the finding that dopamine neurons respond to cues that have acquired the ability the predict the reward indirectly, as is the case in sensory preconditioning. Specifically, as the prediction error has access to prior associative relationships between cues, it can produce an inference that the preconditioned cue is likely to lead to reward due to the previous associative relationship with the cue directly paired with reward. However, these models still propose that the dopamine signal observed in response to the preconditioned cue is a scalar value signal. That is, the dopamine response to the preconditioned cue still reflects the value which has transferred back to the preconditioned cue through this inferred process. This becomes problematic when we consider our recent findings that a preconditioned cue will not support conditioned reinforcement and, therefore, does not possess cached value (Sharpe, Batchelor, et al., 2017). This demonstrates that the dopamine response to the preconditioned cue does not reflect the upcoming scalar value of predicted reward. Rather, the dopamine response must be signalling something that transcends this cached-value prediction.

In addition, these models also cannot easily explain how activation of dopamine neurons is able to support the model-based learning developed during sensory preconditioning. Specifically, the findings reported by Sharpe et al. (2017) demonstrate that dopamine stimulation is capable of facilitating the formation of associations between the two neutral cues in the preconditioning phase (i.e. C → X), subsequently allowing rats to make the novel inference that cue C may lead to food after it's associate X has been paired with reward. If dopamine was functioning during this procedure to endow cue C with a cached value, it would not change behavior in the manner in which we observed. That is, endowing cue C with value would not have facilitated the formation of an association between C and X, subsequently allowing C to enter into a direct association with rewardand produce an increase in devaluation-sensitive magazine entries to cue C. Thus, while the models proposed by Nakahara (2014) and Gershman (2017) expand the ways in which dopamine can influence behavior, they cannot explain findings that are incompatible with an interpretation that dopamine is

signalling a cached-value error (Sadacca et al., 2016; Sharpe et al., 2017) – even if they allow the error computation to have access to associative structures of the world garnered through past experience.

An alternative proposal is that dopamine transients reflect errors in event prediction more generally and that they are also involved in supporting learning about future events whether those events are the delivery of a particular reward, presentation of a neutral stimulus, or even absence of a some stimulus or some other event. This would constitute a return to thinking about the prediction error in associative theory as driving real world associations between events, as described in earlier theories of associative learning (Colwill & Rescorla, 1985; Holland & Rescorla, 1975; Miller & Matzel, 1988; Rescorla, 1973; Rescorla & Wagner, 1972; Wagner & Rescorla, 1972; Wagner, Spear, & Miller, 1981) but somewhat abandoned by the world of neuroscience with the advent of TDRL and the concept of cached-value (Sutton & Barto, 1981, 1987, 1998).

Conceptualising the dopamine prediction error as a signal that detects a discrepancy between expected and actual events make some testable predictions about when phasic activity should be observed. Specifically, the alternative proposal made here suggests that changes in phasic dopaminergic activity should be seen as a result of other changes in the predicted event that do not constitute a shift in value. For example, an increase in dopaminergic signalling should occur in response to a change in the identity of a reward. That is, if a cue previously paired with a particular reward was unexpectedly presented with a different reward that was equally valuable, we would expect to see a prediction error in dopaminergic neurons. And in fact recent evidence has emerged to suggest dopamine does in fact encode such information (Takahashi et al., 2017). Specifically, Takahashi et al. (2017) have shown that dopamine neurons exhibit their classic prediction-error signal to changes in the sensory properties of rewards that are equally prefered. That is, dopamine neurons show errors to the change in reward identity without a change in reward value. These data support the alternative hypothesis suggested here. Namely, that dopamine neurons encode more general violations of expectations whether or not that reflects a change in value.

Future research may also search for the presence of a dopaminergic error signal when a more general associative relationship between neutral stimuli is violated even in the absence of rewards. It is wellestablished that dopamine neurons in the midbrain fire when a novel stimulus is first presented unexpectedly (Schultz, 1998). While this has been interpreted in the literature as a "novelty bonus" (Kakade & Dayan, 2002), it is also possible that this is an error signal in response to the appearance of an unexpected stimulus. Given the apparent role of dopamine transients in supporting preconditioning (Sharpe et al., 2017), it would be valuable to assess in an appropriately controlled environment whether these dopamine signals are seen when the contingency between neutral stimuli is manipulated such that expectation about upcoming stimuli is violated. Such research would support the hypothesis that the dopamine prediction error may reflect a more general signal for detecting the discrepency between actual and expected events. Experiments like these would be useful since positive findings would open up new possibilities for how this biological signal may support associative learning in these and other contexts.

References

Adamantidis, A. R., Tsai, H. C., Boutrel, B., Zhang, F., Stuber, G. D., Budygin, E. A., ... de Lecea, L. (2011). Optogenetic interrogation of dopaminergic modulation of the multiple phases of reward-seeking behavior. *Journal of Neuroscience, 31*(30), 10829–10835.

Aitken, T. J., Greenfield, V. Y., & Wassum, K. M. (2016). Nucleus accumbens core dopamine signaling tracks the need-based motivational value of food-paired cues. *Journal of Neurochemistry, 136*(5), 1026–1036.

Balleine, B., & Dickinson, A. (1991). Instrumental performance following reinforcer devaluation depends upon incentive learning. *The Quarterly Journal of Experimental Psychology, 43*(3), 279–296.

Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward:

Hedonic impact, reward learning, or incentive salience? *Brain Research Reviews, 28*(3), 309–369.

Blundell, P., Hall, G., & Killcross, S. (2001). Lesions of the basolateral amygdala disrupt selective aspects of reinforcer representation in rats. *Journal of Neuroscience, 21*(22), 9018–9026.

Blundell, P., Hall, G., & Killcross, S. (2003). Preserved sensitivity to outcome value after lesions of the basolateral amygdala. *Journal of Neuroscience, 23*(20), 7702–7709.

Brogden, W. (1939). Sensory pre-conditioning. *Journal of Experimental Psychology, 25*(4), 323.

Bromberg-Martin, E. S., & Hikosaka, O. (2009). Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron, 63*(1), 119–126.

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

Colwill, R. M., & Rescorla, R. A. (1985). Postconditioning devaluation of a reinforcer affects instrumental responding. *Journal of Experimental Psychology: Animal Behavior Processes, 11*(1), 120.

Corlett, P. R., Aitken, M. R., Dickinson, A., Shanks, D. R., Honey, G. D., Honey, R. A., ... Fletcher, P. C. (2004). Prediction error during retrospective revaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. *Neuron, 44*(5), 877–888.

Deisseroth, K. (2011). Optogenetics. *Nature Methods, 8*(1), 26–29.

Deisseroth, K., Feng, G., Majewska, A. K., Miesenböck, G., Ting, A., & Schnitzer, M. J. (2006). Next-generation optical technologies for illuminating genetically targeted brain circuits: Soc Neuroscience.

Dickinson, A., & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learn Behav, 22*(1), 1–18.

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science, 299*(5614), 1898–1902.

Gershman, S.J. (2017). Dopamine, Inference, and Uncertainty. *bioRxiv,* 149849.

Hinchy, J., Lovibond, P. F., & Ter-Horst, K. M. (1995). Blocking in human electrodermal conditioning. *The Quarterly Journal of Experimental Psychology, 48*(1), 2–12.

Holland, P. C., & Rescorla, R. A. (1975). The effect of two ways of devaluing the unconditioned stimulus after first-and second-order appetitive conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 1*(4), 355.

Kakade, S., & Dayan, P. (2002). Dopamine: Generalization and bonuses. *Neural Networks, 15*(4), 549–559.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. *Punishment and Aversive Behavior,* 279–296.

Lak, A., Stauffer, W. R., & Schultz, W. (2014). Dopamine prediction error responses integrate subjective value from different reward dimensions. *Proceedings of the National Academy of Sciences, 111*(6), 2343–2348.

Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. *Psychology of Learning and Motivation, 22*, 51–92.

Nakahara, H. (2014). Multiplexing signals in reinforcement learning with internal models and dopamine. *Current Opinion in Neurobiology, 25*, 123–129.

Nakahara, H., Itoh, H., Kawagoe, R., Takikawa, Y., & Hikosaka, O. (2004). Dopamine neurons can represent context-dependent prediction error. *Neuron, 41*(2), 269–280.

Papageorgiou, G. K., Baudonnat, M., Cucca, F., & Walton, M. E. (2016). Mesolimbic dopamine encodes prediction errors in a state-dependent manner. *Cell Reports, 15*(2), 221–228.

Rescorla, R. A. (1970). Reduction in the effectiveness of reinforcement after prior excitatory conditioning. *Learning and Motivation, 1*(4), 372–381.

Rescorla, R. A. (1973). Effects of US habituation following conditioning. *Journal of Comparative and Physiological Psychology, 82*(1), 137.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II: Current Research and Theory, 2*, 64–99.

Sadacca, B. F., Jones, J. L., & Schoenbaum, G. (2016). Midbrain dopamine neurons compute inferred and cached value prediction errors in a common framework. *Elife, 5*, e13665.

Schultz, W. (1986). Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. *Journal of Neurophysiology, 56*(5), 1439–1461.

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

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. *Journal of Neuroscience, 13*(3), 900–913.

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

Sharpe, M. J., Batchelor, H. M., & Schoenbaum, G. (2017). Preconditioned cues have no value. *Elife, 6*, e28362.

Sharpe, M. J., Chang, C. Y., Liu, M. A., Batchelor, H. M., Mueller, L. E., Jones, J. L., ... Schoenbaum, G. (2017). Dopamine transients are sufficient and necessary for acquisition of model-based associations. *Nature Neuroscience*.

Stauffer, W. R., Lak, A., & Schultz, W. (2014). Dopamine reward prediction error responses reflect marginal utility. *Current Biology, 24*(21), 2491–2500.

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.

Sutton, R. S., & Barto, A. G. (1987). *A temporal-difference model of classical conditioning.* Paper presented at the Proceedings of the ninth annual conference of the cognitive science society.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review, 88*(2), 135–170.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction, Vol. 1.* MIT Press Cambridge.

Takahashi, Y. K., Batchelor, H. M., Liu, B., Khanna, A., Morales, M., & Schoenbaum, G. (2017). Dopamine neurons respond to errors in the prediction of sensory features of expected rewards. *Neuron, 95*(6), 1395–1405 e1393.

Takahashi, Y. K., Roesch, M. R., Wilson, R. C., Toreson, K., O'donnell, P., Niv, Y., & Schoenbaum, G. (2011). Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nature Neuroscience, 14*(12), 1590–1597.

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), 1080–1084.

Ungless, M. A. (2004). Dopamine: The salient issue. *Trends in Neurosciences, 27*(12), 702–706.

Wagner, A., & Rescorla, R. (1972). Inhibition in Pavlovian conditioning: Application of a theory. *Inhibition and Learning,* 301–336.

Wagner, A. R., Spear, N., & Miller, R. (1981). SOP: A model of automatic memory processing in animal behavior. *Information Processing in Animals: Memory Mechanisms, 85*, 5–47.

Witten, I. B., Steinberg, E. E., Lee, S. Y., Davidson, T. J., Zalocusky, K. A., Brodsky, M., ... Stuber, G. D. (2011). Recombinase-driver rat lines: Tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron, 72*(5), 721–733.