
PERGAMON

Neural Networks 15 (2002) 603–616

Neural Networks

www.elsevier.com/locate/neunet

2002 Special issue

Opponent interactions between serotonin and dopamine

Nathaniel D. Daw[a,*], Sham Kakade[b], Peter Dayan[b]

[a]*Computer Science Department and Center for the Neural Basis of Cognition, School of Computer Science, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213-3891, USA*

[b]*Gatsby Computational Neuroscience Unit, University College London, London, UK*

Received 5 October 2001; revised 2 April 2002; accepted 2 April 2002

Abstract

Anatomical and pharmacological evidence suggests that the dorsal raphe serotonin system and the ventral tegmental and substantia nigra dopamine system may act as mutual opponents. In the light of the temporal difference model of the involvement of the dopamine system in reward learning, we consider three aspects of motivational opponency involving dopamine and serotonin. We suggest that a tonic serotonergic signal reports the long-run average reward rate as part of an average-case reinforcement learning model; that a tonic dopaminergic signal reports the long-run average punishment rate in a similar context; and finally speculate that a phasic serotonin signal might report an ongoing prediction error for future punishment.

Keywords: Serotonin; Dopamine; Opponency; Reinforcement learning; Punishment; Reward; Solomon-Corbit; Aversion

1. Introduction

From a computational perspective, serotonin (5HT) is the most mysterious of the main vertebrate neuromodulators. Pharmacological investigations reveal that it plays a role in a wide variety of phenomena, including impulsivity, obsessionality, aggression, psychomotor inhibition, latent inhibition, analgesia, hallucinations, eating disorders, attention and mood (Aghajanian & Marek, 1999; Buhot, 1997; De Vry & Schreiber, 2000; Edwards & Kravitz, 1997; Fields, Heinricher, & Mason, 1991; Harrison, Everitt, & Robbins, 1997, 1999; Hollander, 1998; Lesch & Merschdorf, 2000; Masand & Gupta, 1999; Solomon, Nichols, Kiernan, Kamer, & Kaplan, 1980; Soubrié, 1986; Stahl, 2000; Stanford, 1999; Westenberg, den Boer, & Murphy, 1996). However, there are many complexities in these effects. For instance, drugs that take immediate and selective effect on inhibiting serotonin reuptake, and so boost its neural longevity, take two weeks to have an effect on mood. Also, electrophysiological data (Gao, Chen, Genzen, & Mason, 1998; Gao, Kim, & Mason, 1997; Jacobs & Fornal, 1997, 1999) show that serotonin cells do not obviously alter their firing rates in response to the sort of significant stimuli that might be expected to control some of the behaviors described earlier. Thus, the experimental data on the involvement of serotonin are confusing, and this has inevitably impeded the development of computational theory.

In this paper, we focus on one important (though emphatically *not* exclusive) aspect of serotonin suggested by anatomical and pharmacological data, namely an apparent opponent partnership with dopamine (DA, Azmitia, 1978; Azmitia & Segal, 1978; Deakin, 1983, 1996; Fletcher, 1991, 1995; Fletcher & Korth, 1999; Fletcher, Korth, & Chambers, 1999; Kapur & Remington, 1996; Vertes, 1991). Substantial evidence supports the theory that phasic activity of dopamine cells in the ventral tegmental area and substantia nigra pars compacta reports a prediction error for summed future reward (Montague, Dayan, & Sejnowski, 1996; Schultz, 1998; Schultz, Dayan, & Montague, 1997) in the context of a temporal difference (TD) model (Sutton, 1988; Sutton & Barto, 1990) of reinforcement learning (Bertsekas & Tsitsiklis, 1996; Sutton & Barto, 1998). To the extent that serotonin acts as an opponent to dopamine, we can use our understanding of the role of dopamine to help constrain aspects of the role of serotonin. Equally, the TD model of dopamine is based on experiments that only probe a small part of the overall scope of reinforcement learning. Extending the model to cope with theoretical issues such as long-run average rewards (Daw & Touretzky, 2000, 2002) actually leads to the requirement for a signal that acts like an opponent to dopamine. Here, we explore this candidate role for serotonin.

* Corresponding author. Tel.: +1-412-268-2582; fax: +1-412-268-3608.
E-mail address: daw@cs.cmu.edu (N.D. Daw).

0893-6080/02/$ - see front matter
PII: S0893-6080(02)00052-7

Opponency has a venerable history in psychology and neuroscience. In an implementational form (e.g. Grossberg, 1988), it starts from the simple idea of using two systems to code for events (such as affective events), with one system reporting positive excursions from a baseline (appetitive events), the other system reporting negative excursions (aversive events), and mutual inhibition between the systems and/or opposing effects on common outputs. From a physiological perspective, this neatly circumvents the absence of negative firing rates. However, opponency turns out to have some less obvious mathematical and computational properties, which have been given diverse interpretations in everything from affective systems to circadian timing mechanisms (Grossberg, 1984, 2000).

In this paper, we focus on motivational opponency between appetitive and aversive systems. In modeling conditioning, reinforcement learning has largely focused on formalizing a notion of the affective value of stimuli, in terms of the future rewards and punishments their presence implies. Psychologically, this notion of affective value is best thought of as a form of motivational value, and motivational opponency has itself been the focus of substantial experimental study. For instance, following Konorski (1967), Dickinson and Dearing (1979) and Dickinson and Balleine (2002) review evidence (such as transreinforcer blocking, Ganesan & Pearce, 1988) suggesting it is psychologically reasonable to consider just one appetitive and one aversive motivational system, and not either multiple appetitive and multiple aversive systems or just one single combined system. These two systems are motivational opponents; they also have opposing preparatory behavioral effects, with the appetitive system inducing Pavlovian approach, and the aversive system withdrawal. The reinforcement learning model of dopaminergic activity identifies it as the crucial substrate of the appetitive motivational system; here, following, amongst others, Deakin and Graeff (1991), we model serotonergic activity as a crucial substrate of the aversive motivational system.

Psychological and implementational aspects of opponency have been much, and sometimes confusingly, debated. Psychologically, two main forms of opponency have been considered, one associated with the punctate presentation of conditioned and unconditioned stimuli, the other associated with their long-term delivery. For the former, rewarding unconditioned stimuli are assumed able to excite the appetitive system, as are conditioned stimuli associated with reward. Punishing unconditioned stimuli are assumed to excite the aversive system, as are conditioned stimuli associated with punishment. The inhibitory interaction between the two systems can have various consequences. For instance, extinguishing an appetitive conditioned stimulus could equally result from reducing its ability to drive the appetitive motivational system (passive extinction), or increasing its ability to drive the aversive motivational system (active extinction), or both (see, for example, Osgood, 1953). These possibilities have different experimental implications. Another example is that if a conditioned inhibitor for reward acts by exciting the aversive system, then it should be able to block (Kamin, 1969) learning of a conditioned predictor of shock (Dickinson & Dearing, 1979; Goodman & Fowler, 1983), since it will predict away the activation of the aversive motivational system.

Solomon and Corbit (1974) considered an apparently different and dynamic aspect of opponency in the case that one or both of the appetitive or aversive systems are excited for a substantial time. Stopping the delivery of a long sequence of unexpected rewards is aversive (perhaps characterized by frustration); stopping the delivery of a long sequence of unexpected punishments is appetitive (perhaps characterized by relief).

We seek to model both short- and long-term aspects of opponency. One way to proceed would be to build a phenomenological model, such as Solomon and Corbit's (1974), or a mechanistic one, such as Grossberg's (2000) and Grossberg and Schmajuk's (1987). Solomon and Corbit's model (1974) suggests that the long-term delivery of appetitive unconditioned stimuli excites the aversive opponent system at a slower timescale. When the unconditioned stimuli are removed, the opponent system is also slower to lose excitation, and can thus be motivationally dominant for a short while. Grossberg and his colleagues (e.g. Grossberg, 1984, 1988, 2000; Grossberg & Schmajuk, 1987), have extensively discussed an alternative mechanism for this (involving slow adaptation within the system that reports the original unconditioned stimulus rather than the slow build up of the opponent), and have shown how the rich internal dynamics that opponent systems exhibit might themselves be responsible for many otherwise puzzling phenomena.

By contrast with, though not necessarily in contradiction to, these proposals, we seek a computational account. We start by considering long-term aspects, arguing that opponency emerges naturally (Daw & Touretzky, 2000, 2002) from TD learning in the case of predicting long-run average rewards rather than summed future rewards (Mahadevan, 1996; Puterman, 1994; Schwartz, 1993; Tadepalli & Ok, 1998). As will become apparent, this form of TD learning embodies a natural opponency between the existing phasic dopamine signal, and a newly suggested, tonic, signal, which we identify with serotonin. We extend the scope of the model to predictions of summed future punishment, and thereby postulate mirror opponency, between a tonic dopamine signal and a phasic serotonin signal. Short-term aspects of opponency then arise through consideration of the ways that the predictions of future reward and punishment might be represented.

In Section 2, we discuss the various aspects of the data on serotonin that have led us to consider it as being involved in aversive processing in general, and as an opponent to dopamine in particular. Section 3 covers the theoretical background to the TD learning model and the resulting link

to short- and long-term aspects of opponency; Section 4 discusses long-term aspects of opponency; and Section 5 considers the consequences if serotonin exactly mirrors dopamine. The discussion ties together the various strands of our argument.

2. Serotonin in conditioning

As suggested by the vast range of its effects listed earlier, serotonin plays an extremely complicated set of roles in the brain, roles that it is impossible at present to encompass within a single theory. Compared with dopamine, it is anatomically more widespread and behaviorally much more diverse. Further, although the activity of serotonin cells has not been systematically tested in the range of conditioning tasks that has been used to probe dopamine cells (Jacobs & Fornal, 1997, 1999; Schultz, 1998), in those studies that have been performed, it has been hard to find clear correlates for anything other than very general aspects of motor arousal.

Nevertheless, based largely on pharmacological investigations, there have been some valiant attempts to suggest general theories for some aspects of serotonergic functioning. Two that have substantial currency are its involvement with behavioral inhibition, which is discussed extensively and refined by Sourbrié (1986), and its involvement in aversion and punishment, an old idea that has been cogently formulated by Deakin (1983) and Deakin and Graeff (1991). These aspects of serotonin are not strongly in opposition to each other—most of the studies involving behavioral inhibition crucially involve either aversive events such as shocks (e.g. for animals pressing a lever to receive a reward, withholding responding during intervals in which a signal indicates that pressing the lever will also lead to a shock, Geller & Seifter, 1960) or differential reinforcement of low rates of behavior. Importantly, both the aspects of serotonin are in opposition to dopamine, which is involved in approach responses (Everitt et al., 1999; Ikemoto & Panksepp, 1999), has a psychomotor arousing influence (Canales & Iversen, 2000; Waddington, 1989), and, as discussed earlier, is associated with reward processing (Schultz, 1998).

The forebrain serotonin system consists of two nuclei, the dorsal and median raphe nuclei. These nuclei have slightly different anatomical targets (Azmitia, 1978; Azmitia & Segal, 1978; Vertes, 1991; Vertes, Fortin, & Crane, 1999), with the dorsal raphe making connections to those areas also innervated by the dopamine system (such as the amygdala and the striatum), and the median raphe making connections to the hippocampus and septal nuclei, which are not major dopaminergic targets. There are many different types and sub-types of serotonin receptors (e.g. Martin, Eglen, Hoyer, Hamblin, & Yocca, 1998), each with its own particular geographical distribution. The compounds used to probe the workings of the serotonin system (e.g. Stahl, 2000) are either antagonists or agonists of these receptors, which will typically affect different receptors to different degrees, or reuptake inhibitors, which exert a more global influence. Many such pharmacological agents also affect other neuromodulatory systems to some degree. The lack of pharmacological precision is a major hurdle to analyzing the various sub-parts of the serotonin system.

Deakin (1983) used the anatomical separation of dorsal and median nuclei as part of the motivation for his suggestion that the dorsal raphe is the system that opposes dopamine. More broadly, his suggestion is that serotonin is a critical part of the defensive system, which triggers fight/flight responses, and is in general concerned with adaptive responses to aversive events. Dopamine is assumed to promote appetitive behaviors such as approach (Ikemoto & Panksepp, 1999; Panksepp, 1998); the dorsal raphe serotonin projections would oppose these actions and mediate avoidance behavior elicited by aversive incentive stimuli. Deakin (1983) and Deakin and Graeff (1991) suggest that the balance between executing approach and withdrawal or behavioral inhibition is determined by the balance between dopamine and serotonin release in the ventral striatum.

A good deal of the evidence for dopamine/serotonin opponency is indirect, from pharmacological studies showing that dopamine is involved in activating behaviors that serotonin inhibits and vice-versa. Early studies reported a striking similarity between serotonin depletion and amphetamine administration (Lucki & Harvey, 1979; Segal, 1976); other studies have considered reciprocal interactions between dopamine and serotonin (Pucilowski, 1987). More recently, Fletcher and his colleagues have conducted a series of studies into the effects of agonists and antagonists of dopamine and serotonin on unconditioned behaviors such as feeding and conditioned behaviors such as responding for conditioned reward, self-stimulation and conditioned place preferences (Fletcher & Korth, 1999; Fletcher, Ming, & Higgins, 1993; Fletcher, Tampakeras, & Yeomans, 1995). They broadly show that agonizing serotonin opposes conditioned and unconditioned behaviors that are activated by dopamine; agonizing dopamine or antagonizing serotonin has the opposite effect.

These studies are buttressed by some more direct evidence for the opponent interaction of dopamine and serotonin. A range of reports (reviewed, for instance, by Kapur and Remington, 1996) confirms that serotonin antagonizes dopamine function both at the level of the VTA and the substantia nigra and at the terminal sites of the dopamine neurons such as the nucleus accumbens and the striatum. A study by Lorrain, Riolo, Matuszewich, and Hull (1999) combined microdialysis and serotonin administration to show that such an inhibitory action of serotonin on nucleus accumbens dopamine promotes sexual satiety, whereas sexual activity is correlated with an increase in dopamine. Evidence for one mechanism of serotonin's inhibitory effect on dopamine comes from an

electrophysiological study by Jones and Kauer (1999) showing that the excitatory glutamatergic synaptic transmission onto ventral tegmental area neurons is depressed through the activation of serotonin receptors.

It should be noted that although there is substantial suggestive evidence for opponency between dopamine and serotonin, the precise pattern of interactions is far from clear. First, there appears to be much less evidence of the inhibitory action of dopamine on the serotonin system than vice-versa. Second, although Deakin (1983) and Deakin and Graeff (1991) identify the dorsal raphe nucleus as being responsible for the serotonin system that opposes dopamine, data from Fletcher (1995) suggest that median raphe nucleus serotonin also plays an important role, and Imai, Steindler, and Kitai (1986) argue that the distinction between dorsal and median raphe nuclei might not be relevant in any case. Further, it is unlikely that the existing experimental data give a full picture of the interactions between all the different dopamine and serotonin receptor types and subtypes, and particularly the short-, medium-, and long-term dynamics of these interactions.

3. Dopamine and temporal difference learning

Electrophysiological data on the activity of dopamine neurons suggest that they report a TD prediction error for predictions of long-run rewards (Montague et al., 1996; Schultz et al., 1997). The TD learning algorithm (Bertsekas & Tsitsiklis, 1996; Sutton, 1988; Sutton & Barto, 1998) uses samples of received rewards to learn a *value function*, which maps information about the current state of the world (i.e. the current stimuli) to a prediction of the rewards expected in the future. Value functions underlie models of psychological phenomena such as secondary conditioning, and are also useful in instrumental contexts in which actions must be selected to optimize long-run rewards.

In the standard version of TD learning, training is divided into a series of trials, and the value function is defined as the sum of rewards $r(t)$ expected during the remainder of the current trial:

$$V(t) = E\left[\sum_{\tau \geq t}^{\text{trial}} r(\tau)\right] \tag{1}$$

$$V(t) = E[r(t)] + V(t+1) \tag{2}$$

Here the expectation $E[\cdot]$ is over randomness in the choice of actions and delivery of rewards. Eq. (2) follows by a simple recursion.

The TD algorithm is a method for learning an approximation $\hat{V}(t)$ to $V(t)$. TD uses a prediction error signal to improve the estimate in an incremental manner. The TD prediction error $\delta(t)$ is the difference between the two sides of Eq. (2) using the approximated value $\hat{V}(t+1)$ as an estimate of $V(t+1) = E[\sum_{\tau \geq t+1}^{\text{trial}} r(\tau)$. Equivalently, $\delta(t)$ is a measure of the inconsistency between $\hat{V}(t)$ and $\hat{V}(t+1)$ in light of the observed reward $r(t)$.

$$\delta_{\text{p}}(t) = r(t) + \hat{V}(t+1) - \hat{V}(t) \tag{3}$$

We refer to this signal as δ_{p}, the *phasic* component of the error signal, to distinguish it from tonic components we introduce in Section 4. The phasic responses of primate dopamine neurons appear to report $\delta_{\text{p}}(t)$, as the computational signal reproduces several types of burst and pause responses to unexpected events recorded during appetitive conditioning (Schultz, 1998; Schultz et al., 1997; Waelti, Dickinson, & Schultz, 2001). Examples of such modeled phasic characteristics can be seen in the dopamine sections of Figs. 1 and 2. The models assume the neuronal firing rate reflects the TD error plus some constant background firing rate, so that negative error produces a pause in neuronal firing. This device, in itself, does not provide a practical solution to the problem of communicating negative values with a firing rate, since the baseline is extremely weak, and there is no evidence that downstream areas can detect or act on such pauses.

Modulo the constant baseline, which we will omit from all equations, these models strictly identify the rate of dopamine firing with the TD error signal:

$$\delta_{\text{DA}}(t) = \delta_{\text{p}}(t) \tag{4}$$

Here we discuss additional components to the TD error signal that are required to handle long-term predictions and aversive stimuli. We propose that several of these largely negative additions are reported by serotonin, so that the full, augmented error signal $\delta(t)$ is shared between opposing dopaminergic and serotonergic components:

$$\delta(t) = \delta_{\text{DA}}(t) - \delta_{\text{5HT}}(t) \tag{5}$$

The two channels may very well be scaled differently, as well as offset by their own baselines.

4. Long-term opponency

The assumption made in the standard TD model that events are episodic, coming in separated trials, is clearly unrealistic, since most events, even in the context of behavioral experiments, are really ongoing. Treating them as such requires using a different notion of value; in particular, Eq. (1) must be replaced with a return that is not truncated at the end of each trial. Theoretical treatments of this case avoid the possibility of divergence by considering either *discounted values*, in which a reward is worth less the further in the future it is expected, or, to avoid various unsatisfactory aspects of discounting (Mahadevan, 1996), *differential values*, the summed differences between observed rewards and some expected baseline reward $\bar{r}$:

$$V(t) = E\left[\sum_{\tau \geq t}^{\infty} (r(\tau) - \bar{r})\right] \tag{6}$$

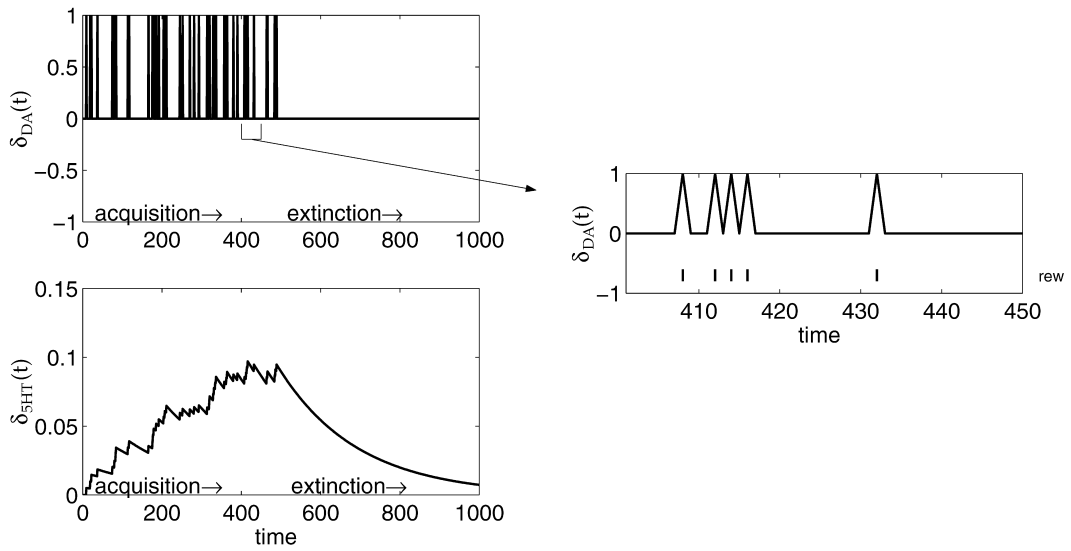

Fig. 1. Unsignaled rewards. (Left) Modeled opposing phasic dopamine (upper) and tonic serotonin (lower) responses during the presence (acquisition: first 500 timesteps) and subsequent absence (extinction: last 500 timesteps) of randomly delivered, unsignaled, rewards. (Right) The dopamine signal shown using an expanded time-scale. In this and subsequent plots, the right hand axis indicates the nature of salient events in the experiment (here, the delivery of reward) whose times of occurrence are shown by the vertical bars.

$$V(t) = E[r(t)] - \bar{r} + V(t+1) \tag{7}$$

This sum converges if the baseline $\bar{r}$ is taken as the long-term average reward per timestep:

$$\lim_{n\to\infty} \frac{1}{n} E\left[\sum_{t=1}^{n} r(t)\right] \tag{8}$$

Under some simplifying assumptions about the structure of the environment (Puterman, 1994), this average has the same value no matter in what state of the world the sum is started.

Reinforcement learning algorithms using this value function are known as average reward RL algorithms, since actions chosen to optimize it will also optimize the long-term average reward $\bar{r}$ received per timestep, instead of the cumulative reward received over a finite time window. A TD algorithm for learning this value function (Mahadevan, 1996; Schwartz, 1993; Tadepalli & Ok, 1998; Tsitsiklis & Van Roy, 1999) uses the error signal

$$\delta(t) = r(t) - \bar{r}(t) + \hat{V}(t+1) - \hat{V}(t) \tag{9}$$

$$\delta(t) = \delta_{\mathrm{p}}(t) - \bar{r}(t) \tag{10}$$

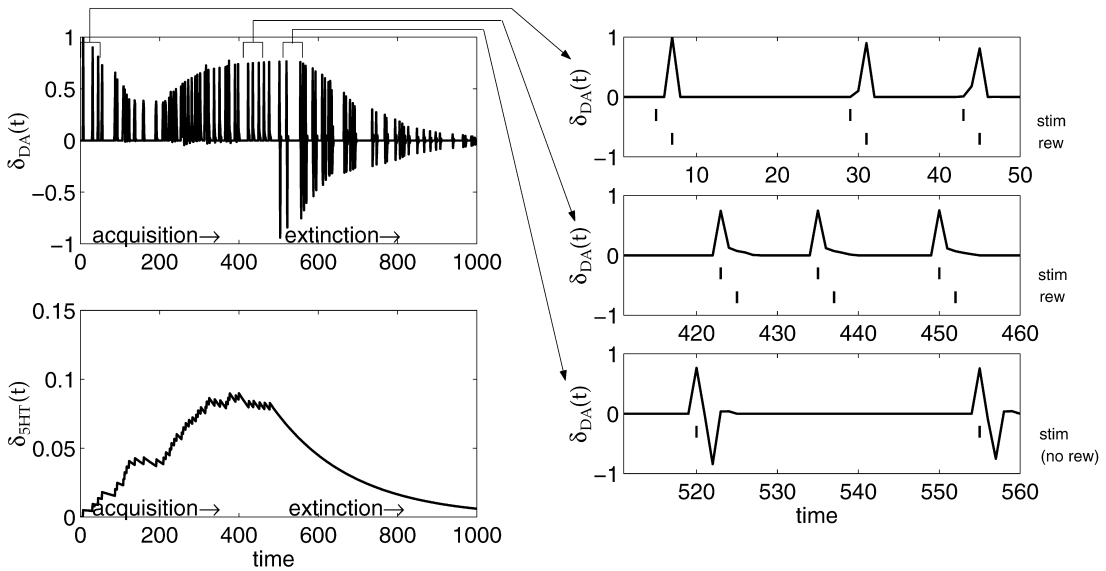

Fig. 2. Signaled rewards. (Left) Modeled opposing phasic dopamine (upper) and tonic serotonin (lower) responses during acquisition and extinction of the delivery of signaled rewards. (Right) The phasic dopamine signals at various points during the experiment. (Top; Middle) The conventional (Montague et al., 1996) movement backwards within each trial of the phasic dopamine signal to the earliest reliable predictor. (Bottom) The conventional below-baseline activity at the time the reward is expected but not delivered during early extinction.

Fig. 3. Schematic of Solomon and Corbit's (1974) opponent process model of motivation. (A) A series of punctate rewards is delivered, and treated (dashed line) as a continuous envelope. (B) The affective response habituates, rebounds, and re-habituates. SC modeled this response as the difference between a fast primary (C) and a slow opponent (D) reaction to the reward.

and requires that the average reward $\bar{r}$ be estimated separately. One way to do this is with an exponentially windowed running average

$$\bar{r}(t) = vr(t) + (1 - v)\bar{r}(t - 1), \tag{11}$$

where the parameter v, which is a form of learning rate, determines the timescale over which the running average is computed.

In this algorithm, immediate phasic reward information $r(t)$ is weighed against a long-term average reward prediction $\bar{r}(t)$ in computing the error signal. The slow-timescale prediction acts as an opponent to the rest of the signal. Here we propose that slow, essentially tonic, fluctuations in dorsal raphe nucleus serotonin activity could be responsible for reporting the average reward prediction:

$$\delta_{5\mathrm{HT}} = \bar{r}(t) \tag{12}$$

while δ_{DA}, reporting the phasic components of the error signal, is as before (Eq. (4)).

Fig. 1 shows the behavior expected of the serotonin and dopamine systems under this model during the acquisition and extinction of unsignaled, randomly delivered rewards. The 5HT signal increases slowly during acquisition and decreases slowly during extinction. As in other TD models (and as in electrophysiological recordings), the DA signal displays phasic responses to the rewards.

Fig. 2 shows what happens if the rewards are signaled by a preceding conditioned stimulus. In this case, the modeled 5HT behavior is unchanged. The phasic DA responses, however, transfer from the time of the reward to the time of the stimulus, as in the standard TD model, given a tapped delay line representation of the time since the presentation of the conditioned stimulus (Schultz et al., 1997). Moreover, since the stimuli are still presented during the extinction phase, the modeled dopamine signal also displays decaying phasic bursts (to the signal) and pauses (at the time of missing rewards) during extinction.

The average reward formulation of TD learning provides a computational account of some of the psychological phenomena identified by Solomon and Corbit (SC, 1974) in their seminal study of opponency. Fig. 3 shows SC's model of affective dynamics, which they applied to a wealth of appetitive and aversive cases.

Consider an example (Fig. 3(A)) in which subjects are made hungry, are then delivered grains of the cereal coco-pops in a manner governed by a constant-rate Poisson process, and are finally extinguished. We will assume that they are never given enough food to become sated. This is exactly analogous to the example of Fig. 1. Fig. 3(B) shows SC's description of the affective dynamics in this case. Note that SC do not focus on the pulsatile nature of the delivery of the coco-pops; rather they imagine that the affective reaction is continuous. The figure shows that the initial

appetitive affective reaction rises to a peak following initiation of the delivery of the coco-pops. This reward is described by subjects as being highly pleasurable. Whilst the provision of reward is maintained, the affective reaction slowly declines to a lower steady level, as subjects report less pleasure. After the termination of the stimulus, a rebound to a negative affective reaction follows, as subjects report being upset. This displeasure quickly peaks and slowly decays to the original affective state. If fewer rewards were given, then the habituation to the reward would be less and the subsequent displeasure upon termination would not be so severe.

Fig. 3(C) and (D) show SC's opponent-process model of these dynamics. A fast primary process for the appetitive affective state is engaged by the reward and maintains its activity as long as the reward persists. In addition, an opponent process having an affective sign opposite to that of the primary process is engaged. This aversive opponent is sluggish in the sense that it slowly climbs to asymptote as long as the stimulus is maintained, and slowly decays back to baseline after the termination of the stimulus. The difference between the outputs of these two processes gives rise to the affective signal observed in Fig. 3(B). Grossberg (1988) suggested a dynamical opponent mechanism to account for the same phenomena.

Fig. 1, from the average-case model, is closely related to Fig. 3(C) and (D). In our model, the serotonin system is the aversive opponent to dopamine, and reports $\bar{r}$. We make the crude assumption that the net affective reaction is $\delta(t) = \delta_{\mathrm{DA}}(t) - \delta_{\mathrm{5HT}}(t)$. When the rewards are first delivered, the prediction of $\bar{r}$ is zero, and so the overall signal $\delta(t)$ is greatest. The increase in $\bar{r}$ to match the average reward rate in the environment while the rewards are still provided captures the habituation of the motivational system in the light of the expectation of reward. $\delta(t)$ is less in this phase, as $\delta_{\mathrm{5HT}}(t)$ grows. Finally, during extinction, the only signal is $\delta_{\mathrm{5HT}}(t)$, which returns to 0 as the expectation of reward $\bar{r}$ returns to 0. Our model better captures the pulsatile nature of the delivery of rewards, which is presumably reflected in motivational state.

SC suggested that their model applies to a very wide variety of different cases of affective opponency, all the way from the short-term effects of shocks to the long-term effects of drugs of addiction. This raises two issues for our model. One, which is the subject of the next section, is the degree of symmetry of the involvement of dopamine and serotonin in affective processing. The other, which we save for the discussion, is the range of timescales involved.

5. Aversive conditioning and mirrored opponency

The key lacuna in the computational model that we have specified is the lack of an account of how the prediction error is reported for aversive events. From a theoretical standpoint, the learning of actions to obtain reward is no different from the learning of actions to avoid punishments. However, as stated earlier, there is both physiological (cells only have positive firing rates) and psychological (data suggest separate appetitive and aversive systems) evidence to support a distinction between the appetitive and aversive learning. Moreover, microdialysis studies reveal evidence of elevated dopamine levels in response to *aversive* stimuli such as footshocks (reviewed by Horvitz, 2000; Salamone, Cousins, & Snyder, 1997), which seems opposite what is expected under a reward prediction model of dopamine. We now suggest how to reconcile our computational model with these data.

We make two rather more speculative hypotheses about the opponency between dopamine and serotonin. First, we suggest that the *phasic* release of serotonin might mirror the phasic release of dopamine, and report a prediction error for future punishment. Second, we suggest that the release of dopamine in aversive conditioning comes from a tonic increase in the activity of dopaminergic cells as a report of the long-term rate of the delivery of punishment. On the first suggestion, we are well aware that there is no evidence of a phasic serotonergic signal from microdialysis, and only a few hints in the spiking rates of serotonergic cells. Microdialysis would not be expected to provide a positive report, since it is not sensitive to the sort of transient fluctuations that we would expect from a phasic signal. This is true even for dopamine. More pertinently, Mason (1997) and collaborators (Gao et al., 1997, 1998) stress that 5HT-containing neurons in the spinal cord-directed serotonin system are relatively unresponsive to noxious stimuli, compared to the dramatic and long-lasting firing rate changes seen in non-serotonergic ON and OFF cells found in the same area. However, the weaker and more transient excitation the authors report for 25–50% of the serotonergic neurons actually somewhat more closely resembles the sort of phasic responses seen in dopamine neurons. Moreover, evidence from the same studies about the mechanism of morphine-based analgesia suggests one reason why more evidence for our proposal has not been forthcoming: the release of serotonin could be decoupled from the activity of serotonergic cells through the action of non-serotonergic cells.

The simplest way to incorporate aversive events $a(t)$ into the TD framework is to treat them as negative rewards. The augmented value function then takes the form:

$$V(t) = E\left[\sum_{\tau \geq t}^{\infty} (r(\tau) - \bar{r} - a(\tau) + \bar{a})\right] \tag{13}$$

with the error signal

$$\delta(t) = r(t) - \bar{r}(t) - a(t) + \bar{a}(t) + \hat{V}(t+1) - \hat{V}(t) \tag{14}$$

The phasic component of the prediction error signal includes the additional component $a(t)$. Thus we can

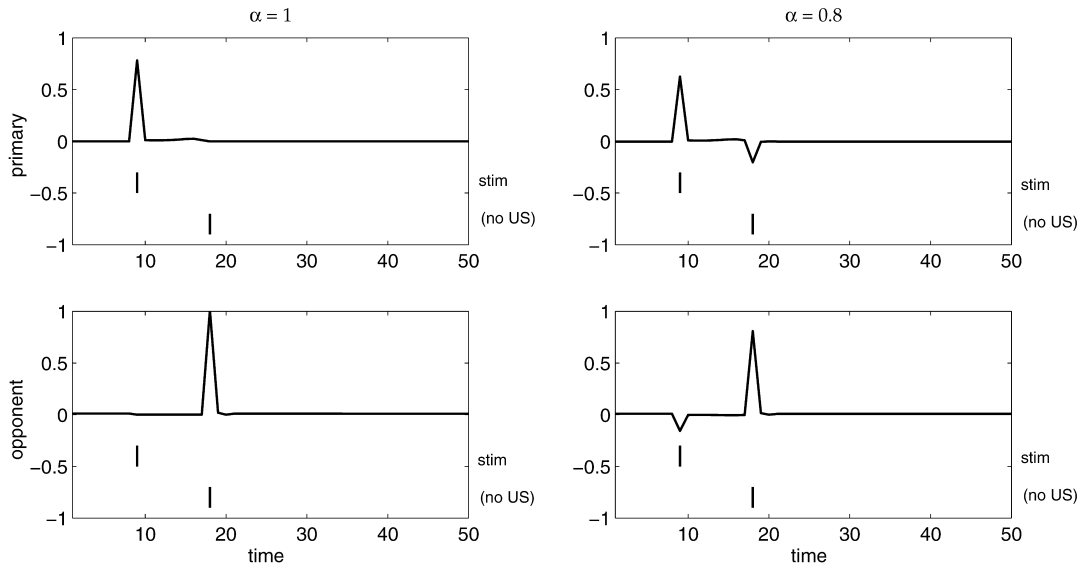

Fig. 4. Responses to probe trial presentations of a conditioned stimulus when the unconditioned stimulus is omitted. Modeled primary are opponent responses are shown when $\alpha = 1$ (left column) and $\alpha = 0.8$ (right column). For appetitive conditioning, the channel labeled 'primary' represents DA and the 'opponent' channel is 5HT. For aversive conditioning, the same results are predicted, but with the roles reversed.

redefine

$$\delta_p(t) = r(t) - a(t) + \hat{V}(t+1) - \hat{V}(t) \tag{15}$$

and rewrite the complete error signal as:

$$\delta(t) = \delta_p(t) - \bar{r}(t) + \bar{a}(t) \tag{16}$$

There are various ways that the full prediction error signal $\delta(t)$ could be apportioned between the modeled dopaminergic and serotonergic opponents. Following the model from Section 4, we attribute the tonic reward signal $\bar{r}$ to 5HT. For symmetry, we assign the new tonic punishment signal $\bar{a}$ to DA. This apportionment is also designed to account for the microdialysis evidence on dopamine activation by aversive stimuli (Horvitz, 2000; Salamone et al., 1997). Such events would increase the average punishment signal $\bar{a}(t)$, in turn increasing $\delta(t)$ and its positive opponent channel $\delta_{DA}(t)$. We would expect a slow ramp-up in tonic dopamine activity exactly analogous to that shown for serotonin in Fig. 1. Though this explanation is consistent with microdialysis data, it is difficult to evaluate it with respect to existing electrophysiological experiments on dopamine in aversive situations (Guaracci & Kapp, 1999; Mirenowicz & Schultz, 1996; Schultz & Romo, 1987), which are in mutual disagreement as to the predominant direction and timescale of dopamine responses, and were not designed as tests of this idea. The general suggestion of our model is that, at short timescales, if they exist at all, dopamine responses should be depressed by aversive stimuli, whereas at longer timescales, they should be excited.

In the absence of relevant recording, or perhaps voltammetric (Garris, Christensen, Rebec, & Wightman, 1997), data, it is particularly difficult to know how to apportion the phasic component to the serotonergic channel. The only *computational* requirement of the model is that the difference between the signals from the two channels (appropriately scaled) be the error signal $\delta(t)$. In this section we investigate one of the simplest models for which this holds true—a perfectly symmetric relationship between DA and 5HT—and we return in the discussion to other options and their implications.

In particular, we split up positive $[\delta_p(t)]_+$ and negative $[\delta_p(t)]_-$ components of the phasic signal, allocating the former mostly to the dopamine system, and the latter mostly to the serotonin system. Formally, we consider

$$\delta_{DA}(t) = \alpha[\delta_p(t)]_+ - (1-\alpha)[\delta_p(t)]_- + \bar{a} \tag{17}$$

$$\delta_{5HT}(t) = \alpha[\delta_p(t)]_- - (1-\alpha)[\delta_p(t)]_+ + \bar{r}, \tag{18}$$

where the parameter α controls the degree to which both negatively and positively rectified information are blended in each signal. When $\alpha = 1$, DA reports exclusively the positive part of the error signal, and 5HT the negative part; with the parameter slightly smaller, each channel has a portion of the oppositely rectified error subtracted from it. The need for $\alpha < 1$ stems from the evidence about the activity of dopamine cells in extinction—they show a phasic depression at the time an expected reward is not delivered (i.e. when the value of $[\delta_p(t)]_-$ is large).

The point about blending is illustrated in Fig. 4, which shows the expected activity in the model when an expected reward or punishment is omitted, as in extinction. First, consider the case of reward, for which dopamine is the primary process and serotonin the opponent. The primary process is activated when the reward is signaled, and the opponent is activated when it fails to arrive. The left and right columns show the behavior expected when $\alpha = 1$ and $\alpha = 0.8$, respectively. With the smaller value of α,

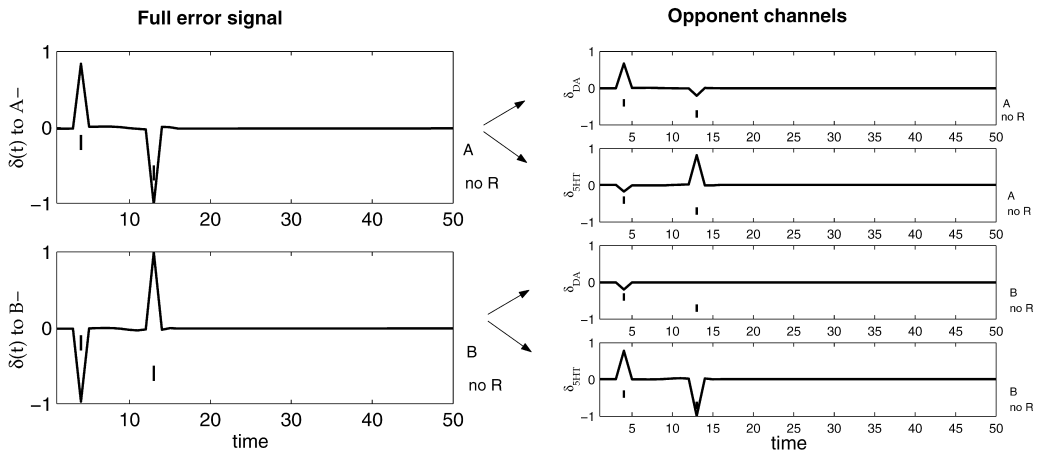

Fig. 5. Responses to probe trial presentations of a conditioned excitor A and a simultaneously trained conditioned inhibitor B, in both cases in extinction. The left graphs show the full phasic TD error signal $\delta_p(t)$; the right graphs show a decomposition of this into dopamine and serotonin opponent channels using Eqs. (17) and (18) with $\alpha = 0.8$ and modified for consistency with experimental results (Tobler et al., 2001).

phasic pauses are seen in the primary channel at the time the unconditioned stimulus was expected, as is seen experimentally for DA in extinction of appetitive associations (Schultz, 1998). Since this version of the model is completely symmetric, the roles of the primary and opponent processes are simply interchanged when conditioning is aversive rather than appetitive. That is, serotonin would be the primary process and dopamine the opponent. Thus, in aversive acquisition, which would precede the traces shown in Fig. 4, the model would suggest that a putative phasic 5HT response to a primary aversive stimulus would transfer, with learning, to a conditioned stimulus that predicts it.

Another conditioning paradigm that may involve both aversive and appetitive opponent channels is conditioned inhibition (Rescorla, 1969). Fig. 5 shows the modeled behavior of the phasic TD error signal $\delta_p(t)$ and the two opponent channels $\delta_{DA}(t)$ and $\delta_{5HT}(t)$ in response to two stimuli A and B that have been used in a conditioned inhibition experiment. Training consisted of alternating trials with A $\rightarrow$ r and A, B $\rightarrow \cdot$; the figure shows the result of probe trials presenting A and B in extinction. The left part of the figure shows that the full phasic TD prediction error signals for the two stimuli are exact opposites—this is because the value prediction occasioned by the conditioned inhibitor B must exactly cancel the value prediction of the conditioned excitor A.

The right part of the figure shows one possible decomposition of each signal into two opponent channels. Many decompositions are consistent with our formal model. In particular, though the temporal locations of phasic error are constrained by the TD model, under a different decomposition, the predicted positive error could appear as either DA excitation or 5HT inhibition (or some combination of both), and vice versa for negative error. Thus, we have decomposed the signal to reflect an additional constraint coming from recordings from dopamine neurons in a conditioned inhibition paradigm (Tobler, Dickinson, & Schultz, 2001). The key aspect of these experiments lies in the response $\delta(t)$ to B−, when the conditioned inhibitor is presented alone. The lower left plot of Fig. 5 shows that $\delta(t)$ is actually positive at this time, because of the violation of the expectation that a reward will be omitted at that point. However, Tobler et al. (2001) shows no evidence of increased dopaminergic activity at this point. Thus, we assume that this error is instead fully conveyed by inhibition in the 5HT channel. This absence of DA activity at the time of the unexpected reward omission may help explain why the power of a stimulus as a conditioned inhibitor does not extinguish when it is repeatedly presented alone (see Detke, 1991; Zimmer-Hart & Rescorla, 1974). Apart from the assignment of this positive error to the 5HT channel, the graphs in the right part of this figure have been generated assuming a decomposition according to the simple scheme above with $\alpha = 0.8$.

Another interesting feature of the decomposition presented here is the activation of the opponent serotonergic channel in response to the presentation of the conditioned inhibitor B. The symmetry of the model implies that a conditioned inhibitor for shock would equivalently activate DA. This may help explain why animals will work to bring about the presentation of such a stimulus (Lolordo, 1969).

6. Discussion

We have suggested various ways in which serotonin, perhaps that released by the dorsal raphe nucleus serotonin system, could act as a motivational opponent system to dopamine in conditioning tasks. We considered how serotonin might report the long-run average reward rate as a tonic opponent to a phasic dopamine signal in the theoretical framework of average-case reinforcement learning. We also considered how dopamine might report the long-run average punishment rate as a tonic opponent to a phasic aversive signal. This suggestion was motivated by

microdialysis and limited electrophysiological data on the involvement of dopamine in aversive circumstances. Finally, we speculated that there might be a phasic component of the release of serotonin, as a more direct mirror of the phasic component of the release of dopamine.

These suggestions go some steps beyond the experimental data. The most critical experiments from the perspective of our model would be to record the phasic and tonic activity of dopamine and serotonin cells and (perhaps using fast voltammetry, Garris et al., 1997), the phasic and tonic concentrations of dopamine and serotonin at their targets, during excitatory and inhibitory, aversive and appetitive, conditioning tasks. Tasks involving secondary associations are particularly interesting, since some aspects of primary appetitive conditioning are unaffected by dopaminergic manipulations (Berridge & Robinson, 1998). It might also be interesting to record from non-serotonergic cells in the raphe nuclei, given evidence (Gao et al., 1997, 1998) about non-serotonergic ON- and OFF-cells in the spinal cord-directed raphe nuclei whose activities are modulated by aversive events, and which might possibly control the release of serotonin at target sites.

Results from Everitt and his colleagues (Wilkinson, Humby, Robbins, & Everitt, 1995) suggest further complexities in the involvement of serotonin in aversive conditioning. In their task, 5HT depletion impairs aversive conditioning to a context, but enhances it to an explicit conditioned stimulus. It might be that learning of the slow timescale average-reward and average-punishment signals is more closely associated with contextual than explicit stimuli—however, an extension to our simple model would be required to suggest why 5HT depletions disrupt learning of $\bar{a}$ as well as $\bar{r}$. The implications for phasic responses are unclear, particularly because the experiments did not involve secondary contingencies.

Our suggestions are also somewhat incomplete. In particular, we have discussed computational aspects of opponency, ignoring the sort of implementational details that have been studied in great theoretical detail in such work as Grossberg (1984, 2000). In practice, it seems that serotonin can affect dopamine at multiple levels, including influencing the activities of dopamine neurons in the midbrain and influencing the effect of dopamine at target structures such as the nucleus accumbens (Kapur & Remington, 1996). There seem to be insufficient data to understand the extent to which these interactions could be inimical to the computational theory; at the very least they will influence issues like the timescale of the sort of dynamical opponency seen in the account of Solomon and Corbit's (1974) suggestion. Also, by no means all experiments support a simple opponent view. For instance, there is evidence (e.g. Parsons & Justice, 1993; Zangen, Nakash, Overstreet, & Yadid, 2001) that increasing the concentration of serotonin in the accumbens increases the concentration of dopamine. Further, we have concentrated on a narrow set of motivational aspects of serotonin. A key direction is to expand the current theory to encompass more of the very wide range of effects of this neuromodulator, partially listed in the introduction.

The theoretical model is also incomplete in three key ways. First, the computational model contains a pair of oversimplifications: it assumes average reward rates are stationary and it does not treat the issues of multiple timescales involved in executing realistic courses of action. Second, we need to consider other ways that the phasic prediction error signal $\delta_p(t)$ might be shared between the opponent systems. Finally, we have not modeled how the predictions associated with the tonic and phasic neuromodulatory signals might themselves be represented.

As described, the model of average-case reinforcement learning assumes that there is a single, fixed, underlying Markov chain, which has a single underlying average reward (or punishment) rate. This simplified model has two key problems—one relating to the problem that contingencies in the world could change and the other, more subtly, relating to how to plan courses of actions at different timescales.

The first oversimplification with the model is that it is rather unrealistic to assume a static model, since the average reward rates are in fact changed during the course of the experiments (as in the examples establishing the relationship to Solomon and Corbit's (1974) theory of opponency). There is a relatively straightforward answer to this concern. As is conventional in engineering models (Anderson & Moore, 1979) and has also been suggested for conditioning models (Dayan, Kakade, & Montague, 2000; Kakade & Dayan, 2000, 2002), the choice of learning rate in a single part of an environment should be tied to how fast the actual reward rate in the world could change. For instance, in our model of Solomon and Corbit (1974) opponency, the timescale at which the opponent decays back to zero after rewards are no longer sustained reflects the timescale at which the average reward decays back to zero. This in turn reflects the expectation of how fast the world could be changing.

The second simplification is that, even within a fixed overall environment, the model only considers two particular timescales at opposite extremes, the shortest possible, associated with immediate changes, giving rise to phasic neuromodulatory signals, and the longest possible, associated with average reward or punishment rates, giving rise to tonic neuromodulatory signals. In reality, multiple timescales are likely to be simultaneously relevant, as animals execute hierarchically structured courses of action with consequences for reward and punishment that play out over different characteristic periods. Despite suggestions in both artificial reinforcement learning (Sutton, Precup, & Singh, 1999) and ethology (Gibbon, 1977), it is not clear how learning can proceed simultaneously at multiple timescales in an integrated manner leading to hierarchically optimal behavior. The neuromodulatory basis underlying the multiple timescales is also not clear—although one can

certainly imagine families of reception mechanisms for neuromodulatory signals, each with different time constants, understanding how the resulting pieces of information might be combined together poses a greater challenge. Note that Doya (2000, 2002) actually makes the issue of setting the timescale of courses of action the defining role for serotonin activity.

After the simplifications of the computational model, the second theoretical issue is the opponent construction of the phasic signal $\delta_p(t)$. Eqs. (17) and (18) formalize the simplest possible symmetric scheme in which serotonin and dopamine are exact, mirrored, opponents, each reporting mostly one motivational channel (aversive or appetitive, respectively) but with a small component (controlled by the value of $1-\alpha$) of the other channel. However, there are obviously many other ways to decompose $\delta_p(t)$ into opposing signals, and experimental evidence such as the recordings of dopamine neurons in conditioned inhibition experiments (Tobler et al., 2001) is hard to reconcile with our simple scheme. In Fig. 5, we used an ad hoc decomposition to capture the tricky aspect of the results of this experiment, namely that dopaminergic excitation is not seen at the time reward is normally omitted after presentation of a conditioned inhibitor. However, a compelling justification for this decomposition is presently lacking.

As another example of the problem with the symmetric model, pauses in dopaminergic activity are seen to negative TD error produced by the non-occurrence of expected reward, but the few available experiments disagree as to whether pauses are commonly seen to other, aversive sources of negative TD error: receipt of primary punishment (Schultz & Romo, 1987; Mirenowicz & Schultz, 1996) or decreases in $\hat{V}(t)$ caused by the onset of a stimulus predicting punishment (Guaracci & Kapp, 1999). These experiments show some combination of excitatory and inhibitory responses, but all differ as to their relative proportions. In the symmetric model, if $\alpha = 1$, pausing would be seen to neither omitted rewards nor unexpected punishments, but with $\alpha < 1$, would be expected to both. Thus, the correct implementation probably does not involve the purely symmetric rectification of a single error signal, but, as in Fig. 5, rather arranges for the various individual components of the full error signal to appear differentially in one channel or the other. The implementational details of this could be significantly clarified by more experimental evidence.

The fact that increases and decreases in $\hat{V}(t)$ seem to affect DA differentially depending on their origin points to the third aspect of our model that we have ignored thus far, namely how the predictions $\hat{V}(t)$ are represented. The standard way to represent value function estimates in TD is to use a single function approximator whose single set of parameters is trained by $\delta(t)$. The opponent architecture we envision here suggests an obvious alternative: there may be both positive and negative parts to the value function prediction, whose difference is the full value function: $\hat{V}(t) = \hat{V}_+(t) - \hat{V}_-(t)$. In this case, each partial prediction can be associated (more or less exclusively) with one of the opponent channels and trained by its partial error signal.

This sort of dual representation, used for instance in the opponent conditioning model of Grossberg (1984), can model a number of behavioral phenomena that are difficult to account for using a more unitary representation of a prediction. It allows for active extinction, in which the extinction of a stimulus-reward association causes the stimulus' contribution to $\hat{V}_-$ to become elevated rather than its contribution to $\hat{V}_+$ to decay. This has the same effect of diminishing the net prediction, but it preserves the predictive association in the positive channel, providing an explanation for phenomena such as the spontaneous recovery of extinguished conditioned responses. In another example, having both positive and negative sets of prediction weights allows the model to explain evidence that conditioned inhibitors can seem simultaneously to carry both positive and negative predictions: that, for instance, attempts to extinguish conditioned inhibitors by presenting them alone can actually enhance their inhibitory properties, presumably by extinguishing their positive associations (Williams & Overmier, 1988). In terms of the DA response data considered here, maintaining two separate partial predictions allows predictions originating from appetitive and aversive events, or from conditioned inhibition and excitation to be separated, so that the modeled DA signal has the ability to react differently to changes in either.

One important spur for this work, which we have nevertheless not directly modeled, is an apparent opponency between DA's involvement in eliciting appetitive approach and psychomotor excitation, and serotonin's involvement in fight and flight behaviors. These automatic (i.e. non-plastic) effects of the neuromodulators on action selection can produce computationally undesirable, but experimentally observable, outcomes such as the inability of animals to learn to get a reward delivered by retreating from a particular stimulus with which the reward is associated. We do not yet have an account of these effects. However, recent work on incentive motivation (Dickinson & Balleine, 2002) is leading to a slightly different picture of how Pavlovian and instrumental conditioning interact, and in this new picture, these effects may be more readily accommodated. An important task is to add opponency to the new model of neurobiological reinforcement learning (Dayan, 2002) that is emerging from this new motivational picture. Such a new model might also be used to capture serotonin's apparent role in withholding inappropriate behaviors. Extending models such as the actor–critic (Barto, Sutton, & Anderson, 1983; Houk, Adams, & Barto, 1995) of DA's involvement in increasing the probability of better-than-expected actions, our model could accommodate a role for 5HT in *learning* to avoid actions whose outcomes were worse than expected, due either to directly aversive events or to frustrative non-reward. However, the behavioral inhibition associated with 5HT can be seen as also having

a more automatic component, and this would naturally fall under the scope of the motivational model.

Given the vast range of phenomena in which it is involved, it is unlikely that there is a simple computational account that can address all aspects of serotonin. We have attempted to use our relatively more comprehensive understanding of the role of dopamine in appetitive conditioning to provide a theoretical window onto a restricted aspect of serotonergic function. The resulting theory makes testable contact with a number of different studies on serotonin, and offers many directions for future investigation.

Acknowledgments

We are very grateful to Bill Deakin, Kenji Doya, Barry Everitt, Zach Mainen, Trevor Robbins, David Touretzky and Jonathan Williams for helpful discussions, and to Kenji Doya for inviting SK to the Metalearning and Neuromodulation workshop. Funding was from the Gatsby Charitable Foundation and the NSF (ND by a Graduate Fellowship and grants IIS-9978403 and DGE-9987588; SK by a Graduate Fellowship).

References

Aghajanian, G. K., & Marek, G. J. (1999). Serotonin and hallucinogens. *Neuropsychopharmacology*, *21*, 16S–23S.

Anderson, B. D. O., & Moore, J. B. (1979). *Optimal filtering*. Englewood Cliffs, NJ: Prentice-Hall.

Azmitia, E. C. (1978). The serotonin-producing neurons of the midbrain median and dorsal raphe nuclei. In L. L. Iversen, S. D. Iversen, & S. Snyder (Eds.), *Handbook of psychopharmacology* (*Vol. 9*) (pp. 233–314). *Chemical pathways in the brain*, New York: Plenum Press.

Azmitia, E. C., & Segal, M. (1978). An autoradiographic analysis of the differential ascending projections of the dorsal and median raphe nuclei in the rat. *Journal of Comparative Neurology*, *179*, 641–667.

Barto, A. G., Sutton, R. S., & Anderson, C. W. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transaction on Systems, Man and Cybernetics*, *SMC-13*, 834–846.

Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Research Reviews*, *28*, 309–369.

Bertsekas, D. P., & Tsitsiklis, J. N. (1996). *Neuro-dynamic programming*. Belmont, MA: Athena Scientific.

Buhot, M. C. (1997). Serotonin receptors in cognitive behaviors. *Current Opinion in Neurobiology*, *7*, 243–254.

Canales, J. J., & Iversen, S. D. (2000). Psychomotor-activating effects mediated by dopamine D-sub-2 and D-sub-3 receptors in the nucleus accumbens. *Pharmacology, Biochemistry and Behavior*, *67*, 161–168.

Daw, N. D., & Touretzky, D. S. (2000). Behavioral considerations suggest an average reward TD model of the dopamine system. *Neurocomputing*, *32*, 679–684.

Daw, N. D., & Touretzky, D. S. (2002). Long-term reward prediction in TD models of the dopamine system. *Neural Computation*, in press.

Dayan, P. (2002). Motivated reinforcement learning. *NIPS*, in press.

Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, *3*, 1218–1223.

De Vry, J., & Schreiber, R. (2000). Effects of selected serotonin 5-HT-sub-1 and 5-HT-sub-2 receptor agonists on feeding behavior: Possible mechanisms of action. *Neuroscience and Biobehavioral Reviews*, *24*, 341–353.

Deakin, J. F. W. (1983). Roles of brain serotonergic neurons in escape, avoidance and other behaviors. *Journal of Psychopharmacology*, *43*, 563–577.

Deakin, J. F. W. (1996). 5-HT, antidepressant drugs and the psychosocial origins of depression. *Journal of Psychopharmacology*, *10*, 31–38.

Deakin, J. F. W., & Graeff, F. G. (1991). 5-HT and mechanisms of defence. *Journal of Psychopharmacology*, *5*, 305–316.

Detke, M. J. (1991). Extinction of sequential conditioned inhibition. *Animal Learning and Behavior*, *19*, 345–354.

Dickinson, A., & Balleine, B. (2002). The role of learning in motivation. In C. R. Gallistel (Ed.), *Stevens' handbook of experimental psychology* (3rd ed) (*Vol. 3*). *Learning, motivation and emotion*, New York: Wiley.

Dickinson, A., & Dearing, M. F. (1979). Appetitive–aversive interactions and inhibitory processes. In A. Dickinson, & R. A. Boakes (Eds.), *Mechanisms of learning and motivation* (pp. 203–231). Hillsdale, NJ: Erlbaum.

Doya, K. (2000). Metalearning, neuromodulation, and emotion. In G. Hatano, N. Okada, & H. Tanabe (Eds.), *Affective minds* (pp. 101–104). Amsterdam: Elsevier Science.

Doya, K (2002). Metalearning and neuromodulation. *Neural Networks*, this issue.

Edwards, D. H., & Kravitz, E. A. (1997). Serotonin, social status and aggression. *Current Opinion in Neurobiology*, *7*, 812–819.

Everitt, B. J., Parkinson, J. A., Olmstead, M. C., Arroyo, M., Robledo, P., & Robbins, T. W. (1999). Associative processes in addiction and reward. The role of amygdala-ventral striatal subsystems. *Annals of the New York Academy of Sciences*, *877*, 412–438.

Fields, H. L., Heinricher, M. M., & Mason, P. (1991). Neurotransmitters in nociceptive modulatory circuits. *Annual Review of Neuroscience*, *14*, 219–245.

Fletcher, P. J. (1991). Dopamine receptor blockade in nucleus accumbens or caudate nucleus differentially affects feeding induced by 8-OH-DPAT injected into dorsal or median raphe. *Brain Research*, *552*, 181–189.

Fletcher, P. J. (1995). Effects of combined or separate 5,7-dihydroxytryptamine lesions of the dorsal and median raphe nuclei on responding maintained by a DRL 20s schedule of food reinforcement. *Brain Research*, *675*, 45–54.

Fletcher, P. J., & Korth, K. M. (1999). Activation of 5-HT1B receptors in the nucleus accumbens reduces amphetamine-induced enhancement of responding for conditioned reward. *Psychopharmacology*, *142*, 165–174.

Fletcher, P. J., Korth, K. M., & Chambers, J. W. (1999). Selective destruction of brain serotonin neurons by 5,7-dihydroxytryptamine increases responding for a conditioned reward. *Psychopharmacology*, *147*, 291–299.

Fletcher, P. J., Ming, Z. H., & Higgins, G. A. (1993). Conditioned place preference induced by microinjection of 8-OH-DPAT into the dorsal or median raphe nucleus. *Psychopharmacology*, *113*, 31–36.

Fletcher, P. J., Tampakeras, M., & Yeomans, J. S. (1995). Median raphe injections of 8-OH-DPAT lower frequency thresholds for lateral hypothalamic self-stimulation. *Pharmacology Biochemistry and Behavior*, *52*, 65–71.

Ganesan, R., & Pearce, J. M. (1988). Effect of changing the unconditioned stimulus on appetitive blocking. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*, 280–291.

Gao, K., Chen, D. O., Genzen, J. R., & Mason, P. (1988). Activation of serotonergic neurons in the raphe magnus is not necessary for morphine analgesia. *Journal of Neuroscience*, *18*, 1860–1868.

Gao, K., Kim, Y. H., & Mason, P. (1997). Serotonergic pontomedullary neurons are not activated by antinociceptive stimulation in the periaqueductal gray. *Journal of Neuroscience*, *17*, 3285–3292.

Garris, P. A., Christensen, J. R. C., Rebec, G. V., & Wightman, R. M. (1997). Real-time measurement of electrically evoked extracellular

dopamine in the striatum of freely moving rats. *Journal of Neurochemistry*, *68*, 152–161.

Geller, I., & Seifter, J. (1960). The effects of mepobromate, barbiturate, *d*-amphetamine and promazine on experimentally induced conflict in the rat. *Psychopharmacology*, *1*, 482–492.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, *84*, 279–325.

Goodman, J. H., & Fowler, H. (1983). Blocking and enhancement of fear conditioning by appetitive CSs. *Animal Learning and Behavior*, *11*, 75–82.

Grossberg, S. (1984). Some normal and abnormal behavioral syndromes due to transmitter gating of opponent processes. *Biological Psychiatry*, *19*, 1075–1118.

Grossberg, S. (Ed.), (1988). *Neural networks and natural intelligence*. Cambridge, MA: MIT press.

Grossberg, S. (2000). The imbalanced brain: From normal behavior to schizophrenia. *Biological Psychiatry*, *48*, 81–98.

Grossberg, S., & Schmajuk, N. A. (1987). Neural dynamics of attentionally modulated Pavlovian conditioning: Conditioned reinforcement, inhibition, and opponent processing. *Psychobiology*, *15*, 195–240.

Guarraci, F. A., & Kapp, B. S. (1999). An electrophysiological characterization of ventral tegmental area dopaminergic neurons during differential Pavlovian fear conditioning in the awake rabbit. *Behavioural Brain Research*, *99*, 169–179.

Harrison, A. A., Everitt, B. J., & Robbins, T. W. (1997). Doubly dissociable effects of median- and dorsal-raphe lesions on the performance of the five-choice serial reaction time test of attention in rats. *Behavioural Brain Research*, *89*, 135–149.

Harrison, A. A., Everitt, B. J., & Robbins, T. W. (1999). Central serotonin depletion impairs both the acquisition and performance of a symmetrically reinforced go/no-go conditional visual discrimination. *Brain Research*, *100*, 99–112.

Hollander, E. (1998). Treatment of obsessive-compulsive spectrum disorders with SSRIs. *British Journal of Psychiatry*, *173*, 7–12.

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

Houk, J. C., Adams, J. L., & Barto, A. G. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 249–270). Cambridge, MA: MIT Press.

Ikemoto, S., & Panksepp, J. (1999). The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. *Brain Research Reviews*, *31*, 6–41.

Imai, H., Steindler, D. A., & Kitai, S. T. (1986). The organization of divergent axonal projections from the midbrain raphe nuclei in the rat. *Journal of Comparative Neurology*, *243*, 363–380.

Jacobs, B. L., & Fornal, C. A. (1997). Serotonin and motor activity. *Current Opinion in Neurobiology*, *7*, 820–825.

Jacobs, B. L., & Fornal, C. A. (1999). Activity of serotonergic neurons in behaving animals. *Neuropsychopharmacology*, *21*, 9S–15S.

Jones, S., & Kauer, J. A. (1999). Amphetamine depresses excitatory synaptic transmission via serotonin receptors in the ventral tegmental area. *Journal of Neuroscience*, *19*, 9780–9787.

Kamin, L. J. (1969). Selective association and conditioning. In N. J. Mackintosh, & W. K. Honig (Eds.), *Fundamental issues in associative learning* (pp. 42–64). Halifax, Canada: Dalhousie University Press.

Kakade, S., & Dayan, P. (2000). Acquisition in autoshaping. In S. A. Solla, T. K. Leen, & K.-R. Muller (Eds.), (*Vol. 12*) (pp. 24–30). *Advances in neural information processing systems*, Cambridge, MA: MIT Press.

Kakade, S., & Dayan, P. (2002). Acquisition and extinction in autoshaping. *Psychological Review*, in press.

Kapur, S., & Remington, G. (1996). Serotonin–dopamine interaction and its relevance to schizophrenia. *American Journal of Psychiatry*, *153*, 466–476.

Konorski, J. (1967). *Integrative activity of the brain: An interdisciplinary approach*. Chicago, IL: University of Chicago Press.

Lesch, K. P., & Merschdorf, U. (2000). Impulsivity, aggression, and serotonin: A molecular psychobiological perspective. *Behavioral Sciences and the Law*, *18*, 581–604.

Lolordo, V. M. (1969). Positive conditioned reinforcement from aversive situations. *Psychological Bulletin*, *72*, 193–203.

Lorrain, D. S., Riolo, J. V., Matuszewich, L., & Hull, E. M. (1999). Lateral hypothalamic serotonin inhibits nucleus accumbens dopamine: Implications for sexual satiety. *Journal of Neuroscience*, *19*, 7648–7652.

Lucki, I., & Harvey, J. A. (1979). Increased sensitivity to *d*- and *l*-amphetamine action after midbrain raphe lesions as measured by locomotor activity. *Neuropharmacology*, *18*, 243–249.

Mahadevan, S. (1996). Average reward reinforcement learning: Foundations, algorithms, and empirical results. *Machine Learning*, *22*, 159–196.

Martin, G. R., Eglen, R. M., Hoyer, D., Hamblin, M. W., & Yocca, F. (Eds.), (1998). *Advances in serotonin receptor research: Molecular biology, signal transduction, and therapeutics*. New York: New York Academy of Sciences.

Masand, P. S., & Gupta, S. (1999). Selective serotonin-reuptake inhibitors: An update. *Harvard Review of Psychiatry*, *7*, 69–84.

Mason, P. (1997). Physiological identification of pontomedullary serotonergic neurons in the rat. *Journal of Neurophysiology*, *77*, 1087–1098.

Mirenowicz, J., & Schultz, W. (1996). Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature*, *379*, 449–451.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*, 1936–1947.

Osgood, C. E. (1953). *Method and theory in experimental psychology*. New York: Oxford University Press.

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.

Parsons, L. H., & Justice, J. B., Jr. (1993). Perfusate serotonin increases extracellular dopamine in the nucleus accumbens as measured by in vivo microdialysis. *Brain Research*, *606*, 195–199.

Pucilowski, O. (1987). Monoaminergic control of affective aggression. *Acta Neurobiologiae Experimentalis*, *47*, 213–238.

Puterman, M. L. (1994). *Markov decision processes: Discrete stochastic dynamic programming*. New York: Wiley.

Rescorla, R. A. (1969). Pavlovian conditioned inhibition. *Psychological Bulletin*, *72*, 77–94.

Salamone, J. D., Cousins, M. S., & Snyder, B. J. (1997). Behavioral functions of nucleus accumbens dopamine: Empirical and conceptual problems with the anhedonia hypothesis. *Neuroscience and Biobehavioral Reviews*, *21*, 341–359.

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

Schultz, W., Dayan, P., & Montague, P. R. (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. *Journal of Neurophysiology*, *57*, 201–217.

Schwartz, A. (1993). A reinforcement learning method for maximizing undiscounted rewards. *Proceedings of the Tenth International Conference on Machine Learning*, San Mateo, CA: Morgan Kaufmann, pp. 298–305.

Segal, M. (1976). 5-HT antagonists in rat hippocampus. *Brain Research*, *103*, 161–166.

Solomon, R. L., & Corbit, J. D. (1974). An opponent-process theory of motivation. I. Temporal dynamics of affect. *Psychological Review*, *81*, 119–145.

Solomon, P. R., Nichols, G. L., Kiernan, J. M., 3rd, Kamer, R. S., & Kaplan, L. J. (1980). Differential effects of lesions in medial and dorsal raphe of the rat: Latent inhibition and septohippocampal serotonin levels. *Journal of Comparative and Physiological Psychology*, *94*, 145–154.

Soubrié, P. (1986). Reconciling the role of central serotonin neurons in

human and animal behavior. *Behavioral and Brain Sciences*, *9*, 319–335.

Stahl, S. M. (2000). *Essential psychopharmacology: Neuroscientific basis and practical applications* (2nd ed). New York: Cambridge University Press.

Stanford, S. C. (Ed.), (1999). *Selective serotonin reuptake inhibitors (SSRIs): Past, present, and future*. Austin, TX: R.G. Landes.

Sutton, R. S. (1988). Learning to predict by the methods of temporal difference. *Machine Learning*, *3*, 9–44.

Sutton, R. S., & Barto, A. G. (1990). Time-derivative models of Pavlovian conditioning. In M. Gabriel, & J. W. Moore (Eds.), *Learning and computational neuroscience* (pp. 497–537). Cambridge, MA: MIT Press.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning*. Cambridge, MA: MIT Press.

Sutton, R. S., Precup, D., & Singh, S. (1999). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, *112*, 181–211.

Tadepalli, P., & Ok, D. (1998). Model-based average reward reinforcement learning. *Artificial Intelligence*, *100*, 177–224.

Tobler, P. N., Dickinson, A., & Schultz, W. (2001). Reward prediction versus attention: The phasic activity of dopamine neurons in a conditioned inhibition task. *Society for Neuroscience Abstracts*, *27*, 421.5.

Tsitsiklis, J. N., & Van Roy, B. (1999). Average cost temporal-difference learning. *Automatica*, *35*, 1799–1808.

Vertes, R. P. (1991). A PHA-L analysis of ascending projections of the dorsal raphe nucleus in the rat. *Journal of Comparative Neurology*, *313*, 643–668.

Vertes, R. P., Fortin, W. J., & Crane, A. M. (1999). AM Projections of the median raphe nucleus in the rat. *Journal of Comparative Neurology*, *407*, 555–582.

Waddington, J. L. (1989). Functional interactions between D-1 and D-2 dopamine receptor systems: Their role in the regulation of psychomotor behaviour, putative mechanisms, and clinical relevance. *Journal of Psychopharmacology*, *3*, 54–63.

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

Westenberg, H. G. M., den Boer, J. A., & Murphy, D. L. (Eds.), (1996). *Advances in the neurobiology of anxiety disorders*. New York: Wiley.

Wilkinson, L. S., Humby, T., Robbins, T. W., & Everitt, B. J. (1995). Differential effects of forebrain 5-hydroxytryptamine depletions on Pavlovian aversive conditioning to discrete and contextual stimuli in the rat. *European Journal of Neuroscience*, *7*, 2042–2052.

Williams, D. A., & Overmier, J. B. (1988). Some types of conditioned inhibitors carry collateral excitatory associations. *Learning and Motivation*, *19*, 345–368.

Zangen, A., Nakash, R., Overstreet, D. H., & Yadid, G. (2001). Association between depressive behavior and absence of serotonin–dopamine interaction in the nucleus accumbens. *Psychopharmacology*, *155*, 434–439.

Zimmer-Hart, C. L., & Rescorla, R. A. (1974). Extinction of Pavlovian conditioned inhibition. *Journal of Comparative and Physiological Psychology*, *86*, 837–845.