Dopamine neuron ensembles signal the content of sensory prediction errors

Thomas A. Stalnaker [1*], James D. Howard [3*], Yuji K. Takahashi [1], Samuel J. Gershman [2], Thorsten Kahnt [3*], and Geoffrey Schoenbaum [1,4.5*]

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

[2] Department of Psychology and Center for Brain Science, Harvard University;

[3] Department of Neurology, Feinberg School of Medicine, Northwestern University;

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

[5] Department of Neuroscience, Johns Hopkins School of Medicine.

*shared first or senior authorship

Correspondence or requests for material should be addressed to T.A.S. (thomas.stalnaker@nih.gov), J.D.H. (james.howard@northwestern.edu), T.K. (thorsten.kahnt@northwestern.edu) or G.S. (geoffrey.schoenbaum@nih.gov).

Abstract

Dopamine neurons respond to errors in predicting value-neutral sensory information. These data, combined with causal evidence that dopamine transients support sensory-based associative learning, suggest that the dopamine system signals a multidimensional prediction error. Yet such complexity is not evident in individual neuron or average neural activity. How then do downstream areas know what to learn in response to these signals? One possibility is that information about content is contained in the pattern of firing across many dopamine neurons. Consistent with this, here we show that the pattern of firing across a small group of dopamine neurons recorded in rats signals the identity of a mis-predicted sensory event. Further, this same information is reflected in the BOLD response elicited by sensory prediction errors in human midbrain. These data provide evidence that ensembles of dopamine neurons provide highly specific teaching signals, opening new possibilities for how this system might contribute to learning.

Introduction

Midbrain dopamine neurons are widely proposed to signal value prediction errors (Mirenowicz and Schultz, 1994). However, the same neurons also respond to errors in predicting the features of rewarding events, even when their value remains unchanged (Howard and Kahnt, 2018; Takahashi et al., 2017). Such sensory prediction errors would be useful for learning detailed information about the relationships between real-world events (Gardner et al., 2018; Howard and Kahnt, 2018; Langdon et al., 2017; Takahashi et al., 2017). Indeed, dopamine transients facilitate learning such relationships, independent of value, when they are appropriately positioned to mimic endogenous errors (Chang et al., 2017; Keiflin et al., 2019; Sharpe et al., 2017). Yet dopaminergic sensory prediction error signals do not seem to encode the content of the mis-predicted event, either at the level of individual neurons or summed across populations (Howard and Kahnt, 2018; Takahashi et al., 2017).

How then do downstream areas that may receive this teaching signal know what to learn? One possibility is that information about the content to be learned might be contained, at least partly, in the pattern of firing across an ensemble of dopamine neurons. It is now widely accepted that information is represented in areas like cortex and hippocampus not by individual neurons, but rather in a distributed fashion in the firing of groups of cells (Gochin et al., 1994; Jennings et al., 2019; Jones et al., 2007; Rich and Wallis, 2016; Rigotti et al., 2013; Schoenbaum and Eichenbaum, 1995; Wikenheiser and Redish, 2015; Wilson and McNaughton, 1993). If this is true for the cortex and hippocampus, then why not for the midbrain dopamine system? Consistent with this, here we show that the pattern of firing across a small group of dopamine neurons recorded in rats contains highly specific information about the content of the event that has been mis-predicted. We further show that this same content-rich signal is evident in the BOLD response elicited by sensory prediction errors in human midbrain. These data provide the first evidence of which we are aware that dopamine neuron ensembles generate unique teaching signals, which not only signal that a prediction error has occurred, but also signal what exactly was mis-predicted. These findings open new possibilities for how this system might contribute to the learning of complex associative information.

Results

To address whether dopamine neurons function as an ensemble to represent sensory prediction errors, we analyzed data from rats trained on a variant of the odor-guided choice task used to demonstrate the joint signaling of value and sensory prediction errors in our prior report (Takahashi et al., 2017) [1]. In the task variant (Figure 1a), two fluid wells delivered either one or three drops of discriminable but equally-preferred solutions of grape or tropical punch Kool Aid. Rats initiated each trial with a nose-poke into an odor port. After a brief delay, one of two odors was presented, indicating that reward would be available in the left or right well on that trial. If the rat responded at the proper fluid well, the reward was delivered. To induce prediction errors to correlate with neural activity, reward number or flavor were manipulated across a series of four transitions between five trial blocks in each recording session. At the first and second transitions, rewards were omitted and delivered unexpectedly, respectively, to allow identification of classic reward prediction errors. At the third and fourth transitions, reward number remained constant, but reward flavor was changed. At one transition, the flavors of all three drops were changed to replicate what was done previously, while at the other, only one drop of the three changed, leaving the others unchanged to provide a control condition to distinguish signaling of flavor errors from signaling of flavor itself.

Neural activity in VTA was recorded using drivable bundles of microelectrodes. During recording, the rats were highly accurate, responding correctly on ~95% of the forced-choice trials, indicating that they had learned the meaning of the odor cues, independent of reward number or flavor (Figure 1b). The rats also exhibited an appreciation of the reward number, responding significantly faster when the 3-drop reward was at stake, an effect that was also independent of the reward flavor (Figure 1c). Indeed, choice latency was similar across the two flavors, even in the behavior of individual rats, suggesting that they valued the two flavors similarly in the task (Figure 1c, lines). This is consistent with preference testing conducted separately after recording, which indicated that individually and as a group the rats had no significant preference between the two flavors of Kool-Aid (Figure 1d).

Using waveform characteristics and firing rate in response to reward as in previous papers (see Methods), we identified 30 putative dopaminergic neurons recorded during these sessions (Figure 1e and 1f). As previously reported (Takahashi et al., 2017, Supplemental Figure 2), the firing of these neurons exhibited classic reward prediction error correlates,

[1] While a limited analysis of a subset of these data were presented in a supplemental section of our prior report, this is the first presentation of the full dataset and its analysis as an ensemble.

decreasing in response to reward omission at the first transition and increasing to unexpected reward at the second transition, and these changes in firing were inversely correlated across neurons (Figure 2a-c). This is as expected based on numerous prior reports that individual dopamine neurons signal bidirectional errors in the prediction of reward, in different species, tasks, and labs (Schultz, 2016).

In addition, however, the same neurons also responded with elevated firing across transitions in which there was a change in reward flavor, combining both the third transition, presented previously (Takahashi et al., 2017, Supplemental Figure 2), and the more selective fourth transition, included here. This change in firing occurred even though the rats' behavior – both in the task and in separate preference testing (Figure 1b-d) – indicated no difference in the subjective value of the two flavors, even for individual subjects. The dopamine neurons increased firing to changes in flavor, and the size of these increases were positively correlated between the two flavor errors (Figure 2d and 2e). Further, individual neurons showed very little difference between initial firing rates in response to the two different flavor errors (Figure 2f). Thus, the activity of these neurons, individually or on average, signaled that something unexpected had happened, but it did not contain any details about that event.

To test whether such information might be available in the pattern of firing across a group or ensemble of dopamine neurons, we aligned activity from all neurons on like trials from each block, and then used a "training set" of trials from each flavor-switch block to identify the ensemble pattern characteristic of the neural response to each flavor. Individual trials left out of this training set were then matched to the two patterns to classify the flavor that had been delivered. To assess the evolution of information coding within and across trials, we used a sliding time window aligned to events in a trial and a sliding window of trials that progressed across each block. The results indicated that the pattern of activity across the ensemble contained information about flavor in both of the flavor-change trial blocks (Figure 3a and 3b). Critically, however, accurate decoding of flavor was observed only for the drops where flavor had changed and then only on trials early in the blocks; accuracy was only seen in epochs immediately after the new drop was delivered and fell to chance later in the block, consistent with representation of the error in predicting the flavor and not representation of the flavor itself.

This impression was confirmed when we formally compared classification accuracy in time windows surrounding drops where the flavor had changed versus windows surrounding drops where the flavor had not changed. Good classification performance was only observed when the drop had changed flavor, and then only in the first 10 trials of these blocks; performance was best in the earliest trials immediately after the transition, fell to chance in the

last 10 trials, and flavors from the early trials did not misclassify with the same flavors in the later trials (Figure 3c and d). The decline in classification accuracy occurred without any gross changes in baseline firing rates across the block (Figure 3d). Thus, the dopamine neuron ensemble was representing not the flavor itself, but flavor when it had been mis-predicted.

Finally, as an additional test of this idea, we next applied a similar approach to examine encoding of the information content of sensory prediction error signals previously reported in fMRI data in the human midbrain (Howard and Kahnt, 2018) [2]. These data were collected from subjects performing a task in which they learned that abstract visual cues predicted the odors of different sweet (SW) and savory (SV) food odor rewards (Figure 4a). The rewarding odors were matched in value, as reflected in both pleasantness ratings acquired before the learning task (Figure 4b) and choices made during the task (Figure 4c). During the fMRI scanning session, the odors associated with the visual cues were switched across blocks of trials (i.e., SW→SV and SV→SW), thereby inducing value-neutral sensory prediction errors similar to those induced by the flavor switches in the rat task described above. Previously it was reported that these switches evoked prediction error-like responses in the BOLD response in human midbrain (Howard and Kahnt, 2018; Suarez et al., 2019). Here we utilized a multivoxel pattern analysis (MVPA) to test whether distributed fMRI activity patterns in the midbrain contain information about the content of the error immediately after a switch and then later after learning.

This task and the analysis were conceptually similar to that applied to the single unit activity described above, and like the ensemble analysis applied to the single unit recording data, the MVPA analysis applied to the fMRI data found that it was possible to decode the identity (SW or SV) of the unexpected odor from the midbrain activity at the time the error was experienced (Figure 4d). Importantly, decoding was significantly above chance only on the trials in which the food odors were mis-predicted, but was at chance on subsequent trials when food odors were delivered as expected (Figure 4d). Follow-up examination of the decoder performance confirmed that decoding was only above chance on the error trial, and that the decoder was not biased towards prediction of a particular odor (Figure 4e). These data show that the ensemble midbrain activity represents the mis-predicted food odors and not the food odors themselves. Thus, the results presented here show that in both rats and humans, sensory prediction errors in the midbrain contain specific information about the features of the mis-predicted event itself, appropriate for instructing or updating representations in downstream brain regions.

[2] While these data were analyzed for sensory errors in our prior report, this is the first presentation of an MVPA analysis of these data to attempt to distinguish the content of the error signal.

Discussion

These results are consistent with the proposal that the midbrain dopamine system signals a generalized prediction error, reflecting a failure to predict features of an unexpected event beyond and even orthogonal to value (Gardner et al., 2018; Howard and Kahnt, 2018; Langdon et al., 2017; Takahashi et al., 2017). Importantly this proposal is not necessarily contrary to current canon; it can account for value errors as a special example of a more general function (Gardner et al., 2018), one readily apparent in the firing of individual neurons perhaps due to the priority given to such information when it is the goal of the experimental subject. However, unlike current canon, this proposal also easily explains why dopamine neurons are often phasically active in settings where value errors were not anticipated *a priori*, at least by the experimenters, such as when novel cues or even information is first presented (Bromberg-Martin and Hikosaka, 2009; Horvitz, 2000; Horvitz et al., 1997; Kakade and Dayan, 2002), or even in response to violations in beliefs or auditory expectations (Glascher et al., 2010; Gold et al., 2019; Iglesias et al., 2013; Schwartenbeck et al., 2016). Further, it provides a neural basis for recent demonstrations that dopamine transients are necessary for learning that cannot be easily accounted for by classic reinforcement learning mechanisms (Chang et al., 2017; Keiflin et al., 2019; Sharpe et al., 2017).

The current findings are critical to the viability of this proposal because they show that the pattern of firing across a relatively small population of dopamine neurons can provide details regarding the mis-predicted event. This is important because otherwise the ability of the dopamine system to convey a generalized error would be quite limited. Specifically, to elicit updates of specific associative information, the dopamine system would have to rely on other actors to provide the key information defining the content of the learning. In this regard, details in the pattern of activity distinguishes error-related activity in these neurons from a permissive signal that can only gate but not inform learning. Interestingly, the idea of a distributed, multidimensional error is key to more advanced computational algorithms, such as the successor representation model (Dayan, 1993), in which the error driving learning is not unitary but rather is represented as a vector. The current results show for the first time that an assembly of dopamine neurons can function in this manner. That the same information is not readily apparent in the activity of individual neurons is in accord with ideas guiding behavioral neurophysiology in other areas (Yuste, 2015), and suggests it is time to consider the functions of the dopamine system across rather than within individual neurons.

Methods

Experiment 1

Subjects: Ten male Long-Evans rats (Charles River Labs, Wilmington, MA), aged approximately 3 months at the start of the experiment, were used in this study. Rats were tested at the NIDA-IRP in accordance with NIH guidelines determined by the Animal Care and Use Committee.

Surgical procedures: All surgical procedures adhered to guidelines for aseptic technique. For electrode implantation, a drivable bundle of eight 25-um diameter NiCr/Formvar wires (A-M Systems, Sequim, WA) 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 800-1000 kOhms. 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 were cut in 40 µm sections and stained with thionin and then examined to determine electrode placement.

Behavioral task: Training and recording was conducted in aluminum chambers approximately 18” on each side with sloping walls narrowing to an area of 12” x 12” at the bottom. A central odor port consisting of a small hemicylinder accessible by nose-poke was located about 2cm above two fluid wells, and higher up on the same wall were mounted two lights. The odor port was connected to an airflow dilution olfactometer to allow the rapid delivery of olfactory cues, which were chosen from compounds obtained from International Flavors and Fragrances (New York, NY). Trial availability was signaled by illumination of the panel lights inside the box. When these lights were on, a nosepoke into the odor port resulted in delivery of the odor cue for 500ms. One of two different odors was delivered to the port on each trial in a pseudorandom order such that in each 50 trials there were 25 of each, and the same odor was never presented for more than three consecutive trials. At odor offset, the rat had 3 seconds to make a response at one of the two fluid wells. One odor indicated that reward would be available at the left well, while the other indicated it would be available at the right well; errors resulted in no reward delivery and the lights turning off (errors occurred on about 5% of trials across all recording sessions; see Figure 1b). On correct trials, lights turned off once rats had finished licking at the

well; the intertrial interval was ~2-3 seconds before the light turned on once again. Once the rats were shaped to respond accurately (at least ~75%) on both odors, we introduced trial-blocks in which the number and flavor of reward drops (one or three drops of Grape or Tropical Punch Kool-Aid solution) were constant within a block but changed between blocks according to the schedule summarized in Figure 1a. The drop volume was ~0.05 ml and multiple drops were delivered 1000ms apart. For each recording session, wells were randomly designated such that in the first trial-block, correct responses at one well resulted in delivery of 3 drops of grape solution while correct responses at the other well resulted in 3 drops of tropical punch solution. In the second trial-block, the number of drops available on both sides changed from three to one, with the flavor remaining the same. In the third trial-block, the number of drops available on both sides changed from one back to three, again with the flavor remaining the same. On the fourth trial-block, the flavor of all three drops on each side were switched to the other flavor. Finally, in the fifth trial-block, the flavor of the second drop on each side was switched to the opposite flavor, with the other two on both sides remaining the same. Thus, in each session, there was one number downshift transition (drop omission), one number upshift transition (new drop deliveries), one flavor transition across all 3 drops, and one flavor transition occurring at only the second drop. In each of the two flavor transitions, one side went from grape to tropical punch, while the other did the opposite.

Flavor preference testing: After the completion of all recording sessions, we conducted two-bottle consumption tests of the Kool-Aid solutions two times over two days for nine of the ten rats. These tests were run in a housing cage different from home-cages and experimental chambers. Tests were 2-min in duration and the location of the bottles was swapped roughly every 20 s to equate time on each side. The flavor and the initial location of the bottles were randomized in rats and swapped between the 1st and 2nd tests.

Single-unit recording: Wires were screened for activity daily; if no isolable single-unit activity was detected, the rat was removed 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 20X 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 50X 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 1-32X. Waveforms (>2.5:1 signal-to-noise) were extracted from active channels and recorded to disk by an associated workstation

Measures and statistical analyses: Average percent correct and choice latency (defined as the time from the end of odor delivery to withdrawal from the odor port on trials resulting in a correct response) were calculated by trial-type (3-drop, 1-drop, grape, tropical punch) across all trials. The flavor of the reward was defined as that of the first drop.

Units were sorted using Offline Sorter software from Plexon Inc (Dallas, TX). Sorted files were then processed and analyzed in Matlab (Natick, MA). Dopamine neurons were identified via a waveform analysis. Briefly, a 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. Neurons were considered putatively dopaminergic if they were in the wide waveform cluster and were also reward-responsive, defined as those that were significant at $p<0.05$ by t-test comparing baseline firing rate with the first 500ms of reward delivery across all rewarded trials. This waveform analysis is based on criteria similar to that typically used to identity dopamine neurons in primate studies (Bromberg-Martin et al., 2010; 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) and isolates neurons in rat VTA whose firing is sensitive to intravenous infusion of apomorphine or quinpirole (Jo et al., 2013; Roesch et al., 2007). Neurons identified in this manner are also selectively eliminated by expression of a Casp3 neurotoxin in TH+ neurons in VTA (by infusion of AAV1-Flex-TaCasp3-TEVp into TH-Cre transgenic rats; (Takahashi et al., 2017).

To calculate difference scores and firing rates for scatter plots, firing rates were aligned to drop delivery and baseline-subtracted using the 500ms immediately before the light-on at the start of the trial. To capture the peak reward-responsive activity, firing rates from 200ms to 700ms after the timestamp for the relevant drop delivery or drop omission were calculated. For number errors, the epochs were aligned to the first omitted drop (at the time the second drop would normally be delivered) in block 2, and the first newly delivered drop (second drop) in block 3. For flavor errors, the epochs were aligned to the first new flavor drop in both blocks 4 and 5. Difference scores were calculated for number transitions as the difference between the average

firing rate on the first three rewarded trials in the relevant block and the last five rewarded trials in the same block and direction, and for flavor transitions as the difference between the average firing rate in the first three rewarded trials in the relevant block and the last five trials in the previous block in the same direction.

For the decoding analyses, we used Matlab code from the Neural Decoding Toolbox (www.readout.info) (Meyers, 2013) to construct pseudoensembles consisting of all 30 putative dopamine neurons as described below. Decoding using pseudoensembles has been found to reveal the information held by the activity of populations of neurons in well-learned tasks such as the one we used here as effectively as analyses of real-time simultaneously recorded ensembles (Rigotti et al., 2013; Schoenbaum and Eichenbaum, 1995). The spike-trains of the 30 neurons were aligned to various trial events (light-on, odor delivery, odor port withdrawal, reward delivery, and light-off), concatenated according to the average time between these events, and then binned into sliding 900ms bins across the resulting spike-trains. All the correct trials from blocks 4 and 5 were labeled according to the flavor delivered on that trial, with trials from block 5 labeled according to the flavor of the second drop (the changed drop). The first ten trials in each block for each flavor were then taken from blocks 4 and 5, resulting in 40 total trials for each neuron. This selection resulted in flavor being fully crossed with side (10 trials from each flavor being left-well rewarded and 10 being right-well rewarded). The trials were then randomly divided into 20 splits, in each of which there was one test trial of each flavor for each neuron and 19 training trials of each flavor for each neuron. For each split, the flavor of each test trial was classified according to which training set had the highest correlation coefficient with it across the 30 neurons. This random split and test procedure was then repeated 500 times for every epoch to yield the average 1-0 accuracy of the classification at that epoch. This entire procedure was then repeated for sliding sets of 10 trials across the blocks (i.e. trials 1-10 of each flavor in each block, trials 2-11 of each flavor in each block, etc., ending with the last 10 trials of each flavor in each block). The 1-0 accuracy was then plotted separately for test trials taken from block 4 and block 5. The one-tailed 95% confidence interval for chance for the first sliding set of trials was calculated by shuffling the flavor labels 100 times and performing the entire analysis on each resulting dataset.

The decoding analysis shown in Figure 3c was similar to that described above, except that only the 900ms epoch beginning 100ms after the first new flavor drop was used, test data from blocks 4 and 5 were included together, and the first ten and last ten trials were labeled separately and both included in the same analysis. The resulting classification accuracy was

compared with a control classification of flavor in which the identical procedure was followed, except that data from the first drop of block 3 and the first drop of block 5 were used. These drops were selected because flavor was unchanged at those drops compared to the previous blocks, because they were part of 3-drop sequences just as in the experimental dataset, and because flavor was crossed with direction just as in the flavor transition analysis. The patterns in the flavor transition vs. flavor unchanged confusion matrices were compared by permutation test in which the flavor labels were shuffled 100 times for each analysis and 100,000 comparisons between the resulting confusion matrices were used to construct a distribution of comparisons. We then calculated the probability that the actual pattern of the two confusion matrices would be observed by chance. That is, we calculated the chance that the differences between flavor transition vs. flavor unchanged in grape early and tropical punch early would be as great as they were in the real data, while the differences in grape late and tropical punch late would be as small as they were in the real data.

The decoding analysis shown in Figure 3d was similar to that described above, except that the decay of decoding accuracy across the block was tested by using a sliding set of trials for both the flavor transition and flavor unchanged analyses. Each curve was then compared to chance by permutation tests with 100 shuffles of the flavor labels each. The accuracy in the unshuffled data was considered significantly greater than chance when it was in the top 5% of the shuffle distribution for five consecutive sliding sets of trials. Average baseline firing rate on the trial-sets included in each of the decoding algorithms was also calculated and shown on Figure 3d.

Experiment 2

Subjects: Twenty three human participants (9 male, ages 19-34, mean ± SD = 25.5 ± 4.1 years) with no history of psychiatric illness gave informed written consent to participate in this study. The study protocol was approved by the Northwestern University Institutional Review Board.

Odor stimuli and presentation: Eight food odors, including four sweet (strawberry, caramel, cupcake, gingerbread) and four savory (potato chips, pot roast, sautéed onions, garlic), were provided by International Flavors and Fragrances (New York, NY). For all experimental tasks, odors were delivered directly to participants' noses using a custom-built computer-controlled olfactometer.

Odor selection and task familiarization: In an initial behavioral testing session, hungry participants (fasted for at least 6 hours) first provided pleasantness ratings of the 8 food odors.

Based on these ratings, one sweet odor and one savory odor were chosen such that they were matched as closely as possible in pleasantness. Next, we acquired pleasantness ratings for the two selected odors across a range of odor concentrations, diluted to varying degrees with odorless air. Based on these ratings, we selected two concentrations for each odor, such that the two low-concentration odors had the same pleasantness and the two high-concentration odors had the same pleasantness.

Participants next completed 84 trials of the instrumental reversal learning task they would eventually complete in the fMRI scanner. For this task, two abstract visual symbols were randomly chosen to serve as conditioned stimuli (CS) throughout the rest of the experiment. Each trial started with either one of the two CS's (indicating it was a forced choice trial) or a question mark (indicating it was a free choice trial) presented for 4 s. Both CS's were then presented on either side of a center crosshair (side fully randomized and counterbalanced) for 1.5 s, during which time participants were instructed to choose via left or right mouse click the CS that appeared alone in the preceding screen (in the case of a forced choice trial), or whichever CS they preferred (in the case of a free choice trial). If no response was made within 1.5 s, "TOO SLOW" appeared on the screen and the next trial was initiated after a variable delay. If a response was made, the odor currently paired with the selected CS was delivered after a 2 s delay. Odor delivery, lasting 3 s, was indicated by changing the color of the center crosshair to blue, informing participants to sniff. Participants then rated either the pleasantness or identity of the received odor (rating type randomized), followed by a 0-2 s inter-trial interval.

Across the 84 trials, the choice task was covertly subdivided into 8 blocks of trials delineated by the specific CS-US associations predetermined for that block. Each block consisted of either 9 or 12 trials, and the length of blocks across the session was pseudorandomized. Within a given block, one of the CS's was paired deterministically with the high concentration of one odor identity (e.g., sweet high: SW_H), while the other CS was paired deterministically with the low concentration of the same odor identity (e.g., sweet low: SW_L). After each block, the CS-US associations were changed without warning, and new blocks always began with two forced choice trials (one for each CS). In the case of flavor reversals, the flavor of the US was changed for both CS's while leaving CS-value associations the same. In the case of reward value reversals, the CS-value association was swapped between the two CS's, while leaving flavor unchanged. Reversals alternated between flavor and value, and there were 7 total reversals across the 84-trial task.

Choice task during fMRI scanning: The fMRI scanning session was conducted within ~10 days (mean ± SD = 10.0 ± 4.4 days) of the initial behavioral session. During scanning, hungry participants (fasted for at least 6 hours) completed 3 runs of the 84-trial reversal learning task described above. Each run lasted ~21 minutes, and the sequence of alternating flavor and value reversals was counterbalanced across subjects.

fMRI data acquisition: MRI data were acquired on a Siemens 3T PRISMA system equipped with a 64-channel head-neck coil. Echo-Planar Imaging (EPI) volumes were acquired with a parallel imaging sequence with the following parameters: repetition time, 2 s; echo time, 22 ms; flip angle, 90°; multi-band acceleration factor, 2; slice thickness, 2mm; no gap; number of slices, 58; interleaved slice acquisition order; matrix size, 104 x 96 voxels; field of view 208 mm x 192 mm. The functional scanning window was tilted ~30° from axial to minimize susceptibility artifacts in OFC(Weiskopf et al., 2006). Each fMRI run consisted of 640 EPI volumes covering all but the dorsal portion of the parietal lobes. To aid in co-registration and normalization of the functional scans, we also acquired 10 EPI volumes for each participant covering the entire brain, with the same parameters as described above except 95 slices and a repetition time of 5.25 s. A 1 mm isotropic T1-weighted structural scan was also acquired for each participant. This image was used for spatial normalization.

fMRI data preprocessing: All image preprocessing and general linear modeling was done using SPM12 software (www.fil.ion.ucl.ac.uk/spm/). To correct for head motion during scanning, for each subject all functional EPI images across the 3 fMRI runs were aligned to the first acquired image. The motion-corrected images were smoothed with a Gaussian kernel at native scan resolution (2 x 2 x 2 mm) to reduce noise but retain potential information content(Gardumi et al., 2016). For reverse normalization of midbrain regions of interest to participant-specific native space, each participant's T1-scan was normalized to Montreal Neurological Institute (MNI) space using the 6-tissue probability map provided by SPM12. The inverse deformation field resulting from this normalization step was then applied for each participant to a region of interest in MNI space defined by spheres of 4-voxel radius centering on the two midbrain coordinates reported to show a significant univariate response to flavor prediction errors (left: x=-16, y=-14, z=-12; right: x=6, y=-14, z=-14) (Howard and Kahnt, 2018).

General linear modeling and MVPA analyses: For the decoding analysis, we constructed independent subject-level event-related general linear models (GLMs) for each fMRI run using finite impulse response (FIR) functions specified over 12 time bins time-locked to the onset of each trial. Nuisance regressors included: normalized respiratory activity traces (measured by MR-safe breathing belts affixed around the torso); the 6 realignment parameters calculated for

each scanned image during motion-correction; the derivative, square, and square of the derivative of each realignment regressor; the absolute signal difference between even and odd slices, and the variance across slices, in each functional volume; additional regressors as needed to censor individual volumes in which particularly strong head motion occurred. Odor onsets corresponding to 13 conditions were specified in each GLM: SV→SW reversals, SW→SV reversals, SW and SV 1, 2, 3, and 4 trials after reversals, SW and SV on the trial immediately preceding reversals, and all other trials. The resulting parameter estimates within a region of interest (ROI) defined by the intersection of an un-normalized anatomical mask of the midbrain and the un-normalized spherical mask described above were extracted for each subject, fMRI run, and condition at the time bin corresponding most closely to odor delivery given hemodynamic lag. Prior to decoding, voxels within each subject's midbrain ROI were sorted according to the difference in responses to flavor transitions on the error trial (combined across SV→SW and SW→SV) and responses on the trial preceding error trials (combined across SW and SV).

The resulting sorted parameter estimates were then submitted to pairwise linear support vector machine decoding analyses using the libsvm implementation (Chang and Lin, 2011). Each pairwise analysis corresponded to the SW and SV conditions at a given trial point (i.e., error trial, error trial +1, error trial +2, etc.), and was conducted using a nested cross-validation approach in which we first performed leave-one-subject-out cross-validation in increasing numbers of voxels within the ROI to determine the number of voxels that most effectively decodes reward flavor in a "training set" of subjects. Leave-one-run-out cross-validated decoding of flavor in the left out subject was then conducted in the number of voxels giving maximal decoding accuracy from the training set of subjects. This process was repeated for each subject, resulting in an independent decoding accuracy value calculated for each subject and decoding pair.

An identical analysis was conducted for value transitions (i.e., flavor unchanged), in which GLM's were specified using the same condition types time locked to these type of reversals: SW and SV at the value error trial, SW and SV at 1, 2, 3, and 4 trials after value reversal and immediately before value reversal, and all other trials. We then implemented the same nested cross-validation method to generate decoding accuracies for pairwise tests at each trial point.

Author Contributions

Experiment 1 in rats: YKT, TAS, and GS designed experiment, YKT conducted the experiment, and TAS analyzed the data, with input on approaches and interpretation from TK, SJG, and GS. Experiment 2 in humans: JDH and TK designed, conducted, and analyzed the experiment, with input on approaches and interpretation from TAS, SJG, and GS. Writing: TAS, JDH, TK and GS wrote the manuscript with input from all of the other authors.

Acknowledgments

This work was supported by the Intramural Research Program at the National Institute on Drug Abuse and National Institute on Deafness and Other Communication Disorders grant R01DC015426 (to TK). The opinions expressed in this article are the authors' own and do not reflect the view of the NIH/DHHS.

Declarations of Interest

The authors declare no competing interests.

References

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

Bromberg-Martin, E.S., Matsumoto, M., Hong, S., and Hikosaka, O. (2010). A pallidus-habenula-dopamine pathway signals inferred stimulus values. Journal of Neurophysiology 104, 1068-1076.

Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 1-27.

Chang, C.Y., Gardner, M., Di Tillio, M.G., and Schoenbaum, G. (2017). Optogenetic blockade of dopamine transients prevents learning induced by changes in reward features. Current Biology 27, 3480-3486.

Dayan, P. (1993). Improving generalization for temporal difference learning: the successor representation. Neural Computation 5, 613-624.

Fiorillo, C.D., Newsome, W.T., and Schultz, W. (2008). The temporal precision of reward prediction in dopamine neurons. Nature Neuroscience 11, 966-973.

Gardner, M.P.H., Schoenbaum, G., and Gershman , S.J. (2018). Rethinking dopamine as generalized prediction error. Proceedings of the Royal Society B 285, 20181645.

Gardumi, A., Ivanov, D., Hausfeld, L., Valente, G., Formisano, E., and Uludag, K. (2016). The effect of spatial resolution on decoding accuracy in fMRI multivariate pattern analysis. Neuroimage 132, 32-42.

Glascher, J., Daw, N., Dayan, P., and O'Doherty, J.P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66, 585-595.

Gochin, P.M., Colombo, M., Dorfman, G.A., Gerstein, G.L., and Gross, C.G. (1994). Neural ensemble coding in inferior temporal cortex. Journal of Neurophysiology 71, 2325-2337.

Gold, B.P., Mas-Herrero, E., Zeighami, Y., Benovoy, M., Dagher, A., and Zatorre, R.J. (2019). Musical reward prediction errors engage the nucleus accumbens and motivate learning. Proceedings of the National Academy of Science 116, 3310-3315.

Hollerman, J.R., and Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. Nature Neuroscience 1, 304-309.

Horvitz, J.C. (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. Neuroscience 96, 651-656.

Horvitz, J.C., Stewart, T., and Jacobs, B.L. (1997). Burst activity of ventral tegmental dopamine neurons is elicited by sensory stimuli in the awake cat. Brain Research 759, 251-258.

Howard, J.D., and Kahnt, T. (2018). Identity prediction errors in the human midbrain update reward-identity expectations in the orbitofrontal cortex. Nature Communications 9, 1-11.

Iglesias, S., Mathys, C., Brodersen, K.H., Kasper, L., Piccirelli, M., den Ouden, H.E., and Stephan, K.E. (2013). Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. Neuron 80, 519-530.

Jennings, J.H., Kim, C.K., Marshel, J.H., Raffiee, M., Ye, L., Quirin, S., Pak, S., Ramakrishnan, C., and Deisseroth, K. (2019). Interacting neural ensembles in orbitofrontal cortex for social and feeding behaviour. Nature 565, 645-649.

Jo, Y.S., Lee, J., and Mizumori, S.J. (2013). Effects of prefrontal cortical inactivation on neural activity in the ventral tegmental area. Journal of Neuroscience 33, 8159-8171.

Jones, L.M., Fontanini, A., and Katz, D.B. (2007). Natural stimuli evoke dynamic sequences of states in sensory cortical ensembles. Proceedings of the National Academy of Science 104, 18772-18777.

Kakade, S., and Dayan, P. (2002). Dopamine: generalization and bonuses. Neural Networks 15, 549-559.

Keiflin, R., Pribut, H.J., Shah, N.B., and Janak, P.H. (2019). Ventral tegmental dopamine neurons participate in reward identity predictions. Current Biology 29, 92-103.

Kobayashi, K., and Schultz, W. (2008). Influence of reward delays on responses of dopamine neurons. Journal of Neuroscience 28, 7837-7846.

Langdon, A.J., Sharpe, M.J., Schoenbaum, G., and Niv, Y. (2017). Model-based predictions for dopamine. Current Opinion in Neurobiology 49, 1-7.

Matsumoto, M., and Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature 459, 837-841.

Meyers, E.M. (2013). The neural decoding toolbox. Frontiers in Neuroinformatics 7, Article 8.

Mirenowicz, J., and Schultz, W. (1994). Importance of unpredictability for reward responses in primate dopamine neurons. Journal of Neurophysiology 72, 1024-1027.

Morris, G., Nevet, A., Arkadir, D., Vaadia, E., and Bergman, H. (2006). Midbrain dopamine neurons encode decisions for future action. Nature Neuroscience 9, 1057-1063.

Rich, E.L., and Wallis, J.D. (2016). Decoding subjective decisions from orbitofrontal cortex. Nature Neuroscience 19, 973-980.

Rigotti, M., Barak, O., Warden, M.R., Wang, X.-J., Daw, N.D., Miller, E.K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. Nature 497, 585-590.

Roesch, M.R., Calu, D.J., and Schoenbaum, G. (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nature Neuroscience 10, 1615-1624.

Schoenbaum, G., and Eichenbaum, H. (1995). Information coding in the rodent prefrontal cortex. II. Ensemble activity in orbitofrontal cortex. Journal of Neurophysiology 74, 751-762.

Schultz, W. (2016). Dopamine reward prediction-error signalling: a two-component response. Nature Reviews Neuroscience 17, 183-195.

Schwartenbeck, P., FitzGerald, T.H.B., and Dolan, R. (2016). Neural signals encoding shifts in beliefs. Neuroimage 125, 578-586.

Sharpe, M.J., Chang, C.Y., Liu, M.A., Batchelor, H.M., Mueller, L.E., Jones, J.L., Niv, Y., and Schoenbaum, G. (2017). Dopamine transients are sufficient and necessary for acquisition of model-based associations. Nature Neuroscience 20, 735-742.

Suarez, J.A., Howard, J.D., Schoenbaum, G., and Kahnt, T. (2019). Sensory prediction errors in the human midbrain signal identity violations independent of perceptual distance. eLIFE 8, e43962.

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

Waelti, P., Dickinson, A., and Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. Nature 412, 43-48.

Weiskopf, N., Hutton, C., Josephs, O., and Deichmann, R. (2006). Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: a whole-brain analysis at 3 T and 1.5 T. Neuroimage 33, 493-504.

Wikenheiser, A.M., and Redish, A.D. (2015). Decoding the cognitive map: ensemble hippocampal sequences and decision making. Current Opinion in Neurobiology 32, 8-15.

Wilson, M.A., and McNaughton, B.L. (1993). Dynamics of the hippocampal ensemble code for space. Science 261, 1055-1058.

Yuste, R. (2015). From the neuron doctrine to neural networks. Naure Reviews Neuroscience 16, 487-497.

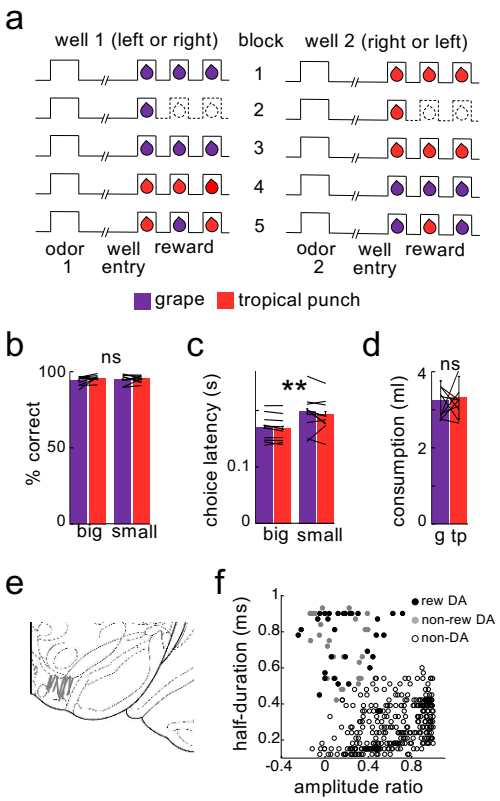

Figure 1: Task design and behavior during recording. Schematic (**a**) illustrates the order of events in trials at each well and the number and type of reward delivered at each well in the five trial-blocks performed in all recording sessions. Dashed lines indicate the omission of drops previously delivered. Rats were highly accurate in choosing the rewarded well during recording (**b**), and accuracy was unaffected by the flavor or number of drops at a particular well, either for the group or for individual subjects (flavor: $F_{1,193}$=1.3, p=0.26; number: $F_{1,193}$=1.0, p=0.32; interactions with subject: F's<=1.0, p's>0.47). Rats were faster to respond for the 3-drop rewards (**c**), and this effect was again unaffected by the flavor of reward, either for the group or for individual subjects (main effect of number: $F_{1,193}$=190, $p<10^{-6}$; main effect of flavor: $F_{1,193}$=1.75, p=0.19; flavor X subject interaction: $F_{9,193}$=0.86, p=0.56). A two-bottle preference test run at the end of the sessions (**d**) also revealed no effect of flavor ($F_{1,9}$=0.17, p=0.69). Data for individual subjects is illustrated by lines; error bars represent standard errors across sessions for percent correct and latency and across rats for the consumption test. Recordings were made in ventral tegmental area (**e**), and dopaminergic neurons (n=30) were identified by waveform cluster analysis (**f**). ** p<0.01. g=grape, tp=tropical punch.

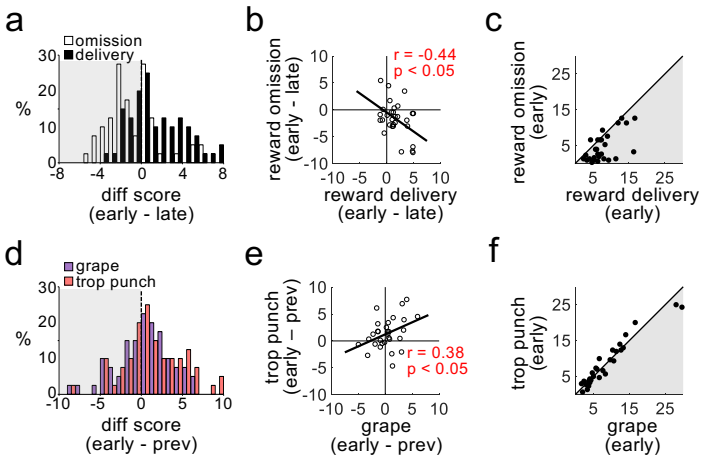

Figure 2. Dopamine neurons do not distinguish the identity of sensory prediction errors. Plots show firing rates of dopamine neurons in response to transitions in number of reward drops (omission or delivery; **a**-**c**) and flavor (grape or tropical punch; **d**-**f**). Changes in firing rate in response to omission (negative errors) and delivery (positive errors) were readily distinguishable (**a;** t_{29}=4.0, $p<10^{-3}$), inversely correlated across neurons (**b**), and firing rates were markedly different after the transition (**c;** t_{29}=5.2, $p<10^{-4}$). The same neurons exhibited increased firing rates in response to transitions in the expected flavor of reward (**d**; t_{29}=2.1, $p<0.05$), but the increases to the two flavors were indistinguishable (t_{29}=-1.95, ns), positively correlated across neurons (**e**), and firing rates after the transition also did not distinguish the two flavor errors (**f;** t_{29}=0.13, ns).

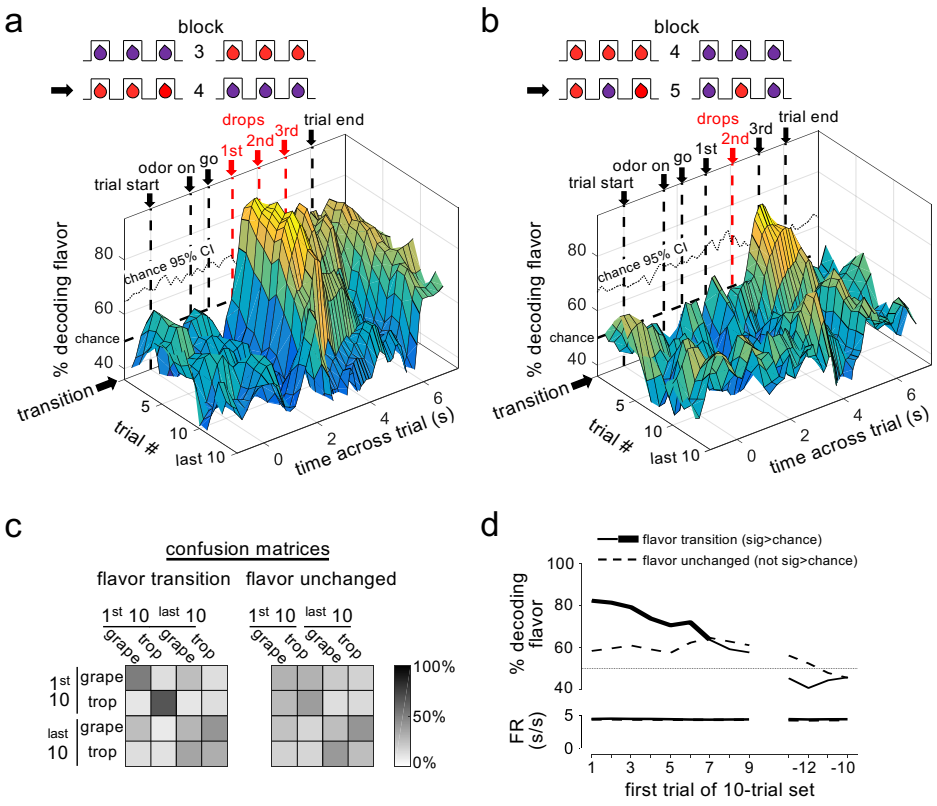

Figure 3. Dopamine ensembles distinguish the identity of sensory prediction errors. Surface plots show decoding of flavor by dopamine neuron ensembles, using data from a sliding window during trials after all three drops changed flavor (**a**) or when only the second drop changed flavor (**b**). Red arrows indicate the time of the new flavor drop delivery. In each case decoding was significantly above chance, at the changed drops, but only early in the block (dotted lines on back walls show one-tailed 95% confidence interval bounds for chance, by permutation tests). This effect was also evident when we collapsed data from the two blocks and compared decoding in epochs capturing firing to the drops where flavor changed versus control epochs capturing firing where flavors had not changed (**c**); flavor could be decoded accurately by dopamine ensembles only immediately after changes in flavor (patterns in confusion matrices were significantly different at $p<10^{-4}$ by permutation test). A more detailed analysis using sliding sets of 10-trials (**d**) showed the decay of flavor decoding as the block progressed (upper plot, solid line), while control decoding of flavor (dotted line) and baseline firing rates in both conditions (lower plot) were unchanged across the block. Thick line in the upper plot shows significance compared to chance ($p<0.05$ for at least 5 significant trial sets by permutation test). Thin dotted line in upper plot shows chance decoding level.

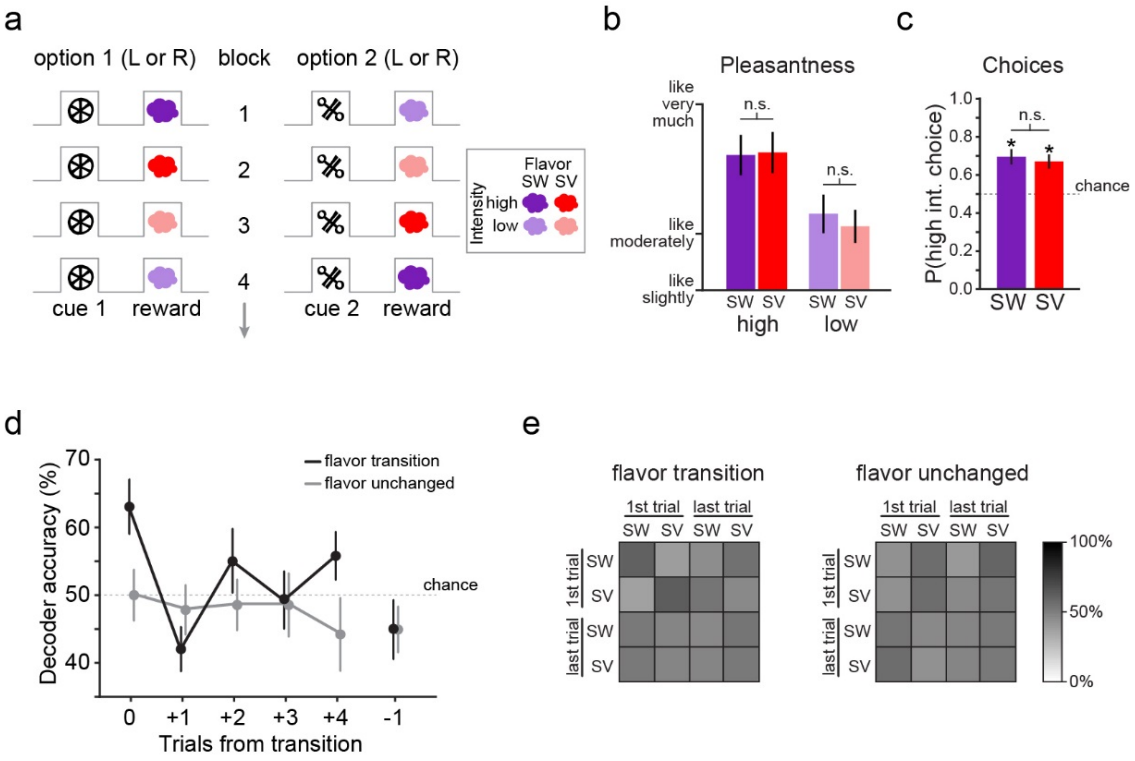

Figure 4. Human midbrain distinguishes the identity of sensory prediction errors. a) The reversal learning task involved binary choices between two abstract visual cues to receive either a high or low concentration of one of two food odor rewards (one sweet [SW] and one savory [SV]). The associations were covertly changed throughout the task to induce either sensory prediction errors (i.e. transition from block 1 to block 3) or value prediction errors (i.e. transition from block 2 to block 3). **b)** Sweet and savory food odors were matched for pleasantness within each odor concentration (SW high vs. SV high: $t_{(22)}$ = 0.18, p = 0.86; SW low vs. SV low: $t_{(22)}$ = 1.16, p = 0.26). Error bars depict within-subject s.e.m. **c)** On free choice trials, the cue associated with the high-concentration odor was chosen significantly above chance (50%) for both odor identities (SW: $t_{(22)}$ = 4.03, p = 2.83 x 10^{-4}; SV: $t_{(22)}$ = 4.20, p = 1.83 x 10^{-4}) and did not differ ($t_{(22)}$ = 0.71, p = 0.48). Error bars depict within-subject s.e.m. **d)** Decoding accuracy of SW vs. SV was significantly above chance on the error trial of flavor transitions ($t_{(22)}$ = 3.22, p = 0.004), but not for subsequent trials or the trial preceding error trials (p's > 0.12). Decoding accuracy of SW vs. SV was at chance for the error trial, subsequent trials, and the trial preceding value transitions (p's > 0.15). Error bars depict within-subject s.e.m. **e)** Confusion matrices show the decoding accuracy for individual conditions within the decoding analyses. Within the top left quadrant of the flavor transition matrix (i.e. training and testing the classifier on the error trial of flavor transitions), across all subjects and iterations, accuracy was at 63.3% for SW predictions and 63.8% for SV predictions. All other comparisons for flavor transitions and all comparisons for value transitions were at chance.