

A Model of Antipsychotic Action in Conditioned Avoidance: A Computational Approach

Andrew Smith¹, Ming Li², Sue Becker¹ and Shitij Kapur^{*,3}

¹Psychology Department, McMaster University, Hamilton, Ontario, Canada; ²Centre for Addiction and Mental Health, Toronto, Ontario, Canada;

³Canada Research Chair and Professor of Psychiatry, University of Toronto, Vice President Research, Centre for Addiction and Mental Health, Toronto, Ontario, Canada

The selective ability of antipsychotic drugs (APDs) to attenuate conditioned avoidance responding (CAR) has been recognized for over 50 years. However, most efforts to account for this finding have been either neurochemically oriented (focusing on the neuromodulator dopamine) or behavioral, with little effort invested in uniting the two within a computational model. In this paper we propose a computational model, based on concepts from formal reinforcement learning theory, which accounts for the basic finding that noncataleptic doses of APDs disrupt avoidance without disrupting escape. The model formally separates out sensory, motor, and reward processes, and makes novel predictions pertaining to the dose- and time-dependent effects of APDs on response latencies—predictions which we verified in experimental studies using four different APDs (haloperidol, chlorpromazine, risperidone, and clozapine). The APD action in this model is most consistent with an effect on ‘expected future reward’—an idea closely linked to motivational drives and consistent with several leading theories of dopamine action.

Neuropsychopharmacology (2004) **29**, 1040–1049, advance online publication, 3 March 2004; doi:10.1038/sj.npp.1300414

Keywords: dopamine; schizophrenia; antipsychotics; modeling; neural networks; motivation; conditioned avoidance

INTRODUCTION

Conditioned avoidance response (CAR) is one of most important preclinical animal models in the study of antipsychotic drugs (APDs) (Kilts, 2001; Wadenberg and Hicks, 1999). In a typical CAR experiment, a rat is placed in a two-compartment shuttle box and presented with a neutral conditioned stimulus (CS) such as a light or tone, followed after a short delay by an aversive unconditioned stimulus (US), such as a foot-shock. The animal may *escape* the US when it arrives by running from one compartment to the other. However, after several presentations of the CS–US pair, the animal typically runs during the CS and before the onset of the US, thereby *avoiding* the US altogether. Animals treated with low (noncataleptic) doses of APDs fail to perform avoidance responses to the CS, even though their escape response to the shock itself is relatively unaffected (Ader and Clink, 1957; Arnt, 1982; Cook and Catania, 1964; Cook and Weidley, 1957; Courvoisier, 1956; Davidson and Weidley, 1976; Ponsluns, 1962). This selective

disruption of avoidance is characteristic of *all* APDs, but neither anxiolytics nor antidepressants show this effect (Courvoisier, 1956; Morpurgo, 1965; Reynolds and Czudek, 1995). Furthermore, the ability of an APD to suppress CAR has been shown to be closely correlated with its clinical potency (with respect to its treatment of psychosis) (Arnt, 1982; Janssen *et al*, 1965). Therefore, suppression of avoidance in CAR is correlated with the specific antipsychotic action of APDs.

From a neurochemical perspective, it has been established that the blockade of the dopamine D2 receptor is strongly implicated APD-induced disruption of avoidance (Wadenberg *et al*, 2000, 2001). However, despite the fact that CAR has been used to study APDs for nearly 50 years, no consensus has been reached regarding the underlying *behavioral* or *psychological* processes. Early work proposed a role for APDs in inhibiting internal ‘fear’ or ‘anxiety’ (Cook and Weidley, 1957; Davis *et al*, 1961; Hunt, 1956; Miller *et al*, 1957). Later work shifted the focus to the motor impairment effect of APDs (Beninger *et al*, 1980a,b; Cook and Catania, 1964; Fibiger *et al*, 1976; Grilly *et al*, 1984; Morpurgo, 1965; Ponsluns, 1962), and indeed the currently dominant explanation of APD-induced avoidance deficits is still the ‘motor initiation’ hypothesis (Aguilar *et al*, 2000; Ogren and Archer, 1994). Other suggestions have included an APD-induced reduction in responsiveness to external stimuli (Dews and Morse, 1961), a decrease in sensory afferent stimulation (Irwin, 1958; Key, 1961), a loss of

*Correspondence: Dr S Kapur, Canada Research Chair and Professor of Psychiatry, University of Toronto, Vice President Research, Centre for Addiction and Mental Health, 250 College Street, Toronto, Ontario, M5T 1R8 Canada, Tel: +1 416 979 6890, Fax: +1 416 260 4206, E-mail: shitij_kapur@camh.net

Received 30 October 2003; accepted 06 January 2004

Online publication: 24 January 2004 at <http://www.acnp.org/citations/Npp01210403504/default.pdf>

attention or arousal (Low *et al*, 1966), and a decrease in incentive motivation (Beninger, 1989).

In this paper, we attempt to understand the role of APDs in CAR by presenting a simple computational model. Our model not only simulates the selective, dose-dependent effects of ADPs on avoidance but also makes a number of novel predictions pertaining to the effect of APDs on avoidance latency. These predictions are tested and verified using experimental studies.

THE MODEL

The model is based upon the assumption that during conditioning an animal builds an explicit internal model of its environment as suggested in Figure 1. Although the types of representations used by animals are many and varied (Balleine *et al*, 1995; Berridge and Robinson, 1998; Cardinal *et al*, 2002; Dickinson, 1980), the evidence that animals internally represent action-outcome relationships (among others) is compelling (Balleine *et al*, 1995; Cardinal *et al*, 2002; Dickinson, 1980, 1987; Dickinson *et al*, 1983). These representations form the basis of our model. In Figure 1, the inclusion of intermediate 'wait' states between the CS and the US allows for an abstract representation of the passage of time. The fact that animals are able to represent time as part of the conditioning process, and also that dopamine modulation interacts with this representation is suggested by studies such as Richards *et al* (1999) and Wade *et al* (2000).

We will use these representational conveniences to capture two important features of dopamine manipulation. Firstly, dopamine manipulation can apparently affect the *expression* of behaviors independently of their acquisition. For example (Cousins *et al*, 1996), trained rats in a T-maze in which the two arms of the maze contained different amounts of food. The arm containing the greater quantity of food was obstructed by a wall, yet was normally preferred by rats because of the larger payoff. Dopamine blockade administered after training produced a switch in preference from one arm to the other. Other examples of dopamine manipulations affecting the *expression* of previously acquired behaviors are found in Berridge and Robinson (1998), Cousins *et al* (1996), Dickinson *et al* (2000), Fowler *et al* (1986), Grilly *et al* (1984), Heyden and Bradford (1988), Maffii (1959), Rolls *et al* (1974), Salamone *et al* (1991, 1993)

and Wadenberg *et al*, 2001). Secondly, dopamine manipulation appears to influence behavior differently depending on that behavior's temporal/instrumental relationship to the outcome. The T-maze experiment of Cousins *et al* (1996) is also a good example of this phenomena. The differential effect of APDs on primary vs secondary conditioned avoidance (Maffii, 1959) is another good example which we consider shortly. Other examples are found in Heyden and Bradford (1988), Richards *et al* (1999), Rolls *et al* (1974), Salamone *et al* (1991, 1994) and Wade *et al* (2000).

Although abstract, the model still bears some similarity to neurophysiological reality. For example, the states can be interpreted as abstract neurons or ensembles of neurons and the transitions between states as connections between neurons. The effects of dopamine will be simulated by temporarily modulating the strength of those connections as activation is passed from one state to another. This builds on previous suggestions that dopamine acts as a *gain modulator* of expressed behavior (Braver *et al*, 1995; Servan-Schreiber and Blackburn, 1995; Servan-Schreiber *et al*, 1990).

Assuming that an internal model of the task has been constructed as in Figure 1, it can be used to motivate behavior in the following way. Let us say, for example, that a hypothetical rat is presented with the CS. The 'CS' state in the model is then given an *activation* of 1 in order to represent this fact. We write this: $A(\text{CS}) = 1$, where A is called the activation function. The model can now be used to generate the *expected future reward* associated with each of the two available actions (Run and Do Nothing) by hypothetically playing through the consequences of those actions. In the case of evaluating the 'Run' action, this involves propagating the activation, $A(\text{CS})$, to the 'Safety' state. The amount of activation that actually reaches the 'Safety' state is directly proportional to the strength of the transition connection between the 'CS' state and the 'Safety' state. We write the strength of this connection, $T(\text{CS}, \text{Run})$, where T is called the transition function. Now, the activation of the 'Safety' state can be written as

$$A(\text{Safety}) = A(\text{CS}) \times T(\text{CS}, \text{Run}) \quad (1)$$

The idea is that $T(\text{CS}, \text{Run})$ reflects the probability of 'Safety' indeed being the consequence of taking the 'Run' action when presented with the 'CS'. For our purposes, all transition connections have a value of 1, but the model can be generalized to partial reinforcement schedules. Now we

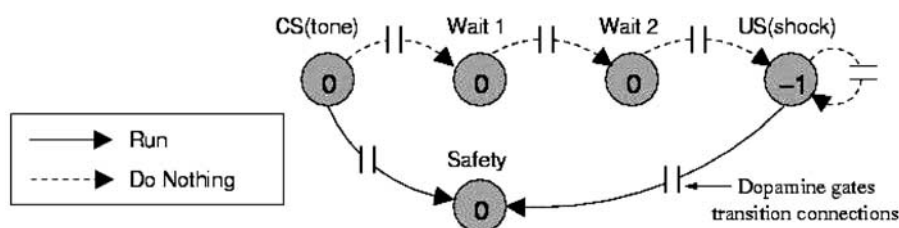


Figure 1 We propose that an animal builds an internal model of its environment, through trial and error interaction with the CAR task. Such a model comprises three core components: states, rewards, and transitions. The circles represent states that the animal can be in, the number inside each state represents the amount of reward (or in this case punishment) received in that state, and the arrows denote the consequence of taking each action (bottom left) in each state. These arrows represent the *transition function*, which is modulated (or gated) by dopamine (see vertical bars). The wait states are internal states for which there is no external cue, and allow the model to represent the delay between CS onset and the shock. The 'Safety' state is a terminal state at which the trial is ended. Following formal reinforcement learning methods we interpret punishment as negative reward.

propose a role for dopamine in modulating the efficacy of these transition connections. This is easily achieved by replacing equation (1) with

$$A(\text{Safety}) = A(\text{CS}) \times T(\text{CS}, \text{Run}) \times D \quad (2)$$

where D represents the global availability of dopamine. We assume that normal levels of dopamine receptor occupancy are represented by setting $D=1$, and complete dopamine blockade by setting $D=0$.

Since the 'Safety' state marks the end of the trial, the total expected future reward of taking the 'Run' action when in the 'CS' state can be calculated as

$$\text{Future_Reward}(\text{CS}, \text{Run}) = A(\text{Safety}) \times \text{reward}(\text{Safety}) \quad (3)$$

where $\text{reward}(\text{Safety})$ is the value inside the 'Safety' state in Figure 1. The idea is that the consequences of taking an action in a state can be calculated by hypothetically traversing an internal representation of the environment, and that this process is achieved by propagating activation from one abstract neural representation to another. In the case of evaluating the 'Run' action in the 'CS' state, this can be achieved by first applying equation (2), and then applying equation (3). The result is simply that $\text{Future_Reward}(\text{CS}, \text{Run}) = 1 \times 1 \times 1 \times 0 = 0$.

A similar process can be performed to calculate the expected future reward of doing nothing in the 'CS' state. This time, the activation has to be propagated through the two wait states. First we calculate the activation of 'Wait 1':

$$A(\text{Wait 1}) = A(\text{CS}) \times T(\text{CS}, \text{Do nothing}) \times D \quad (4)$$

Then this activation is propagated to 'Wait 2':

$$A(\text{Wait 2}) = A(\text{Wait 1}) \times T(\text{Wait 1}, \text{Do nothing}) \times D \quad (5)$$

And finally, to the 'US' state:

$$A(\text{US}) = A(\text{Wait 2}) \times T(\text{Wait 2}, \text{Do nothing}) \times D \quad (6)$$

Then, in a similar manner to equation (3), the expected future reward of taking the 'Do Nothing' action when in the 'CS' state can be calculated by

$$\begin{aligned} \text{Future_Reward}(\text{CS}, \text{Do Nothing}) \\ = A(\text{US}) \times \text{reward}(\text{US}) \end{aligned} \quad (7)$$

The result is that $\text{Future_Reward}(\text{CS}, \text{Do Nothing}) = 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times -1 = -1$. All the 1s represent the cumulative effect of the appropriate transition connections, along with their modulation by D , as activation is propagated through the internal model. By comparing the results of equation (3) with equation (7), the action with the highest expected future reward can then be selected for execution. Expected future reward is the central component of all formal reinforcement learning techniques¹ because of its pivotal role in action selection (Sutton and Barto, 1998). This value is a natural analogy of motivation in animals (see (McClure et al, 2003) for example) providing we assume

¹ Such techniques are used to train machines through an artificial analogy of animal reinforcement learning. See Sutton and Barto (1998) for the definitive review, and Crites and Barto (1996), Mahadevan and Connell (1991) and Tesauro (1992, 1994) for celebrated applications (including training a computer to play backgammon at the highest human level, the solution of a complex elevator scheduling problem, and a robot learning application).

that animals are motivated to achieve future reward, and avoid future punishment.

Once an action has been selected, a new state in the model will be activated to reflect the change in the animal's environment as a result of that action. If the hypothetical animal receives a shock, then the 'US' state will be activated. At this point, the model is faced with another choice, and the whole process can be repeated, but this time with activation originating in the 'US' state and passing to all its subsequent states. In the case of the 'US' state, the only subsequent states are the 'US' itself and the 'Safety' state (depending on which action is being evaluated). The proximity of these states (in terms of the single transition required to get to them) will make the expected future reward associated with the 'US' more robust to APD-induced devaluation.

MODELING THE BASIC FUNCTION OF APDS IN CAR: DIFFERENTIAL EFFECTS ON PRIMARY AVOIDANCE, SECONDARY AVOIDANCE, AND ESCAPE

The most reliable finding pertaining to APDs in CAR is their ability to selectively disrupt the avoidance response. As an exemplar of this finding, we consider data from one of the earliest CAR studies (Maffii, 1959) (reviewed along with other classic APD studies in Dews and Morse, 1961). Maffii's training procedure consisted of presenting rats with a tone followed by a shock, where the appropriate avoidance response involved jumping out of the conditioning box onto a pole. However, not only did the rats learn to jump onto the pole in response to the tone, but after sufficient training they jumped onto the pole as soon as they were placed in the box, and before the tone was even presented. He termed this the *secondary* conditioned response (elicited by the environmental cue of the box itself), and climbing on the pole in response to the tone the *primary* avoidance response. The *escape* response refers to the rat climbing onto the pole when presented directly with the shock itself. When orally administered with various doses of chlorpromazine, dose-dependent decreases in the primary, secondary and escape responses were all observed. However, Maffii found that the doses required to disrupt secondary avoidance were significantly lower than those required to disrupt primary avoidance, which in turn were significantly lower than those required to disrupt escape response (see Figure 3 left).

Figure 2 shows two alternatives for the way in which an animal might internally represent this task. In one case the secondary stimulus (the environmental box cues) enters into a direct relationship with the shock, while in the other case the relationship is only indirect. In either case, the model suggests that the escape response is less susceptible to dopamine blockade than the primary avoidance response, and that the primary avoidance response is itself less susceptible to dopamine blockade than the secondary avoidance response. This is due to the increasing distance of the respective CS from the shock itself in the internal representation. For example, in order to evaluate the consequence of the 'Do Nothing' action in the 'Environmental Cue' state, activity must be passed over four (dopamine modulated) transition connections, while the

consequences of 'Do Nothing' in the 'auditory stimulus' and 'shock' states can be generated using only two transitions or one transition (respectively). Figure 3 demonstrates that by

assuming a typically shaped relationship between drug dose and dopamine receptor blockade, the model's predictions capture the qualitative nature of Maffii's results.

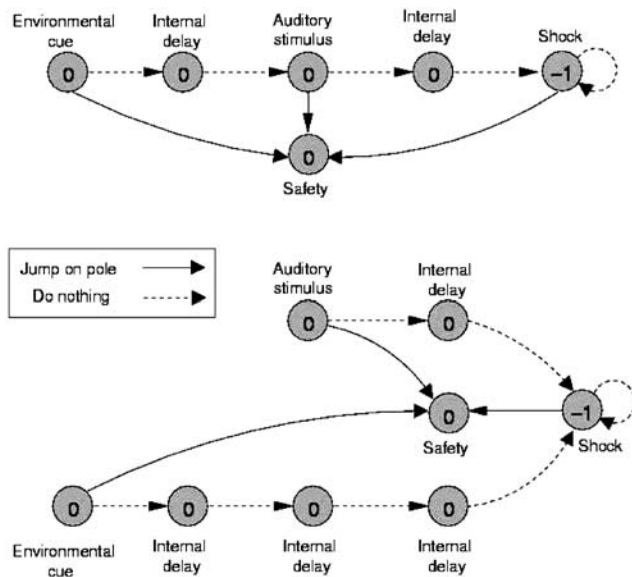


Figure 2 Two alternative suggestions for an animal's approach to modeling the conditioning experiment of Maffii (1959). (Top) The secondary CS (environment cues) enters into a relationship with the US (shock) via the primary CS (auditory stimulus). (Bottom) The secondary stimulus enters into a direct relationship with the US. We show both to demonstrate that the qualitative performance of the model is not sensitive to the finer arguments regarding the animal's perception of cause and effect. Also, the proposed role for dopamine is not linked to any particular model, but rather to a general process that acts on a task-specific representation. Arguments pertaining to the appropriateness of a specific representation (i.e. top or bottom) can be separated from those pertaining to the proposed dopaminergic process itself.

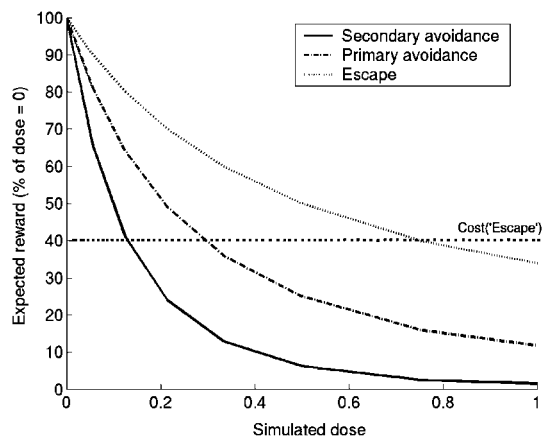
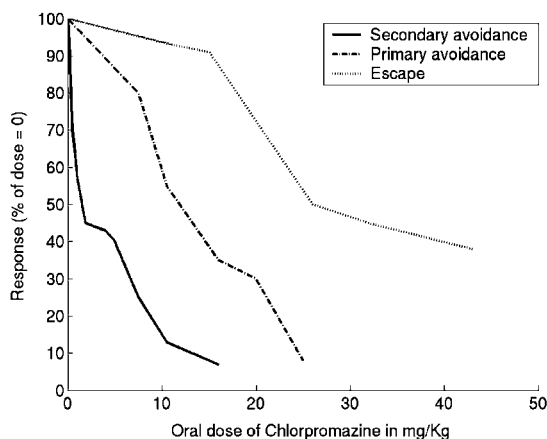


Figure 3 (Left) Number of secondary avoidance responses (climbing pole on being placed in cage before the onset of the tone), primary avoidance responses (climbing pole on presentation of the tone), and escape responses (climbing pole on shock) under increasing doses of chlorpromazine, as a percentage of the number of responses without the drug. Adapted from Maffii (1959). (Right) The qualitative nature of Maffii's results can be captured by the model. Expected future reward is calculated from the three important states: 'environment cue' (solid line), 'auditory stimulus' (dashed line), and 'shock' (dotted), under the 'Do Nothing' action, for decreasing values of D . This gives us a measure of the *incentive salience* of the relevant stimulus, and therefore also of the motivation to act. Although the relationship between D and the 'simulated dose' was hand picked to best fit the data, this relationship was fixed across the three curves and assumed the shape of a typical antipsychotic dose/ D_2 receptor occupancy curve. Since D is an abstract representation of dopamine, and the model does not attempt to address the underlying neurochemical processes, it is important to emphasize that it is the *qualitative* but not the *quantitative* performance of the model that is of interest. The horizontal line suggests an example escape cost. The model could be made to abolish a particular response when the respective curve falls below this threshold, capturing the ubiquitous experimental finding that avoidance is disrupted before escape.

NOVEL RESPONSE LATENCY PREDICTIONS UNDER APDS

We have suggested a simple computational model that, by maintaining an internal representation of its environment, is able to account for the basic CAR result that APDs disrupt avoidance before escape. To do this, we used the abstract notion of internal delay states to represent the temporal relationship between the CS and US. However, the model was only given the opportunity to produce an action in the externally activated states—that is, 'CS' and 'US'. A natural question arises as to how the model would perform if allowed to produce a response in *any* state, including the internal delay states.

To answer this question, we consider a slightly more complex, but generalized version of the basic model of Figure 1 in which we assume a distinct delay state for each second of time between the CS and US (10 s being a typical CS duration). Furthermore, we allow the model to make an action selection at each of these internally measured intervals (see Figure 4). The number of delay states used to represent the CS-US interval is arbitrary, and the qualitative nature of the results presented below could be achieved with a range of such states. The same mechanism as before is used for generating the expected future reward of each behavior in each state, except that now this value is interpreted as the *probability* of acting in that state. For example, if $D=1$ and all the transition connections are equal to 1, then the expected future reward associated with 'Do Nothing' in any state will be -1 , and the model will

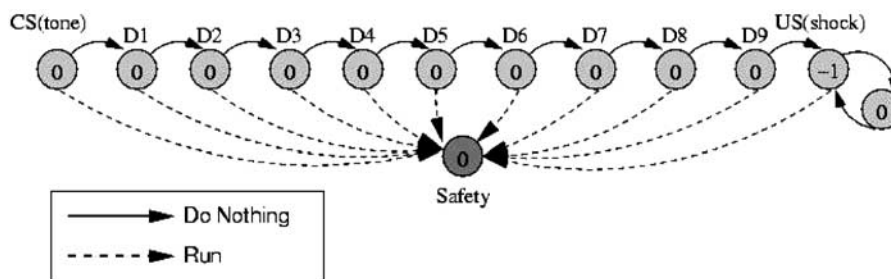


Figure 4 A generalized version of the model of Figure 1. We now assume that a distinct delay state is perceived for each second that elapses between the onset of the CS and the arrival of the US. Furthermore, the model is able to choose whether or not to act in each of these internal states. At any point during the simulation, one of the states is the current state. The onset of the tone makes the 'CS' state current, the presence of the shock makes the 'US' state current, and in the absence of either of these conditions the appropriate internal delay state (D1–D9) becomes current. For any current state, the expected future reward of either behavior ('Run' or 'Do Nothing') can be calculated using the method described previously. We include an additional delay state between the shock and itself in order to make a better quantitative account of the experimental data considered below.

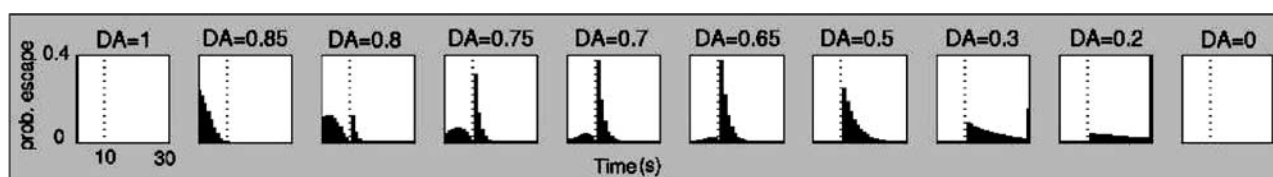


Figure 5 The performance of the model (with all transition connections = 1) under decreasing levels of dopamine (left to right). The CS onset occurs at the far left of each figure (0s), and the vertical line denotes the onset of the shock (10s). The trial is ended after a maximum of 30s (20s of shock). Each figure shows the probability of the model producing the 'Run' action during each 1s interval for the 30s of the trial. This value is determined for each second by first calculating the expected future reward associated with doing nothing in the current state (yielding the probability of producing the 'Run' response in this state), and then by multiplying this value by the probability of not having produced the 'Run' response in any of the previous intervals. The result is that the areas under the graphs sum to one, and each figure can be interpreted as a normalized frequency graph (with one second bins). Note that when $D = 1$, all the responses occur in the first second. Note also that the 30-s bin is used to catch all 'would be' responses from 30s onwards (simulating escape failures caused by trial termination at 30s in the experiment).

therefore choose to 'Run' with probability = 1. If the expected future reward is only -0.5 (because $D < 1$ for example), then the model will escape with probability 0.5, etc. Thus, expected future reward is now being used as a probabilistic interpretation of motivation.

We make one additional assumption. If the 'CS' state is current (onset of the CS has just occurred), but the expected future reward associated with doing nothing is close to zero (ie $\text{Future_Reward}(\text{'CS', 'Do Nothing'}) \approx 0$, perhaps because $D < 1$), then the CS is simply ignored. Furthermore, the CS is ignored with increasing probability as this value becomes closer to zero. The consequence of ignoring the CS is that the model will not activate the internal delay states and will remain dormant until the external stimulus of the shock itself arrives. In this instance, expected future reward is being interpreted as salience, and the intuitive principle is being invoked that if a stimulus is perceived as nonsalient, then there is no need to waste resources keeping track of the time since its onset. This assumption will allow the model to use the dotted line from Figure 3 (right) to produce a smooth shift from avoidance to escape under increasing levels of dopamine blockade.

This generalized model allows us to predict the detailed effect of APDs on the *pattern* of both avoidance and escape, particularly with respect to response latencies (Figure 5). For example, when $D = 1$, the model escapes in the first second of the trial with probability 1 (an unnatural state of affairs). As dopamine is reduced, three effects are observed.

Firstly, a smooth transition from avoidance (1–10s, before US onset) to escape (10–30s, after US onset) is observed. Secondly, and most significantly, the peak latency to produce an avoidance response (if one is produced at all) increases. This is particularly pronounced for $D = 0.75$ and 0.7. Thirdly, irrespective of the dopamine level, the peak latency to escape while being shocked always occurs on the first second of the shock (ignoring the escape failure effect at 30s). Hence, the model not only predicts an increase in the *mean* latency of both avoidance and escape (a standard finding), but more specifically it predicts a detailed change in the *pattern* of the two responses. Our model suggests that the difference in pattern between avoidance and escape is due to the interaction of APD induced dopamine blockade with the internal delay states.

Predicting Dose-Dependent Effects of APDs on Avoidance and Escape Response Latencies

We were unable to find any detailed published experimental data on the response latencies of APD-treated animals since most studies traditionally report only avoidance percentages or mean avoidance latencies. While our group has previously published several CAR experiments, we had also adhered to this tradition, and had not evaluated the data specifically for response latencies. Therefore, in order to test the novel predictions of the model, we reanalyzed previously collected behavioral data. Although these data

were collected for a different study (currently unpublished), the experimental method was the same as that used in a previous study (Wadenberg *et al*, 2001), with the exception of some minor procedural differences (intertrial interval, number of training trials, etc). These discrepancies would not be expected to affect the qualitative nature of the results. A summary of these methods is provided below. Four different drugs were used: a typical antipsychotic that is specific in its dopamine blockade (haloperidol), a typical antipsychotic that blocks several neurochemical systems including dopamine, serotonin, and the adrenergic system (chlorpromazine), and two atypical antipsychotics (risperidone and clozapine), where the latter is known not to induce either catalepsy or extrapyramidal side effects. For each drug, three different doses were used—these doses were chosen on the basis of previous experience to provide a range of actions on avoidance/escape/failure indices.

For each drug tested, 20 naive Sprague–Dawley rats were trained (drug free) over 4 days (40 trials per day) in a two-way avoidance paradigm. At this point, between seven and 10 animals had reached an 80% avoidance criterion and were selected for subsequent APD testing under various doses which were administered subcutaneously. Each trial (during both training and APD-testing) consisted of the presentation of a white noise for 10 s, followed immediately by a continuous scrambled foot-shock for up to 20 s. If a rat changed compartments during the white noise then no shock was administered and an avoidance response was recorded. If the animal changed compartments during the shock then the shock was terminated and an escape response recorded.

Recall that the main prediction from Figure 5 is that there should be a dose-dependent effect of APDs on peak latency to avoid, but not on peak latency to escape. These predictions are qualitatively validated by the comparison between model performance and experimental data shown in Figure 6. Not only does increasing APD dose increase the peak avoidance latency, but even under high doses, where avoidance is abolished and escape itself is significantly affected, the peak latency to escape (10–30 s) still occurs during the first few seconds following shock onset. This helps to rule out the explanation that the increase in avoidance latency is caused by a drug-induced fixed motoric cost.

Predicting the Time Course of APD Effects on Latency

The model also predicts that as APDs take effect, there will be a shift from left to right along Figure 5—an effect that will be smoothly reversed as the drug wears off. This is a straightforward consequence of the assumption that D2-receptor blockade will slowly increase following APD administration up to a maximum occupancy level, followed by a slow reverse of this process. Since there are no published data on this matter, these model-driven predictions were tested by analyzing the same data as described above. Figure 7 shows how the distributions of avoidance and escape responses vary with time since (subcutaneous) administration for both typical and atypical APDs. Since the drugs are most effective at blocking the avoidance response at 20 and 90 min following administration (Wadenberg *et al*, 2001), we assume that dopamine is maximally blocked at

these times (see also Wadenberg *et al*, 2000). Although there is considerable variance across the four drugs, the important observation is that as the APDs take effect, the peak latency of the remaining avoidance responses (1–10 s) increases, and this effect is reversed as the drug wears off. In contrast, this shift in peak latency is not observed for the escape response, validating the predictions of the model.

DISCUSSION

We have presented a model that was originally developed to account for the classic effects of APDs on avoidance *vs* escape. However, the model was also able to generate testable predictions pertaining to response latency profiles, and these predictions were empirically verified for a range of typical and atypical APDs.

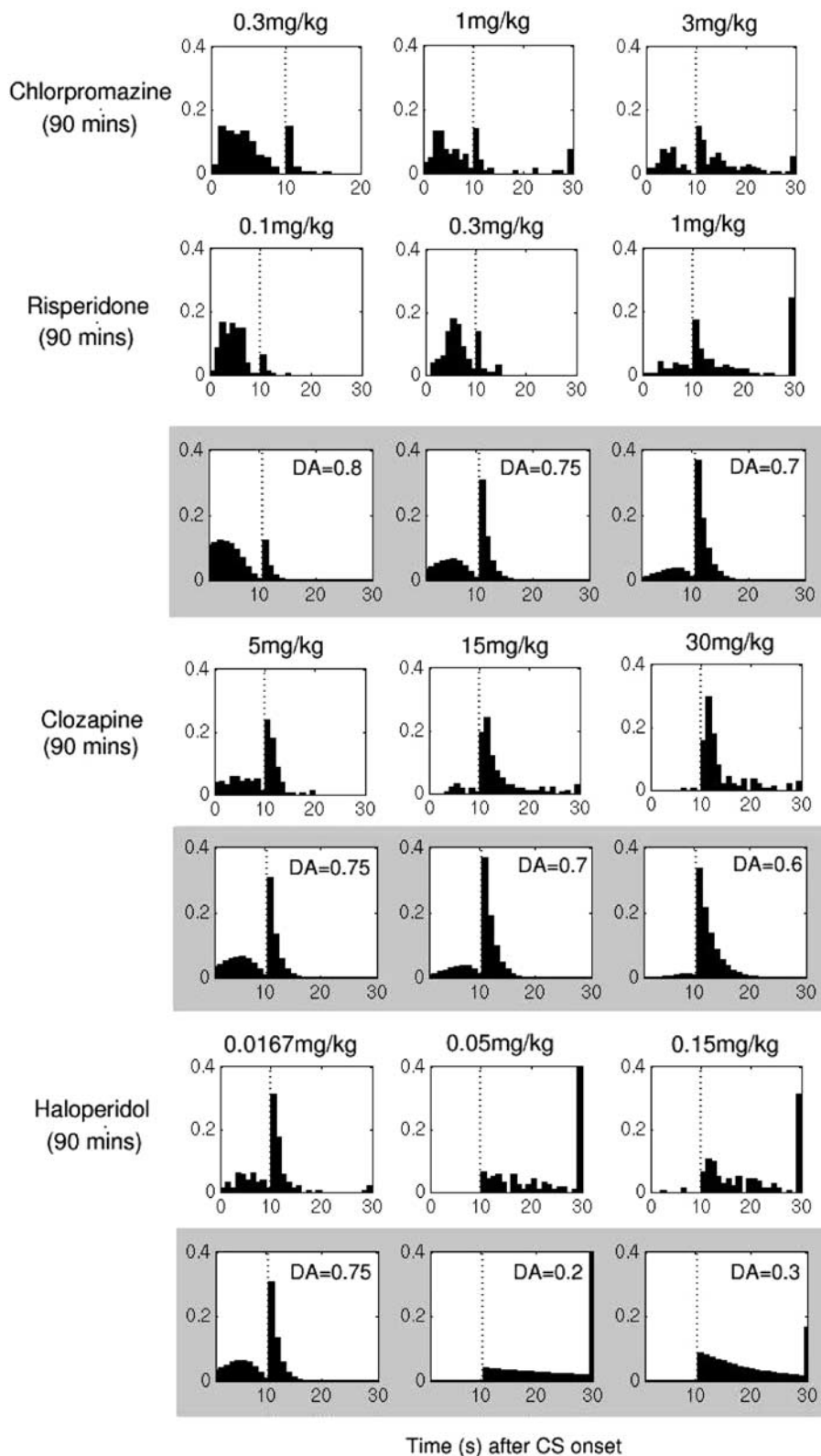
While our model is an abstract one, several elements of it can be related to what physiologists and psychologists would recognize as sensory, motoric, and reward processes. Sensory processes can be equated with the model's ability to identify and activate the representation of the current state. Motoric processes can be equated with the model's ability to take a prescribed action. While reward processes are complex and increasingly seen as a multidimensional construct (Berridge and Robinson, 2003), the ability of the model to perceive the actual reward value associated with each state can be equated with 'liking' or hedonia (Berridge and Robinson, 1998). However, the model proposes that APDs are acting not in any of these processes, but in the generation of expected future reward. Expected future reward is used by all formal reinforcement learning methods to drive action selection, and is therefore a natural analogy of both motivation and incentive salience (or 'wanting' in the terminology of Berridge and Robinson (1998); Mc Clure *et al* (2003)). Our model of APD action in CAR is therefore consistent with the claim that *the selective disruption of the avoidance response is due to the APD-induced impairment of the motivational processes of the animal via blockade of dopamine neurotransmission*.

Our approach is different to the Temporal Difference (TD) prediction-error hypothesis (Houk *et al*, 1995; Schultz *et al*, 1997), which suggests that the phasic dopamine response signals the *difference* ('error') between the future reward predicted by the animal and the actual reward received by the animal. This error is then used exclusively to drive the learning process forwards in a biologically plausible manner (Waelti *et al*, 2001). In contrast, our role for dopamine is in the generation of expected future reward completely independently of the acquisition process. The advantages of our approach are that we can model not only the effect of dopamine manipulation on the *expression* of previously acquired behavior but also the sensitivity of this effect to the relationship between CS (or action) and US (or outcome).

The disadvantage of our approach is that we have not modeled the acquisition process itself (ie the *construction* of Figures 1, 2, and 4), in which dopamine is likely to play an important role. We therefore suggest that by combining the prediction error hypothesis of dopamine with our proposed gating role for dopamine, it might be possible to address a wider range of behavioral data pertaining to both

acquisition and expression. Towards this end, a number of dopamine models have been suggested that combine TD-based representations with explicit internal model representations (Daw *et al*, 2003; Dayan, 2002; Dayan and Balleine, 2002; Suri and Schultz, 1998; Suri, 2001, 2002;

Suri *et al*, 2001). However, a significant challenge remains in bridging the gap between models of dopamine neuron firing, and models of behavioral and psychological phenomena in which dopamine may play a pivotal role.



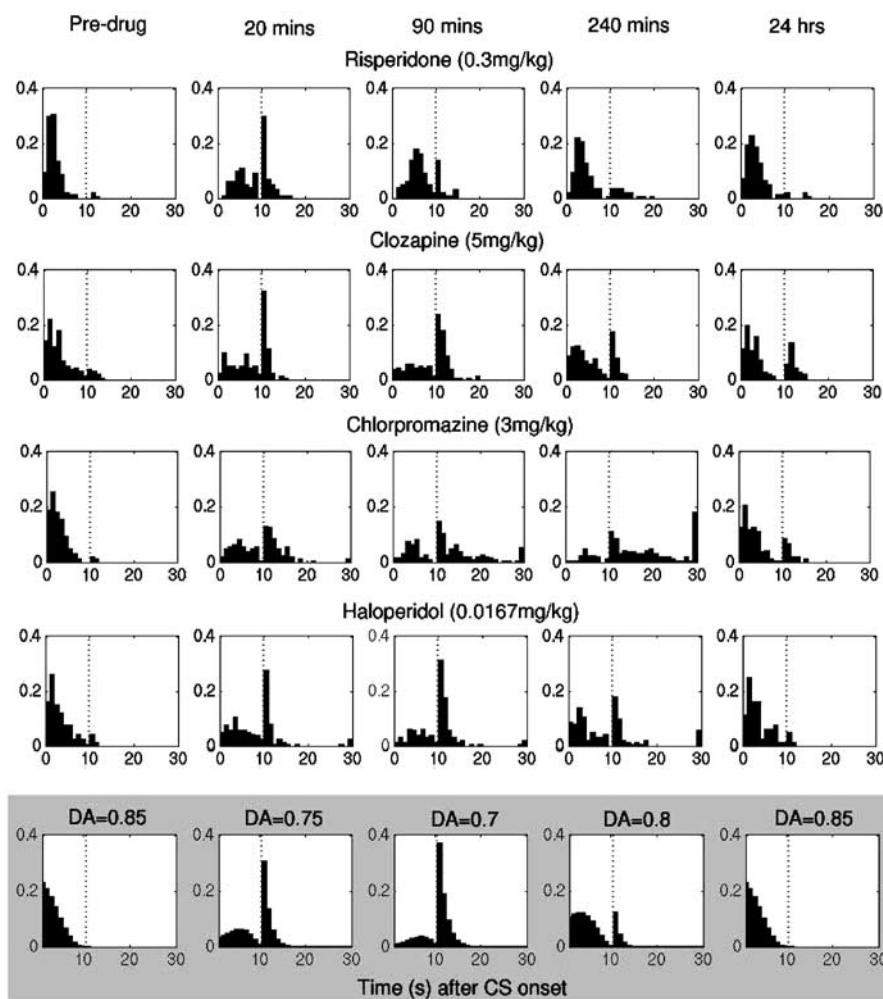


Figure 7 A comparison of animal performance (white background) at various intervals following APD administration with model performance (gray background) for various values of $D < 1$ (selected from Figure 5). Doses were selected from the available data so that avoidance was impaired but not abolished allowing analysis of avoidance latencies. Experimental method as for Figure 6.

By hypothesizing that dopamine modulates the efficacy of internal transition connections between environmental states, we have been able to produce a model that extends the performance of an existing model of neuroleptic action in CAR (Servan-Schreiber and Blackburn, 1995) to account for secondary avoidance and to predict novel latency data (Figures 6 and 7). However, apart from predicting novel data and aiding our understanding of psychological processes, the other desirable property of a computational model is an ability to unite existing cognitive hypotheses

within a formal framework. For example, Berridge and Robinson (1998) have suggested that dopamine mediates the *wanting* component of reward as distinct from the *liking*, and in our model, following McClure *et al* (2003), we interpret expected future reward as precisely this *incentive salience* or *wanting* (although 'reversed' for the purposes of an aversive paradigm). Also, based on anatomical and neurophysiological data, Horvitz (2000, 2002) have suggested that dopamine may play a *gating* role between the sensory, motor and reward-based afferents that converge on

Figure 6 A comparison of model performance (gray background) for various values of $D < 1$ (selected from Figure 5) with animal performance (white background) 90 min after the administration of various doses of typical and atypical APDs (4 days of drug-free acquisition had previously taken place). Experimental procedure identical to Wadenberg *et al* (2001) except where described differently in the text. The animal graphs show the frequency (1 s bins) of avoidance/escape summed over between 7 and 10 rats (different groups of rats were used for each of the four drugs). Each animal graph consists of data from 20 consecutive trials with a random intertrial interval. Each graph is therefore constructed from between $7 \times 20 = 140$ and $10 \times 20 = 200$ individual trials. The bars represent the number of responses in each whole second following the onset of the CS, normalized so that the total height of the bars = 1. This provides a discrete approximation to a probability density function, which can be compared with the performance of the model. Escape failures caused by trial termination at 30 s are simply added to the 30 s bin. Note that an unexpected large number of avoidance and escape failures are observed for the 0.05 mg/kg dose of Haloperidol, and these even exceed those observed at 0.15 mg/kg. We do not have a simple explanation for this anomaly. Whatever the cause for this quantitative anomaly (misdosing, higher bioavailability, extra individual sensitivity), the qualitative correspondence between the model's predictions and experimental data can still be obtained by assuming a higher level of blockade at the 0.05 mg/kg dose.

the D2 receptor-rich striatal region of the brain. Our model formally captures one interpretation of such gating. Finally, Salamone *et al* (1997) observe that it is in the areas of overlap between sensory, motor, and reward processes that dopamine's role is expressed, leading to the proposition that 'accumbens dopamine is important for responding to stimuli that are spatially and temporally distant from the organism' (p 353). This statement summarizes the psychological value of dopamine in our model, since the greater the number of intervening 'states' between action and reward outcome in the animal's internal world model, the more susceptible to APD-induced devaluation that action or stimulus will be. We are currently examining how this principle can be used to account for a range of additional data from appetitive paradigms in which dopamine manipulation is shown to selectively influence motivation towards stimuli that are spatially, temporally, or instrumentally distant from the animal (Cousins *et al*, 1996; Richards *et al*, 1999; Rolls *et al*, 1974; Salamone *et al*, 1991, 1993, 1994; Wade *et al*, 2000). We are also currently considering the implications of this model for a psychological theory of APDs in psychosis and also in ADHD.

ACKNOWLEDGEMENTS

This work was primarily supported by an OMHF Special Initiative grant and a NET grant from the Canadian Institutes of Health Research. SK is additionally supported by a Canada Research chair.

REFERENCES

- Ader R, Clink DW (1957). Effects of chlorpromazine on the acquisition and extinction of an avoidance response in the rat. *J Pharmacol Exp Ther* **131**: 144–148.
- Aguilar MA, Mari-Sanmillan MI, Morant-Deusa JJ, Minarro J (2000). Different inhibition of conditioned avoidance response by clozapine and DA D1 and D2 antagonists in male mice. *Behav Neurosci* **114**: 389–400.
- Arnt J (1982). Pharmacological specificity of conditioned avoidance response inhibition in rats: inhibition by neuroleptics and correlation to dopamine receptor blockade. *Acta Pharmacol Toxicol (Copenh)* **51**: 321–329.
- Balleine BW, Garner C, Gonzalez F, Dickinson A (1995). Motivational control of heterogeneous instrumental chains. *J Exp Psychol: Anim Behav Process* **21**: 203–217.
- Beninger RJ (1989). The role of serotonin and dopamine in learning to avoid aversive stimuli. In: Archer T, Nilsson L (eds). *Aversion, Avoidance, and Anxiety: Perspective on Aversively Motivated Behavior*. Lawrence Erlbaum Associates: Hillsdale, NJ. pp 265–284.
- Beninger RJ, Mason ST, Phillips AG, Fibiger HC (1980a). The use of conditioned suppression to evaluate the nature of neuroleptic-induced avoidance deficits. *J Pharmacol Exp Ther* **213**: 623–627.
- Beninger RJ, Mason ST, Phillips AG, Fibiger HC (1980b). The use of extinction to investigate the nature of neuroleptic-induced avoidance deficits. *Psychopharmacology* **69**: 11–18.
- Berridge K, Robinson TE (2003). Parsing reward. *Trends Neurosci* **26**: 507–513.
- Berridge KC, Robinson TE (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res Rev* **28**: 309–369.
- Braver TS, Cohen JD, Servan-Schreiber D (1995). A computational model of prefrontal cortex function. *Adv Neural Inform Process Systems* **7**: 141–148.
- Cardinal RN, Parkinson JA, Hall J, Everitt BJ (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci Biobehav Rev* **26**: 321–352.
- Cook L, Catania AC (1964). Effects of drugs on avoidance and escape behavior. *Fed Proc* **23**: 818–835.
- Cook L, Weidley E (1957). Behavioral effects of some psychopharmacological agents. *Ann NY Acad Sci* **66**: 740–752.
- Courvoisier S (1956). Pharmacodynamic basis for the use of chlorpromazine in psychiatry. *Quart Rev Psychiatry Neurol* **17**: 25–37.
- Cousins MS, Atherton A, Turner L, Salamone JD (1996). Nucleus accumbens dopamine depletions alter relative response allocation in a T-maze cost/benefit task. *Behav Brain Res* **74**: 189–197.
- Crites RH, Barto AG (1996). Improving elevator performance using reinforcement learning. *Neural Inform Process Systems* **8**: 1017–1023.
- Davidson AB, Weidley E (1976). Differential effects of neuroleptic and other psychotropic agents on acquisition of avoidance in rats. *Life Sci* **18**: 1279–1284.
- Davis WM, Capehart J, Llewellyn WL (1961). Mediated acquisition of a fear-motivated response and inhibitory effects of chlorpromazine. *Psychopharmacologia* **2**: 268–276.
- Daw ND, Courville AC, Touretzky DS (2003). Timing and partial observability in the dopamine system. *Adv Neural Inform Process Systems* **16**: (in press).
- Dayan P (2002). Motivated reinforcement learning. In: Ghahramani TGDaSBaZ (ed). *Advances in Neural Information Processing System*. MIT Press: Cambridge, MA.
- Dayan P, Balleine BW (2002). Reward, motivation and reinforcement learning. *Neuron* **36**: 285–298.
- Dews PB, Morse WH (1961). Behavioral pharmacology. *Annu Rev Pharmacol* **1**: 145–174.
- Dickinson A (1980). *Contemporary Animal Learning Theory*. Cambridge University Press: Cambridge.
- Dickinson A (1987). Instrumental performance following saccharin pre-feeding. *Behav Process* **14**: 147–154.
- Dickinson A, Nicholas DJ, Adams CD (1983). The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *Quart J Exp Psychol* **35B**: 35–51.
- Dickinson A, Smith J, Mirennowicz J (2000). Dissociation of Pavlovian and instrumental incentive learning under dopamine antagonists. *Behav Neurosci* **40**: 468–483.
- Fibiger HC, Carter DA, Phillips AG (1976). Decreased intracranial self-stimulation after neuroleptics of 6-hydroxydopamine: evidence for mediation by motor deficits rather than by reduced reward. *Psychopharmacology* **47**: 21–27.
- Fowler SC, LaCerra MM, Ettenberg A (1986). Effects of haloperidol on the biophysical characteristics of operant responding: implications for motor and reinforcement processes. *Pharmacol, Biochem Behav* **25**: 791–796.
- Grilly DM, Johnson SK, Minardo R, Jacoby D, LaRiccia J (1984). How do tranquilizing agents selectively inhibit conditioned avoidance responding. *Psychopharmacology* **84**: 262–267.
- Heyden JAMvd, Bradford LD (1988). A rapidly acquired one-way conditioned avoidance procedure in rats as a primary screening test for antipsychotics: influence of shock intensity on avoidance performance. *Behav Brain Res* **31**: 61–67.
- Horvitz JC (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience* **96**: 651–656.
- Horvitz JC (2002). Dopamine gating of glutamatergic sensorimotor and incentive motivational input signals to the striatum. *Behav Brain Res* **137**: 65–74.

- Houk JC, Adams JL, Barto AG (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In: Beiser JCHAJLDA DG (ed). *Models of Information Processing in the Basal Ganglia*. MIT Press: Cambridge, MA. pp 249–270.
- Hunt HF (1956). Some effects of drugs on classical (type S) conditioning. *Ann NY Acad Sci* 65: 258–267.
- Irwin S (1958). Factors influencing acquisition of avoidance behavior and sensitivity to drugs. *Fed Proc* 17: 380.
- Janssen PAJ, Niemegeers CJE, Schellekens KHL (1965). Is it possible to predict the clinical effects of neuroleptic drugs (major tranquilizers) from animal data? *Arzneimittelforschung* 15: 104–117.
- Key BJ (1961). The effects of drugs on discrimination and sensory generalization of auditory stimuli in cats. *Psychopharmacologia* 2: 352–363.
- Kilts CD (2001). The changing roles and targets for animal models of schizophrenia. *Biol Psychiatry* 50: 845–855.
- Low LA, Eliasson M, Kornetsky C (1966). Effects of chlorpromazine on avoidance acquisition as a function of CS-US interval length. *Psychopharmacologia* 10: 148–154.
- Maffii G (1959). The secondary conditioned response of rats and effects of some psychopharmacological agents. *J Pharmacy Pharmacol* 11: 129–139.
- Mahadevan S, Connell J (1991). *Automatic Programming of Behaviour-Based Robots Using Reinforcement Learning Proceedings of the Ninth International Conference on Artificial Intelligence (AAAI 91)*. Anaheim, CA. pp 768–773.
- McClure SM, Daw N, Montague PR (2003). A computational substrate for incentive salience. *Trends Neurosci* 26: 423–428.
- Miller RE, Murphy JV, Mirsky A (1957). Persistent effect of chlorpromazine on extinction of an avoidance response. *AMA Arch Neurol Psychiatry* 78: 526–530.
- Morpurgo C (1965). Drug-induced modifications of discriminated avoidance behavior in rats. *Psychopharmacologia* 8: 91–99.
- Ogren SO, Archer T (1994). Effects of typical and atypical antipsychotic drugs on two-way active avoidance. Relationship to DA receptor blocking profile. *Psychopharmacology (Berlin)* 114: 383–391.
- Ponsluns D (1962). An analysis of chlorpromazine-induced suppression of the avoidance response. *Psychopharmacologia* 3: 361–373.
- Reynolds GP, Czudek C (1995). New approaches to the drug treatment of schizophrenia. *Adv Pharmacol* 32: 461–503.
- Richards JB, Sabol KE, Wit Hd (1999). Effects of methamphetamine on the adjusting amount of procedure, a model of impulsive behavior in rats. *Psychopharmacology* 146: 432–439.
- Rolls ET, Rolls BJ, Kelly PH, Shaw SG, Wood RJ, Dale R (1974). The relative attenuation of self-stimulation, eating and drinking produced by dopamine-receptor blockade. *Psychopharmacology* 38: 219–230.
- Salamone JD, Cousins MS, Bucher S (1994). Anhedonia or anergia? Effects of haloperidol and nucleus accumbens dopamine depletion on instrumental response selection in a T-maze cost/benefit procedure. *Behav Brain Res* 65: 221–229.
- Salamone JD, Cousins MS, Snyder BJ (1997). Behavioural functions of nucleus accumbens dopamine: empirical and conceptual problems with the anhedonia hypothesis. *Neurosci Biobehav Rev* 21: 341–359.
- Salamone JD, Kurth PA, McCullough LD, Sokolowski JD, Cousins MS (1993). The role of brain dopamine in response initiations: effects of haloperidol and regionally-specific dopamine depletions on the local rate of instrumental responding. *Brain Res* 628: 218–226.
- Salamone JD, Steinpreis RE, McCullough LD, Smith P, Grebel D, Mahan K (1991). Haloperidol and nucleus accumbens dopamine depletion suppress lever pressing for food but increase free food consumption in a novel food-choice procedure. *Psychopharmacology* 104: 515–521.
- Schultz W, Dayan P, Montague PR (1997). A neural substrate of prediction and reward. *Science* 275: 1593–1599.
- Servan-Schreiber D, Blackburn JR (1995). Neuroleptic effects on acquisition and performance of learned behaviors: a reinterpretation. *Life Sci* 56: 2239–2245.
- Servan-Schreiber D, Printz H, Cohen JD (1990). A network model of catecholamine effects: gain, signal to noise ratio, and behavior. *Science* 249: 892–895.
- Suri R, Schultz W (1998). Learning of sequential movements by neural network model with dopamine-like reinforcement signal. *Exp Brain Res* 121: 350–354.
- Suri RE (2001). Anticipatory responses of dopamine neurons and cortical neurons reproduced by internal model. *Exp Brain Res* 140: 234–240.
- Suri RE (2002). TD models of reward predictive responses in dopamine neurons. *Neural Networks: Special Issue on Computational Models of Neuromodulation* 15: 523–533.
- Suri RE, Bargas J, Arbib MA (2001). Modeling functions of striatal dopamine modulation in learning and planning. *Neuroscience* 103: 65–85.
- Sutton RS, Barto AG (1998). *Reinforcement Learning*. MIT Press: Cambridge, MA.
- Tesauro GJ (1992). Practical issues in temporal difference learning. *Machine Learning* 8: 257–277.
- Tesauro GJ (1994). TD-Gammon, a self-teaching backgammon program, achieves master-level play. *Neural Comput* 6: 215–219.
- Wade TR, Wit Hd, Richards JB (2000). Effects of dopaminergic drugs on delayed reward as a measure of impulsive behavior in rats. *Psychopharmacology* 150: 90–101.
- Wadenberg ML, Hicks PB (1999). The conditioned avoidance response test re-evaluated: is it a sensitive test for the detection of potentially atypical antipsychotics? *Neurosci Biobehav Rev* 23: 851–862.
- Wadenberg ML, Kapur S, Soliman A, Jones C, Vaccarino F (2000). Dopamine D2 receptor occupancy predicts catalepsy and the suppression of conditioned avoidance response behavior in rats. *Psychopharmacology (Berl)* 150: 422–429.
- Wadenberg M-LG, Soliman A, Vanderspek SC, Kapur S (2001). Dopamine D2 receptor occupancy is a common mechanism underlying animal models of antipsychotics and their clinical effects. *Neuropsychopharmacology* 25: 633–641.
- Waelti P, Dickinson A, Schultz W (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature* 412: 43–48.