TRANSIENT ACTIVATION OF MIDBRAIN DOPAMINE NEURONS BY REWARD RISK

C. D. FIORILLO[a,b*]

aDepartment of Bio and Brain Engineering, Korean Advanced Institute of Science and Technology, Daejeon, Republic of Korea

bDepartment of Neurobiology, Stanford University, Stanford, CA, USA

Abstract—Dopamine neurons of the ventral midbrain are activated transiently following stimuli that predict future reward. This response has been shown to signal the expected value of future reward, and there is strong evidence that it drives positive reinforcement of stimuli and actions associated with reward in accord with reinforcement learning models. Behavior is also influenced by reward uncertainty, or risk, but it is not known whether the transient response of dopamine neurons is sensitive to reward risk. To investigate this, monkeys were trained to associate distinct visual stimuli with certain or uncertain volumes of juice of nearly the same expected value. In a choice task, monkeys preferred the stimulus predicting an uncertain (risky) reward outcome. In a Pavlovian task, in which the neuronal responses to each stimulus could be measured in isolation, it was found that dopamine neurons were more strongly activated by the stimulus associated with reward risk. Given extensive evidence that dopamine drives reinforcement, these results strongly suggest that dopamine neurons can reinforce risk-seeking behavior (gambling), at least under certain conditions. Risk-seeking behavior has the virtue of promoting exploration and learning, and these results support the hypothesis that dopamine neurons represent the value of exploration. © 2011 IBRO. Published by Elsevier Ltd. All rights reserved.

Key words: dopamine, reward, risk, uncertainty, exploration, reinforcement learning.

Following the onset of stimuli that predict future reward, the firing rate of midbrain dopamine neurons is increased for about 200 ms. In a typical experiment, such as those described here, a conditioned visual stimulus predicts that liquid will be delivered 1 s later. The amplitude of the response to such a stimulus appears to increase monotonically with the animal's expectation of future reward value (e.g. liquid volume) (Fiorillo et al., 2003; Tobler et al., 2005; Bayer and Glimcher, 2005; Morris et al., 2006; Roesch et al., 2007). This finding supports the proposed role of dopamine neurons in driving reinforcement (Wise, 2004) in accord with reinforcement learning models (Schultz et al., 1997).

The reward prediction in reinforcement learning models of dopamine function is often conceived to be synony- mous with expected reward value. However, an animal can never be completely certain about future reward, and thus the prediction of reward is best described by a probability distribution of potential reward outcomes (reward magnitudes). Uncertainty refers to the width of the probability distribution. In the context of reward, uncertainty is often called "risk" and it is known to influence decision making and behavior (e.g. Kahneman and Tversky, 1979; Kacelnik and Brito e Abreu, 1998; McCoy and Platt, 2005; Hayden and Platt, 2007; Hayden et al., 2008; So and Stuphorn, 2010). The firing rate of dopamine neurons gradually increases prior to uncertain reward outcomes (during the "delay period" following a conditioned stimulus) (Fiorillo et al., 2003), and there is also evidence that the transient response of dopamine neurons may be scaled by prior uncertainty about reward value, through what could be called "divisive normalization" (Tobler et al., 2005). The present work provides the first direct evidence that the transient response of dopamine neurons depends on uncertainty (risk) in the prediction of future reward.

EXPERIMENTAL PROCEDURES

Animals

Three rhesus macaques (*Macaca mulatta*) were studied. Monkeys F (5.5 kg) and L (7.0 kg) were male, and monkey O (10.5 kg) was female. Procedures complied with guidelines established by the National Institutes of Health, and were overseen locally by the Stanford University Animal Care and Use Committee.

Choice task

Expo software (written by Peter Lennie and modified by Julian Brown) was used to run experiments and to collect data. Two visual stimuli (Fig. 1A) were simultaneously presented on a computer monitor. Each stimulus spanned 4 degrees of visual angle, and was centered 6 degrees to the left or right of the center of the monitor. The positions of the two stimuli varied randomly between left and right across trials. Eye position was monitored with an infrared eye-tracking system (Eyelink II from SR Research of Toronto, Canada), and continuous fixation on either of the two stimuli for 500 ms was immediately followed by a sound (72 dB) that signaled correct performance. The sound was identical for stimuli U and C. One second after the sound, the reward outcome was delivered. On rewarded trials, apple juice (diluted to two-thirds of total volume by addition of water) was delivered from a spout that was placed inside the monkey's mouth. Stimulus C was followed by 125 μl of juice delivered over 150 ms. Stimulus U was followed by 240 μl of juice delivered over 250 ms on a pseudo-randomly selected 50% of trials, and nothing on the remaining trials. Regardless of the animal's choice, both visual stimuli were extinguished at the same time (synchronous with offset of juice delivery on trials in which juice was delivered).

Note that the subjective expected value of stimulus U should be slightly less than stimulus C because stimulus U was associ-

*Correspondence to: C. D. Fiorillo, Department of Bio and Brain Engineering, KAIST, 335 Gwahangno, Building E16, Yuseong-gu, Daejeon 305-701, Republic of Korea. Tel: +82-42-350-4326.
E-mail address: fiorillo@kaist.ac.kr (C. D. Fiorillo).
Abbreviations: SN, substantia nigra; VTA, ventral tegmental area.
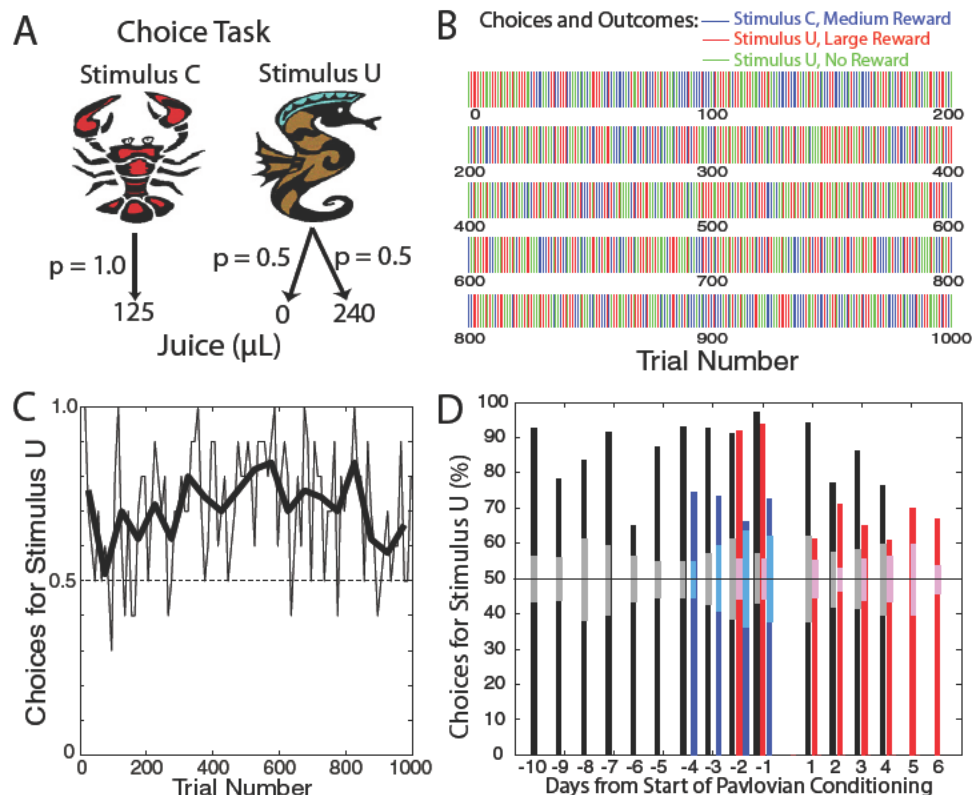
Fig. 1. Risk preference in a choice task. (A) Two images were presented simultaneously. Fixation of gaze on stimulus C for 0.5 s resulted in certain delivery of 125 µl of juice after a further delay of 1.0 s, whereas fixation on stimulus U resulted in 240 µl of juice on a pseudorandomly selected 50% of trials, and no juice on the remaining trials. These two images were used in all 3 monkeys, but were reversed in meaning between monkeys O and F (the case of monkey F is illustrated here) Distinct and novel stimuli were used in some cases, with similar results (not shown). (B) Choice behavior in a single session of 1000 trials that lasted for 70 min. Each vertical line represents a single trial, starting with the first trial in the top left corner and ending with the last trial in the bottom right corner. Choices for stimulus C are in blue. Choices for stimulus U that resulted in large reward outcome are in red, and those that resulted in no reward are in green. (C) Fraction of choices for stimulus U over time during the same session illustrated in (B). The thick line represents bins of 50 trials, whereas the thin line represents bins of 10 trials. (D) Percentage of choices for stimulus U by day. Day 1 is the first day on which choice behavior was measured after the start of Pavlovian conditioning. 70–1000 trials were performed each day. The lighter area of each bar represents ±2 standard deviations from the mean of the binomial distribution with $P=0.5$ and "n" equal to the total number of choice trials performed. Monkey O in black, monkey F in red, and monkey L in blue. The data shown in (B) and (C) correspond to the second day after the start of Pavlovian conditioning in monkey F.

ated with a slightly smaller average volume of liquid (120 versus 125 µl) and because of the temporal discounting associated with the delayed delivery of liquid (since the additional amount of juice provided in the large drop did not begin to flow until 1.150 s after stimulus onset).

Liquid delivery followed presentation of stimulus U in a pseudorandom manner. In every "block" of 10 consecutive choices of stimulus U (disregarding any intervening choices for stimulus C), a randomly selected five choices were followed by juice (240 µl). The start of a block was not signaled to the monkey (a block was only defined for the purpose of randomization). The intertrial interval from reward outcome to presentation of visual stimuli on the next trial was random between 0.5 and 1.5 s. If fixation on one of the stimuli was not acquired within 1.0 s, the stimuli disappeared and another intertrial interval began.

The behavioral and neuronal results presented here were all obtained using the two visual stimuli shown in Fig 1A. To confirm that the preference for stimulus U over C was not related to the particular visual features of the stimuli rather than their associated reward uncertainty, choice behavior was also examined using novel pairs of stimuli in some sessions (of 100–200 choice trials), with similar results. Significant preferences for stimulus U were evident within the first 100 trials of the choice task when using

novel visual stimuli (not shown). In addition, the two primary images used for stimuli U and C (Fig. 1A) were reversed in meaning between monkeys O and F.

Pavlovian task

On each trial, one of the two visual stimuli (Fig. 1A) appeared in the center of the monitor. Reward outcome was delivered 1.0 s after onset of the stimulus. The stimulus remained present until the reward outcome had been delivered, so that stimulus offset occurred at the same time that juice stopped flowing (on trials in which juice was delivered). In a block of 20 sequential trials, 10 were randomly selected to be stimulus U, and the other 10 were stimulus C. Of the 10 trials of stimulus U within one block, five were randomly selected to deliver juice and the others delivered no juice. The onset of a block was not signaled to the monkey. The intertrial interval (interval between reward outcome on one trial and stimulus onset on the next trial) was random between 1 and 4 s.

Electrophysiological recordings were performed during the Pavlovian task. Recordings began after at least 2 days of training in the choice task, followed by at least 2 days of training in the Pavlovian task, with at least 100 trials of each stimulus on each

day. The Pavlovian task was performed over the course of a month in each monkey, for a total of about 2000 trials of each stimulus.

Recording and localization of dopamine neurons

Glass-insulated tungsten electrodes were purchased from Alpha-Omega (Nazareth, Israel), and subsequently plated with gold and then platinum, as previously described (Merrill and Ainsworth, 1972). Plating reduced impedance from ~2 MΩ before plating to ~1 MΩ after plating. The tungsten core of the shaft of the electrode had a diameter of 125 μm, and with glass insulation, the outer diameter of the shaft was 300 μm. Thus these electrodes were relatively inflexible and unlikely to bend during tissue penetration. A single electrode was lowered vertically each day through a cylinder centered on the midline approximately 7.5 mm anterior to the interaural line. Dopamine neurons were distinguished from other neurons in the region by their extracellularly recorded discharge characteristics, including long, multiphasic waveforms (2.0–5.0 ms when high-pass filtered at 100 Hz) and slow, fairly regular basal firing rates (0.1–10.0 Hz), as previously described (Schultz and Romo, 1987). Consistent with previous studies, 82% of presumed dopamine neurons were significantly activated by delivery of juice following a long and variable interval outside of any task (not shown). However, activation by juice was not used as a criterion for inclusion in this study.

Midbrain dopamine neurons of substantia nigra (SN) and ventral tegmental area (VTA) were localized with the aid of physiologically identified landmarks. The somatosensory representation of the face in ventral posterior medial thalamus lies dorsal of SN (Paxinos et al., 2000). It was identified in each monkey by manual stimulation of the face, under mild anesthesia with ketamine and xylazine. The oculomotor nucleus was found in each monkey by monitoring eye position while searching for dopamine neurons. It is centered on the midline at the same depth as the more dorsal dopamine neurons, and it extends only about 1 mm lateral of the midline (Fig. 3A, left) (Paxinos et al., 2000). Neurons were identified as being in the left oculomotor nucleus as they responded with both phasic and tonic components to saccades and smooth pursuit eye movements. Many single neurons were activated preferentially in response to eye movements to the right, some neurons preferred upward movements, and some preferred downward movements. No neurons were observed to be activated in response to eye movements to the left. Responses during eye movements persisted in darkness. The positions at which the left oculomotor nucleus was recorded were used to adjust coordinates that were otherwise based solely upon stereotaxy (by 0.5–2.0 mm). The position of the dorsal extent of the thalamus (with respect to the micromanipulator) was measured for each electrode penetration. The distance from the top of thalamus to the oculomotor nucleus was found in each monkey to be about 10 mm, whereas the atlas of Paxinos and colleagues (2000) indicates about 12 mm. The dorsal-ventral position of neurons in atlas coordinates was therefore estimated by rescaling based upon distance from the top of thalamus. Nonetheless, the error in the estimates of dorsal-ventral position is likely to be relatively high, which probably contributes to the apparent localization of presumed dopamine neurons in non-dopaminergic nuclei in Fig 3A.

Further evidence that the recorded neurons were dopamine neurons derives from the similarity of neuronal responses in the present study to those observed in previous studies of dopamine neurons. Monkey O in the present study was the same as monkey B in the study of Fiorillo et al. (2008). No histological results were obtained from this monkey. However, histological results were obtained from monkey A in the study by Fiorillo et al. (2008), which was identical to monkey A in the study by Fiorillo et al. (2003). In the supplemental materials of the latter, it is shown that the region of recorded neurons (marked with electrolytic lesions) matches the region of dopamine neurons in the ventral midbrain (as deter-

mined with cresyl violet and tyrosine hydroxylase staining) (Fig. S2 in the study by Fiorillo et al. (2003)). Fiorillo et al. (2008) recorded putative dopamine neurons both in that monkey, and in monkey O of the present study, in a variety of tasks involving temporal variations between conditioned stimuli and juice reward. The response properties of the neurons in each of the two monkeys were qualitatively identical (Fiorillo et al., 2008). Similarly, the responses of the recorded neurons to reward outcomes in the present study varied depending on the prediction of reward (Fig. 7), in precisely the same manner that has been observed in many previous studies of dopamine neurons (e.g. Ljungberg et al., 1992; Fiorillo et al., 2003; Bayer and Glimcher, 2005; Tobler et al., 2005; Bromberg-Martin and Hikosaka, 2009).

Data analysis

Data were analyzed using Matlab software (MathWorks, Natick, MA, USA). For each neuron, about 60 trials were recorded for each trial type (stimulus U and stimulus C). For statistical analyses and to make Figs 3C, 5B, and 6, firing rates were measured in a window of 120−400 ms after stimulus onset in each neuron. For statistical comparisons within a single neuron, firing rates across all trials following stimulus U were compared with rates following stimulus C using an unpaired t-test; $P<0.05$ was taken to be significant, without any correction for the fact that the same test was performed separately on all neurons. For comparisons across the whole population of recorded neurons, the mean firing rate across trials was calculated for each condition in each neuron, and these mean firing rates were then compared between conditions across the population of neurons using a paired t-test.

RESULTS

Behavior

Monkeys were conditioned with two visual stimuli, first in a choice task (Fig. 1A) and subsequently in a Pavlovian task. In each task, Stimulus C (for "certain") was followed by 0.125 ml of juice on every trial, and stimulus U (for "uncertain") was followed by 0.240 ml of juice on a pseudorandomly selected 50% of trials, and by no juice on the remaining trials. Thus the expected liquid volume associated with stimulus U (0.120 ml) was just slightly less (and slightly more delayed; see Experimental procedures) than that associated with stimulus C, but stimulus U was associated with a higher level of subjective uncertainty about reward outcome relative to stimulus C.

Three monkeys performed the choice task in which stimuli U and C were presented simultaneously. Eye position was monitored, and fixation on either of the visual stimuli resulted in delivery of its associated outcome. Each of the three monkeys preferred stimulus U. Preference for stimulus U appeared to be quite stable across trials within individual sessions. Figure 1B shows the full sequence of choices and their associated outcomes (juice or no juice) during a single session of 1000 trials (within a period of about 70 min), whereas Fig 1C shows the percentage of choices for stimulus U in the same session in bins of 10 and 50 trials. All three monkeys chose stimulus U on 61–98% of trials across all daily sessions (Fig. 1D). Thus the preference for stimulus U appeared to be present within each period of 50 trials (Fig. 1C), to be stable over the course of a day (Fig. 1C), and to be stable over the several weeks during which the experiments were per-
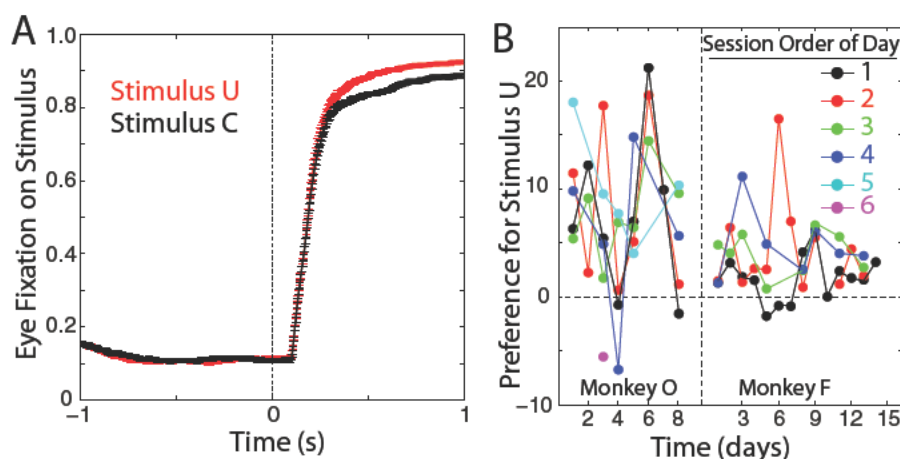
Fig. 2. Behavioral preference for stimulus U during Pavlovian conditioning. (A) Animals viewed stimulus U on a higher percentage of trials than stimulus C. Eye position was measured every 5 ms, and the fraction of trials (mean±SEM) for each 5 ms bin in which gaze was within±2 degrees from the center of the image was calculated for about 5000 trials of each type (collected across all sessions in which a neuron was recorded). Stimulus onset was at time "zero." (B) Preference for stimulus U over C persisted for multiple sessions within a day and across days. Each point represents the preference for stimulus U in a single session of about 60 trials of each type. The sessions shown are only those in which a dopamine neuron was simultaneously recorded (1–6 dopamine neurons per day). The preference was calculated as the average percentage of trials spent viewing stimulus U in the 0.5-s period before reward outcome (as shown in (A)) minus the average percentage of trials spent viewing stimulus C. In the first sessions of a day, monkeys usually viewed each stimulus on every trial, producing a "ceiling effect" that may have obscured any preference. No clear relationship was observed between this measure of viewing preference and neuronal preference (not shown).

formed (Fig. 1D). The preference was not related to the particular characteristics of the visual stimuli (see Experimental procedures).

To assess dopamine responses to each stimulus in isolation, monkeys O and F were subsequently trained in a Pavlovian version of the same task in which the same two stimuli were presented separately on pseudorandomly interleaved trials. During Pavlovian conditioning, monkeys spent more time viewing stimulus U than stimulus C (Fig. 2A) (mean±SEM of 91.0±0.4% of time for stimulus U versus 86.6±0.5 for stimulus C in the last 500 ms before delivery of reward outcome). The preference for stimulus U was quite consistent over days, and over multiple sessions in a single day (Fig. 2B). The greater viewing time for stimulus U is consistent with its greater reward value as measured in the choice task. Stronger orienting responses are known to accompany stimuli associated with greater reward uncertainty (e.g. Pearce and Hall, 1980).

Following experiments with Pavlovian conditioning, each monkey still preferred stimulus U in subsequent sessions of the choice task (Fig. 1D). Thus the preference for stimulus U persisted in each monkey even after experience of over 2000 choice trials and 2000 Pavlovian trials of each stimulus. The risk-seeking behavior observed here appears to be the same as that observed previously in rhesus macaques under laboratory conditions (McCoy and Platt, 2005; Hayden and Platt, 2007; Hayden et al., 2008; So and Stuphorn, 2010). In the studies of Platt and colleagues (McCoy and Platt, 2005; Hayden and Platt, 2007; Hayden et al., 2008), the programmed value of a stimulus was changed after a relatively small number of trials (50 trials in the case of McCoy and Platt (2005)). Thus the risk preference was observed in a relatively dynamically varying environment, and it was not clear whether the risk

preference would be maintained over an extended period of conditioning. Figs. 1, 2 demonstrate that risk preference is stable even when the same stimuli and probabilities are maintained over days and weeks of conditioning.

Responses of dopamine neurons

Extracellular electrophysiological recordings of single dopamine neurons were performed during the Pavlovian version of the task. The Pavlovian task allowed the responses of dopamine neurons to each stimulus (option) to be measured in isolation, as opposed to the simultaneous presentation of two stimuli in the choice task. In addition, responses of dopamine neurons in a Pavlovian task are easier to interpret than those in an instrumental task, since the experimenter fully controls the contingency between stimulus and reward.

Neurons were recorded in both medial and lateral regions of the ventral midbrain, corresponding to VTA and SN (Fig. 3A, B). Of 74 neurons recorded in the task, the firing rates of 61 neurons (82%) increased significantly (relative to baseline) following delivery of juice outside of the task (not shown), consistent with other studies of dopamine neurons. A somewhat lower percentage (~65%) was significantly activated within the task by stimuli U and C. Fig. 4 shows the responses of a single neuron that was more strongly activated by stimulus U than stimulus C. To quantify the difference in firing rates between the two conditions, firing rates in each neuron were measured within a window of 120−400 ms after stimulus onset. It was found that the total population of recorded dopamine neurons in each monkey (33 in monkey O and 41 in monkey F) showed a stronger activation to stimulus U than stimulus C (Fig. 5A) ($P<0.001$ in each monkey, paired t-tests across

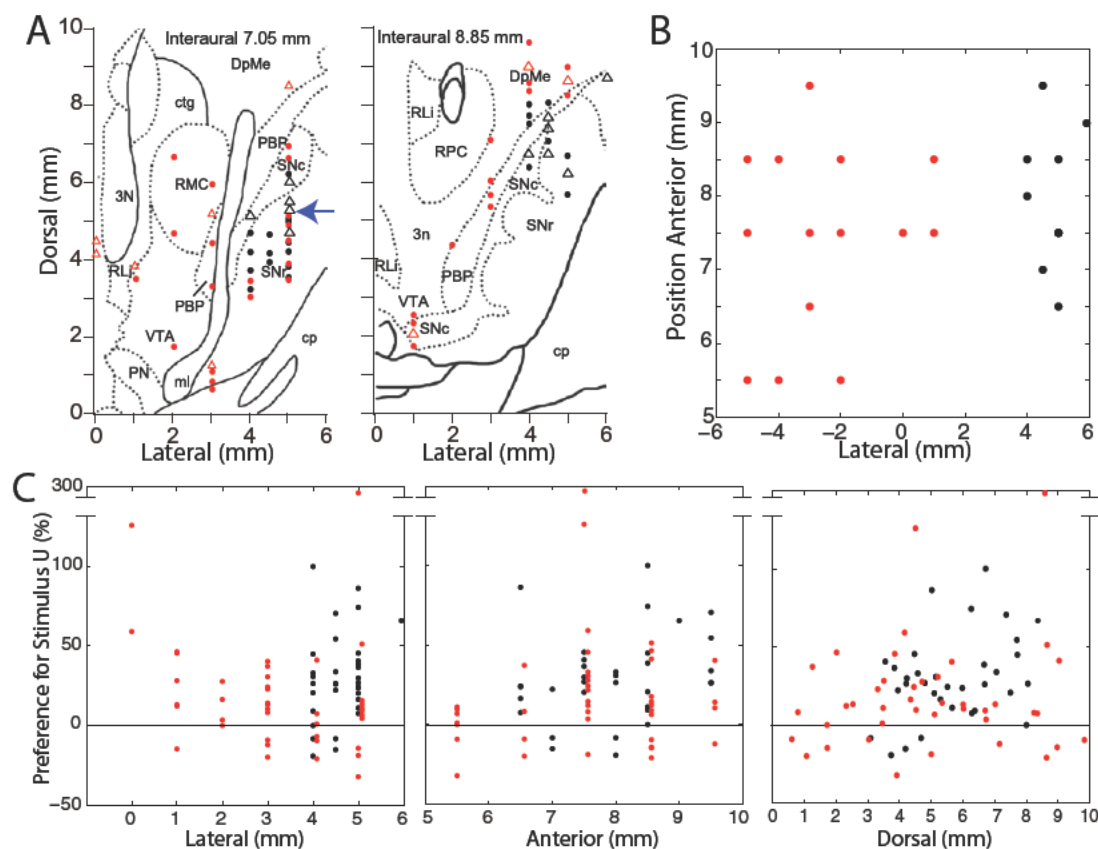C. D. Fiorillo / Neuroscience 197 (2011) 162–171

Fig. 3. Positions of recorded neurons in relation to neuronal risk preference. Monkey F in red, monkey O in black. (A) Estimated positions of neurons in drawings of coronal sections at 7.05 and 8.85 mm anterior of the interaural line (based upon the atlas of Paxinos and colleagues (2000)). Triangles indicate neurons that displayed statistically significant discrimination of stimulus U versus stimulus C. The blue arrow indicates the anatomical location of the neuron for which data are displayed in Fig 4. All neurons from each hemisphere are shown as though in one hemisphere. The more anterior section (8.85 mm) displays all neurons recorded between 8.0 and 9.5 mm anterior of the interaural line, whereas the posterior section (7.05 mm) includes all neurons recorded between 5.5 and 7.5 mm. According to the classification scheme depicted here, dopamine neurons would be expected to be found within VTA, parabrachial plexus (PBP), and paranigral nucleus (PN) (all classified as "A10" dopamine neurons), as well SN pars compacta (SNc) (A9). Some of the neurons recorded at more posterior levels (less than approximately 6 mm anterior to the interaural line) may have been in the retrorubral field (A8) (not shown here). Positions of neurons were estimated as described in Experimental procedures. The fact that some neurons appear in this figure to lie outside of dopaminergic regions is most likely due to errors in measuring the positions of neurons and in translating those positions to the atlas. The errors are likely to be greater in the dorsal-ventral than medial-lateral dimension. (B) The position of electrode tracks lateral of the midline and anterior of the interaural line. Several neurons were typically recorded on a single track in a single day (Fig. 2B). (C) Neuronal preferences for stimulus U over C as a function of neuronal position lateral from the midline (without regard to hemisphere) (left), anterior of the interaural line (middle), and dorsal of the interaural line (right). Neuronal preference was calculated as the percentage by which the response (firing rate) to stimulus U was greater than stimulus C (firing rate following onset of stimulus U minus stimulus C, divided by firing rate to stimulus C and multiplied by 100). Note that some neurons found to have very large preferences also had very low firing rates (spontaneously and in response to stimulus C; see Fig. 5B).

neurons). The difference in firing rates between stimuli U and C was significantly greater in monkey O than monkey F during the chosen temporal window ending at 400 ms ($P<0.05$, unpaired t-test), but the stronger activation to stimulus U appeared to extend beyond 400 ms in monkey F but not monkey O (Fig. 5A). Of the 74 neurons, 60 showed larger responses to stimulus U than to stimulus C, and 20 of these reached statistical significance (Fig. 5B) ($P<0.05$, unpaired t-tests across trials). No cells showed a significant difference in the opposite direction (Fig. 5B).

Inspection of scatter plots (Fig. 5B) provides no evidence of statistically discrete populations of dopamine neurons, consistent with past neuronal recordings in behaving animals that have also failed to identify discrete populations (e.g. Fig. 3D of Fiorillo et al., (2003)). Further-

more, there was no clear correlation across neurons between firing rates in response to conditioned stimuli and whether or not the firing rates displayed significant discrimination between stimulus U and stimulus C (Fig. 5B). Similarly, there was no clear correlation between the extent to which neurons preferred stimulus U over stimulus C (as calculated in the manner described in Fig. 3C) and the firing rate in response to unexpected juice delivery (not shown). The firing rate in response to unexpected juice delivery significantly exceeded baseline firing rate in 61 of the 74 neurons (82%), and 16 of these 61 neurons (26%) displayed a significant preference for stimulus U over stimulus C. Of the 13 of 74 neurons in which the response to unexpected juice did not reach statistical significance, 4 (31%) displayed a significant preference for stimulus U
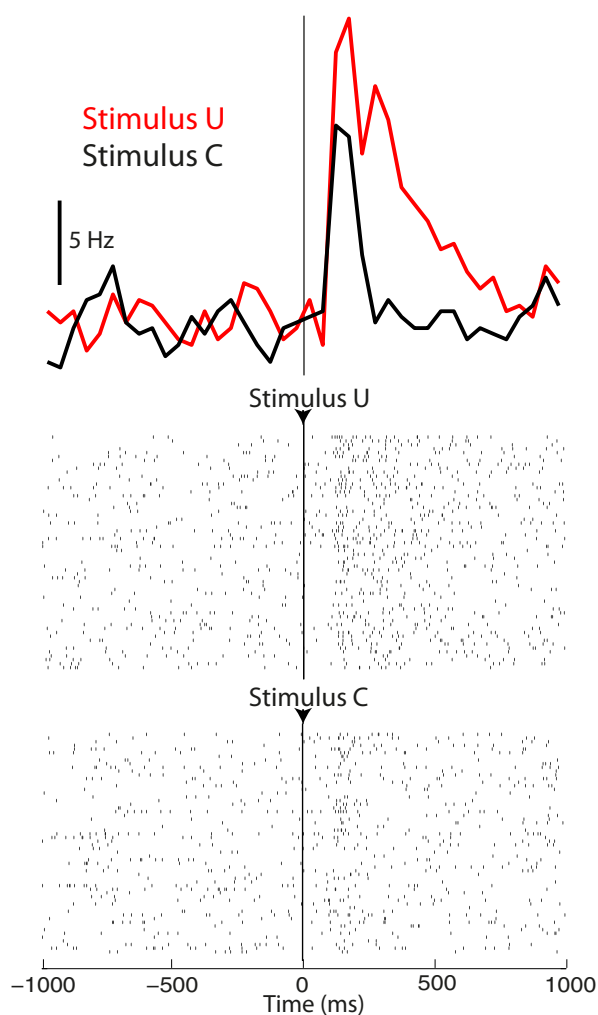
Fig. 4. Responses to stimuli U and C within a single dopamine neuron. Top, peristimulus time histograms, demonstrating a stronger mean response to stimulus U than to stimulus C. Bin size=50 ms. Bottom two panels, rasters of 60 trials each of stimulus U (middle) and stimulus C (bottom), arranged in each panel in chronological order from top to bottom.

over stimulus C. Thus there appears to be a single population of dopamine neurons, at least with respect to the present phenomenon, that is preferentially activated by stimulus U over stimulus C.

The activation by stimulus U was 23±4% (mean±SEM) (n=74) greater than the activation by stimulus C (the percentage increase in firing rate was calculated in each cell, and then averaged across cells). However, the magnitude of the difference should be interpreted with caution, since it is presumably highly sensitive to the animal's context-dependent expectation at the time of stimulus onset, and thus is likely to depend on factors such as the intertrial interval (Tobler et al., 2005).

The stronger responses to stimulus U did not appear to depend on the anatomical location of recorded neurons. It was observed in both hemispheres, and did not display a clear correlation with medial-lateral, anterior-posterior, or

dorsal-ventral coordinates (Fig. 3C). It was observed both in the more medial neurons of VTA as well as the more lateral neurons of SN.

The subjective reward value of stimulus U would be expected to vary depending on the outcomes of recent trials. If the responses of dopamine neurons to stimulus U varied considerably across trials depending on the recent reward history of stimulus U, then the average responses shown in Figs. 4 and 5 might be misleading (since the functional effect of dopamine may be a nonlinear function of firing rate). To examine this issue, responses to stimulus U were sorted depending on whether or not the preceding trial of stimulus U was or was not followed by juice reward. No difference was found (Fig. 6). In both cases, the firing rate across all 74 neurons was 5.6±0.5, with two neurons being significantly more activated by stimulus U when the preceding trial was rewarded, and three neurons showing the opposite preference (Fig. 6, right). Thus the responses of dopamine neurons to stimulus U did not depend substantially on the outcome of the preceding trial. Although subjective reward values in this sort of task fluctuate depending on recent reward history (McCoy and Platt, 2005), the extended period of training with stimulus U may have caused the monkeys' subjective valuation of stimulus U to have become quite stable, changing only very slightly from trial to trial according to reward outcomes.

Based on the results described above, it appears that stimulus U had greater subjective reward value than stimulus C (on average), as reflected in behavior as well as the activation of dopamine neurons at the time of stimulus onset. According to theories of dopamine function (e.g. Schultz et al., 1997), one factor that might influence the reward value of conditioned stimuli is the history of their association with dopamine release triggered by unconditioned stimuli. Thus the dopamine response to juice delivery (or omission) in the present study could have contributed to the learning of the values of stimuli U and C. It is therefore interesting to compare the neuronal responses at the time of juice delivery (or omission) following stimulus U versus stimulus C.

The delivery of well predicted juice (0.125 ml), following stimulus C, induced a small and transient activation of dopamine neurons (Fig. 7). Delivery of a larger volume of juice (0.240 ml) following stimulus U caused a larger activation, whereas the omission of juice following stimulus U caused a suppression of firing rate below baseline (Fig. 7). These results qualitatively match expectations based upon previous studies (Fiorillo et al., 2003; Tobler et al., 2005). It would be expected that the larger activation to juice following stimulus U than stimulus C was due primarily to the greater uncertainty in the prediction of juice, rather than the greater juice volume (Tobler et al., 2005). More critical to the present topic is whether the average activation at the time of the reward outcomes differed between stimuli U and C. For this comparison, firing rates were measured in a window of 80–300 ms following juice onset (or stimulus offset in the case of trials in which juice was omitted). This window was chosen to cover both the period of activation to juice delivery as well as inhibition to juice
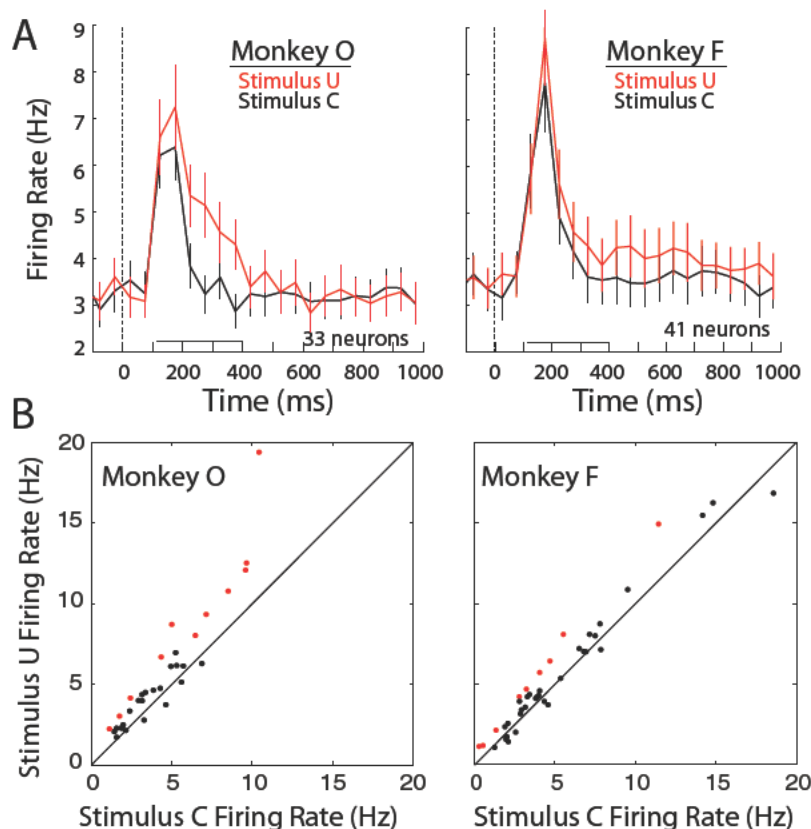
Fig. 5. Dopamine neurons are more strongly activated by a conditioned stimulus ("U") that predicts an uncertain reward outcome than by a stimulus ("C") that predicts a certain reward outcome. (A) Peristimulus time histograms of firing rate for the entire population of individually recorded dopamine neurons (mean±SEM) following onset of conditioned stimuli at time "zero" (stimulus U in red, stimulus C in black; 33 neurons in monkey O at left, 41 neurons in monkey F at right). To further characterize responses, firing rate was measured in each neuron in a window of 120–400 ms following stimulus onset, as indicated by the bars shown beneath the histograms. (B) Scatter plot of firing rates following stimuli U and C. Each point represents a single neuron. The 20 points in red represent neurons in which responses to stimulus U were significantly greater that to stimulus C ($P<0.05$, unpaired t-tests across trials).

omission (Fig. 7). The average firing rate across all neurons ($n=74$) was slightly but significantly greater in the

case of reward outcomes following stimulus U (4.81±0.43 Hz) than stimulus C (4.24±0.41 Hz) ($P=0.002$, paired
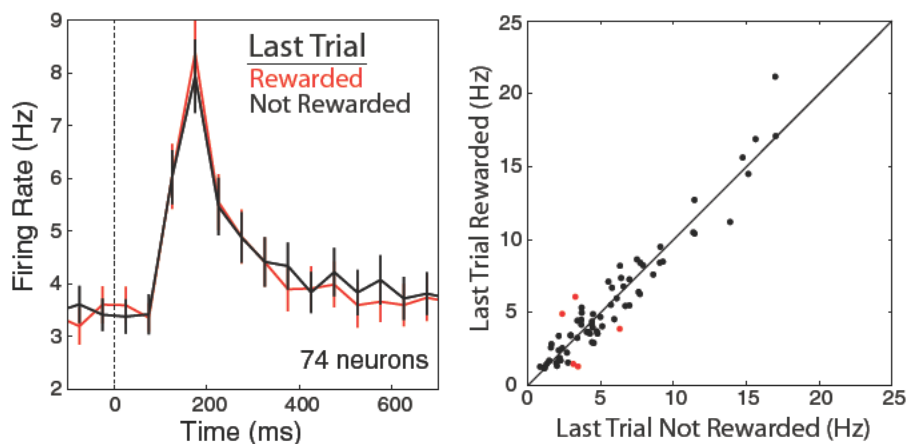

Fig. 6. The neuronal response to stimulus U did not depend on the reward outcome of the previous trial. Responses to stimulus U were segregated into two groups depending on whether or not the most recent presentation of stimulus U was or was not followed by juice reward. Left, peristimulus time histograms (stimulus U in red, stimulus C in black). Stimulus onset occurred at time "zero." Right, scatter plot of responses in individual neurons. The five points in red signify neurons in which responses were significantly larger, or smaller, depending on the outcome of the previous trial (without correction for multiple comparisons).
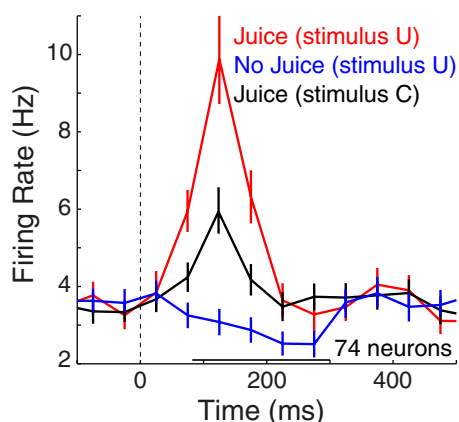
Fig. 7. Average neuronal responses to reward outcomes. Juice onset (or offset of stimulus U on trials in which juice was omitted) occurred at time "zero." Consistent with previous studies, when the delivery of juice is uncertain following stimulus U, its delivery causes a strong activation (red), whereas its omission causes a suppression of firing rate below the baseline level (blue). When the delivery of juice is nearly certain following stimulus C, its delivery causes only a small activation (black). The black horizontal line between 80 and 300 ms indicates the period in which firing rates were measured for comparison, as described in the results.

t-test). Among individual neurons, 16 of 74 neurons had a significantly higher firing rate in response to the reward outcomes following stimulus U versus C, and 3 of 74 showed the opposite relationship ($P<0.05$; unpaired *t*-tests across trials). Of the 16 of 74 neurons that had significantly higher firing rates during the reward outcome period following stimulus U, only four neurons also showed significantly greater activation in response to stimulus U versus stimulus C immediately following stimulus onset. Thus there was no apparent relationship across neurons between preferential responding for stimulus U at the time of conditioned stimulus onset, and preferential responding at the time of reward outcome. However, these measures of average responding at the time of reward outcome should be interpreted with caution, since the effects of changes in dopamine concentration in terminal regions are likely to be a highly nonlinear function of firing rate. Thus it may be misleading to average small decreases in firing rate following omission of reward with large increases following delivery of reward.

The analyses presented above concern the brief, "phasic" activation of dopamine neurons following shortly after the onset of stimuli. It was previously found that the firing rate of primate dopamine neurons is increased near the end of a 2-s delay period between onset of Pavlovian conditioned stimuli and juice delivery when there is uncertainty about the amount of juice to be delivered (Fiorillo et al., 2003). It would therefore be expected that in the present study the firing rate at the end of the 1-s delay period would be greater following stimulus U than following stimulus C. This was found to be true across the population of neurons in monkey F ($P=0.003$, paired *t*-test), but not in monkey O ($P=0.44$), for the period of 0.8–1.0 s after stimulus onset. Detailed analysis of neural activity during

the delay period is planned for presentation in a later paper.

DISCUSSION

Although animals avoid variability in reward outcomes under some conditions, monkeys were risk seeking in the present and previous studies (McCoy and Platt, 2005; Hayden and Platt, 2007; Hayden et al., 2008; So and Stuphorn, 2010). The transient responses of dopamine neurons corresponded to the behavioral preference of the animals, with stronger activation by the stimulus associated with greater uncertainty about reward outcome. It is widely believed that this transient dopamine activation drives positive reinforcement of immediately preceding stimuli and actions (e.g. Schultz et al., 1997). If that is correct, then the present results provide strong though indirect evidence that dopamine would reinforce risk-seeking behavior in a context similar to the present experiments, at least in these subjects. We cannot say whether dopamine contributed to causing the observed risk preference, and the conditioned stimuli that were studied here were not tested as reinforcers (through higher-order conditioning). However, we would expect that any stimuli or actions that were to precede the larger dopamine response to the "risky" stimulus would gain in value to a greater extent than any stimuli or actions that were to precede the smaller dopamine response to the "safe" stimulus. Thus dopamine would reinforce risk seeking. Since the task studied here resembles conditions in a casino, the present results support the suggestion that dopamine promotes gambling behavior (e.g. Ambermoon et al., 2011; Fiorillo et al., 2003).

The values that contribute directly to decision-making are necessarily subjective (without any connotation of consciousness), insofar as they are directly conditional upon internal neuronal states rather than external stimuli. By manipulating the mean and variance of liquid volume, we may influence a monkey's subjective expected reward value and uncertainty, respectively. However, if variance is observed to influence behavior or neuronal responses, this is not necessarily through an effect on subjective uncertainty. Rather, variance could influence expected value. Indeed, in their famous analysis of "decision under risk," Kahneman and Tversky (1979) explained risk aversion in this manner. The potential for variance to influence expected reward value can result from a nonlinear relationship between physical quantities (e.g. liquid volume) and subjective reward value. A nonlinearity could result from distorted (nonlinear) perception of liquid volume (e.g. Kacelnik and Brito e Abreu, 1998), or from a nonlinear relationship between perceived liquid volume and its subjective value (or utility) to the animal (a nonlinear "value function") (e.g. Kahneman and Tversky, 1979).

In contrast to the concave relationships proposed to explain risk aversion (Kahneman and Tversky, 1979; Kacelnik and Brito e Abreu, 1998), explanations of risk seeking that do not involve subjective reward uncertainty require that subjective reward value is a convex function of

liquid volume, such that doubling volume more than doubles value. Hayden and colleagues (2008) have provided evidence that a nonlinear value function is unable to explain risk seeking in monkeys under conditions similar to those studied here. They argued instead in favor of an explanation that depends on distorted perception, in which more attention is paid to larger reward outcomes. According to this view, the animal essentially calculates the expected volume (and hence, expected reward value) by giving greater weight to the large volume and lesser weight to the small volume, so that perceived volume is a convex function of actual volume and the monkey overestimates the value of the risky option. In this way, variance in reward outcomes could explain risk seeking without reference to subjective reward uncertainty. An alternative explanation of risk seeking is that subjective uncertainty about reward has reward value (drives positive reinforcement) in and of itself.

Regardless of whether risk seeking is caused by the reward value of subjective reward uncertainty, or by an inflated expected reward value, it has the beneficial effect of promoting exploration and learning. A particularly interesting aspect of the present results is their relevance to the distinction between exploration and exploitation value. The "exploitation value" of a stimulus or action refers to the value of primary rewards (e.g. water) that it predicts. It is essentially what people usually mean when they refer simply to "reward value." The goal of animals should be to maximize exploitation value in the long term. However, if animals were to always choose the stimulus or action associated with the greatest exploitation value, they would forego the opportunity to explore and thereby acquire new information. Thus they would fail to identify "new" stimuli and actions associated with larger rewards (greater exploitation value). In making decisions, it is best to choose the option with the greatest total reward value, corresponding to the sum of exploration and exploitation values. In principle, exploitation value should correspond to expected reward value, and exploration value should correspond to reward uncertainty. However, as implied above, a certain nonlinear relationship between subjective expected reward value and "true value" might be effective in promoting exploration, even without any influence of subjective reward uncertainty.

Although some have suggested that dopamine may represent (and reinforce) total reward value (Kakade and Dayan, 2002; Fiorillo et al., 2003), others have speculated that dopamine represents only exploitation value and that neurons in other brain regions may represent exploration value (Daw et al., 2006; McClure et al., 2006; Frank et al., 2009). The dopamine responses shown here appear to correspond to total reward value, since they are larger for the stimulus that has greater exploration value and is preferred in a choice task. Other studies have also found that the activity of dopamine neurons corresponds more closely to subjective reward value and choice behavior rather than liquid volume, which is more consistent with a representation of total reward value rather than solely exploitation value (Morris et al., 2006; Bromberg-Martin and

Hikosaka, 2009). The general hypothesis that dopamine neurons represent total reward value is also supported by the finding that dopamine neurons are transiently activated by novel sensory stimuli (Ljungberg et al., 1992), which have been shown to have reward value in a laboratory setting (Blatter and Schultz, 2006). Novel stimuli are inherently associated with subjective reward uncertainty and therefore have exploration value (Kakade and Dayan, 2002). In addition, activation of dopamine neurons by exploration value can explain the observation that "dopamine neurons signal preference for advance information about upcoming rewards" (Bromberg-Martin and Hikosaka, 2009). The latter study can be understood as a variant of the present study, but employing higher-order conditioning. Exploration value derives from the value of reward information in general, and the value of "advance information" reflects the fact that it is better for information to come sooner rather than later (this can be understood by applying the concept of temporal discounting to exploration value).

The value of exploration may explain risk seeking as an adaptive behavior. Natural environments are rich in information, and thus information that an animal lacks is likely to be present in the environment in the form of an as yet unidentified stimulus or action that, once discovered, could be used to better predict and maximize reward. It is rarely if ever the case that subjective uncertainty derives from an inherently random process in nature. The present experiments are unnatural and atypical because, as in a casino, the probabilities are essentially fixed and subjective uncertainty cannot be further reduced through exploration and learning. Risk-seeking behavior in this context is not useful. Nonetheless, it can be understood as rational given an animal's knowledge of the natural environment (and its ignorance of the artificial environment), much as sensory illusions have come to be understood as rational inferences based upon a person's information about natural statistical patterns (e.g. Weiss et al., 2002; Yang and Purves, 2003; Niemeier et al., 2003). Just as an artificial sweetener "tricks" the brain into the unconscious belief that it will soon receive the exploitation value of valuable calories, a gamble in a casino may be unconsciously driven by the illusory promise of the exploration value of reward information.

REFERENCES

Ambermoon P, Carter A, Hall WD, Dissanayaka NN, O'Sullivan JD (2011) Impulse control disorders in patients with Parkinson's disease receiving dopamine replacement therapy: evidence and implications for the addictions field. Addiction 106:283–293.

COM.0025.0001.0153

Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47:129–141.

Blatter K, Schultz W (2006) Rewarding properties of visual stimuli. Exp Brain Res 168:541–546.

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

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. Nature 441:876–879.

Fiorillo CD, Newsome WT, Schultz W (2008) The temporal precision of reward prediction in dopamine neurons. Nat Neurosci 11:966–973.

Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. Science 299:1898–1902.

Frank MJ, Doll BB, Oas-Terpstra J, Moreno F (2009) Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. Nat Neurosci 12:1062–1068.

Hayden BY, Heilbronner SR, Nair AC, Platt ML (2008) Cognitive influences on risk-seeking by rhesus macaques. Judgm Decis Mak 3:389–395.

Hayden BY, Platt ML (2007) Temporal discounting predicts risk sensitivity in rhesus macaques. Curr Biol 17:49–53.

Kacelnik A, Brito e Abreu F (1998) Risky choice and Weber's Law. J Theor Biol 194:289–298.

Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. Econometrica 47:263–292.

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

Ljungberg T, Apicella P, Schultz W (1992) Responses of monkey dopamine neurons during learning of behavioral reactions. J Neurophysiol 67:145–163.

McClure SM, Gilzenrat MS, Cohen JD (2006) An exploration-exploitation model based on norepinepherine and dopamine activity. In: Advances in Neural Information Processing Systems, Vol. 18 (Weiss Y, Sholkopf B, Platt J, eds), pp 867–874. Cambridge: MIT Press.

McCoy AN, Platt ML (2005) Risk-sensitive neurons in macaque posterior cingulate cortex. Nat Neurosci 8:1220–1227.

Merrill EG, Ainsworth A (1972) Glass-coated platinum-plated tungsten microelectrodes. Med Biol Eng 10:662–672.

Morris G, Nevet A, Arkadir D, Vaadia E, Bergmann H (2006) Midbrain dopamine neurons encode decisions for future action. Nat Neurosci 9:1057–1063.

Niemeier M, Crawford JD, Tweed D (2003) Optimal transsaccadic integration explains distorted spatial perception. Nature 422:76–80.

Paxinos G, Huang XF, Toga AW (2000) The Rhesus Monkey Brain in Stereotaxic Coordinates. San Diego: Academic Press.

Pearce JM, Hall G (1980) A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol Rev 87:532–552.

Roesch MR, Calu DJ, Schoenbaum G (2007) Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nat Neurosci 10:1615–1624.

Schultz W, Dayan P, Montague RR (1997) A neural substrate of prediction and reward. Science 275:1593–1599.

Schultz W, Romo R (1987) Responses of nigrostriatal dopamine neurons to high-intensity somatosensory stimulation in the anesthetized monkey. J Neurophysiol 57:201–217.

So NY, Stuphorn V (2010) Supplementary eye field encodes option and action value for saccades with variable reward. J Neurophysiol 104:2634–2653.

Tobler PN, Fiorillo CD, Schultz W (2005) Adaptive coding of reward value by dopamine neurons. Science 307:1642–1645.

Weiss Y, Simoncelli EP, Adelson EH (2002) Motion illusions as optimal percepts. Nat Neurosci 5:598–604.

Wise RA (2004) Dopamine, learning and motivation. Nat Rev Neurosci 5:483–494.

Yang Z, Purves D (2003) A statistical explanation of visual space. Nat Neurosci 6:632–640.