Dopamine neurons respond to errors in the prediction of sensory features of expected rewards

Yuji K. Takahashi[1], **Hannah M. Batchelor**[1], **Bing Liu**[1], **Akash Khanna**[1,3], **Marisela Morales**[1], and **Geoffrey Schoenbaum**[1,2,3,*]

[1]Intramural Research program of the National Institute on Drug Abuse, NIH

[2]Department of Anatomy and Neurobiology, University of Maryland School of Medicine

[3]Department of Neuroscience, Johns Hopkins School of Medicine

Summary

Midbrain dopamine neurons have been proposed to signal prediction errors as defined in model-free reinforcement learning algorithms. While these algorithms have been extremely powerful in interpreting dopamine activity, these models do not register any error unless there is a difference between the value of what is predicted and what is received. Yet learning often occurs in response to changes in the unique features that characterize what is received, sometimes with no change in its value at all. Here, we show that classic error-signaling dopamine neurons also respond to changes value-neutral sensory features of an expected reward. This suggests that dopamine neurons have access to a wider variety of information than contemplated by the models currently used to interpret their activity and that, while their firing may conform to predictions of these models in some cases, they are not restricted to signaling errors in the prediction of value.

Keywords

dopamine; prediction error; learning; single-unit; rodent

Introduction

Midbrain dopamine neurons have been proposed to signal the reward prediction errors defined in model-free reinforcement learning algorithms (Schultz et al., 1997; Sutton, 1988). This proposal was initially based on observations that these neurons fired more strongly to unpredicted than to predicted reward, suppressed firing on omission of a predicted reward,

Correspondence or requests for material should be addressed to Y.T. (yuji.takahashi@nih.gov) or G.S. (geoffrey.schoenbaum@nih.gov).
*lead contact

and developed firing to reward-paired cues with learning (Mirenowicz and Schultz, 1994). Further work has shown that dopaminergic activity obeys formal predictions for model-free error signals under more complex conditions (Bayer and Glimcher, 2005; Hart et al., 2014; Hollerman and Schultz, 1998; Lak et al., 2014; Pan et al., 2005; Tobler et al., 2003; Waelti et al., 2001), including in tasks such as blocking and conditioned inhibition, in which experimental conditions are arranged to distinguish between prediction error signals and other possible explanations of such activity. Further, artificially stimulating or inhibiting dopamine neurons for very brief periods is sufficient to mimic the effects of endogenous positive or negative prediction errors (Chang et al., 2016; Steinberg et al., 2013). This body of work leaves little doubt that brief phasic changes in the activity of dopamine neurons can act like reward prediction errors, at least in some downstream targets and behavioral paradigms (Glimcher, 2011; Schultz, 2002; Schultz, 2016).

But are these neurons restricted to signaling the cached-value prediction error proposed to drive model-free reinforcement learning (Clark et al., 2012; Glimcher, 2011; Niv and Schoenbaum, 2008; Schultz, 2002) or might they signal errors in event prediction more broadly? This is an important question because the cached-value errors in model-free reinforcement learning algorithms support only a relatively limited subset of error-driven learning. For example, cached value errors would not occur in situations in which value remains the same but the specific properties of the predicted event change. Such learning occurs commonly in our daily lives and can be isolated in experimental procedures such as identity or transreinforcer unblocking and sensory preconditioning (Brogden, 1939; Burke et al., 2008; Jones et al., 2012; Rescorla, 1999), the latter recently shown to depend on dopamine transients (Sharpe et al., 2017). The existence of these value-neutral forms of learning (and the fact that they can be blocked) suggests that there must be some mechanism for signaling the underlying sensory or identity prediction errors. Here we tested whether the dopamine neurons might signal such errors and whether such signals would be dissociable from classic value-based prediction errors at the level of individual dopamine neurons.

Results

We recorded single-unit activity from the ventral tegmental area (VTA) of awake, behaving rats. Dopamine neurons were identified by means of a cluster analysis based on spike duration and amplitude ratio, features similar to those used to distinguish dopamine neurons in primates (Bromberg-Martin et al., 2010b; Fiorillo et al., 2008; Hollerman and Schultz, 1998; Kobayashi and Schultz, 2008; Matsumoto and Hikosaka, 2009; Mirenowicz and Schultz, 1994; Morris et al., 2006; Waelti et al., 2001). The cluster analysis identified 81 of 473 neurons recorded in the VTA of the experimental group as dopaminergic (n = 8, Fig. 1a). These neurons had a significantly lower amplitude ratio (Fig. 1b) and longer spike duration (Fig. 1c) than neurons in other clusters (amplitude ratio, t-test, $t_{471} = 15.84$, p < 0.01; spike duration, t-test, $t_{471} = 24.55$, p < 0.01). This cluster analysis has been shown previously to isolate wide-waveform neurons in rat VTA whose firing is sensitive to intravenous infusion of apomorphine or quinpirole (Jo et al., 2013; Roesch et al., 2007), and nigral neurons identified by a similar cluster analysis in mice are selectively activated by optical stimulation in tyrosine hydroxylase channelrhodopsin-2 mutants and show reduced bursting in tyrosine hydroxylase striatal-specific NMDAR1 knockouts (Xin and Costa,

2010). These features suggest that the wide waveform neurons identified in this manner are dopamine neurons, an assertion further confirmed here by the selective elimination of nearly all wide waveform neurons from recordings made in a separate group of TH-Cre rats that received ipsilateral infusions of a Casp3 neurotoxin (AAV1-Flex-taCasp3-TEVp) into VTA immediately prior to electrode implantation to destroy TH+ neurons (n = 5, Figs. 1a-c, red).

Neurons in the experimental group (Fig. 1a) were recorded during performance of an odor-guided choice task in which the rats sampled one of three different odor cues at a central port and then had to respond at either a left or right well to receive a big (three drop) or a small (one drop) amount of one of two equally preferred flavored milk solutions (chocolate or vanilla, Figs. 2a-c). One odor signaled reward in the left well (forced choice left), a second odor signaled reward in the right well (forced choice right), and a third odor signaled the availability of reward at either well (free choice). Rats were trained on the task prior to the start of recording, and during the recording sessions, we independently manipulated value or identity across blocks of trials by switching either the number or flavor of the reward delivered at each well (Fig. 2b). The rats responded to these manipulations by selecting the high value reward more often on free choice trials (Figs. 2d-e) and by responding faster and more accurately when the high value reward was at stake on forced choice trials (Figs. 2f-g).

Notably, these large changes in behavior were observed irrespective of reward identity. Accordingly, ANOVAs revealed main effects of changes in number on each of these measures (Fs > 91.9, ps < 0.0001), but there were no main effects nor any interactions with changes in flavor (Fs < 2.1, ps > 0.15). We also monitored licking behavior during reward delivery, which did not differ based on number or flavor (Fig. 2h).

Given the apparent lack of any effect of reward identity on the rats' behavior, we also examined these measures immediately before versus after changes in reward to make certain that the rats were attending to the feature (number or flavor) that was shifted (see methods for details). Neural data acquired in blocks in which none of these measures showed any significant change in the initial trials after a shift were excluded from the subsequent analysis. This amounted to 10 number shift blocks and 100 identity shift blocks. Importantly, the value-based behaviors illustrated in Fig 2a did not differ in these blocks (Fig. S1).

Dopamine neurons respond to errors in the prediction of reward value

Sixty neurons out of 81 putative dopamine neurons in the experimental group increased firing to reward. As in prior studies (Roesch et al., 2007; Takahashi et al., 2016; Takahashi et al., 2011), classic value-based prediction error signaling was observed almost exclusively in these reward-responsive putative dopamine neurons. These reward-responsive neurons (n = 60) changed firing in response to shifts in reward number at the start of blocks 2 and 4, increasing specifically to unexpected delivery of the 2nd drop of reward at one well (Fig. 3a) and decreasing to omission of the expected 2nd drop of reward at the other well (Fig. 3b). Changes in firing were not observed to the delivery or omission of the 3rd drop of reward. This is consistent with signaling of the underlying prediction error, since in this experiment (unlike our prior studies) the 3rd drop was always delivered after the 2nd drop, so its

occurrence or omission was fully predicted by the presence or absence of the 2nd drop. To quantify these changes, we computed difference scores for each neuron, comparing the average firing at the beginning versus the end of the number shift blocks at the time of delivery or omission of each potential drop of reward. Distributions of these scores were shifted above zero when an unexpected 2nd reward was delivered (Fig. 3c, middle) and below zero when an expected 2nd reward was omitted (Fig. 3d, middle), but remained unchanged in the time windows related to the 1st (Figs, 3c-d, left) and 3rd drops of reward (Figs. 3c-d, right). Changes in firing to 2nd reward were maximal at the beginning of the block and then diminished with learning (Fig. 3e, ANOVA's F's > 2.00, p's < 0.05), and the changes in firing to unexpected delivery and omission of the 2nd drop were inversely correlated in individual neurons (Fig. 3f). There were no changes in firing across trials at the time of delivery or omission of the other drops of reward (Fig. 3e, ANOVA's, F's < 1.80, p's > 0.05).

Dopamine neurons respond to errors in the prediction of reward features

The above findings are consistent with current proposals that dopamine neurons signal a bidirectional prediction error reflecting changes in value (Fig. 5a). Yet clearly many other features that are independent of value change when a new reward is delivered or an expected reward is omitted. To test whether dopamine neurons might also reflect errors in the prediction of these other features, we analyzed activity in the reward-responsive dopamine neurons in response to shifts in reward identity in blocks 3 and 5 (Fig. 2b). We quantified changes in firing to the identity shifts by again computing difference scores for each neuron, in this case comparing firing in response to the new reward at the start of the new block versus the old reward at the end of the prior block. This analysis revealed marked changes in the population responses, which increased to the new rewards (Figs. 4a-b), and in the distributions of scores comparing firing before and after the identity shift, which were shifted significantly above zero for both the 1st and 2nd drops of the big reward (Fig. 4c, left and middle) and for the single drop of the small reward (Fig. 4d, left). These increases were generally maximal immediately after the identity shift, and diminished rapidly on subsequent trials (Fig. 4e, black line, ANOVA, F = 2.40, p = 0.01), as would be expected if the firing changes reflected errors in prediction of the new rewards. Notably there was no change in firing to the 3rd drop of the big reward (Fig. 4c, right), which again was fully predicted in both is value and identity by the preceding 2nd drop.

What are the dopamine neurons signaling?

What do these changes mean? Are dopamine neurons signaling something other than cached-value errors or is there another explanation for the apparent error signals in the identity blocks in the current design? One alternative explanation is that the signals reflect salience or perhaps a value "bonus" due to the novelty of the new reward (Bromberg-Martin et al., 2010a; Horvitz, 2000; Kakade and Dayan, 2002; Matsumoto and Hikosaka, 2009; Schultz, 2016). If the new flavor is novel or highly salient compared to the old reward, then each drop would induce a cached-value error or salience signal (Fig. 5b). Yet on its face, this does not seem likely. The rats have had extensive experience with the two rewards in each well, the blocks are short and interspersed daily, and the rewards are available on interleaved trials within each block. Thus, it seems unlikely that such large differences in firing (as large

or larger as seen to entire additional rewards) would be driven by such trivial differences in moment-to-moment exposure. In addition, at least two pieces of evidence contradict this account. First, while we did observe increased firing to the small reward and to the 1st and 2nd drops of the big reward, we did not observe increased firing to the 3rd drop of the big reward. This drop should be equally imbued with novelty or salience due to flavor, so the data do not completely match the predictions of this account. Second, the changes in firing in the identity shift blocks were correlated with the reward-prediction error signals in the number shift blocks (Fig. 4f). This relationship – especially the inverse correlation with the negative error signals (Fig. 4f, red) - is not consistent with the proposal that these neurons are signaling salience in response to changes in reward identity, since reward omission is also a salient event.

A second and more interesting alternative explanation for the apparent error signals in the identity blocks is that the rats might be using the identity of the rewards themselves as predictors of the number of drops that will be delivered. While this is not possible when different amounts of the same reward are used at both wells, when two different rewards are used, then the identity of the initial drop can be used to predict whether subsequent drops are likely to follow (n.b. this mechanism does not apply if the two wells deliver similar amounts of different rewards, and as illustrated in Fig. S2, dopamine neurons also exhibited correlated changes in firing in response to number and identity shifts under these conditions). If the rats attend to this information, then by the end of a trial block, the flavor of the initial drop of the big reward should have a higher value than the flavor of the initial drop of the small reward, because the former predicts 2 more rewarding drops, while the latter signals the absence of any further drops of reward. Any difference in value carried by the initial drop of the two rewards would induce cached value errors (Fig. 5c), which would confound the identity errors that were the *a priori* target of the current design. This mechanism would explain the increased firing to the small reward after an identity shift as unexpected increase in its value (Fig. 5c). The lack of suppression of firing thereafter to the missing subsequent drops could reflect the relative weakness of the underlying prediction and the difficulty measuring suppression of firing in neurons with such low firing rates.

However, by the same logic, the 1st drop of the big reward should initially be less valuable, because its flavor predicted no further reward in the prior block (Fig. 5c, red box). Thus, a negative value-prediction error should occur in response to the 1st drop, followed by positive errors to the subsequent, unexpected drops. The firing of the dopamine neurons complied with these predictions for the 2nd drop but not for the 1st drop. Instead, rather than showing suppression, as these same neurons did when reward was omitted in the number shift blocks, they did the opposite in identity shift blocks, showing a dramatic increase in firing to the critical 1st drop of the new big reward (Figs. 4a and c). Indeed, the increase in firing to the first drop of the small reward was positively correlated with the increase in firing to the first drop of the big reward for individual neurons in individual blocks (r = 0.35, p < 0.01).

While increased firing to the 1st drop of the big reward is the opposite of what is predicted for a cached-value error computed using outcome flavor as a predictor of subsequent reward, it is exactly what is expected for a prediction error based on changes in identity or sensory information. In fact, the 1st drop of the large reward turns out to be the only place in the

design where the predictions for a cached-value error unambiguously diverge from those for a sensory or identity prediction error (Fig. 5d, red box). Elsewhere, the predictions of the two accounts are arguably similar if not identical. For example, in the identity shift blocks, both a sensory-based and a value-based account predict increased firing when the small reward is presented and when the 2nd drop of the big reward is delivered, because in both cases there is novel sensory information and excess, unpredicted value. Both also predict no change in firing to the 3rd drop, as its flavor and value are as expected based on the preceding drops. The predictions of the two accounts are also aligned in the number shift blocks; each predicts increased firing to the additional 2nd drop of reward in these blocks, since again there is excess value but also substantial unpredicted sensory information. Further, both predict decreased firing upon omission of the 2nd drop of reward since there is less value than expected but also a loss of expected sensory input that is not offset by the addition of a new reward.

Discussion

Midbrain dopamine neurons have been proposed to signal prediction errors as defined in model-free reinforcement learning algorithms (Schultz et al., 1997; Sutton, 1988). While these algorithms have been extremely powerful in interpreting dopaminergic activity (Bayer and Glimcher, 2005; Hart et al., 2014; Hollerman and Schultz, 1998; Lak et al., 2014; Mirenowicz and Schultz, 1994; Pan et al., 2005; Tobler et al., 2003; Waelti et al., 2001), the errors are exclusively value-based. Here we have shown that dopamine neurons that exhibit classic error-like responses to changes in reward value show similar signals to changes in its sensory features. The changes in firing to value and identity did not occur in unique populations of neurons, nor were the identity-based changes in firing easily explicable as reflecting novelty or salience. Further, in the one area of the task design in which the identity and value accounts cannot be reconciled, the 1st drop of the large reward (Fig. 5c-d, red boxes), the firing of the dopamine neurons did the opposite of what a cached-value hypothesis would predict. Thus, at a minimum, these results contradict the proposal that dopamine errors only reflect cached value errors, even if one amends this idea to include modulation by salience or novelty. In fact, value alone, whether cached or model-based, cannot explain the increased firing to the 1st drop of the big reward, given the rats' behavior on these trials. The most parsimonious account is that the neurons are responding to the unexpected sensory features of the reward, at least in some trial blocks.

Before considering the possible implications of these findings, there are several important caveats to point out. First, our analysis has focused on classic wide waveform dopamine neurons. While these are the dopamine neurons thought to signal value errors in primates (Bromberg-Martin et al., 2010b; Fiorillo et al., 2008; Hollerman and Schultz, 1998; Kobayashi and Schultz, 2008; Matsumoto and Hikosaka, 2009; Mirenowicz and Schultz, 1994; Morris et al., 2006; Waelti et al., 2001), evidence from a variety of sources suggest there may be TH+ neurons in the other populations (Margolis et al., 2006; Ungless and Grace, 2012). These neurons do not signal errors in the current task in appreciable numbers (Roesch et al., 2007; Takahashi et al., 2016; Takahashi et al., 2011), but it is possible they do so in other settings in ways we cannot address here (Cohen et al., 2012). Second recordings here were made in rats and in VTA, whereas much of the data showing that dopamine

signaling complies with predictions for a cached-value error comes from recordings made in primate substantia nigra (Glimcher, 2011; Schultz, 2002; Schultz, 2016). It is possible that the correlates isolated here reflect a specialized function of rat midbrain generally or of VTA in particular. Finally, we have considered single unit firing patterns of VTA neurons without knowledge of their inputs or downstream targets. While we did not observe unique populations of value- and identity-error responsive dopamine neurons, it is possible that if the dopamine neurons were segregated by their inputs or outputs, subpopulations could be resolved. In this regard, it is worth noting that identity errors have not been observed in studies using fast scan cyclic voltammetry to monitor dopamine release in nucleus accumbens (Collins et al., 2016; Papageorgiou et al., 2016).

Caveats aside, how can these data be reconciled with existing evidence that dopamine neurons respond to changes in reward value? One possibility is that dopamine neurons are signaling multiple forms of information, performing different functions in different situations and at different time scales, perhaps for different downstream targets. The general idea that other signals coexist in the activity of dopamine neurons alongside cached value errors has already been suggested for functions as diverse as reward rate (Niv et al., 2007), incentive motivation (Saddoris et al., 2015), value (Hamid et al., 2016), salience (Bromberg-Martin et al., 2010a; Schultz, 2016), and information (Bromberg-Martin and Hikosaka, 2009). The current findings may simply to add another component –identity or sensory prediction errors - to this multiplexed signal.

However, another possibility, made explicit by the close correspondence between the predictions of the two models (Fig. 5c-d), is that value errors may be a special case of a more general function in which dopamine neurons signal errors in event prediction. One might think of this error signal as equivalent to displacement in physics, except in a high-dimensional feature space in which expected and actual events are positions represented by neural activity. Value would be only one dimension (or combination of dimensions) along which actual events might differ from expectations. The distance between the actual and expected positions in feature space would be represented by the magnitude of the phasic dopamine signal, with its sign perhaps signaling the presence or absence of information. This formulation is similar in some regards to the aforementioned proposal that dopamine transients reflect information (Bromberg-Martin and Hikosaka, 2009). However rather than reflecting information's value (or salience), these responses may be evidence of a richer function in which the neurons are signaling the displacement in a high-dimensional neural space that the unexpected information represents – whether it be information about an event's value or its number, probability, timing, temperature, texture, location, color, flavor, or other attribute. This would provide a parsimonious explanation for the current data as well as most prior reports, including those showing that dopamine neurons respond to completely neutral cues in some situations (Horvitz, 2000; Schultz, 2016).

But if dopamine neurons signal errors in the prediction of events that are not tied to value, why has this not been evident previously? One reason is that few attempts have been made to manipulate features to isolate errors in their prediction that are orthogonal to value. Instead, as a rule, most studies manipulate value-relevant features, like size, probability, and timing, to demonstrate value errors, while ignoring the confounding changes in the

underlying sensory features. By contrast, we held the variable proposed to be encoded by dopaminergic errors (e.g. value) constant while overtly manipulating other information not thought to be registered by this signal (e.g. sensory or other predicted features).

Rarely have single unit recording studies manipulated the value of different rewards. In the two cases of which we are aware, there were no effects of identity on dopaminergic errors (Lak et al., 2014; Stauffer et al., 2014). In one exemplar (Lak et al., 2014), monkeys were presented with cues that predicted fixed or variable amounts of a single or two different juice rewards. The monkeys exhibited stable, ordered preferences among the predictive cues when allowed to make choices between them, and the activity of dopamine neurons at the time of cue presentation and reward delivery on forced choice trials varied linearly with errors in the prediction of utility or economic value. There was no difference in the dopamine response based which juice was predicted or delivered, unless the monkey preferred one of the two juices.

While this sounds like strong evidence against identity or sensory prediction error signaling, the identity predictions were never systematically violated to reveal sensory errors. For example, firing at the time of reward was examined when a larger or a smaller than expected amount of two different juices was delivered, and activity reflected the value error, independent of juice type (Lak et al., 2014, see Fig. 5a). While this may show that the phasic response does not encode juice type, it does not rule out encoding of *errors* in the prediction of juice type, since the juice type expectation was not violated. Similarly when the predicted juice type was uncertain, the juice-evoked response of the dopamine neurons did not differ according to which juice was delivered (Lak et al., 2014, see Fig. 5b). Here again, the data do not rule out a juice type prediction error, since the uncertainty in the expected sensory features of the juice did not differ depending on which juice was delivered. Additionally, cues predictive of multiple outcomes may actually represent each outcome quite strongly (Esber and Haselgrove, 2011), so sensory errors may be difficult to demonstrate using this approach.

These studies also failed to observe differences in dopaminergic activity to the cues predicting different rewards, unless the predicted values were different. While these data show that dopamine transients do not encode the actual specifics of the prediction (i.e. the prediction that chocolate will be delivered), they do not rule out encoding of errors in the prediction, since any sensory prediction error (i.e. the difference between the prediction before versus after the cue appears) would be the same for each cue versus an average baseline prediction. Notably this is also true in the current study. As expected, the dopamine neurons fired to the cues on the forced choice trials, and this response was stronger for the high value cue than to the low value cue, whereas there was no difference in firing to the cues that predicted the same amount of different rewards, chocolate versus vanilla (see Fig. S3).

The similar firing to these two cues is what one would expect, since the size of this error for a chocolate-predictive cue should not be significantly different than its size for a vanilla-predictive cue.

Lastly failure to observe sensory prediction errors in these previous studies may simply reflect relative inattention to reward identity in the face of large value differences. Value is overwhelming. This is evident in identity unblocking, which is weaker when rats are trained with different amounts of the two rewards than when they receive only a single amount (McDannald et al., 2011). Indeed, in the current study, the rats sometimes failed to show even small, transient changes in their established behavior (choice, % correct, reaction time, or licking) when we switched the chocolate and vanilla milk rewards. Interestingly, when the rats failed to acknowledge the change in reward identity, dopamine neurons did not show changes in firing to the identity changes (see Fig. S4).

In conclusion, the current findings show that reward-prediction error signaling dopamine neurons respond to errors in the prediction of sensory features of expected rewards. These responses cannot be easily explained as reflecting value or salience or any other construct that is currently applied to explain phasic changes in the firing of dopamine neurons. One possible interpretation of these findings is that dopamine neurons serve a more general function of signaling errors in event prediction in which value is just one dimension. This reconceptualization would be an important extension of our current understanding of these important signals because it would open the potential for them to support a much wider array of learning than is possible within the current model-free reinforcement learning algorithms applied to interpret them.

STAR Methods Text

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, GS (geoffrey.schoenbaum@nih.gov).

Experimental Model and Subject Details—Eight male, experimentally-naïve, Long-Evans rats (Charles River Labs, Wilmington, MA) and five male transgenic rats that carried a TH-dependent Cre expression system in a Long-Evans background (NIDA animal breeding facility) were used in this study. Rats were of normal immune function and aged approximately 3 months at the start of the study (175-200 g). Rats were singly housed in a facility accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC), and maintained on a 12-h light–dark cycle. All experimental and animal care procedures complied with US National Institutes of Health guidelines and were approved by that National Institutes on Drug Abuse Intramural Research Program Animal Care and Use Committee.

Method Details

<u>Surgical procedures:</u> All surgical procedures adhered to guidelines for aseptic technique. For electrode implantation, a drivable bundle of eight 25-um diameter FeNiCr wires (Stablohm 675, California Fine Wire, Grover Beach, CA) chronically implanted dorsal to VTA in the left or right hemisphere at 5.2 mm posterior to bregma, 0.7 mm laterally, and 7.5 mm ventral to the brain surface at an angle of 5° toward the midline from vertical. Wires were cut with surgical scissors to extend ~ 2.0 mm beyond the cannula and electroplated with platinum (H_2PtCl_6, Aldrich, Milwaukee, WI) to an impedance of 300~800 kOhms.

Five TH-Cre transgenic rats received unilateral infusion of AAV1-Flex-taCasp3-TEVp into the VTA at 5.5 mm posterior, 0.2 mm lateral and 8.35 mm ventral from bregma as the same hemisphere in which an electrode was to be implanted. Cephalexin (15 mg/kg p.o.) was administered twice daily for two weeks post-operatively

Histology: All rats were perfused with phosphate-buffered saline (PBS) followed by 4% paraformaldehyde (Santa Cruz Biotechnology Inc., CA). Brains that received only electrode implantation were cut in 40 μm sections and stained with thionin. Brains from the 5 Casp3 rats that had received viral infusions were processed for TH immunolabeling as described previously. Coronal free-floating sections (14 μm in thickness) were rinsed with 0.1 M phosphate buffer (PB, pH 7.4) and incubated for 1 h in PB supplemented with 4% bovine serum albumin (BSA) and 0.3% Triton X-100. This was followed by the overnight incubation at 4°C with an anti-TH mouse monoclonal antibody (1:500). After rinsing 3×10 min in PB, sections were processed with an ABC kit (Vector Laboratories, Burlingame, CA). The material was incubated for 1 h at room temperature in a 1:200 dilution of the biotinylated secondary antibody, rinsed with PB, and incubated with avidin-biotinylated horseradish peroxidase for 1 h. Sections were rinsed and the peroxidase reaction was then developed with 0.05% DAB and 0.003% H_2O_2. Free-floating sections were mounted on coated slides.

Odor-guided choice task: Recording was conducted in aluminum chambers approximately $18''$ on each side with sloping walls narrowing to an area of $12'' \times 12''$ at the bottom. A central odor port was located above two fluid wells (Fig. 2a). Two lights were located above the panel. The odor port was connected to an air flow dilution olfactometer to allow the rapid delivery of olfactory cues. Odors where chosen from compounds obtained from International Flavors and Fragrances (New York, NY). Trials were signaled by illumination of the panel lights inside the box. When these lights were on, nosepoke into the odor port resulted in delivery of the odor cue to a small hemicylinder located behind this opening. One of three different odors was delivered to the port on each trial, in a pseudorandom order. At odor offset, the rat had 3 seconds to make a response at one of the two fluid wells. One odor instructed the rat to go to the left to get reward, a second odor instructed the rat to go to the right to get reward, and a third odor indicated that the rat could obtain reward at either well. Odors were presented in a pseudorandom sequence such that the free-choice odor was presented on 7/20 trials and the left/right odors were presented in equal numbers. In addition, the same odor could be presented on no more than 3 consecutive trials. Once the rats were shaped to perform this basic task, we introduced blocks in which we independently manipulated the size of the reward or the identity of the reward (Fig. 2b). For recording, one well was randomly designated as big reward condition in which three boli of chocolate or vanilla milk were delivered, and the other as small reward condition in which one bolus of the other flavored milk was delivered at the start of the session (Block 1, Fig. 2b). The bolus size was ~0.05 ml and between boli was 500 ms apart. In the second block of trials the reward value was switched by changing the number of reward without changing flavors. (Block 2, Fig. 2b). In the third block of trials, the reward identity was switched by changing the flavor of the reward without changing the number of reward (Block 3, Fig. 2b). The value and identity changes were repeated one more time in Block 4 and Block 5,

respectively (Fig. 2b). The block 1 was 30-50 trial long and all subsequent blocks were 60-100 trials long. Block switches occurred when rats chose high value side more than 60% in last 10 free-choice trials.

Flavor preference testing: After all recording sessions, 2 sessions of consumption tests were given in the rats with experimental group. We compared consumption of two flavored milks in a housing cage separate from their home cage and experimental chamber, using procedures similar to those normally used to give water to the rats after training. In these tests, the rats were allowed to adapt the testing cage for 3-5 minutes. Then, bottles containing chocolate or vanilla flavored milk solution identical to that used in the recording sessions were placed onto the wire lid so that the spouts were accessible to the rat for 2 minutes. The initial location of the bottles (left or right) was randomized across rats and switched between the consumption tests. In addition, the location of the bottles was swapped roughly every 20s during each test to equate time on each side. The total amount of consumption was measured and averaged across the two tests.

Single-unit recording: Wires were screened for activity daily; if no activity was detected, the rat was removed, 4 and the electrode assembly was advanced 40 or 80 μm. Otherwise active wires were selected to be recorded, a session was conducted, and the electrode was advanced at the end of the session. Neural activity was recorded using Plexon Multichannel Acquisition Processor systems (Dallas, TX). Signals from the electrode wires were amplified 20× by an op-amp headstage (Plexon Inc, HST/8o50-G20-GR), located on the electrode array. Immediately outside the training chamber, the signals were passed through a differential pre-amplifier (Plexon Inc, PBX2/16sp-r-G50/16fp-G50), where the single unit signals were amplified 50× and filtered at 150-9000 Hz. The single unit signals were then sent to the Multichannel Acquisition Processor box, where they were further filtered at 250-8000 Hz, digitized at 40 kHz and amplified at 1-32×. Waveforms (>2.5:1 signal-to-noise) were extracted from active channels and recorded to disk by an associated workstation

Quantification and Statistical Analysis—All data were analyzed using Matlab. Instances of multiple comparisons were corrected for with the Benjamini-Hochberg procedure. Error bars in figures denote the standard error of the mean. The number of subjects were chosen based on previous similar single-unit recording studies in rats.

Data analysis: Units were sorted using Offline Sorter software from Plexon Inc (Dallas, TX). Sorted files were then processed and analyzed in Neuroexplorer and Matlab (Natick, MA). Dopamine neurons were identified via a waveform analysis. Briefly cluster analysis was performed based on the half time of the spike duration and the ratio comparing the amplitude of the first positive and negative waveform segments. The center and variance of each cluster was computed without data from the neuron of interest, and then that neuron was assigned to a cluster if it was within 3 s.d. of the cluster's center. Neurons that met this criterion for more than one cluster were not classified. This process was repeated for each neuron. The putative dopamine neurons that showed increase in firing to reward compared to baseline (400ms before reward) were further classified as reward-responsive (t-test, $p < 0.05$).

Prior to analyzing single unit activity for error correlates, we also eliminated blocks in which there was no effect of the changes in reward on behavior on the initial trials. For this, we examined % correct, reaction time, and licking behavior on forced choice trials, comparing values in each well on the 10 trials immediately before and after the change in reward. We also calculated the probability of selecting the large reward on free choice trials in these trial blocks. If none of these comparisons yielded a significant effect across the shift ($p < 0.05$), the block was excluded. This analysis eliminated 10 out of 204 number shift blocks and 100 out of 204 blocks identity shift blocks.

To analyze neural activity to reward, we examined firing rate in the 400 ms beginning 100 ms after reward delivery. Reward activity was normalized by subtracting average baseline firing (400 ms before light on). Difference scores for each neuron in the number shift blocks in Fig. 3 were computed by subtracting average firing during 100 – 500 ms after reward or omission on the last 5 trials from that on the first 5 trials. Difference scores for each neuron in the identity shift blocks in Fig. 4 were computed by subtracting average firing during 100 – 500 ms on the last 5 trials before identity shift from that on the first 5 trials after identity shift.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Bayer HM, Glimcher P. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron. 2005; 47:129–141. [PubMed: 15996553]

Brogden WJ. Sensory pre-conditioning. Journal of Experimental Psychology. 1939; 25:323–332.

Bromberg-Martin ES, Hikosaka O. Midbrain dopamine neurons signal preference for advance information about upcoming rewards. Neuron. 2009; 63:119–126. [PubMed: 19607797]

Bromberg-Martin ES, Matsumoto M, Hikosaka O. Dopamine in motivational control: rewarding, aversive and alerting. Neuron. 2010a; 68:815–834. [PubMed: 21144997]

Bromberg-Martin ES, Matsumoto M, Hong S, Hikosaka O. A pallidus-habenula-dopamine pathway signals inferred stimulus values. Journal of Neurophysiology. 2010b; 104:1068–1076. [PubMed: 20538770]

Burke KA, Franz TM, Miller DN, Schoenbaum G. The role of the orbitofrontal cortex in the pursuit of happiness and more specific rewards. Nature. 2008; 454:340–344. [PubMed: 18563088]

Chang CY, Esber GR, Marrero-Garcia Y, Yau HJ, Bonci A, Schoenbaum G. Brief optogenetic inhibition of VTA dopamine neurons mimics the effects of endogenous negative prediction errors during Pavlovian over-expectation. Nature Neuroscience. 2016; 19:111–116. [PubMed: 26642092]

Clark JJ, Hollon NG, Phillips PEM. Pavlovian valuation systems in learning and decision making. Current Opinion in Neurobiology. 2012; 22:1054–1061. [PubMed: 22749132]

Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature. 2012; 482:85–88. [PubMed: 22258508]

Collins AL, Greenfield VY, Bye JK, Linker KE, Wang AS, Wassum KM. Dynamic mesolimbic dopamine signaling during action sequence learning and expectation violation. Scientific Reports. 2016; 6:20231. [PubMed: 26869075]

Esber GR, Haselgrove M. Reconciling the influence of predictiveness and uncertainty on stimulus salience: a model of attention in associative learning. Proceedings of the Royal Society B. 2011; 278:2553–2561. [PubMed: 21653585]

Fiorillo CD, Newsome WT, Schultz W. The temporal precision of reward prediction in dopamine neurons. Nature Neuroscience. 2008; 11:966–973. [PubMed: 18660807]

Glimcher PW. Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. Proceedings of the National Academy of Science. 2011; 108:15647–15654.

Hamid AA, Pettibone JR, Mabrouk OS, Hetrick VL, Schmidt R, VAnder Weele CM, Kennedy RT, Aragona BJ, Berke JD. Mesolimbic dopamine signals the value of work. Nature Neuroscience. 2016; 19:117–126. [PubMed: 26595651]

Hart AS, Rutledge RB, Glimcher PW, Phillips PE. Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. Journal of Neuroscience. 2014; 34:698–704. [PubMed: 24431428]

Hollerman JR, Schultz W. Dopamine neurons report an error in the temporal prediction of reward during learning. Nature Neuroscience. 1998; 1:304–309. [PubMed: 10195164]

Horvitz JC. Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. Neuroscience. 2000; 96:651–656. [PubMed: 10727783]

Jo YS, Lee J, Mizumori SJ. Effects of prefrontal cortical inactivation on neural activity in the ventral tegmental area. Journal of Neuroscience. 2013; 33:8159–8171. [PubMed: 23658156]

Jones JL, Esber GR, McDannald MA, Gruber AJ, Hernandez G, Mirenzi A, Schoenbaum G. Orbitofrontal cortex supports behavior and learning using inferred but not cached values. Science. 2012; 338:953–956. [PubMed: 23162000]

Kakade S, Dayan P. Dopamine: generalization and bonuses. Neural Networks. 2002; 15:549–559. [PubMed: 12371511]

Kobayashi K, Schultz W. Influence of reward delays on responses of dopamine neurons. Journal of Neuroscience. 2008; 28:7837–7846. [PubMed: 18667616]

Lak A, Stauffer WR, Schultz W. Dopamine prediction error responses integrate subjective value from different reward dimensions. Proceedings of the National Academy of Science. 2014; 111:2342–2348.

Margolis EB, Lock H, Hjelmstad GO, Fields HL. The ventral tegmental area revisited: Is there an electrophysiological marker for dopaminergic neurons? Journal of Physiology. 2006; 577:907–924. [PubMed: 16959856]

Matsumoto M, Hikosaka O. Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature. 2009; 459:837–841. [PubMed: 19448610]

McDannald MA, Lucantonio F, Burke KA, Niv Y, Schoenbaum G. Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. Journal of Neuroscience. 2011; 31:2700–2705. [PubMed: 21325538]

Mirenowicz J, Schultz W. Importance of unpredictability for reward responses in primate dopamine neurons. Journal of Neurophysiology. 1994; 72:1024–1027. [PubMed: 7983508]

Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H. Midbrain dopamine neurons encode decisions for future action. Nature Neuroscience. 2006; 9:1057–1063. [PubMed: 16862149]

Niv Y, Daw ND, Joel D, Dayan P. Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology. 2007; 191:507–520. [PubMed: 17031711]

Niv Y, Schoenbaum G. Dialogues on prediction errors. Trends in Cognitive Sciences. 2008; 12:265–272. [PubMed: 18567531]

Pan WX, Schmidt R, Wickens JR, Hyland BI. Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. Journal of Neuroscience. 2005; 25:6235–6242. [PubMed: 15987953]

Papageorgiou GK, Baudonnat M, Cucca F, Walton ME. Mesolimbic dopamine encodes prediction errors in a state-dependent manner. Cell Reports. 2016; 15:221–228. [PubMed: 27050518]

Rescorla RA. Learning about qualitatively different outcomes during a blocking procedure. Animal Learning and Behavior. 1999; 27:140–151.

Roesch MR, Calu DJ, Schoenbaum G. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nature Neuroscience. 2007; 10:1615–1624. [PubMed: 18026098]

Saddoris MP, Cacciapaglia F, Wightman RM, Carelli RM. Differential dopamine release dynamics in the nucleus accumbens core and shell reveal complimentary signals for error prediction and incentive motivation. Journal of Neuroscience. 2015; 35:11572–11582. [PubMed: 26290234]

Schultz W. Getting Formal with Dopamine and Reward. Neuron. 2002; 36:241–263. [PubMed: 12383780]

Schultz W. Dopamine reward prediction-error signalling: a two-component response. Nature Reviews Neuroscience. 2016; 17:183–195. [PubMed: 26865020]

Schultz W, Dayan P, Montague PR. A neural substrate for prediction and reward. Science. 1997; 275:1593–1599. [PubMed: 9054347]

Sharpe MJ, Chang CY, Liu MA, Batchelor HM, Mueller LE, Jones JL, Niv Y, Schoenbaum G. Dopamine transients are sufficient and necessary for acquisition of model-based associations. Nature Neuroscience. 2017; 20:735–742. [PubMed: 28368385]

Stauffer WR, Lak A, Schultz W. Dopamine reward prediction error responses reflect marginal utility. Current Biology. 2014; 24:2491–2500. [PubMed: 25283778]

Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH. A causal link between prediction errors, dopamine neurons and learning. Nature Neuroscience. 2013; 16:966–973. [PubMed: 23708143]

Sutton RS. Learning to predict by the method of temporal difference. Machine Learning. 1988; 3:9–44.

Takahashi YK, Langdon AJ, Niv Y, Schoenbaum G. Temporal specificity of reward prediction errors signaled by putative dopamine neurons in rat VTA depends on ventral striatum. Neuron. 2016; 91:182–193. [PubMed: 27292535]

Takahashi YK, Roesch MR, Wilson RC, Toreson K, O'Donnell P, Niv Y, Schoenbaum G. Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. Nature Neuroscience. 2011; 14:1590–1597. [PubMed: 22037501]

Tobler PN, Dickinson A, Schultz W. Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. Journal of Neuroscience. 2003; 23:10402–10410. [PubMed: 14614099]

Ungless MA, Grace AA. Are you or aren't you? Challenges associated with physiologically identifying dopamine neurons. Trends in Neurosciences. 2012; 35:422–430. [PubMed: 22459161]

Waelti P, Dickinson A, Schultz W. Dopamine responses comply with basic assumptions of formal learning theory. Nature. 2001; 412:43–48. [PubMed: 11452299]

Xin J, Costa RM. Start/stop signals emerge in nigrostriatal circuits during sequence learning. Nature. 2010; 466:457–462. [PubMed: 20651684]
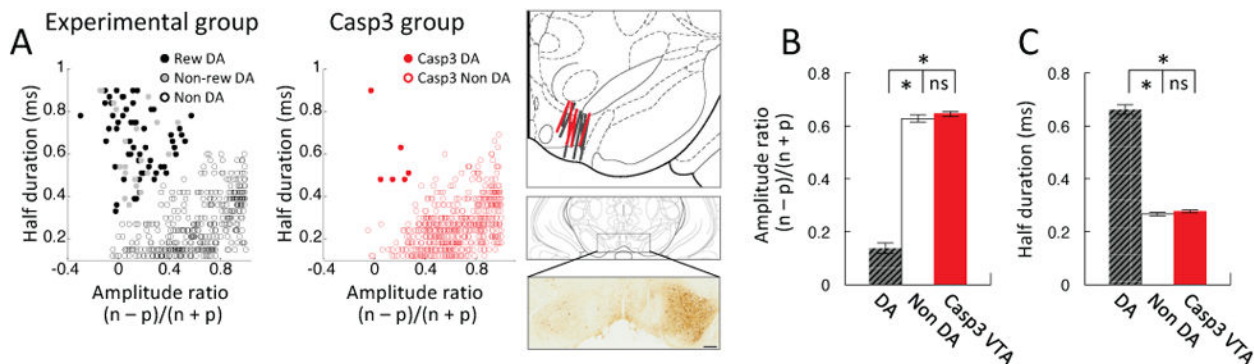
Figure 1. Identification, waveform features of putative dopamine neurons

(A) Result of cluster analysis based on the half time of the spike duration and the ratio comparing the amplitude of the first positive and negative waveform segments $((n – p) / (n + p))$. Data on the left show VTA neurons (n = 473) from the experimental group, plotted as reward-responsive (filled black circles) and nonresponsive dopamine neurons (filed gray circles), and neurons that classified with other clusters, no clusters or more than on cluster (open circles). Data on the right show VTA neurons (n = 556) from rats that received AAV1-Flex-taCasp3-TEVp infusions, plotted as dopamine (filled red circles) or non dopamine neurons (open red circles) based on whether they were assigned to the dopamine cluster from the experimental data. Drawings on right show electrode tracks in experimental (n = 8, gray) and Casp3 groups (n = 5, red) and a representative coronal brain slices showing unilateral loss of TH-positive neurons in VTA in a Casp3-infused rat (left hemisphere, avg loss vs intact side 76.5%, range 65-85% over all 5 rats). **(B)** Bar graph indicating average amplitude ratio of putative dopamine neurons, non dopamine neurons recorded from experimental rats, and VTA neurons recorded from Casp3-infused rats. **(C)** Bar graph indicating average half duration of putative dopamine, non dopamine neurons recorded from experimental rats, and VTA neurons recorded in Casp3-infused rats. Error bars, S.E.M. The average amplitude ratio and spike duration of the neurons recorded in the Casp3 rats did not differ from neurons in non-dopamine clusters recorded in the experimental rats (amplitude ratio, t-test, $t_{946} = 0.99$, p = 0.32; spike duration, t-test, $t_{946} = 1.40$, p = 0.16). Indeed only 6 of the neurons recorded in these rats (filled red circles in A) classified as the putative dopamine neurons (1% vs 17%, $X^2 = 87.7$, p < 0.0000001). This amounted to an average of 0.06 putative dopamine neurons per session in the Casp3 group versus 0.79 neurons per session observed in the experimental rats (t-test, p < 0.01). The prevalence of the non-dopamine neurons actually increased from 3.84 neurons per session in the experimental rats to 5.44 neurons per session in the Casp3 group (t-test, p < 0.01), perhaps due to network effects of the loss of the TH+ neurons.
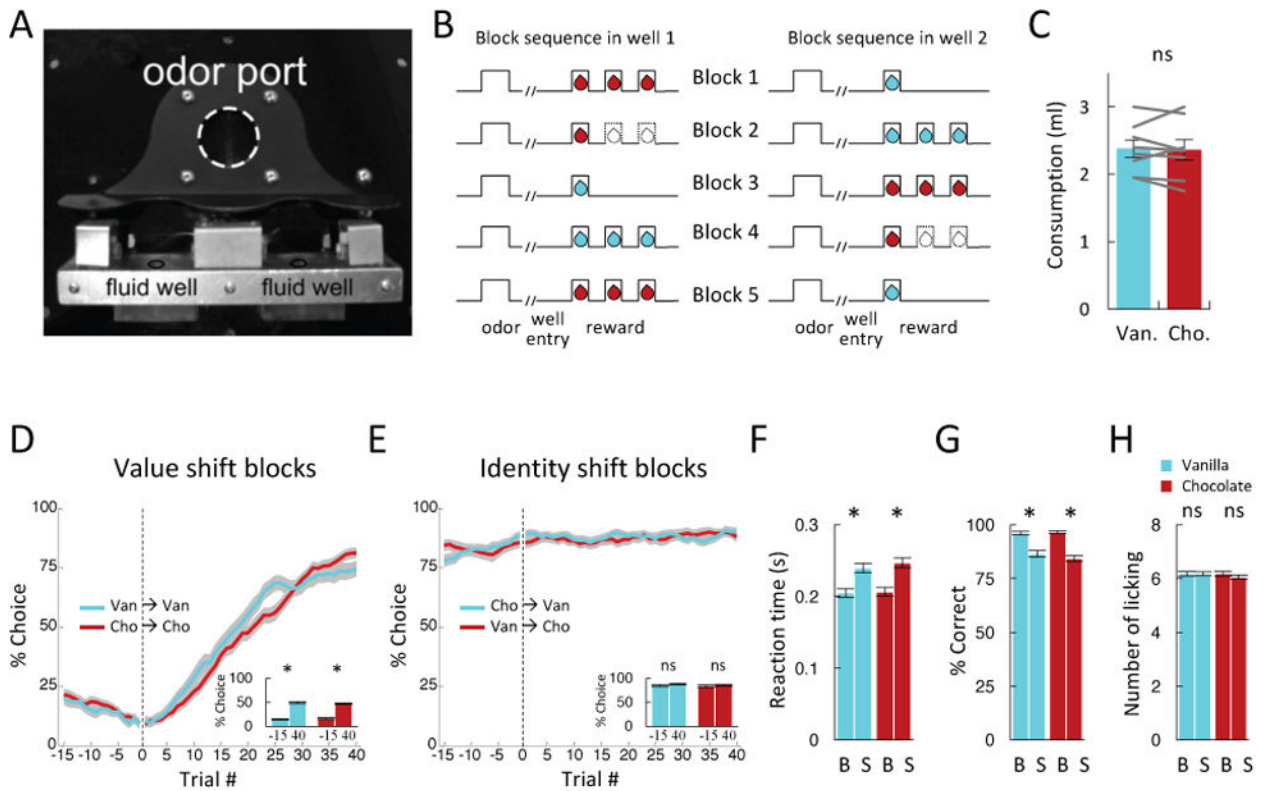
Figure 2. Task design and behavior

(**A**) Picture of apparatus used in the task, showing the odor port (~2.5 cm diameter) and two fluid wells. (**B**) Line deflections indicate the time course of stimuli (odors and rewards) presented to the animal on each trial. Dashed lines show when a reward was omitted, and solid lines show when reward was delivered. At the start of each recording session one well was randomly designated to deliver the big reward, which consisted of 3 drops of flavored milk (chocolate or vanilla). One drop of the other flavored milk was delivered in the other well (block 1). In the second and fourth blocks, number of drops delivered in the two wells were switched without changing the flavors (value shift). In the third and fifth blocks, the flavors delivered in the two wells were switched without changing the number of drops (identity shift). (**C**) Chocolate and vanilla flavored milk were equally preferred in 2-min consumption tests conducted at the end of some sessions. Gray lines indicate data from individual rats. (**D – E**) Choice rates in last 15 trials before and first 40 trials after a switch in reward number (D) or flavor (E). Y-axis indicates percent choice of side designated as big reward after block switch. Inset bar graphs show average choice rates in the last 15 before and first 40 trials after the switch. (**F**) Reaction times on the last 10 forced-choice trials in response to big and small amounts of each flavor. (**G**) Percentage correct on the last 10 forced-choice trials in response to big and small amounts of each flavor. (**H**) Number of licks in 500 ms after 1st drop of reward on the last 10 trials in response to big and small amounts of each flavor. B, big; S, small. Error bars, S.E.M.
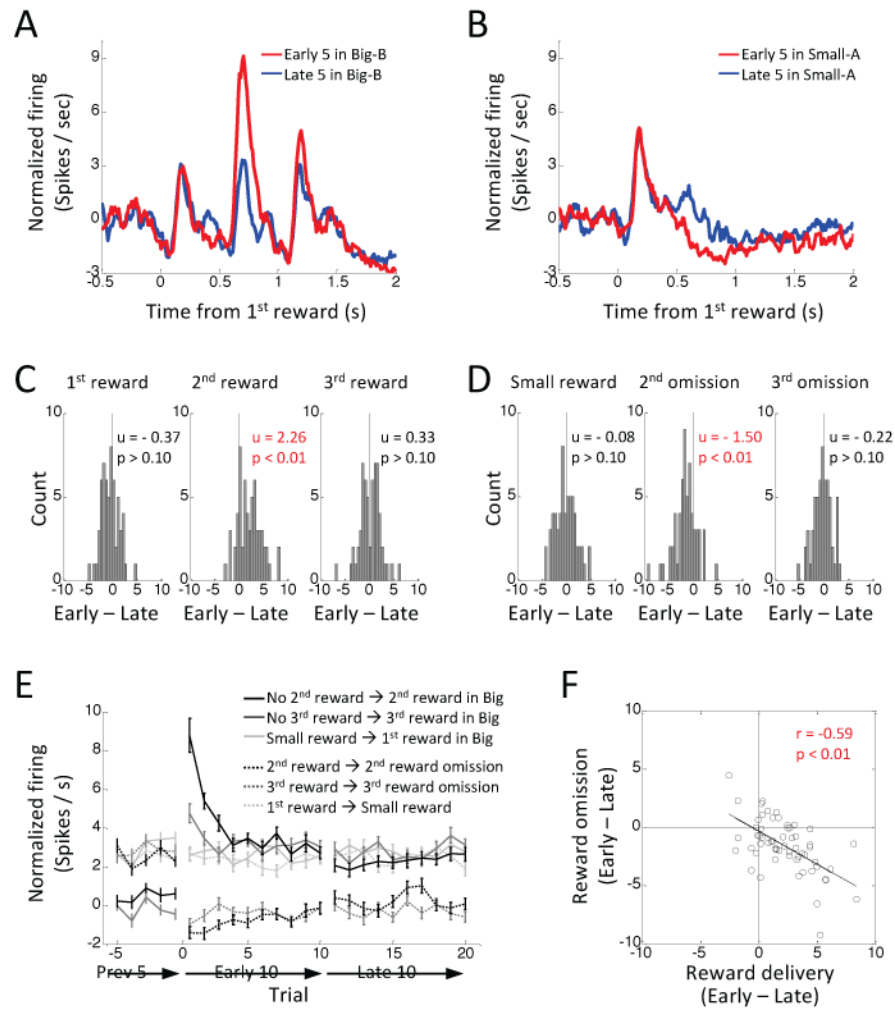
Figure 3. Changes in reward-evoked activity of reward-responsive dopamine neurons (n = 60) to changes in reward number

(A – B) Average firing on first 5 (red) and last 5 (blue) trials after a shift in reward number. (A) shows firing to the 3 drops of the big reward when the small reward had been expected, and (B) shows firing to the single drop of the small reward when the big reward had been expected. Big-B, 3 drops of reward B; small-A, one drop of reward A. **(C)** Distributions of difference scores comparing firing to 1st (left), 2nd (middle) and 3rd drops (right) of the big reward in the first 5 versus last 5 trials in a number shift block. **(D)** Distributions of difference scores comparing firing to the single drop of the small reward (left), and omissions of 2nd (middle) and 3rd drops (right) of the big reward in the first 5 versus last 5 trials in a number shift block. Difference scores were computed from the average firing rate of each neuron. The numbers in each panel indicate results of Wilcoxon signed-rank test (p) and the average difference score (u). **(E)** Changes in average firing before and after reward number shift. Light-gray, black and dark-gray solid lines indicate firing at the time of the 1st, 2nd, and 3rd drop of reward on big trials. Light-gray, black and dark-gray dashed-lines indicate firing at the time of the small reward, and omissions of the 2nd and 3rd drops thereafter on small trials. Error bars, S.E.M. **(F)** Correlation between differences scores representing changes in firing to delivery and omission of the 2nd drop of the big reward.
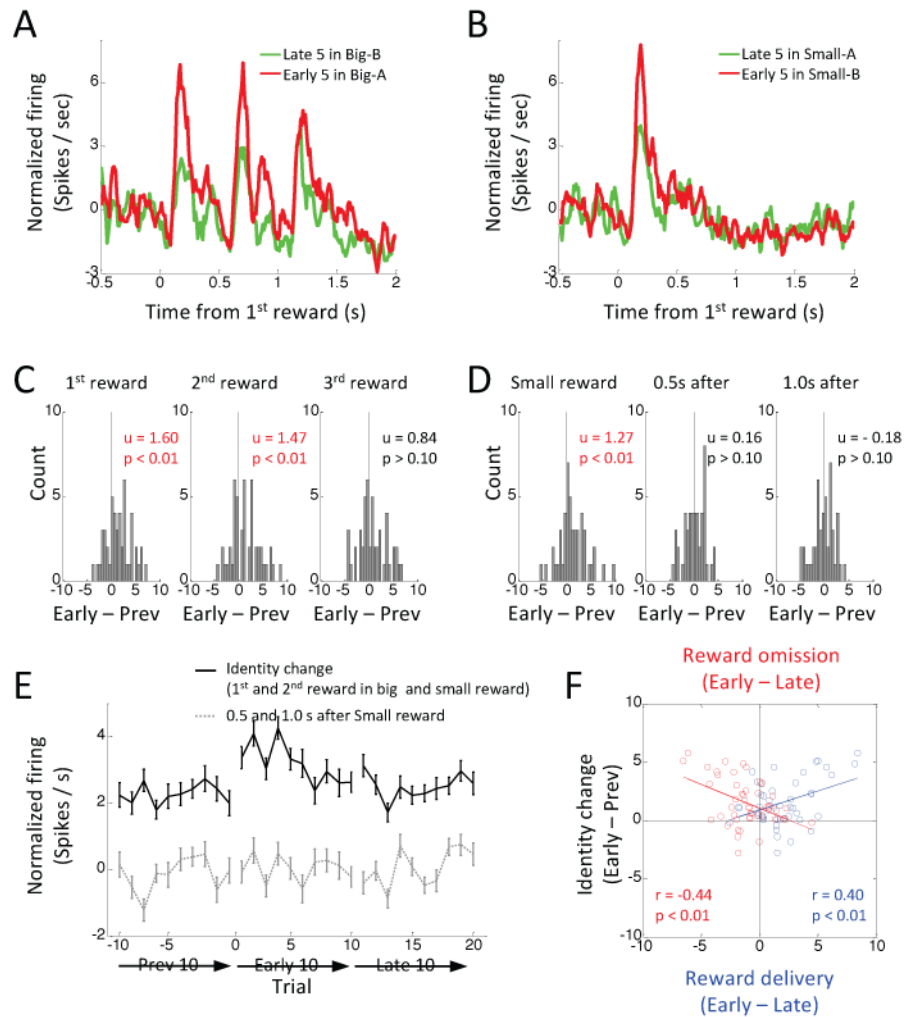
Figure 4. Changes in reward-evoked activity of reward-responsive dopamine neurons (n = 48) to changes in reward identity

(A – B) Average firing on last 5 trials before (green) and first 5 trials after (red) a shift in reward identity for the big (A) and small (B) rewards. Big-A, 3 drops of reward A; Big-B, 3 drops of reward B; small-A, one drop of reward A; small-B, one drop of reward B. **(C)** Distributions of difference scores comparing firing to 1st (left), 2nd (middle) and 3rd drops (right) of the big reward in the last 5 versus first 5 trials before and after identity shift. **(D)** Distributions of difference scores comparing firing to the single drop of the small reward (left), and omissions of 2nd (middle) and 3rd drops (right) of the big reward in the last 5 versus first 5 small trials before and after an identity shift. Difference scores were computed from the average firing rate of each neuron. The numbers in each panel indicate results of Wilcoxon signed-rank test (p) and the average difference score (u). **(E)** Changes in average firing before and after reward identity shift. Black line indicates average firing at the time of the 1st and 2nd drops of the big reward and the small reward. Gray dashed line indicates average firing 0.5 s and 1.0 s after small reward. Error bars, S.E.M. **(F)** Correlation between changes in firing to shifts in reward identity, shown here, and changes in firing to delivery (blue dots) or omission (red dots) of the 2nd drop of the big reward, shown in Fig 3.
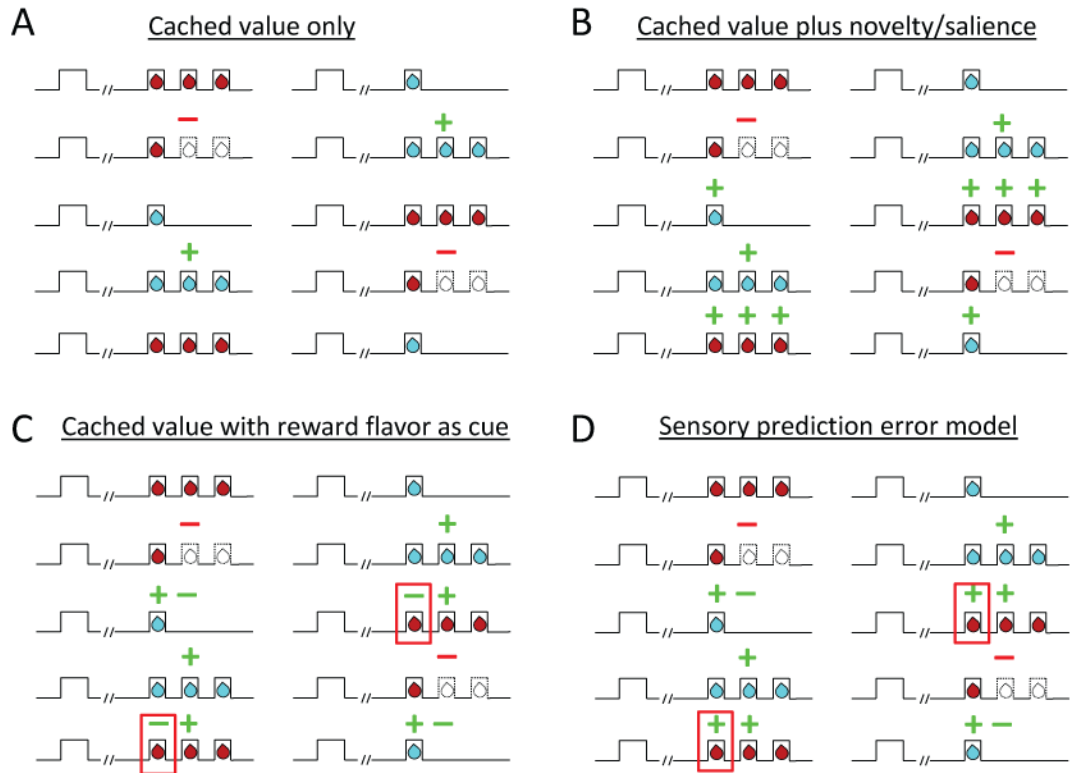
Figure 5. Schematic illustrating different ways that error signals might appear in response to changes in reward number and identity in the behavioral design reproduced from Fig 2, depending on what information dopaminergic errors reflect

(**A**) If dopamine neuron firing reflects errors in cached value only, then the conventional prediction would be increased firing only to the second drop of the big reward and decreased firing only on omission of this second drop of reward, since these are the only places in the design where cached value predictions are clearly violated. (**B**) If dopamine neuron firing reflects errors in cached value plus a novelty bonus or salience, then the prediction is for the same cached value errors shown in panel A on number shifts, plus increased firing to each drop of reward after an identity shift, assuming the unexpected flavor of each drop is salient or novel. (**C**) If dopamine neuron firing reflects errors in cached value, based on both the odor cues and also the sensory features of the first drop of each reward, then the prediction is for the same cached value errors shown in panel A for number shifts, plus a mixture of increased and decreased firing to identity shifts dependent on cached value accrued by the first drop's flavor in the prior block. For example, when chocolate was the small reward previously and becomes the large reward, one would expect the first drop to evoke decreased firing because it would be less valuable than expected (since its flavor predicts no more drops), followed by increased firing to subsequent drops because they would be unexpected. This again assumes the rat is using the flavor of the first drop to make predictions about subsequent drops. Note this does not apply if the two wells deliver similar amounts of different rewards, and as illustrated in Fig. S2, dopamine neurons also exhibited errors in response to identity shifts under these conditions. (D) Finally if dopamine neuron firing reflects errors in the prediction of sensory information or features, either instead of or in addition to cached value errors, then the predictions are for increased firing to unexpected

events generally and decreased firing to their omission. See text for full description. (green = positive errors, red = negative errors, red boxes highlight only place of divergence of predictions between C and D)