A Computational Role for Dopamine Delivery in Human Decision-Making

Egelman, D.M., Person, C., Montague, P.R.

Division of Neuroscience, Baylor College of Medicine, 1 Baylor Plaza, Houston, TX, 77030
email: {david, read}@dirac.bcm.tmc.edu

Published in 1998 in ***Journal of Cognitive Neuroscience*. 10(5): 623-630.**

Introduction

Recent work suggests that fluctuations in dopamine delivery at target structures represent an evaluation of future events that can be used to direct learning and decision making. To examine the behavioral consequences of this interpretation, we gave simple decision making tasks to 66 human subjects and to a network based on a predictive model of mesencephalic dopamine systems. The human subjects displayed behavior similar to the network behavior in terms of choice allocation and the character of deliberation times. The agreement between human and model performances suggests a direct relationship between biases in human decision strategies and fluctuating dopamine delivery. We also show that the model offers a new interpretation of deficits that result when dopamine levels are increased or decreased through disease or pharmacological interventions. The bottom-up approach presented here also suggests that a variety of behavioral strategies may result from the expression of relatively simple neural mechanisms in different behavioral contexts.

Decision Making

Even for the simplest creatures, there are vast complexities inherent in any decision-making task. Nonetheless, any creature has limited available time in which to arbitrate decisions. Decision-making is likely to possess automatic components which may possess direct relationships to the underlying neural mechanisms. Previously, decision-making theories have been based on formal, top-down approaches that produced *normative*

strategies for decision makers, i.e., they prescribed strategies that *ought* to be followed under a predetermined notion of the goal (Bernoulli, 1738; Von Neumann and Morgenstern, 1947; Luce and Raiffa, 1957) (see endnote 1). Although normative accounts may produce functional descriptions of behavior that match experimental data, they do not yield a well-specified and testable relationship to potential neural substrates. Recent work suggests the existence of covert neural mechanisms that automatically and unconsciously bias decision-making in human subjects (Bechara, 1997). Consonant with this latter work, recent work on midbrain dopaminergic neurons suggests that their activity may participate in the construction of such covert signals, and thereby provide a more bottom-up explanation for decision-making strategies employed by animals (Montague, et al., 1995; Egelman, et al., 1995; Montague, et al., 1996; Schultz, et al., 1997; Egelman, et al., 1998).

Specifically, studies on neuromodulator delivery in behaving animals (Wise, 1980; Wise and Bozarth, 1984; Romo and Schultz, 1990; Schultz, 1992; Ljunberg, et al., 1992; Aston-Jones, 1994; Mirenowicz and Schultz, 1996) suggest that changes in dopamine delivery represent *errors in predictions* of the time and amount of future rewarding stimuli (Montague, et al., 1996). Models based on this interpretation account for physiological recordings from dopamine neurons in behaving primates (Montague, et al., 1996; Schultz, et al., 1997), and capture foraging behavior of bees (Montague, et al., 1995). This computational interpretation suggests that a behavioral meaning may be associated with dopamine delivery: increases from baseline release mean the current state is ‘better than expected’ and decreases mean the current state is ‘worse than expected’ (Quartz, 1992; Egelman, et al., 1995; Montague, et al., 1995; Montague, et al., 1996). In this paper, we explore the hypothesis that this behavioral interpretation of fluctuating dopamine delivery provides one simple bottom-up model of how dopaminergic (or related) projections implement general constraints that influence ongoing decision-making in humans. Such a model provides useful meeting grounds for the psychology and neurobiology underlying human decision making.

Methods

As described in the legend of Figure 1, our model is based on a simplified anatomy of the mesencephalic dopamine systems. We begin with the hypothesis that such an anatomy comes with commensurate computational principles (Quartz et al, 1992; Egelman et al, 1995; Montague et al, 1996; see also papers on temporal difference algorithms, e.g., (Sutton, 1987; Sutton, 1988; Sutton, et al., 1990)) (endnote 2). Specifically, we note that the rich arborizations of midbrain dopaminergic axons could deliver a global, scalar prediction error to the cortex. The cortex, driven by incoming polysensory information, could construct and deliver convergent neuronal activity to midbrain nuclei in the form of a temporal derivative. The output of a midbrain neuron is used in dual roles: (1) to update synaptic weights after each selection, and (2) to bias the *process* of making a selection. In other words, each option the model ``looks at'' has a commensurate pattern of cortical activity (which is filtered through associated weights); simply "considering" the choice (not selecting it) will generate the $\delta(t)$ signal, and such a signal is used to commit to decisions (see full description, Figure 1).

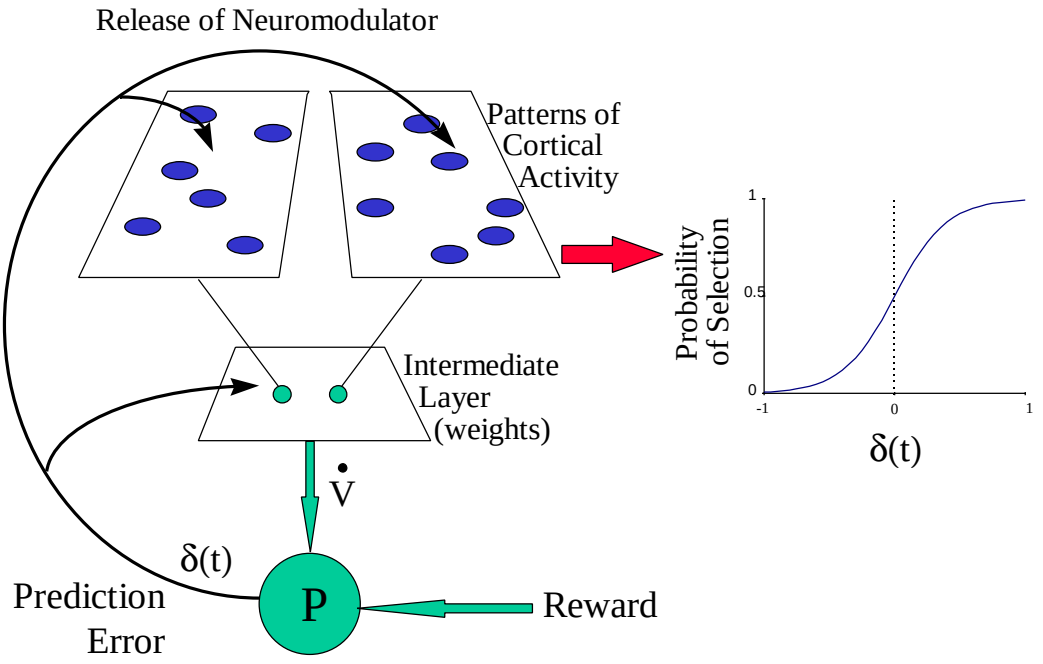

Figure 1. Bottom-up interpretation of decision-making model.
Choices A and B are represented by separate patterns of cortical activity, each associated with a modifiable weight *w(i,t)*, where ***i*** indexes A or B. In the figure, *w*s are represented by the 2 circles in the intermediate layer. ***P*** is a linear unit representing a midbrain neuron with output: $\delta(t) = r(t) + \dot{V}(t) + b(t)$

r(t) is input from pathways representing rewarding stimuli (marked "Reward" in figure), $\dot{V}$ (t) represents a scalar surprise signal which arrives from the cortex in the form of a temporal derivative of net excitatory activity, *b(t)* is *P*'s baseline activity level which is set to 0. Here, $\dot{V}$ (t) is taken as a one time-step difference V(t) - V(t-1) where $V(t) = \sum_i x(i,t)w(i,t)$, and *x(i,t)* is the activity associated with choice *i* at time *t*. In this case, there are only 2 *x*s, each representing one of the choices, and each using a binary activity level: 1 when a choice was being "considered", 0 otherwise. δ(t) is a signed quantity which we interpret as fluctuations in dopamine delivery to targets above (δ(t) > 0) and below (δ(t) < 0) baseline levels (see Montague et al, 1995, 1996). In this form, δ(t) is interpreted as an ongoing prediction error between the amount of reward expected and the amount actually received (Sutton, 1987; Sutton, 1988; Sutton, et al., 1990). This prediction error is used to direct selections and to update the weights *w(i,t)* (the internal model).

Making selections using ongoing prediction error. The model chooses among alternatives by making random transitions from one alternative to another which induces fluctuations in the output, δ(t), of neuron P. The output δ(t) controls the probability p_s of making a selection on a given transition (see endnotes 4 and 5):

$$p_s = \frac{1}{1 + \exp(-m\delta(t) + b)}$$

Updating the internal model. Weights *w* associated with each alternative *i* are updated (after a selection) according to the Hebbian correlation of *P*'s output with cortical activity: $w(i)_{new} = w(i)_{old} + \lambda x(i, t-1)\delta(t)$ where λ is the learning rate. Varying the network's parameters had little effect on the final behavioral outcome (endnote 4). The model relies on a linear predictor; however, it obtains a stochastic component to its decision behavior through the function p_s. A simple model suffices here because its basic principles are robust.

To highlight the behavioral consequences of such an interpretation of dopamine delivery, we designed variations of a two-choice decision task (Vaughan and Herrnstein, 1987; Herrnstein, 1990; Herrnstein, 1991) which was given to human subjects and to the network. The humans were required to select between two large buttons, labeled A and B, displayed on a computer screen. After each selection (with a mouse pointer), a vertical, red slider bar indicated the amount of reward obtained. Subjects were instructed to maximize their long-term return over 250 selections. There was no time limit for making choices. The reward earned at each selection was a function of past selections. Specifically, the computer kept track of the subject's last 40 choices, and the relative fraction of those choices (e.g., the percentage of selections that went to choice A) determined the amount of reward earned at the next selection of A or B. Shown in Figure 2 is the fraction of choices to A (of the last 40 selections) versus the reward to be earned if the next choice is A or B. Thus, each task amounted to a game wherein the subject's 'opponent' (the reward functions) employed a fixed strategy. The speed with which the 'opponent' responded to the subject's choices was defined by the window size over which the fraction of choices from button A was computed.

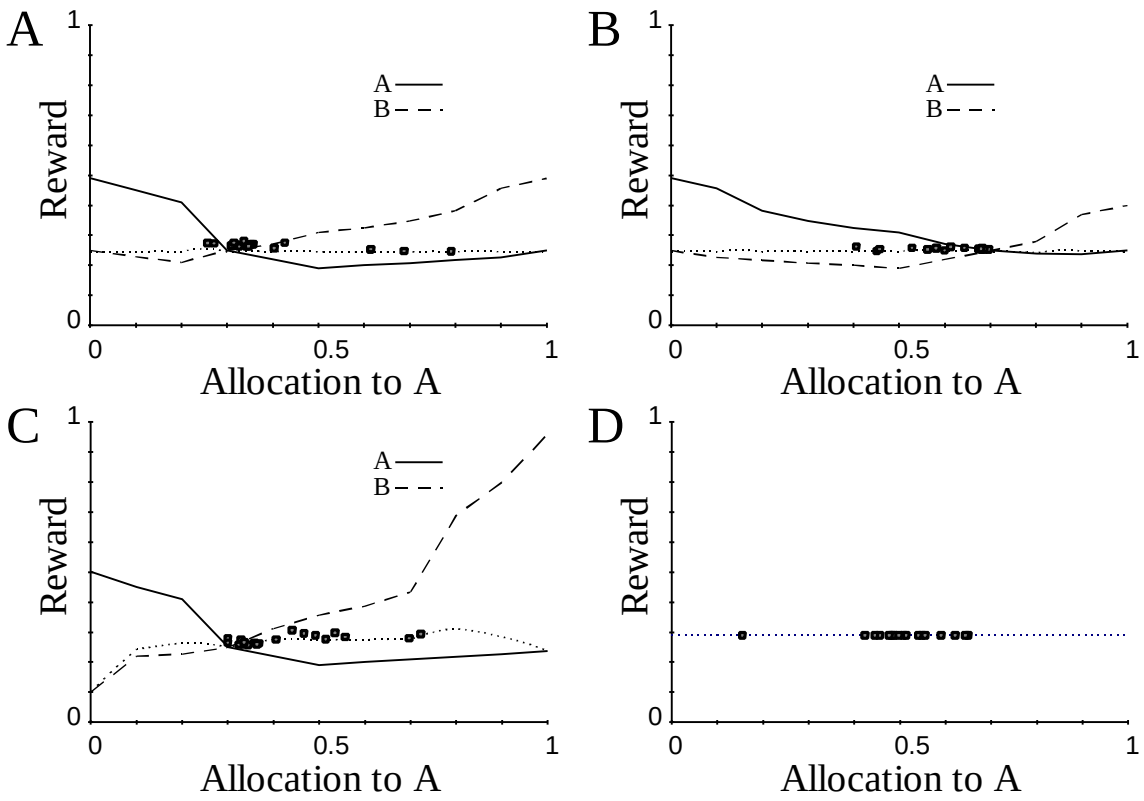

Figure 2. Four Reward Distributions (with no clear optimum).

Subjects were instructed to maximize long-term return in all four tasks (panels A, B, C, D). The reward given after each selection is a function of (1) the button selected, and (2) the subject's fraction of choices allocated to button A over the past 40 choices. In all four panels, the lines with diamonds show the reward from a selection of button A at a given choice allocation; the crosses show the reward earned from selecting button B. The unmarked line indicates the expected value of the reward for a fixed allocation to button A. For each subject, the square marks the average allocation and average earned reward after a trial of 250 selections.

(A) In this reward paradigm, the expected value of the earned reward is the same regardless of choice allocation. Subjects' average allocations lie just to the right of the crossing point of the functions (mean allocation: human=0.411±0.003, network=0.380±0.001; n=18).

(B) Reward functions reflected around the crossing point. Subjects cluster at a higher allocation to A, suggesting that the attractant is the crossing point and not some local features experienced as the crossing point is approached. This point is further strengthened in Figure 3. (mean allocation: human=0.605±0.002, network=0.596±0.001; n=19).

(C) The grouping of subjects near the crossing point is generally unaffected by local features such as the larger differentials in reward for allocations to A between 0.7 and 1.0. (mean allocation: human=0.430±.003, network=0.374±0.001; n=19).

(D) Pseudo-random reward paradigm. Subjects receive a fixed, pseudo-randomized sequence of reward yielding a mean close to 0.3. Subjects display a mean allocation of 0.501±0.002 (n=19), confirming a central tendency in these two-choice tasks. Network mean allocation=0.498±0.007, n=19. These reward functions were chosen loosely for their general shape; our observations indicate that the overall shape, but not the finer details, influences the general behavior displayed by subjects.

Results

The experiments shown in Figure 2 assay choice behavior under conditions where every allocation strategy earns the same long-term return. The primary difference among the tasks is the local structure in the reward functions. In the tasks displayed in Figure 2(A,B,C), humans and networks converge quickly to a stable strategy, making choices that tend to equalize the return from the two alternatives. Such behavior is described as event-matching (endnote 3). The mean allocation to choice A settled close to the crossing points in the reward functions, with a slight central tendency. The existence of the central tendency was confirmed using a randomly distributed reward schedule (Figure 2D): under these random returns, both humans and networks equalized their allocations to A and B

To spotlight how a simple underlying mechanism can appear to express different behaviors in different contexts, we engineered two more choice tasks (Figure 3). In the first, the optimal strategy lies at the crossing point of the reward functions; in the second, an allocation at the crossing point is highly *sub*optimal. Figure 3A quantifies the subjects' behavior on the first task: most subjects (18 of 24) maximized their long-term return. However, in the second context (Figure 3B), the same attraction to the crossing point blinds them to higher long-term profit: over half (14 of 25) of the subjects converged to the crossing point even when other allocations yielded much higher return. As shown, higher allocations to A yield increasing reward. The result demonstrates the strong influence of the crossing of the reward functions since both the optimal allocation point and the central tendency point lie to the right of the crossing point. The histograms in Figure 3 show the results of the network on the same tasks. Given the simplicity of the model and the many levels of human strategies, we are not surprised to find differences in the histogram, such as the rightward tails in the human data. However, the result is instructive in the character of the match: the majority of subjects allocated their behavior at the crossing point of the reward functions, which, in Figure 3B, is highly suboptimal.

Variation in the free parameters over an extremely broad range does not qualitatively change the behavior of the network (see endnote 4).

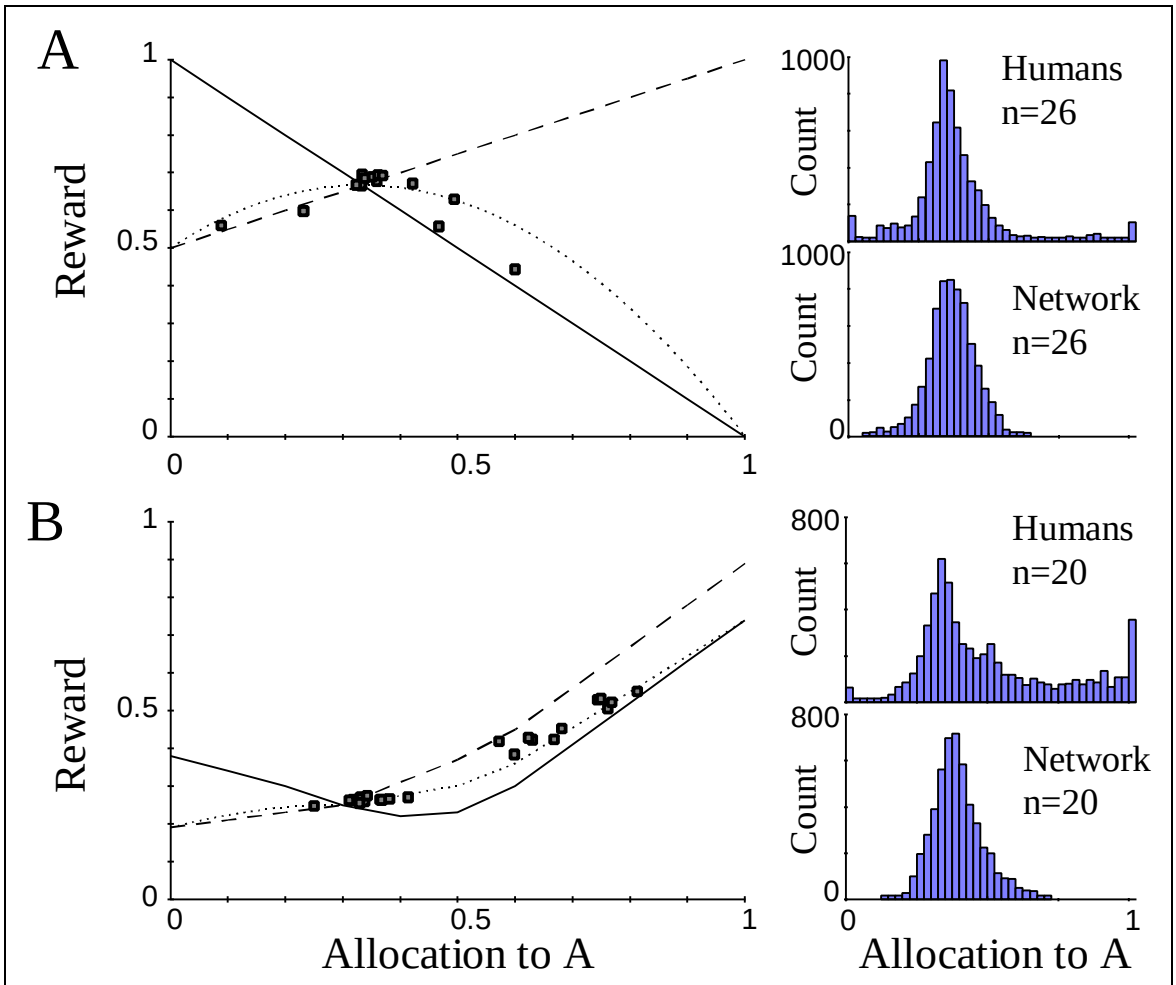

Figure 3. Context dependence of strategy selection for model and human.
Subjects and networks pursue optimal or suboptimal strategies depending on the context of the task. Lines with diamonds show the reward from a selection of button *A*, crosses show the reward earned from button *B*, and the unmarked line indicates the expected value of the reward for a given allocation.
(A) In this reward paradigm, the optimal allocation to A is the same as the crossing point of the reward functions (0.35). Subjects approximately maximize their reward on this task (mean allocation=0.366±0.002, n=26). Cumulative allocation histograms from humans and networks show that both groups stabilize around an allocation to A just to the right of the crossing point of the reward functions (network mean = 0.383±0.0009, n=26).
(B) This reward paradigm demonstrates that over half the subjects (14 of 25) settle into a stable behavior at the crossing point even when such a strategy is vastly sub-optimal. Here the most profitable strategy is total allocation to A. Subjects are drawn to the crossing point even when it lies to one side of both the optimal allocation and central tendency allocation.

The results of Figure 2 and Figure 3 can be understood by noting that the network tends to implement a greedy decision-making strategy, and that the cost functions associated with these tasks possess global minima at the crossing point of the reward functions. In a greedy strategy, the decision-maker compares the expected returns from alternative choices, and then selects the one that is likely to be most profitable (see endnote 5). On a task such as the one pictured in Figure 3A, greedy strategies will converge quickly to the crossing point of the reward functions (Borgstrom, 1993; Kilian, 1994). For the task shown in Figure 3A, a strategy converging to the crossing point will be called "optimal", whereas in Figure 3B it may be called "risk-averse". Such observations verify that different behaviors can be expressed by a simple underlying mechanism expressed in different behavioral contexts.

The model captures not only human allocation behavior, but also the deliberation times between choices. In all tasks, human subjects had no time pressure between selections. In spite of the broad range of interselection delays (mean = 0.793 sec, sd = 2.01 sec), human subjects demonstrated stable choice-dependent dynamics, i.e., choice allocation was independent of deliberation time.

Figure 4 shows some typical examples of the interselection delays for the task shown in Figure 3A. Note that while the subjects' allocations to button A fluctuate smoothly around the crossing point in the reward functions (0.33), the delays are uncorrelated (average correlation coefficient = -0.2). Such data suggest that subjects update their internal models at the time of each button choice, in a fashion independent of the delay between choices. The network model, updating its weights only at each choice, captures the delay-independent dynamics of the humans.

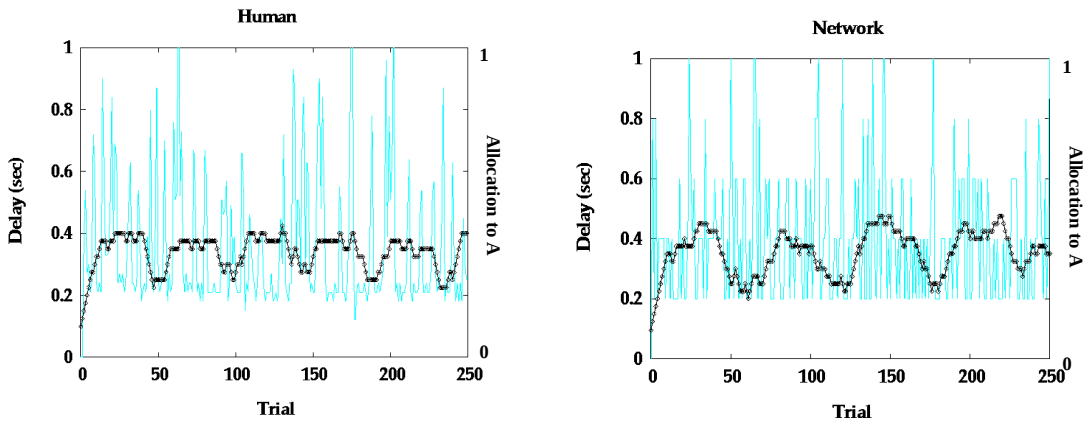

Figure 4. Decision dynamics are independent of delays in both humans and networks. (A) The inter-selection delays and allocation to A are plotted for a typical subject performing the "matching shoulders" task (seen in Figure 3A; so named because of the matched peaks of the reward functions). Note that the delays vary widely, even while the allocation to A smoothly oscillates around the crossing point (0.35) of the reward functions. In the model, the oscillations come about because the difference in rewards from A and B becomes reflected in the difference of weights w_A and w_B. This in turn drives the system toward more selections from the choice with higher weight. The averaged allocation is thus driven to the other side of the crossing point, where the same process begins again. The correlation of delay times with the subjects allocation to A, over all subjects, yielded a correlation coefficient of 0.0117.
(B) Network's behavior on the same task. The network's delays are defined as being proportional to the number of transitions made between alternatives before the model committed to a selection (see definition of decision function p_s in legend of Figure 1). Network correlation coefficient = 0.0137. The constant of proportionality that relates a network transition to seconds is taken to be on the order of 200msec, which is the physiologically characterized time of dopamine transients in alert monkeys performing similar tasks (Romo and Schultz, 1990).

Traditional decision-making theories (especially those following the tradition of expected utility theory) are deterministic, i.e., preference of A over B is either true or false. Such theories have consistently fallen short in explanations of observed human decision making, both in terms of choices and the distribution of deliberation times (Busemeyer, 1993). To date, delay distributions have only been successfully captured by non-deterministic models (Carpenter and Williams, 1995). It may be the case that preference and deliberation times cannot be studied separately; our model addresses both properties of decision-making by appealing to a common underlying mechanism.

The relationship of choice preference and delay behavior suggests some new interpretations of lesion, disease, and drug effects on dopaminergic systems. We begin by simulating a blunting of the dopamine neuron's output signal $\delta(t)$: such a blunting might be expected following a blockade of dopamine receptors. Results are shown in Figure 5A, where the model is presented with the decision task from Figure 2A, but with a 90% reduction in the magnitude of $\delta(t)$. The mean allocation to button A shifts from the crossing point (0.35) to random (0.5) with no concomitant change in inter-selection delays.

Figure 5B tests the model on the same task, but with a nonspecific decrease in the average amount of dopamine delivered to targets; the baseline (average) of $\delta(t)$ is reduced with no change in its sign or magnitude (Figure 1). The result is a dramatic increase in delay times with no change in choice allocation. The model follows its usual strategy, however, it takes a prohibitively long time to make a choice. Observers of such a symptom in a patient might interpret this change as a motor deficit, or "sluggishness". Such a non-specific baseline reduction in dopamine levels and the ensuing increase in the time-to-selection is reminiscent of symptoms associated with Parkinson's disease. This disease is characterized biologically by degeneration of dopamine cells in the substantia nigra (see endnote 6) and typically includes a slowing in the initiation and execution of voluntary movements and motor sequences.

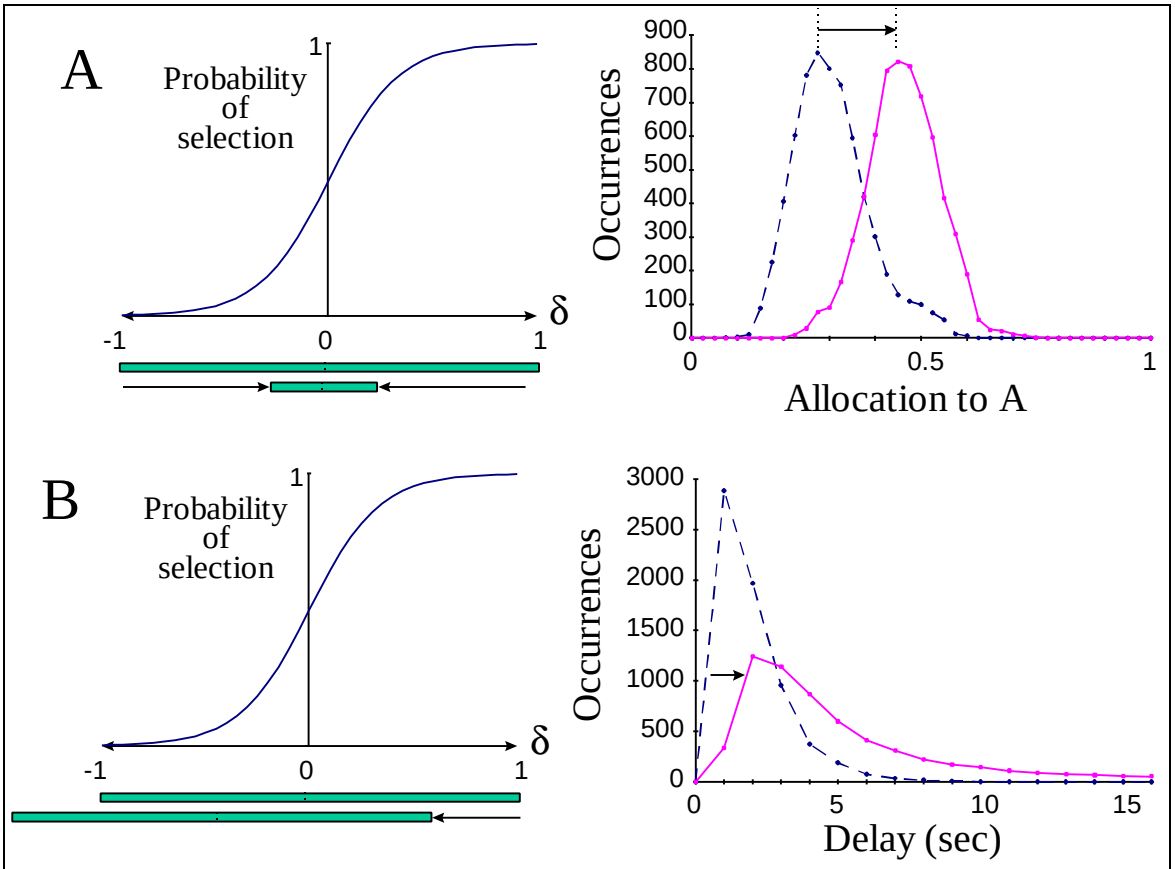

Figure 5. Some predicted effects of pharmacology, lesion, and disease.

(A) The blockade of dopamine receptors (as by neuroleptics) is modeled by blunting the effects of dopamine release at the target regions. Using the task in Fig 3A, the average choice allocation shifted from 0.35 allocation to A to random (0.5 allocation to A). Interselection delays (not shown) were unaffected.

(B) The destruction of input to dopamine neurons (as by a lesion in the cortex), or the degeneration of dopaminergic neurons (as in nigro-striatal pathway loss in Parkinson's disease) is represented by shifting $\delta(t)$ to a lower baseline level. Thus, while dopamine neuron output continues to fluctuate appropriately, the reduced baseline component leads to a dramatic increase of interselection delays. Choice strategies are unaltered but are interpreted as being prohibitively long (mean allocation to A = 0.369±0.001 in normal network, 0.347±0.002 in diseased network, n=26). Mean delay in the normal network = 1.99±0.002 seconds, in the diseased network mean delay = 7.29±0.012 seconds. Graph is shown only to 16 seconds; maximum delays reached 151 seconds in the 'diseased' network.

The results in Figure 5B suggest that Parkinson's patients may retain the ability to construct appropriate error signals to influence ongoing decision-making---however, the dramatic decrease in average baseline dopamine levels prevents the proper use of this information at the level of target structures. In other words, the non-specific decreases in

baseline dopamine levels could result in dramatic changes in motor behavior: while the plans remain intact, the time to arbitrate a selection among plans increases. See (Berns, 1996) for a similar interpretation of sequence selection.

Accordingly, the model predicts that a return to normal baseline dopamine levels, which would return fluctuations of neuromodulator release to an appropriate operational range, would restore selection times to normal. This interpretation is consistent with the systematic and highly successful use of L-dopa (dopamine precursor) with Parkinsonian patients (Hornykiewicz and Kish, 1987; Agid, 1989).

A reduction in the baseline (average) of $\delta(t)$ might also result from damage to prefrontal cortex. Humans with damage to the ventromedial sector of the prefrontal lobes present with deficits in decision-making and planning skills (Eslinger, 1985; Damasio, et al., 1990; Bechara, 1994; Damasio, 1994; Bechara, 1995). Patients can be well aware of contingencies of the decision and can enumerate differences between choices, but have difficulty concluding with a decision. In the model, as before, such a lesion to the frontal lobes might be represented by a sustained decrease in the baseline (average) of $\delta(t)$ because of the lack of cortical influence on the output of midbrain dopamine neurons. This change would lower significantly the probability of making a choice independent of the capacity to categorize or assess the value of the choice.

Conclusions

The results verify that for simple decision-making tasks, especially when information about the task is impoverished, human choice behavior is capable of being characterized by a simple neural model based on anatomical arrangements, physiological data, and a set of well-understood computational principles. The mesencephalic dopaminergic system fulfills the requirements of the model; however, we note that related projections (such as the cerulean noradrenergic system) may fulfill or contribute to the same roles. We have engineered choice tasks that highlight certain behaviors of this system (such as suboptimal choice allocation), and presented the task to 66 human subjects. The close match of the human and model data supports a direct relationship between biases in human decision

strategies and fluctuating neuromodulator delivery. Although humans surely have sufficient memory capacity to learn long-term strategies, their mechanisms appear to be tuned to use short-term information to arbitrate decisions under rapidly changing reward contingencies. This latter property is reminiscent of the behavior of honeybees on similar decision-making tasks (Montague et al, 1995). The bottom-up approach presented here suggests that a variety of behavioral strategies may result from the expression of relatively simple neural mechanisms in different behavioral contexts. Further, the approach suggests that certain motor deficits may share the same underlying cause as deficits of decision-making.

Notes

1. The first decision-making theories were normative, meaning they prescribed what strategies humans *ought* to follow under a predetermined notion of the goals. Such theories, e.g., utility theory (Bernoulli, 1738; Von Neumann and Morgenstern, 1947) held long influence on economic theory. However, the systematic study of decision-making has exposed sets of reproducible behaviors that cannot be fit into traditional normative frameworks of rational choice (Kahneman and Tversky, 1984). This has given rise to descriptive theories, some of which are more axiomatic in nature (e.g., prospect theory, (Kahneman and Tversky, 1979)), and some of which suggest architectural components that could implement the theories (Grossberg, 1987). However, no approaches thus far yield a well-specified and testable relationship to potential neural substrates.
2. The goal of temporal difference methods is to learn a function V(t) that anticipates (predicts) the sum of future rewards. As demonstrated in Montague et al (1996), this simple computational theory captures a wide range of physiological recordings from midbrain dopamine neurons in alert primates.

3. Event matching is a well-described behavior displayed by both animals and humans in choice situations. It is defined by the "matching" of behavioral investments to the return on those investments, expressed concisely by:

$$\frac{B_j}{\sum_i B_i} = \frac{Y_j}{\sum_i Y_i} \tag{1}$$

where Y_j is the yield (return) earned from any given behavioral investment, B_j. Whereas matching behavior is not always optimal, it is generally adaptive (Herrnstein, 1990; Herrnstein, 1991).

4. Initial starting points along the x-axis were varied from 0.0 to 0.95. The learning rate λ was varied from 0.1 to 0.9. The slope *m* in Fig. 2 was varied from 3 to 50. The offset *b* was varied from 0.0 to 1.0. Such variations modified the size of the basin of attraction, the dynamics of the approach, and the character of the delays. However, the convergence to the crossing points was unchanged (but see Figure 5).

5. While a decision is being arbitrated, $\delta(t)=V(t)-V(t)$ (see legend of Figure 1). To illustrate, when the model `looks' from choice A to choice B, $\delta(t)=w_B-w_A$, allowing the probability of selection to be written:

$$p_s = \frac{1}{1+e^{m(w_B-w_a)+b}} \tag{2}$$

or (setting b=0),

$$p_s = \frac{e^{-mw_B}}{e^{-mw_B}+e^{-mw_A}} \tag{3}$$

which relates our model to a Boltzmann (or 'soft-max') choice mechanism, wherein the probability of making a selection is a function of the changing weights. Since the weights will be maximally influenced by the most recent rewards, and the probability

of selection will be highest for the larger weight, this mechanism engenders a greedy decision-making strategy.

6. There are dopamine cells in the substantia nigra that also appear to report prediction errors in future appetitive stimuli, suggesting that the model may explain some aspects of the deficits involved in losing the majority of these cells in Parkinson's disease (Schultz, et al., 1993).

REFERENCES

Agid, Y., Cervera, P., Hirsch, E., Javoy-Agid, F., Lehericy, S., Raisman, R., Ruberg, M. (1989). Biochemistry of Parkinson's disease 28 years later: a critical review. *Movement Disorders* **4 Suppl 1,** S126-144.

Aston-Jones, G., Rajkowski, J., Kubiak, P., Alexinsky, T. (1994). Locus coeruleus neurons in monkey are selectively activated by attended cues in a vigilance task. *Journal of Neuroscience* **14,** 4467.

Bechara, A., Damasio, A.R., Damasio, H., Anderson, S.W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* **50,** 7-15.

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., Damasio, A.R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science* **269,** 1115-8.

Bechara, A., Damasio, H., Tranel, D., Damasio, A.R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science* **275**.

Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis. *Commentarii academiae scientiarum imperalis Petropolitanae* **5,** 175-192.

Berns, G. S., Sejnowski, T.J. (1996). How the basal ganglia make decisions, *Neurobiology of decision-making*, Springer-Verlag, pp. 101-113.

Borgstrom, R. S., Kosaraju, S.R. (1993). *Proceedings, ACM Symposium on Theory of Computing* **25,** 130.

Busemeyer, J. R., Townsend, J.T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychol. Rev* **100,** 432.

Carpenter, R. H. S. and Williams, M. L. L. (1995). Neural computations of log likelihood in control of saccadic eye movements. *Nature* **377,** 59-62.

Damasio, A. R. (1994). *Descartes' Error*. Putnam.

Damasio, A. R., Tranel, D. and Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioral Brain Research* **41,** 81-94.

Egelman, D. M., Person, C. and Montague, P. R. (1995). A predictive model of diffuse systems matches human choice behavior on simple decision-making tasks. *Soc. Neurosci. Abstr.* **21,** 2087.

Egelman, D. M., Person, C. and Montague, P. R. (1998). A computational role of dopamine delivery in human decision making. *Journal of Cognitive Neuroscience* **In press**.

Eslinger, P. J. a. D., A.R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR. *Neurology* **35,** 1731-1741.

Grossberg, S., Gutowski, W.E. (1987). Neural dynamics of decision making under risk: affective balance and cognitive-emotional interactions. *Psychological Review* **94,** 300-318.

Herrnstein, R. J. (1990). Rational choice theory: necessary but not sufficient. *American Psychologist* **45,** 356.

Herrnstein, R. J. (1991). Experiments on stable suboptimality in individual behavior. *AEA Papers and Proceedings* **81,** 360-364.

Hornykiewicz, O. and Kish, S. J. (1987). Biochemical pathophysiology of Parkinson's disease. *Advances in Neurology* **45,** 19-34.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An analysis of decision under risk. *Econometrica* **47,** 263.

Kahneman, D. and Tversky, A. (1984). Choices, values and frames. *American Psychologist* **39,** 341-350.

Kilian, J., Land, K.J., Pealmutter, B.A. (1994). Playing the matching-shoulders lob-pass game with logarithmic regret. *7th Annual ACM Workshop on Computer Learning Theory*, pp. 159.

Ljunberg, T., Apicella, P. and Schultz, W. (1992). Responses of monkey dopamine neurons during learning of behavioral reactions. *Journal of Neurophysiology* **67,** 145-163.

Luce, R. and Raiffa, H. (1957). *Games and decisions: Introduction and Critical Survey*. Dover.

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

Montague, P. R., Dayan, P., Person, C. and Sejnowski, T. J. (1995). Bee foraging in uncertain environments using predictive hebbian learning. *Nature* **377,** 725-728.

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

Quartz, S. R., Dayan, P., Montague, P.R., Sejnowski, T.J. (1992). Expectation learning in the brain using diffuse ascending projections. *Society for Neuroscience Abstracts* **18,** 1210.

Romo, R. and Schultz, W. (1990). Dopamine neurons of the monkey midbrain: contingencies of responses to active touch during self-initiated arm movements. *Journal of Neurophysiology* **63,** 592-606.

Schultz, W. (1992). Activity of dopamine neurons in the behaving primate. *Semin. Neurosci* **4,** 129-138.

Schultz, W., Apicella, P. and Ljunberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *Journal of Neuroscience* **13,** 900-913.

Schultz, W., Dayan, P. and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science* **275,** 1593-1598.

Sutton, R. S., Barto, A.G. (1987). A temporal-difference model of classical conditioning. *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*.

Sutton, R. S. (1988). Learning to predict by the method of temporal differences. *Machine Learning* **3,** 9-44.

Sutton, R. S., Barto, A.G. (1990) Time-derivative models of Pavlovian, Neuroscience}, r. I. i. L. a. C. and (Gabriel M., M. J., eds). pp. 497-537. Cambridge: MIT (1990). Time-derivative models of Pavlovian reinforcement in Gabriel M., M. J. (Ed), *Learning and Computational Neuroscience*, MIT, pp. 497-537.

Vaughan, W. and Herrnstein, R. J. (1987). Stability, melioration, and natural selection in Green, L. and J.H.Kagel (Eds), *Advances in Behavioral Economics*, Ablex, pp. 185-215.

Von Neumann, J. and Morgenstern, O. (1947). *Theory of Games and Economic Behavior*. Princeton University Press.

Wise, R. A. (1980). Action of drugs of abuse on brain reward systems. *Pharmacology, Biochemistry and Behavior* **13 Suppl 1,** 213-223.

Wise, R. A. and Bozarth, M. A. (1984). Brain reward circuitry: four circuit elements "wired" in apparent series. *Brain Res. Bull.* **12,** 203.