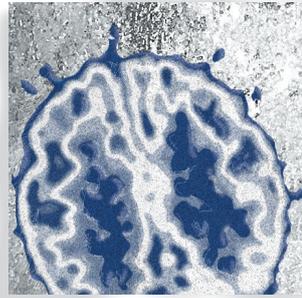


Dopamine reward prediction error coding

Wolfram Schultz, MD, FRS



Reward prediction errors consist of the differences between received and predicted rewards. They are crucial for basic forms of learning about rewards and make us strive for more rewards—an evolutionary beneficial trait. Most dopamine neurons in the midbrain of humans, monkeys, and rodents signal a reward prediction error; they are activated by more reward than predicted (positive prediction error), remain at baseline activity for fully predicted rewards, and show depressed activity with less reward than predicted (negative prediction error). The dopamine signal increases nonlinearly with reward value and codes formal economic utility. Drugs of addiction generate, hijack, and amplify the dopamine reward signal and induce exaggerated, uncontrolled dopamine effects on neuronal plasticity. The striatum, amygdala, and frontal cortex also show reward prediction error coding, but only in subpopulations of neurons. Thus, the important concept of reward prediction errors is implemented in neuronal hardware.

© 2016, AICH – Servier Research Group

Dialogues Clin Neurosci. 2016;18:23-32.

Keywords: *neuron; substantia nigra; ventral tegmental area; striatum; neurophysiology; dopamine; reward; prediction*

Introduction

I am standing in front of a drink-dispensing machine in Japan that seems to allow me to buy six different types of drinks, but I cannot read the words. I have a low expectation that pressing a particular button will deliver my preferred blackcurrant juice (a chance of one in six). So I just press the second button from the right, and then a blue can appears with a familiar logo that happens to be exactly the drink I want. That is a pleasant surprise, better than expected. What would I do the next time I want the same blackcurrant juice from the machine? Of course, press the second button from the right. Thus, my surprise directs my behavior to a specific button. I have learned something, and I will keep pressing the same button as long as the same can comes out. However, a couple of weeks later, I press that same button again, but another, less preferred can appears. Unpleasant surprise, somebody must have filled the dispenser differently. Where is my preferred can? I press another couple of buttons until my blue can comes out. And of course I will press that button again the next time I want that blackcurrant juice, and hopefully all will go well.

Author affiliations: Department of Physiology, Development and Neuroscience, University of Cambridge, United Kingdom

Address for correspondence: Wolfram Schultz, Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3DY, United Kingdom
(email: ws234@cam.ac.uk)

Basic research

What happened? The first button press delivered my preferred can. This pleasant surprise is what we call a positive reward prediction error. “Error” refers to the difference between the can that came out and the low expectation of getting exactly that one, irrespective of whether I made an error or something else went wrong. “Reward” is any object or stimulus that I like and of which I want more. “Reward prediction error” then means the difference between the reward I get and the reward that was predicted. Numerically, the prediction error on my first press was 1 minus 1/6, the difference between what I got and what I reasonably expected. Once I get the same can again and again for the same button press, I get no more surprises; there is no prediction error, I don’t change my behavior, and thus I learn nothing more about these buttons. But what about the wrong can coming out 2 weeks later? I had the firm expectation of my preferred blackcurrant juice but, unpleasant surprise, the can that came out was not the one I preferred. I experienced a negative prediction error, the difference between the nonpreferred, lower valued can and the expected preferred can. At the end of the exercise, I have learned where to get my preferred blackcurrant juice, and the prediction errors helped me to learn where to find it. Even if all this sounds arcane, it is the formal description of my Japanese experience.

This is what this article is about: how our brains process reward prediction errors to help us get our drinks, and all the other rewards and good things in life.

Reward prediction errors for learning

Rewards produce learning. Pavlov’s dog hears a bell, sees a sausage, and salivates. If done often enough, the dog will salivate merely on hearing the bell.¹ We say that the bell predicts the sausage, and that is why the dog salivates. This type of learning occurs automatically, without the dog doing anything except being awake. Operant conditioning, another basic form of learning, requires the animal’s participation. Thorndike’s cat runs around a cage until it happens to press a latch and suddenly gets out and can eat.² The food is great, and the cat presses again, and again. Operant learning requires the subject’s own action, otherwise no reward will come and no learning will occur. Pavlovian and operant learning constitute the building blocks for behavioral reactions to rewards.

Both learning forms involve prediction errors.³ To understand prediction errors, we distinguish between

a prediction about a future reward, or no prediction (which is also a prediction, but a poorly defined one), and the subsequent reward. Then we compare the reward with the prediction; the reward is either better than, equal to, or worse than than its prediction. The future behavior will change depending on the experienced difference between the reward and its prediction, the prediction error (*Figure 1*). If the reward is different from its prediction, a prediction error exists, and we should update the prediction and change our behavior (red). Specifically, if the reward is better than predicted (positive prediction error), which is what we all want, the prediction becomes better and we will do more of the behavior that resulted in that reward. If the reward is worse than predicted (negative prediction error), which nobody wants, the prediction becomes worse and we will avoid this the next time around. In both cases, our prediction and behavior changes; we are learning. By contrast, if the reward is exactly as predicted (blue), there is no prediction error, and we keep our prediction and behavior unchanged; we learn nothing. The intuition behind prediction error learning is that we often learn by making mistakes. Although mistakes are usually poorly regarded, they nevertheless help us to get a task right at the end and obtain a reward. If no further error occurs, the behavior will not change until the next error. This applies to learning for obtaining rewards as well as it does for learning movements.

The whole learning mechanism works because we want positive prediction errors and hate negative prediction errors. This is apparently a mechanism built in by evolution that pushes us to always want more and never want less. This is what drives life and evolution, and makes us buy a bigger car when our neighbors

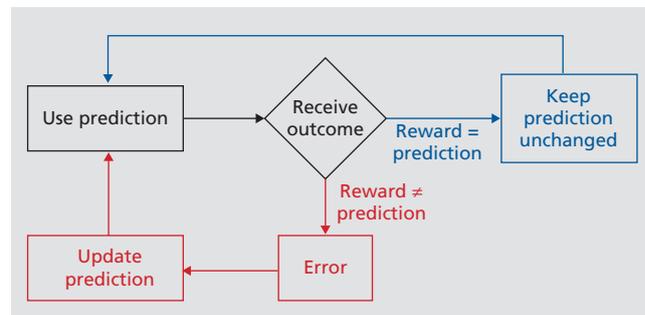


Figure 1. Scheme of learning by prediction error. Red: a prediction error exists when the reward differs from its prediction. Blue: no error exists when the outcome matches the prediction, and the behavior remains unchanged.

muscle up on their cars (the neighbors' average car size serves as a reference that is equivalent to a prediction). Even a buddhist, who hates wanting and craving for material goods and unattainable goals, wants more happiness rather than less. Thus, the study of reward prediction errors touches the fundamental conditions of life.

Reward is in the brain

The study of reward processing in the brain started when Olds and Milner⁴ introduced electrodes into the brains of rats and subjected them to small electric currents when they pressed a lever. Such currents elicit action potentials in thousands of neurons within a millimeter around the electrode. Olds and Milner placed electrodes in different brain regions. In some of these regions they found a remarkable effect. The rats pressed more to get more electric shocks to their brains. The animals were so fascinated by the lever pressing that they forgot to eat and drink for a while. Not even a female rat could distract a male. It seemed that there was nothing better than this brain stimulation. The surge of action potentials in neurons incited the animals again and again to press the lever, a typical manifestation of the function of rewards in learning and approach behavior. Olds and Milner had found a physical correlate for reward in the brain!

Subsequent studies showed that about half of the effective locations for electrical self-stimulation are connected with dopamine neurons.⁵ Dopamine neurons are located in the midbrain, just behind the mouth, between the ears, about a million in humans, 200 000 in monkeys, and 20 000 in rats. They extend their axons several millimeters into the striatum, frontal cortex, amygdala, and several other brain regions. The self-stimulation data demonstrate that the action potentials of dopamine neurons induce learning and approach behavior, thus linking brain function in a causal way to behavior.

But do dopamine neurons generate action potentials when a reward is encountered, without being stimulated by electric currents? The answer is yes.⁶ Showing a human, a monkey, or a rat money, food, or liquid makes the large majority of their dopamine neurons produce similar action potentials to those that occur during electrical self-stimulation. The higher the reward, the stronger the dopamine response. A closer look reveals that the dopamine neurons not only respond when the animal receives a reward but also when a stimulus, such

as a light, picture, or sound predicts a reward. Such reward-predicting stimuli are conditioned rewards and have similar effects on learning and approach behavior to real rewards. Dopamine neurons treat reward predictors and real rewards in a similar way, as events that are valuable for the individual. In addition, predictive stimuli allow animals to plan ahead and make informed decisions. Thus, in signaling both rewards and reward-predicting stimuli, dopamine neurons provide information about past and future rewards that is helpful for learning and decision-making.

Reward prediction errors in dopamine neurons

However, the dopamine response shows something else. The response to the reward itself disappears when the reward is predicted. But if more than the predicted reward occurs, the dopamine neurons show stronger responses. By contrast, their activity decreases if no, or less than predicted, reward occurs. The dopamine response thus reflects a reward prediction error and can be described by the simple difference between obtained and predicted reward (*Figure 2*). When we look at the time of the reward, more reward than predicted induces a positive dopamine response (excitation or activation, top), as much reward as expected induces no response (middle), and less than predicted reward leads to a negative response (depression of activity, bottom).^{7,8} These responses exist not only in monkeys, but are also found in dopamine neurons in humans⁹ and rodents.^{10,11} Thus, dopamine neurons don't just respond to any old reward: they respond only to rewards that differ from their prediction. But that's not all. The dopamine response is transferred to the next preceding reward-predicting stimulus, and ultimately to the first predictive stimulus (*Figure 3A*).^{7,12} The longer the time between the first stimulus and the final reward, the smaller the dopamine response, as subjective reward value becomes lower with greater delays, a phenomenon known as temporal discounting; dopamine responses decrease in specific temporal discounting tests.¹³ The response to the reward-predicting stimulus itself depends on the prediction of that stimulus, in the same way as the response to the reward. Thus, dopamine neurons respond to reward-predicting stimuli in the same way as to rewards, only with slightly less intensity, which allows them to use predictive informa-

Basic research

tion for teaching even earlier stimuli and actions. In this way, dopamine signals can be useful for learning long chains of events. Processing of prediction errors rather than full information about an environmental event saves neuronal information processing¹⁴ and, in the case of rewards, excites neurons with larger-than-predicted rewards.

The dopamine reward prediction error response occurs pretty much in the way used in the Rescorla-Wagner model³ that conceptualizes reward learning by prediction errors. In addition, the dopamine prediction error signal with reward-predicting stimuli corresponds well to the teaching term of temporal difference (TD) learning, a derivative of the Rescorla-Wagner model.¹⁵ Indeed, dopamine responses in simple and complex tasks correlate very well with formal TD models (Figure 3B).^{8,16} The existence of such neuronal error signals suggests that some brain processes operate on the principle of error learning. The dopamine error signal could be a teaching signal that affects neuronal plasticity in brain structures that are involved in reward learning, including the striatum, frontal cortex, and amygdala. The error signal serves also an important function in economic decisions because it helps to update the value signals for the different choice options.

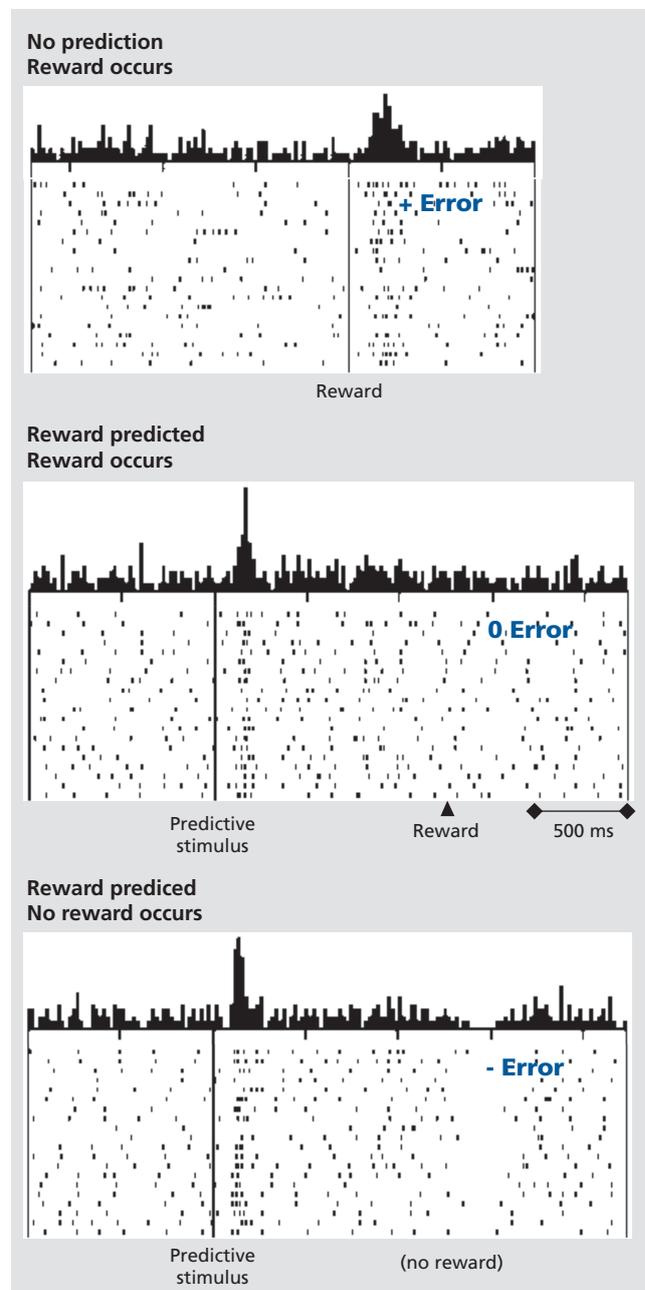
Multiple components in dopamine responses

If we look closely, the form of the dopamine response can be quite tricky, not unlike that of other neuronal responses in the brain. With careful analysis,¹⁷ or when demanding stimuli require extended processing time,¹⁸ two response components become visible (Figure 4). An initial, unselective response component registers any environmental object that occurs in the environment, including a reward. This salience response occurs with all kinds of stimuli, including punishers and

Figure 2. Reward prediction error responses at the time of reward (right) and reward-predicting visual stimuli (left in bottom two graphs). The dopamine neuron is activated by the unpredicted reward eliciting a positive reward prediction error (blue, + error, top), shows no response to the fully predicted reward eliciting no prediction error (0 error, middle), and is depressed by the omission of predicted reward eliciting a negative prediction error (- error, bottom).

Reproduced from ref 8: Schultz W, Dayan P, Montague RR. A neural substrate of prediction and reward. *Science* 1997;275:1593-1599. Copyright © American Association for the Advancement of Science 1997

neutral stimuli, and seems to simply alert the neurons of possible rewards in the environment. It subsides in a few tens to hundreds of milliseconds when the neurons identify the object and its reward value properly. Thus, the neurons code salience only in a transitory manner. Then a second, selective response component becomes identifiable, which reflects only the reward



information (as reward prediction error). From this point on, the dopamine neurons represent only reward information.

With these two components, the dopamine neurons start processing the encountered stimulus or object before they even know whether it is a reward, which gives them precious time to prepare for a potential behavioral reaction; the preparation can be cancelled if the object turns out not to be a reward. Also, the attentional characteristics of the initial response enhance the subsequent processing of reward information. This mechanism affords overall a gain in speed and accuracy without major cost.

Before this two-component response structure had been identified, some interpretations considered only the attentional response, which we now know concerns only the initial response component. Without looking at the second, reward response component, the whole dopamine response appeared to be simply a salience signal,¹⁹ and the function in reward prediction error coding was missed. Experiments on aversive stimuli reported some dopamine activations,^{20,21} which were later found to reflect the physical rather than aversive stimulus components.¹⁷ Together, these considerations led to assumptions of primarily salience coding in dopamine neurons,²² which can probably now be laid to rest.

Risky rewards, subjective value, and formal economic utility

Rewards are only sure in the laboratory. In real life, rewards are risky. We go to the pub expecting to meet friends and have a pint of beer. But we don't know whether the friends will actually be there this evening, or whether the pub might have run out of our favorite beer. Risk is usually considered as something bad, and in the case of rewards, the risk is associated with the possibility of not getting the best reward we expect. We won't give animals beer in the laboratory, but we can nicely test reward risk by going to economic theory. The most simple and confound-free risky rewards can be tested with binary gambles in which either a large or a small reward occurs with equal probability of $P=0.5$ ²³: I get either the large or the small reward with equal chance, but not both. Larger risk is modeled by increasing the large reward and reducing the small reward by equal amounts, thus keeping the mean reward constant. Monkeys like fruit juices and prefer such risky rewards over safe rewards with the same mean when the juice amount is low; they are risk seekers. Accordingly, they prefer the more widely spread gamble over the less spread one; thus they satisfy what is referred to as second-order stochastic dominance, a basic econom-

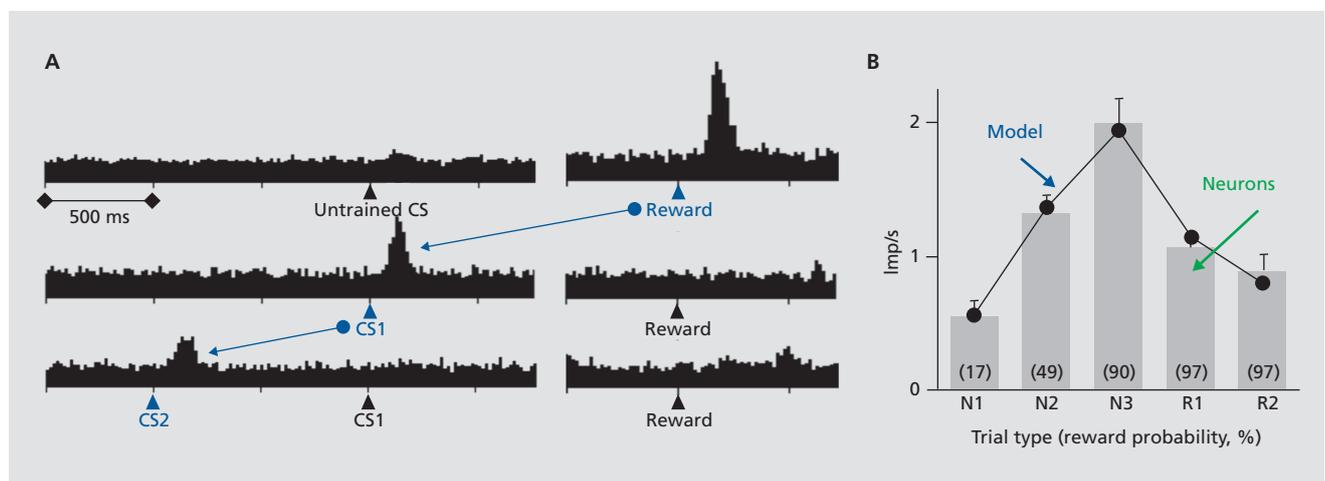


Figure 3. (A) Stepwise transfer of dopamine response from reward to first reward-predicting stimulus. CS2: earliest reward-predicting stimulus, CS1: subsequent reward-predicting stimulus. From Schultz 12. **(B)** Reward prediction error responses in sequential stimulus-reward task (gray bars) closely parallel prediction errors of formal temporal difference (TD) reinforcement model (black lines). Averaged population responses of 26 dopamine neurons follow the reward probabilities at each sequence step (numbers indicate reward probabilities in %) and match the time course of TD prediction errors.

Reproduced from ref 16: Enomoto K, Matsumoto N, Nakai S, et al. Dopamine neurons learn to encode the long-term value of multiple future rewards. *Proc Natl Acad Sci U S A*. 2011;108:15462-15467. Copyright © National Academy of Sciences 2011

Basic research

ic construct for assessing the integration of risk into subjective value. However, with larger juice amounts, monkeys prefer the safe reward over the gamble; they are risk avoiders, just as humans often are with larger rewards (*Figure 5A*). Thus, monkeys show meaningful choices of risky rewards.

The binary reward risk is fully characterized by the statistical variance. However, as good a test as it is for decisions under risk, it does not fully comprise everyday risks which often include asymmetric probability distributions and thus skewness risk. Why do gamblers often have health insurance? They prefer positive skewness risk, a small but possible chance of winning large amounts, while avoiding negative skewness risk, the small but possible chance of a costly medical treatment. Future risk tests should include skewness risk to model a more real-life scenario for risk.

The study of risky rewards allows us to address two important questions for dopamine neurons, the incorporation of risk into subjective reward value, and the construction of a formal, mathematical economic utility function, which provides the most theory-constrained definition of subjective reward value for economic de-

isions.^{24,25} Dopamine neurons show larger responses to risky compared with safe rewards in the low range, in a similar direction to the animal's preferences; thus dopamine neurons follow second-order stochastic dominance. Formal economic utility can be inferred from risky gambles and constitutes an internal measure of reward value for an individual; it is measured in utils, rather than milliliters or pounds, euros or dollars. It can be measured from choices under risk,²⁶ using the fractile chaining procedure.²⁷ Monkeys show nonlinear utility functions that are compatible with risk seeking at small juice amounts and risk avoiding at larger amounts. Importantly, dopamine neurons show the same nonlinear response increases with unpredicted rewards and risky gambles (*Figure 5B*). Thus, dopamine responses signal formal economic utility as the best characterized measure of reward value; the dopamine reward prediction error response is in fact a utility prediction error response. This is the first utility signal ever observed in the brain, and to the economist it identifies a physical implementation of the artificial construct of utility.

Dopamine reward prediction errors in human imaging

Dopamine reward prediction error signals are not limited to animals, and occur also in humans. Besides the mentioned electrophysiological study,⁹ hundreds of human neuroimaging studies demonstrate reward prediction error signals in the main reward structures,²⁸ including the ventral striatum.^{29,30} The signal reflects the dopamine response³¹ and occurs in striatal and frontal dopamine terminal areas rather than in midbrain cell body regions, presumably because it reflects summed postsynaptic potentials. Thus, the responses in the dopamine-receiving striatum seen with human neuroimaging demonstrate the existence of neural reward prediction error signals in the human brain and their convenient measurement with noninvasive procedures.

General consequences of dopamine prediction error signaling

We know that dopamine stimulation generates learning and approach behavior.⁴ We also know that encountering a better-than-predicted reward stimulates dopamine neurons. Thus, the dopamine stimulation arising from a natural reward may directly induce behavioral learning

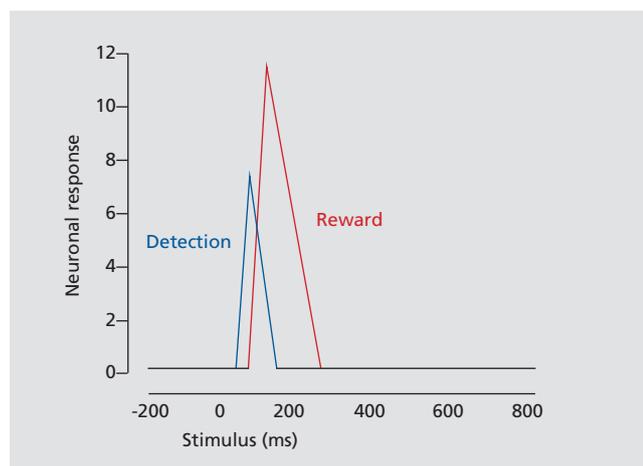


Figure 4. Schematic of the two phasic dopamine response components. The initial component (blue) detects the event before having identified its value. It increases with sensory impact (physical salience), generalization to rewarded stimuli, reward context and novelty (novelty/surprise salience). The second component (red) codes reward value (as reward utility prediction error). The two components become more distinct with more demanding stimuli. Adapted from data graph in ref 17: Fiorillo CD, Song MR, Yun SR. Multiphasic temporal dynamics in responses of midbrain dopamine neurons to appetitive and aversive stimuli. *J Neurosci.* 2013;33:4710–4725. Copyright © Society for Neuroscience 2013

and actions. Every time we see a reward, the responses of our dopamine neurons affect our behavior. They are like “little devils” in our brain that drive us to rewards! This becomes even more troubling because of the particular dopamine response characteristics, namely the positive dopamine response (activation) to positive prediction errors: the dopamine activation occurs when we get more reward than predicted. But any reward we receive automatically updates the prediction, and the previously larger-than-predicted reward becomes the norm and no longer triggers a dopamine prediction error surge. The next same reward starts from the higher prediction and hence induces less or no prediction error response. To continue getting the same prediction error, and thus the same dopamine stimulation, requires getting a bigger reward every time. The little devil not only drives us towards rewards, it drives us towards ever-increasing rewards.

The dopamine prediction error response may belong to a mechanism that underlies our drive for always wanting more reward. This mechanism would explain why we need ever higher rewards and are never satisfied with what we have. We want another car, not only because the neighbors have one but because we have become accustomed to our current one. Only a better, or at least a new, car would lead to a dopamine response, and that drives us to buy one. I have enough of my old clothes, even if they are still in very good condition, and therefore I go shopping. What the neighbors have, I want also, but better. The wife of 7 years is no longer good enough, so we need a new one, or at least another mistress. The house needs to be further decorated or at least rewallpapered, or we just buy a bigger one. And we need a summer house. There is no end to the ever-increasing needs of rewards. And all that because of the little dopamine neurons with their positive reward prediction error responses!

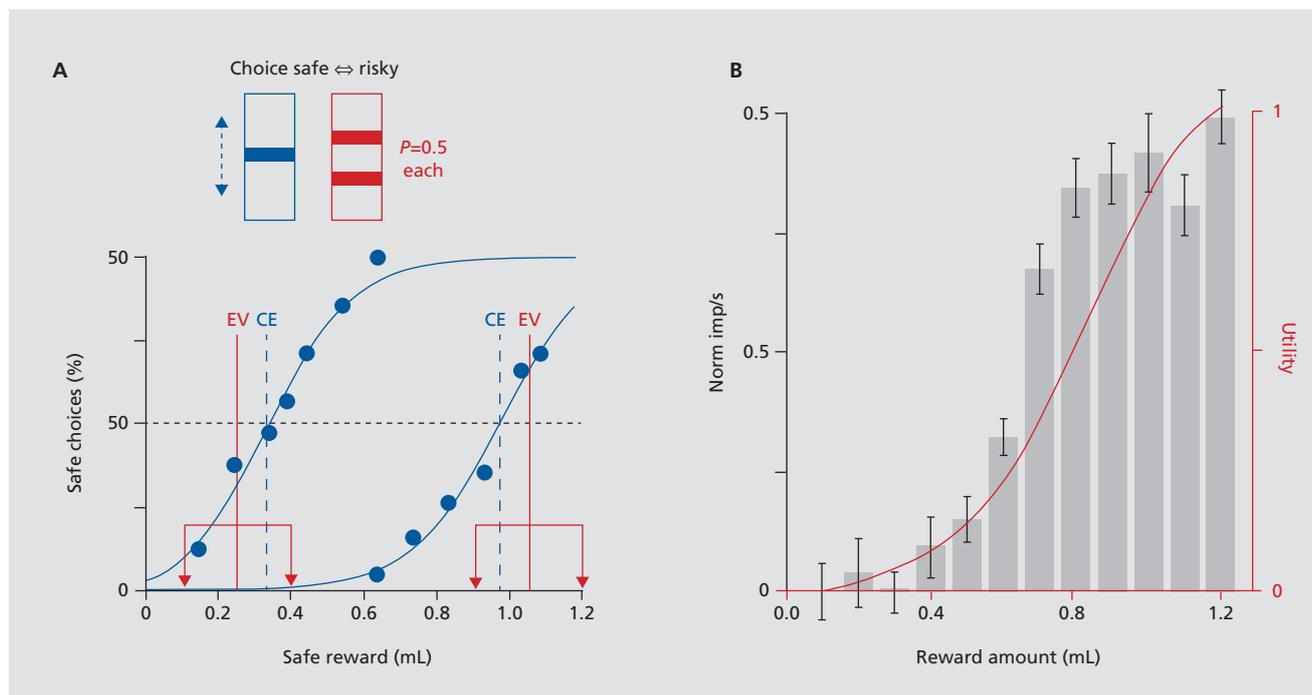


Figure 5. (A) Top: testing risky rewards: an animal chooses between a safe reward whose amount is adjusted by the experimenter (left) and a fixed, binary, equiprobable gamble (right). Height of each bar indicates juice volume; two bars indicate occurrence of each indicated amount with $P=0.5$ (risky reward). Bottom: Two psychophysical assessments of risk attitude in a monkey. CE indicates amount of safe reward at choice indifference against gamble (certainty equivalent). EV = expected value of gamble. A $CE > EV$ suggests risk seeking (subjective gamble value CE exceeds objective gamble value EV, left), $CE < EV$ indicates risk avoidance (right). (B) Positive utility prediction error responses to unpredicted juice rewards. Red: utility function derived from binary, equiprobable gambles. Black: corresponding, nonlinear increase of population response ($n=14$ dopamine neurons) in the same animal.

A and B reproduced from ref 25: Stauffer WR, Lak A, Schultz W. Dopamine reward prediction error responses reflect marginal utility. *Curr Biol.* 2014;24:2491-2500. Copyright © Cell Press 2014

Basic research

Dopamine mechanism of drug addiction

Dopamine neurons are even more devilish than explained so far. They are at the root of addictions to drugs, food, and gambling. We know, for example, that dopamine mechanisms are overstimulated by cocaine, amphetamine, methamphetamine, nicotine, and alcohol. These substances seem to hijack the neuronal systems that have evolved for processing natural rewards. Only this stimulation is not limited by the sensory receptors that process the environmental information, because the drugs act directly on the brain via blood vessels. Also, the drug effects mimic a positive dopamine reward prediction error, as they are not compared against a prediction, and thus induce continuing strong dopamine stimulation on their postsynaptic receptors, whereas the evolving predictions would have prevented such stimulation.³² The overstimulation resulting from the unfiltered impact and the continuing positive prediction error-like effect is difficult to handle for the neurons, which are not used to it from their evolution, and some brains cannot cope with the overstimulation and become addicted. We have less information about the mechanisms underlying gambling and food addiction, but we know that food and gambling, with their strong sensory stimulation and prospect of large gains, activate dopamine-rich brain areas in humans and dopamine neurons in animals.

Non-dopamine reward prediction errors

The dopamine reward prediction error signal is propagated along the widely divergent dopamine axons to the terminal areas in the striatum, frontal cortex, and amygdala, where they innervate basically all (striatum) or large proportions of postsynaptic neurons. Thus, the rather homogeneous dopamine reward signal influences heterogeneous postsynaptic neurons and thus affects diverse postsynaptic functions. However, reward prediction error signals occur also in other reward structures of the brain. Lateral habenula neurons show

bidirectional reward prediction error signals that are sign inverted to dopamine responses and may affect dopamine neurons via inhibitory midbrain reticular neurons.³³ Select groups of phasically and tonically firing neurons in the striatum and globus pallidus code positive and negative reward prediction errors bidirectionally.³⁴⁻³⁶ Some neurons in the amygdala display separate, bidirectional error coding for reward and punishment.³⁷ In the cortex, select neurons in anterior cingulate^{38,39} and supplementary eye field⁴⁰ code reward prediction errors. All of these reward prediction error signals are bidirectional; they show opposite changes to positive versus negative prediction errors. The reward prediction error responses in these subcortical and cortical neurons with their specific neuronal connections are unlikely to serve reinforcement processes via divergent anatomical projections; rather they would affect plasticity at specifically connected postsynaptic neurons.

Conclusions

The discovery that the most powerful and best characterized reward signal in the brain reflects reward prediction errors rather than the simple occurrence of rewards is very surprising, but is also indicative of the role rewards play in behavior. Rather than signaling every reward as it appears in the environment, dopamine responses represent the crucial term underlying basic, error-driven learning mechanisms for reward. The existence of the error signal validates error-driven learning rules by demonstrating their implementation in neuronal hardware. The additional characteristic of economic utility coding conforms to the most advanced definition of subjective reward value and suggests a role in economic decision mechanisms. Having a neuronal correlate for a positive reward prediction error in our brain may explain why we are striving for ever-greater rewards, a behavior that is surely helpful for surviving competition in evolution, but also generates frustrations and inequalities that endanger individual well-being and the social fabric. □

REFERENCES

1. Pavlov PI. *Conditioned Reflexes*. London, UK: Oxford University Press; 1927.
2. Thorndike EL. *Animal Intelligence: Experimental Studies*. New York, NY: MacMillan; 1911.
3. Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Black AH, Prokasy WF, eds. *Classical Conditioning II: Current Research and Theory*. New York, NY: Appleton Century Crofts; 1972:64-99.
4. Olds J, Milner P. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J Comp Physiol Psychol*. 1954;47:419-427.

Codificación del error de predicción de la recompensa dopaminérgica

Los errores en la predicción de la recompensa se deben a las diferencias entre las recompensas recibidas y predichas. Ellos son cruciales para las formas básicas de aprendizaje acerca de las recompensas y nos hacen esforzarnos por más recompensas, lo que constituye un rasgo evolucionario beneficioso. La mayoría de las neuronas dopaminérgicas en el mesencéfalo de los humanos, monos y roedores dan información sobre el error en la predicción de la recompensa; ellas son activadas por más recompensa que la predicha (error de predicción positivo); se mantienen en una actividad basal para el total de las recompensas predichas, y muestran una actividad disminuida con menos recompensa que la predicha (error de predicción negativo). La señal de dopamina aumenta de forma no lineal con el valor de la recompensa y da claves sobre la utilidad económica formal. Las drogas adictivas generan, secuestran y amplifican la señal de recompensa dopaminérgica e inducen efectos dopaminérgicos exagerados y descontrolados en la plasticidad neuronal. El estriado, la amígdala y la corteza frontal también codifican errores en la predicción de la recompensa, pero solo en ciertas subpoblaciones de neuronas. Por lo tanto, el concepto importante de errores en la predicción de recompensa está implementado en el hardware neuronal.

Codage de l'erreur de prédiction de la récompense dopaminérgique

Les erreurs de prédiction de la récompense consistent en différences entre la récompense reçue et celle prévue. Elles sont déterminantes pour les formes basiques d'apprentissage concernant la récompense et nous font lutter pour plus de récompense, une caractéristique bénéfique de l'évolution. La plupart des neurones dopaminérgiques du mésencéphale des humains, des singes et des rongeurs indiquent une erreur de prédiction de la récompense ; ils sont activés par plus de récompense que prévu (erreur de prédiction positive), restent dans l'activité initiale pour une récompense complètement prévue et montrent une activité diminuée en cas de moins de récompense que prévu (erreur de prédiction négative). Le signal dopaminérgique augmente de façon non linéaire avec la récompense et code l'utilité économique formelle. Les médicaments addictifs génèrent, détournent et amplifient le signal de la récompense dopaminérgique et induisent des effets dopaminérgiques exagérés et non contrôlés sur la plasticité neuronale. Le striatum, l'amygdale et le cortex frontal manifestent aussi le codage erroné de la prédiction de la récompense, mais seulement dans des sous-populations de neurones. L'important concept d'erreurs de prédiction de la récompense est donc mis en œuvre dans le matériel neuronal.

5. Corbett D, Wise RA. Intracranial self-stimulation in relation to the ascending dopaminergic systems of the midbrain: A moveable microelectrode study. *Brain Res.* 1980;185:1-15.
6. Schultz W. Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. *J Neurophysiol.* 1986;56:1439-1462.
7. Schultz W, Apicella P, Ljungberg T. Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J Neurosci.* 1993;13:900-913.
8. Schultz W, Dayan P, Montague RR. A neural substrate of prediction and reward. *Science.* 1997;275:1593-1599.
9. Zaghoul KA, Blanco JA, Weidemann CT, et al. Human substantia nigra neurons encode unexpected financial rewards. *Science.* 2009;323:1496-1499.
10. Pan W-X, Schmidt R, Wickens JR, Hyland BI. Dopamine cells respond to predicted events during classical conditioning: Evidence for eligibility traces in the reward-learning network. *J Neurosci.* 2005;25:6235-6242.
11. Cohen JY, Haesler S, Vogt L, Lowell BB, Uchida N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature.* 2012;482:85-88.
12. Schultz W. Predictive reward signal of dopamine neurons. *J Neurophysiol.* 1998;80:1-27.
13. Kobayashi S, Schultz W. Influence of reward delays on responses of dopamine neurons. *J Neurosci.* 2008;28:7837-7846.
14. Rao RPN, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci.* 1999;2:79-87.
15. Sutton RS, Barto AG. Toward a modern theory of adaptive networks: expectation and prediction. *Psychol Rev.* 1981;88:135-170.
16. Enomoto K, Matsumoto N, Nakai S, et al. Dopamine neurons learn to encode the long-term value of multiple future rewards. *Proc Natl Acad Sci U S A.* 2011;108:15462-15467.
17. Fiorillo CD, Song MR, Yun SR. Multiphasic temporal dynamics in responses of midbrain dopamine neurons to appetitive and aversive stimuli. *J Neurosci.* 2013;33:4710-4725.
18. Nomoto K, Schultz W, Watanabe T, Sakagami M. Temporally extended dopamine responses to perceptually demanding reward-predictive stimuli. *J Neurosci.* 2010;30:10692-10702.
19. Redgrave P, Prescott TJ, Gurney K. Is the short-latency dopamine response too short to signal reward? *Trends Neurosci.* 1999;22:146-151.
20. Mirenovic J, Schultz W. Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature.* 1996;379:449-451.
21. Matsumoto M, Hikosaka O. Two types of dopamine neuron distinctively convey positive and negative motivational signals. *Nature.* 2009;459:837-841.
22. Kapur S. Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *Am J Psychiatry.* 2003;160:13-23.
23. Rothschild M, Stiglitz JE. Increasing risk: I. A definition. *J Econ Theory.* 1970;2:225-243.
24. Lak A, Stauffer WR, Schultz W. Dopamine prediction error responses integrate subjective value from different reward dimensions. *Proc Natl Acad Sci U S A.* 2014;111:2343-2348.
25. Stauffer WR, Lak A, Schultz W. Dopamine reward prediction error responses reflect marginal utility. *Curr Biol.* 2014;24:2491-2500.
26. von Neumann J, Morgenstern O. *The Theory of Games and Economic Behavior.* Princeton, NJ: Princeton University Press; 1944.

Basic research

27. Machina MJ. Choice under uncertainty: problems solved and unsolved. *J Econ Perspect*. 1987;1:121-154.
28. Thut G, Schultz W, Roelcke U, Nienhusmeier M, Maguire RP, Leenders KL. Activation of the human brain by monetary reward. *NeuroReport*. 1997;8:1225-1228.
29. O'Doherty J, Dayan P, Friston K, Critchley H, Dolan RJ. Temporal difference models and reward-related learning in the human brain. *Neuron*. 2003;28:329-337.
30. McClure SM, Berns GS, Montague PR. Temporal prediction errors in a passive learning task activate human striatum. *Neuron*. 2003;38:339-346.
31. Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*. 2006;442:1042-1045.
32. Redish AD. Addiction as a computational process gone awry. *Science*. 2004;306:1944-1947.
33. Matsumoto M, Hikosaka O. Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature*. 2007;447:1111-1115.
34. Hong S, Hikosaka O. The globus pallidus sends reward-related signals to the lateral habenula. *Neuron*. 2008;60:720-729.
35. Apicella P, Deffains M, Ravel S, Legallet E. Tonicly active neurons in the striatum differentiate between delivery and omission of expected reward in a probabilistic task context. *Eur J Neurosci*. 2009;30:515-526.
36. Ding L, Gold JI. Caudate encodes multiple computations for perceptual decisions. *J Neurosci*. 2010;30:15747-15759.
37. Belova MA, Paton JJ, Morrison SE, Salzman CD. Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. *Neuron*. 2007;55:970-984.
38. Seo H, Lee D. Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J Neurosci*. 2007;27:8366-8377.
39. Kennerley SW, Behrens TEJ, Wallis JD. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat Neurosci*. 2011;14:1581-1589.
40. So N-Y, Stuphorn V. Supplementary eye field encodes reward prediction error. *J Neurosci*. 2012;32:2950-2963.