Physiol Rev 95: 853–951, 2015
Published June 24, 2015; doi:10.1152/physrev.00023.2014

NEURONAL REWARD AND DECISION SIGNALS: FROM THEORIES TO DATA

Wolfram Schultz

Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, United Kingdom

Schultz W. Neuronal Reward and Decision Signals: From Theories to Data. *Physiol Rev* 95: 853–951, 2015. Published June 24, 2015; doi:10.1152/physrev.00023.2014.—Rewards are crucial objects that induce learning, approach behavior, choices, and emotions. Whereas emotions are difficult to investigate in animals, the learning function is mediated by neuronal reward prediction error signals which implement basic constructs of reinforcement learning theory. These signals are found in dopamine neurons, which emit a global reward signal to striatum and frontal cortex, and in specific neurons in striatum, amygdala, and frontal cortex projecting to select neuronal populations. The approach and choice functions involve subjective value, which is objectively assessed by behavioral choices eliciting internal, subjective reward preferences. Utility is the formal mathematical characterization of subjective value and a prime decision variable in economic choice theory. It is coded as utility prediction error by phasic dopamine responses. Utility can incorporate various influences, including risk, delay, effort, and social interaction. Appropriate for formal decision mechanisms, rewards are coded as object value, action value, difference value, and chosen value by specific neurons. Although all reward, reinforcement, and decision variables are theoretical constructs, their neuronal signals constitute measurable physical implementations and as such confirm the validity of these concepts. The neuronal reward signals provide guidance for behavior while constraining the free will to act.

I. INTRODUCTION

Rewards are the most crucial objects for life. Their function is to make us eat, drink, and mate. Species with brains that allow them to get better rewards will win in evolution. This is what our brain does, acquire rewards, and do it in the best possible way. It may well be the reason why brains have evolved. Brains allow multicellular organisms to move about the world. By displacing themselves they can access more rewards than happen to come along by chance, thus enhancing their chance of survival and reproduction. However, movement alone does not get them any food or mating partners. It is necessary to identify stimuli, objects, events, situations, and activities that lead to the best nutrients and mating partners. Brains make individuals learn, select, approach, and consume the best rewards for survival and reproduction and thus make them succeed in evolutionary selection. To do so, the brain needs to identify the reward value of objects for survival and reproduction, and then direct the acquisition of these reward objects through learning, approach, choices, and positive emotions. Sensory discrimination and control of movements serve this prime role of the brain. For these functions, nature has endowed us with explicit neuronal reward signals that process all crucial aspects of reward functions.

Rewards are not defined by their physical properties but by the behavioral reactions they induce. Therefore, we need behavioral theories that provide concepts of reward functions. The theoretical concepts can be used for making testable hypotheses for experiments and for interpreting the results. Thus the field of reward and decision-making is not only hypothesis driven but also concept driven. The field of reward and decision-making benefits from well-developed theories of behavior as the study of sensory systems benefits from signal detection theory and the study of the motor system benefits from an understanding of mechanics. Reward theories are particularly important because of the absence of specific sensory receptors for reward, which would have provided basic physical definitions. Thus the theories help to overcome the limited explanatory power of physical reward parameters and emphasize the requirement for behavioral assessment of the reward parameters studied. These theories make disparate data consistent and coherent and thus help to avoid seemingly intuitive but paradoxical explanations.

Theories of reward function employ a few basic, fundamental variables such as subjective reward value derived from measurable behavior. This variable condenses all crucial factors of reward function and allows quantitative formal-

0031-9333/15 Copyright © 2015 the **American Physiological Society**

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

ization that characterizes and predicts a large variety of behavior. Importantly, this variable is hypothetical and does not exist in the external physical world. However, it is implemented in the brain in various neuronal reward signals, and thus does seem to have a physical basis. Although sophisticated forms of reward and decision processes are far more fascinating than arcane fundamental variables, their investigation may be crucial for understanding reward processing. Where would we be without the discovery of the esoteric electron by J. J. Thompson 1897 in the Cambridge Cavendish Laboratory? Without this discovery, the microprocessor and the whole internet would be impossible. Or, if we did not know about electromagnetic waves, we might assume a newsreader sitting inside the radio while sipping our morning coffee. This review is particularly concerned with fundamental reward variables, first concerning learning and then related to decision-making.

The reviewed work concerns primarily neurophysiological studies on single neurons in monkeys whose sophisticated behavioral repertoire allows well detailed, quantitative behavioral assessments while controlling confounds from sensory processing, movements, and attention. Thus I am approaching reward processing from the point of view of the tip of a microelectrode, one neuron at a time, thousands of them over the years, in rhesus' brains with more than two billion neurons. I apologize to the authors whose work I have not been able to cite in full, as there is a large number of recent studies on the subject and I am selecting these studies by their contribution to the concepts being treated here.

II. REWARD FUNCTIONS

A. Proximal Reward Functions Are Defined by Behavior

We have sensory receptors that react to environmental events. The retina captures electromagnetic waves in a limited range. Optical physics, physical chemistry, and biochemistry help us to understand how the waves enter the eye, how the photons affect the ion channels in the retinal photoreceptors, and how the ganglion cells transmit the visual message to the brain. Thus sensory receptors define the functions of the visual system by translating the energy from environmental events into action potentials and sending them to the brain. The same holds for touch, pain, hearing, smell, and taste. If there are no receptors for particular environmental energies, we do not sense them. Humans do not feel magnetic fields, although some fish do. Thus physics and chemistry are a great help for defining and investigating the functions of sensory systems.

Rewards have none of that. Take rewarding stimuli and objects: we see them, feel them, taste them, smell them, or hear them. They affect our body through all sensory systems, but there is not a specific receptor that would capture the particular motivational properties of rewards. As reward functions cannot be explained by object properties alone, physics and chemistry are only of limited help, and we cannot investigate reward processing by looking at the properties of reward receptors. Instead, rewards are defined by the particular behavioral reactions they induce. Thus, to understand reward function, we need to study behavior. Behavior becomes the key tool for investigating reward function, just as a radio telescope is a key tool for astronomy.

The word reward has almost mystical connotations and is the subject of many philosophical treatises, from the ethics of the utilitarian philosophy of Jeremy Bentham (whose embalmed body is displayed in University College London) and John Stuart Mill to the contemporary philosophy of science of Tim Schroeder (39, 363, 514). More commonly, the man on the street views reward as a bonus for exceptional performance, like chocolate for a child getting good school marks, or as something that makes us happy. These descriptions are neither complete nor practical for scientific investigations. The field has settled on a number of well-defined reward functions that have allowed an amazing advance in knowledge on reward processing and have extended these investigations into economic decision-making. We are dealing with three, closely interwoven, functions of reward, namely, learning, approach behavior and decision-making, and pleasure.

1. Learning

Rewards have the potential to produce learning. Learning is Pavlov's main reward function (423). His dog salivates to a bell when a sausage often follows, but it does not salivate just when a bell rings without consequences. The animal's reaction to the initially neutral bell has changed because of the sausage. Now the bell predicts the sausage. No own action is required, as the sausage comes for free, and the learning happens also for free. Thus Pavlovian learning (classical conditioning) occurs automatically, without the subject's own active participation, other than being awake and mildly attentive. Then there is Thorndike's cat that runs around the cage and, among other things, presses a lever and suddenly gets some food (589). The food is great, and the cat presses again, and again, with increasing enthusiasm. The cat comes back for more. This is instrumental or operant learning. It requires an own action; otherwise, no reward will come and no learning will occur. Requiring an action is a major difference from Pavlovian learning. Thus operant learning is about actions, whereas Pavlovian learning is about stimuli. The two learning mechanisms can be distinguished schematically but occur frequently together and constitute the building blocks for behavioral reactions to rewards.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

Rewards in operant conditioning are positive reinforcers. They increase and maintain the frequency and strength of the behavior that leads to them. The more reward Thorndike's cat gets, the more it will press the lever. Reinforcers do nt only strengthen and maintain behavior for the cat but also for obtaining stimuli, objects, events, activities, and situations as different as beer, whisky, alcohol, relaxation, beauty, mating, babies, social company, and hundreds of others. Operant behavior gives a good definition for rewards. Anything that makes an individual come back for more is a positive reinforcer and therefore a reward. Although it provides a good definition, positive reinforcement is only one of several reward functions.

2. Approach behavior and decision-making

Rewards are attractive. They are motivating and make us exert an effort. We want rewards; we do not usually remain neutral when we encounter them. Rewards induce approach behavior, also called appetitive or preparatory behavior, and consummatory behavior. We want to get closer when we encounter them, and we prepare to get them. We cannot get the meal, or a mating partner, if we do not approach them. Rewards usually do not come alone, and we often can choose between different rewards. We find some rewards more attractive than others and select the best reward. Thus we value rewards and then decide between them to get the best value. Then we consume them. So, rewards are attractive and elicit approach behavior that helps to consume the reward. Thus any stimulus, object, event, activity, or situation that has the potential to make us approach and consume it is by definition a reward.

3. Positive emotions

Rewards have the potential to elicit positive emotions. The foremost emotion evoked by rewards is pleasure. We enjoy having a good meal, watching an interesting movie, or meeting a lovely person. Pleasure constitutes a transient response that may lead to the longer lasting state of happiness. There are different degrees and forms of pleasure. Water is pleasant for a thirsty person, and food for a hungry one. The rewarding effects of taste are based on the pleasure it evokes. Winning in a big lottery is even more pleasant. But many enjoyments differ by more than a few degrees. The feeling of high that is experienced by sports people during running or swimming, the lust evoked by encountering a ready mating partner, a sexual orgasm, the euphoria reported by drug users, and the parental affection to babies constitute different forms (qualities) rather than degrees of pleasure (quantities).

Once we have experienced the pleasure from a reward, we may form a desire to obtain it again. When I am thirsty or hungry and know that water or food helps, I desire them. Different from such specific desire, there are also desires for imagined or even impossible rewards, such as flying to Mars, in which cases desires become wishes (514). Desire requires a prediction, or at least a representation, of reward and constitutes an active process that is intentional [in being about something (529)]. Desire makes behavior purposeful and directs it towards identifiable goals. Thus desire is the emotion that helps to actively direct behavior towards known rewards, whereas pleasure is the passive experience that derives from a received or anticipated reward. Desire has multiple relations to pleasure; it may be pleasant in itself (I feel a pleasant desire), and it may lead to pleasure (I desire to obtain a pleasant object). Thus pleasure and desire have distinctive characteristics but are closely intertwined. They constitute the most important positive emotions induced by rewards. They prioritize our conscious processing and thus constitute important components of behavioral control. These emotions are also called liking (for pleasure) and wanting (for desire) in addiction research (471) and strongly support the learning and approach generating functions of reward.

Despite their immense power in reward function, pleasure and desire are very difficult to assess in an objectively measurable manner, which is an even greater problem for scientific investigations on animals, despite attempts to anthropomorphize (44). We do not know exactly what other humans feel and desire, and we know even less what animals feel. We can infer pleasure from behavioral responses that are associated with verbal reports about pleasure in humans. We could measure blood pressure, heart rate, skin resistance, or pupil diameter as manifestations of pleasure or desire, but they occur with many different emotions and thus are unspecific. Some of the stimuli and events that are pleasurable in humans may not even evoke pleasure in animals but act instead through innate mechanisms. We simply do not know. Nevertheless, the invention of pleasure and desire by evolution had the huge advantage of allowing a large number of stimuli, objects, events, situations, and activities to be attractive. This mechanism importantly supports the primary reward functions in obtaining essential substances and mating partners.

4. Potential

Rewards have the potential to produce learning, approach, decisions, and positive emotions. They are rewards even if their functions are not evoked at a given moment. For example, operant learning occurs only if the subject makes the operant response, but the reward remains a reward even if the subject does not make the operant response and the reward cannot exert its learning function. Similarly an object that has the potential to induce approach or make me happy or desire it is a reward, without necessarily doing it every time because I am busy or have other reasons not to engage. Pavlovian conditioning of approach behavior, which occurs every time a reward is encountered as long as it evokes at least minimal attention, nicely shows this.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

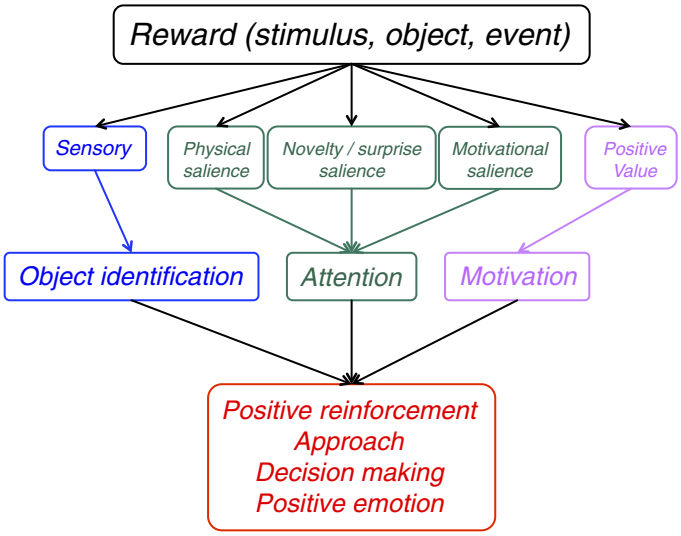


FIGURE 1. Reward components and their functions. The sensory component reflects the impact of environmental stimuli, objects, and events on the organism (blue). Pleasurable activities and situations belong also in this sensory component. The three salience components eliciting attentional responses (green) derive from the physical impact (*left*), novelty (*middle*), and commonly from reward and punishment (*right*). The specific positively motivating function of rewards derives from the value component (pink). Value does not primarily reflect physical parameters but the brain's subjective assessment of the usefulness of rewards for survival and reproduction. These reward components are either external (sensory, physical salience) or internal (generated by the brain; value, novelty/surprise salience, motivational salience). All five components together ensure adequate reward function.

5. Punishment

The second large category of motivating events besides rewards is punishment. Punishment produces negative Pavlovian learning and negative operant reinforcement, passive and active avoidance behavior and negative emotions like fear, disgust, sadness, and anger (143). Finer distinctions separate punishment (reduction of response strength, passive avoidance) from negative reinforcement (enhancing response strength, active avoidance).

6. Reward components

Rewarding stimuli, objects, events, situations, and activities consist of several major components. First, rewards have basic sensory components (visual, auditory, somatosensory, gustatory, and olfactory) (**FIGURE 1,** *left*), with physical parameters such as size, form, color, position, viscosity, acidity, and others. Food and liquid rewards contain chemical substances necessary for survival such as carbohydrates, proteins, fats, minerals, and vitamins, which contain physically measurable quantities of molecules. These sensory components act via specific sensory receptors on the brain. Some rewards consist of situations, which are detected by cognitive processes, or activities involving motor processes, which also constitute basic components analogous to sensory ones. Second, rewards are salient and thus elicit attention, which are manifested as orienting responses (**FIGURE 1,** *middle*). The salience of rewards derives from three principal factors, namely, their physical intensity and impact (physical salience), their novelty and surprise (novelty/surprise salience), and their general motivational impact shared with punishers (motivational salience). A separate form not included in this scheme, incentive salience, primarily addresses dopamine function in addiction and refers only to approach behavior (as opposed to learning) and thus to reward and not punishment (471). The term is at odds with current results on the role of dopamine in learning (see below) and reflects an earlier assumption of attentional dopamine function based on an initial phasic response component before distinct dopamine response components were recognized (see below). Third, rewards have a value component that determines the positively motivating effects of rewards and is not contained in, nor explained by, the sensory and attentional components (**FIGURE 1,** *right*). This component reflects behavioral preferences and thus is subjective and only partially determined by physical parameters. Only this component constitutes what we understand as a reward. It mediates the specific behavioral reinforcing, approach generating, and emotional effects of rewards that are crucial for the organism's survival and reproduction, whereas all other components are only supportive of these functions.

The major reward components together ensure maximal reward acquisition. Without the sensory component, reward discrimination would be difficult; without the attentional components, reward processing would be insufficiently prioritized; and without valuation, useless objects would be pursued. In practical reward experiments, the value component should be recognized as a distinct variable in the design and distinguished and uncorrelated from the sensory and attentional components.

The reward components can be divided into external components that reflect the impact of environmental stimuli, objects and events on the organism, and internal components generated by brain function. The sensory components are external, as they derive from external events and allow stimulus identification before evaluation can begin. In analogy, the external physical salience components lead to stimulus-driven attention. The foremost internal component is reward value. It is not inherently attached to stimuli, objects, events, situations, and activities but reflects the brain's assessment of their usefulness for survival and reproduction. Value cannot be properly defined by physical reward parameters but is represented in subjective preferences that are internal, private, unobservable, and incomparable between individuals. These preferences are elicited by approach behavior and choices that can be objectively measured. The internal nature of value extends to its associated motivational salience. Likewise, reward predictors are not

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

hardwired to outside events but require neuronal learning and memory processes, as does novelty/surprise salience which relies on comparisons with memorized events. Reward predictors generate top-down, cognitive attention that establishes a saliency map of the environment before the reward occurs. Further internal reward components are cognitive processes that identify potentially rewarding environmental situations, and motor processes mediating intrinsically rewarding movements.

B. Distal Reward Function Is Evolutionary Fitness

Modern biological theory conjectures that the currently existing organisms are the result of evolutionary competition. Advancing the idea about survival of the fittest organisms, Richard Dawkins stresses gene survival and propagation as the basic mechanism of life (114). Only genes that lead to the fittest phenotype will make it. The phenotype is selected based on behavior that maximizes gene propagation. To do so, the phenotype must survive and generate offspring, and be better at it than its competitors. Thus the ultimate, distal function of rewards is to increase evolutionary fitness by ensuring survival of the organism and reproduction. Then the behavioral reward functions of the present organisms are the result of evolutionary selection of phenotypes that maximize gene propagation. Learning, approach, economic decisions, and positive emotions are the proximal functions through which phenotypes obtain the necessary nutrients for survival, mating, and care for offspring.

Behavioral reward functions have evolved to help individuals to propagate their genes. Individuals need to live well and long enough to reproduce. They do so by ingesting the substances that make their bodies function properly. The substances are contained in solid and liquid forms, called foods and drinks. For this reason, foods and drinks are rewards. Additional rewards, including those used for economic exchanges, ensure sufficient food and drink supply. Mating and gene propagation is supported by powerful sexual attraction. Additional properties, like body form, enhance the chance to mate and nourish and defend offspring and are therefore rewards. Care for offspring until they can reproduce themselves helps gene propagation and is rewarding; otherwise, mating is useless. As any small edge will ultimately result in evolutionary advantage (112), additional reward mechanisms like novelty seeking and exploration widen the spectrum of available rewards and thus enhance the chance for survival, reproduction, and ultimate gene propagation. These functions may help us to obtain the benefits of distant rewards that are determined by our own interests and not immediately available in the environment. Thus the distal reward function in gene propagation and evolutionary fitness defines the proximal reward functions that we see in everyday behavior. That is why foods, drinks, mates, and offspring are rewarding.

The requirement for reward seeking has led to the evolution of genes that define brain structure and function. This is what the brain is made for: detecting, seeking, and learning about rewards in the environment by moving around, identifying stimuli, valuing them, and acquiring them through decisions and actions. The brain was not made for enjoying a great meal; it was made for getting the best food for survival, and one of the ways to do that is to make sure that people are attentive and appreciate what they are eating.

C. Types of Rewards

The term *reward* has many names. Psychologists call it positive reinforcer because it strengthens behaviors that lead to reward, or they call it outcome of behavior or goal of action. Economists call it a good or commodity and assess the subjective value for the decision maker as utility. We now like to identify the kinds of stimulus, object, event, activity, and situation that elicit the proximal functions of learning, approach, decision-making, and positive emotions and thus serve the ultimate, distal reward function of evolutionary fitness.

1. Primary homeostatic and reproductive rewards

To ensure gene propagation, the primary rewards mediate the survival of the individual gene carrier and her reproduction. These rewards are foods and liquids that contain the substances necessary for individual survival, and the activities necessary to mate, produce offspring, and care for the offspring. They are attractive and the main means to achieve evolutionary fitness in all animals and humans. Primary food and liquid rewards serve to correct homeostatic imbalances. They are the basis for Hull's drive reduction theory (242) that, however, would not apply to rewards that are not defined by homeostasis. Sexual behavior follows hormonal imbalances, at least in men, but is also strongly based on pleasure. To acquire and follow these primary alimentary and mating rewards is the main reason why the brain's reward system has evolved in the first place. Note that "primary" reward does not refer to the distinction between unconditioned versus conditioned reward; indeed, most primary rewards are learned and thus conditioned (foods are primary rewards that are typically learnt).

2. Nonprimary rewards

All other rewards serve to enhance the function of primary alimentary and mating rewards and thus enhance the chance for survival, reproduction, and evolutionary selection. Even though they are not homeostatic or reproductive rewards, they are rewards in their own rights. These nonprimary rewards can be physical, tangible objects like money, sleek cars, or expensive jewelry, or material liquids like a glass of wine, or particular ingredients like spices or alcohol. They can have particular pleasant sensory properties

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

like the visual features of a Japanese garden or a gorgeous sunset, the acoustic beauty of Keith Jarrett's Cologne Concert, the warm feeling of water in the Caribbean, the gorgeous taste of a gourmet dinner, or the irresistible odor of a perfume. Although we need sensory receptors to detect these rewards, their motivating or pleasing properties require further appreciation beyond the processing of sensory components **(FIGURE 1)**. A good example is Canaletto's Grand Canal **(FIGURE 2)** whose particular beauty is based on physical geometric properties, like the off-center Golden Ratio position reflecting a Fibonacci sequence (320). However, there is nothing intrinsically rewarding in this ratio of physical proportions. Its esthetic (and monetary) value is entirely determined by the subjective value assigned by our brain following the sensory processing and identification of asymmetry. Although we process great taste or smell as sensory events, we appreciate them as motivating and pleasing due to our subjective valuation. This rewarding function, cultivated in gourmet eating, enhances the appreciation and discrimination of high-quality and energy-rich primary foods and liquids and thus ultimately leads to better identification of higher quality food and thus higher survival chances (as gourmets are usually not lacking food, this may be an instinctive trait for evolutionary fitness). Sexual attraction is often associated with romantic love that, in contrast to straightforward sex, is not required for reproduction and therefore does not have primary reward functions. However, love induces attachment and facilitates care for offspring and thus supports gene propagation. Sexual rewards constitute also the most straightforward form of social rewards. Other social rewards include friendship, altruism, general social encounters, and societal activities that promote group coherence, cooperation, and competition which are mutually beneficial for group members and thus evolutionarily advantageous.

Nonphysical, nonmaterial rewards, such as novelty, gambling, jokes, suspense, poems, or relaxation, are attractive but less tangible than primary rewards. These rewards have no homeostatic basis and no nutrient value, and often do not promote reproduction directly. We may find the novelty of a country, the content of a joke, or the sequence of words in a poem more rewarding than the straightforward physical aspects of the country or the number of words in the joke or poem. But novelty seeking, and to some extent gambling, may help to encounter new food sources. Jokes, suspense, poems, and relaxation may induce changes of viewpoints and thus help to understand the world, which may help us to consider alternative food sources and mating partners, which is helpful when old sources dry up. Although these rewards act indirectly, they increase evolutionary fitness by enhancing the functions of primary alimentary and reproductive rewards.

Rewards can also be intrinsic to behavior (31, 546, 547). They contrast with extrinsic rewards that provide motivation for behavior and constitute the essence of operant behavior in laboratory tests. Intrinsic rewards are activities that are pleasurable on their own and are undertaken for their own sake, without being the means for getting extrinsic rewards. We may even generate our own rewards through internal decisions. Mice in the wild enter wheels and run on them on repeated occasions without receiving any other reward or benefit, like the proverbial wheel running hamster (358). Movements produce proprioceptive stimulation in muscle spindles and joint receptors, touch stimulation on the body surface, and visual stimulation from seeing the movement, all of which can be perceived as pleasurable and thus have reward functions. Intrinsic rewards are genuine rewards in their own right, as they induce learning, approach, and pleasure, like perfectioning, playing, and enjoying the piano. Although they can serve to condition higher order rewards, they are not conditioned, higher order rewards, as attaining their reward properties does not require pairing with an unconditioned reward. Other examples for intrinsic rewards are exploration, own beauty, gourmet eating, visiting art exhibitions, reading books, taking power and control of people, and investigating the natural order of the world. The pursuit of intrinsic rewards seems private to the individual but may inadver-

FIGURE 2. Subjective esthetic reward value derived from objective physical properties. The beauty of the Canaletto picture depends on the Golden Ratio of horizontal proportions, defined as $(a + b)/a = a/b \sim 0.618$; a and b for width of image. The importance of geometric asymmetry becomes evident when covering the left part of the image until the distant end of the canal becomes the center of the image: this increases image symmetry and visibly reduces beauty. However, there is no intrinsic reason why physical asymmetry would induce subjective value: the beauty appears only in the eye of the beholder. (Canaletto: The Upper Reaches of the Grand Canal in Venice, 1738; National Gallery, London.)

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

tently lead to primary and extrinsic rewards, like an explorer finding more food sources by venturing farther afield, a beauty queen instinctively promoting attractiveness of better gene carriers, a gourmet improving food quality through hightened culinary awareness, an artist or art collector stimulating the cognitive and emotional capacities of the population, a scholar providing public knowledge from teaching, a politician organizing beneficial cooperation, and a scientist generating medical treatment through research, all of which enhance the chance of survival and reproduction and are thus evolutionary beneficial. The double helix identified by Watson and Crick for purely scientific reasons is now beneficial for developing medications. The added advantage of intrinsic over solely extrinsic rewards is their lack of narrow focus on tangible results, which helps to develop a larger spectrum of skills that can be used for solving wider ranges of problems. Formal mathematical modeling confirms that systems incorporating intrinsic rewards outperform systems relying only on extrinsic rewards (546). Whereas extrinsic rewards such as food and liquids are immediately beneficial, intrinsic rewards are more likely to contribute to fitness only later. The fact that they have survived evolutionary selection suggests that their later benefits outweigh their immediate costs.

D. What Makes Rewards Rewarding?

Why do particular stimuli, objects, events, situations, and activities serve as rewards to produce learning, approach behavior, choices, and positive emotions? There are four separate functions and mechanisms that make rewards rewarding. However, these functions and mechanisms serve the common proximal and distal reward functions of survival and gene propagation. Individuals try to maximize one mechanism only to the extent that the other mechanisms are not compromised, suggesting that the functions and mechanisms are not separate but interdependent.

1. Homeostasis

The first and primary reward function derives from the need of the body to have particular substances for building its structure and maintaining its function. The concentration of these substances and their derivatives is finely regulated and results in homeostatic balance. Deviation from specific set points of this balance requires replenishment of the lost substances, which are contained in foods and liquids. The existence of hunger and thirst sensations demonstrates that individuals associate the absence of necessary substances with foods and liquids. We obviously know implicitly which environmental objects contain the necessary substances. When the blood sodium concentration exceeds its set point, we drink water, but depletion of sodium leads to ingestion of salt (472).

Two brain systems serve to maintain homeostasis. The hypothalamic feeding and drinking centers together with intestinal hormones deal with immediate homeostatic imbalances by rapidly regulating food and liquid intake (24, 46). In contrast, the reward centers mediate reinforcement for learning and provide advance information for economic decisions and thus are able to elicit behaviors for obtaining the necessary substances well before homeostatic imbalances and challenges arise. This preemptive function is evolutionarily beneficial as food and liquid may not always be available when an imbalance arises. Homeostatic imbalances are the likely source of hunger and thirst drives whose reduction is considered a prime factor for eating and drinking in drive reduction theories (242). They engage the hypothalamus for immediate alleviation of the imbalances and the reward systems for preventing them. The distinction in psychology between drive reduction for maintaining homeostasis and reward incentives for learning and pursuit may grossly correspond to the separation of neuronal control centers for homeostasis and reward. The neuroscientific knowledge about distinct hypothalamic and reward systems provides important information for psychological theories about homeostasis and reward.

The need for maintaining homeostatic balance explains the functions of primary rewards and constitutes the evolutionary origin of brain systems that value stimuli, objects, events, situations, and activities as rewards and mediate the learning, approach, and pleasure effects of food and liquid rewards. The function of all nonprimary rewards is built onto the original function related to homeostasis, even when it comes to the highest rewards.

2. Reproduction

In addition to acquiring substances, the other main primary reward function is to ensure gene propagation through sexual reproduction, which requires attraction to mating partners. Sexual activity depends partly on hormones, as shown by the increase of sexual drive with abstinence in human males. Many animals copulate only when their hormones put them in heat. Castration reduces sexual responses, and this deficit is alleviated by testosterone administration in male rats (146). Thus, as with feeding behavior, hormones support the reward functions involved in reproduction.

3. Pleasure

Pleasure is not only one of the three main reward functions but also provides a definition of reward. As homeostasis explains the functions of only a limited number of rewards, the prevailing reason why particular stimuli, objects, events, situations, and activities are rewarding may be pleasure. This applies first of all to sex (who would engage in the ridiculous gymnastics of reproductive activity if it were not for the pleasure) and to the primary homeostatic rewards of food and liquid, and extends to money, taste, beauty, social encounters and nonmaterial, internally set, and intrinsic

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

rewards. Pleasure as the main effect of rewards drives the prime reward functions of learning, approach behavior, and decision making and provides the basis for hedonic theories of reward function. We are attracted by most rewards, and exert excruciating efforts to obtain them, simply because they are enjoyable.

Pleasure is a passive reaction that derives from the experience or prediction of reward and may lead to a longer lasting state of happiness. Pleasure as hallmark of reward is sufficient for defining a reward, but it may not be necessary. A reward may generate positive learning and approach behavior simply because it contains substances that are essential for body function. When we are hungry we may eat bad and unpleasant meals. A monkey who receives hundreds of small drops of water every morning in the laboratory is unlikely to feel a rush of pleasure every time it gets the 0.1 ml. Nevertheless, with these precautions in mind, we may define any stimulus, object, event, activity, or situation that has the potential to produce pleasure as a reward.

Pleasure may add to the attraction provided by the nutrient value of rewards and make the object an even stronger reward, which is important for acquiring homeostatically important rewards. Once a homeostatic reward is experienced, pleasure may explain even better the attraction of rewards than homeostatis. Sensory stimuli are another good example. Although we employ arbitrary, motivationally neutral stimuli in the laboratory for conditioning, some stimuli are simply rewarding because they are pleasant to experience. The esthetic shape, color, texture, viscosity, taste, or smell of many rewards are pleasant and provide own reward value independently of the nutrients they contain (although innate and even conditioned mechanisms may also play a role, see below). Examples are changing visual images, movies, and sexual pictures for which monkeys are willing to exert effort and forego liquid reward (53, 124), and the ever increasing prices of paintings (fame and pride may contribute to their reward value). Not surprisingly, the first animal studies eliciting approach behavior by electrical brain stimulation interpreted their findings as discovery of the brain's pleasure centers (398), which were later partly associated with midbrain dopamine neurons (103, 155) despite the notorious difficulties of identifying emotions in animals.

4. Innate mechanisms

Innate mechanisms may explain the attractions of several types of reward in addition to homeostasis, hormones, and pleasure. A powerful example is parental affection that derives from instinctive attraction. Ensuring the survival of offspring is essential for gene propagation but involves efforts that are neither driven by homeostasis nor pleasure. As cute as babies are, repeatedly being woken up at night is not pleasurable. Generational attraction may work also in the other direction. Babies look more at human faces than at scrambled pictures of similar sensory intensity (610), which might be evolutionary beneficial. It focuses the baby's attention on particularly important stimuli, initially those coming from parents. Other examples are the sensory aspects of rewards that do not evoke pleasure, are nonnutritional, and are not conditioned but are nevertheless attractive. These may include the shapes, colors, textures, viscosities, tastes, or smells of many rewards (539), although some of them may turn out to be conditioned reinforcers upon closer inspection (640).

5. Punisher avoidance

The cessation of pain is often described as pleasurable. Successful passive or active avoidance of painful events can be rewarding. The termination or avoidance might be viewed as restoring a "homeostasis" of well being, but it is unrelated to proper vegetative homeostasis. Nor is avoidance genuine pleasure, as it is built on an adverse event or situation. The opponent process theory of motivation conceptualizes the reward function of avoidance (552), suggesting that avoidance may be a reward in its own right. Accordingly, the simple timing of a conditioned stimulus relative to an aversive event can turn punishment into reward (584).

E. Rewards Require Brain Function

1. Rewards require brains

Although organisms need sensory receptors to detect rewards, the impact on sensory receptors alone does not explain the effects of rewards on behavior. Nutrients, mating partners, and offspring are not attractive by themselves. Only the brain makes them so. The brain generates subjective preferences that reflect on specific environmental stimuli, objects, events, situations, and activities as rewards. These preferences are elicited by choices and quantifiable from behavioral reactions, typically choices but also reaction times and other measures. Reward function is explained by assuming the notion of value attributed to individual rewards. Value is not a physical property but determined by brain activity that interprets the potential effect of a reward on survival and reproduction. Thus rewards are internal to the brain and based entirely on brain function (547).

2. Explicit neuronal reward signals

Information processing systems work with signals. In brains, the signals that propagate through the circuits are the action potentials generated by each neuron. The output of the system is the observable behavior. In between are neurons and synapses that transmit and alter the signals. Each neuron works with thousands of messenger molecules and membrane channels that determine the action potentials. The number of action potentials, and somewhat their

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

pattern, varies monotonically with sensory stimulation, discrimination, and movements (3, 385). Thus the key substrates for the brain's function in reward are specific neuronal signals that occur in a limited number of brain structures, including midbrain dopamine neurons, striatum, amygdala, and orbitofrontal cortex **(FIGURE 3)**. Reward signals are also found in most component structures of the basal ganglia and the cerebral cortical areas, often in association with sensory or motor activity. The signals can be measured as action potentials by neurophysiology and are also reflected in transmitter concentrations assessed by electrochemistry (638) and as synaptic potentials detected by magnetic resonance imaging in blood oxygen level dependent (BOLD) signals (328). Whereas lesions in humans and animals demonstrate necessary involvements of specific brain structures in behavioral processes, they do not inform about the way the brain processes the information underlying these processes. Electric and optogenetic stimulation evokes action potentials and thus helps to dissect the influence of individual brain structures on behavior, but it does not replicate the natural signals that occur simultaneously in several interacting structures. Thus the investigation of neuronal signals is an important method for understanding crucial physiological mechanisms for the survival and reproduction of biological organisms.

3. Reward retina

Neuronal signals in sensory systems originate in specific receptors that define the signal content. However, rewards have no dedicated receptors. Neuronal processing would benefit from an explicit signal that identifies a reward irrespective of sensory properties and irrespective of actions required to obtain it. The signal might be analogous to visual responses of photoreceptors in the retina that constitute the first processing stage for visual perception. To obtain an explicit reward signal, the brain would extract the rewarding component from heterogeneous, polysensory environmental objects and events. A signal detecting the reward properties of an apple should not be concerned with its color unless color informs about reward properties of the fruit. Nor should it code the movement required to obtain the apple, other than assessing the involved effort as economic cost. External visual, somatic, auditory, olfactory, and gustatory stimuli predicting original, unconditioned rewards become conditioned rewards through Pavlovian conditioning. The issue for the brain is then to extract the reward information from the heterogeneous responses to the original and conditioned rewards and generate a common reward signal. Neurons carrying such a signal would constitute the first stage in the brain at which the reward property of environmental objects and events would be coded and conveyed to centers engaged in learning, approach, choice, and pleasure. Such abstract reward neurons would be analogous to the retinal photoreceptors as first visual processing stage (519).

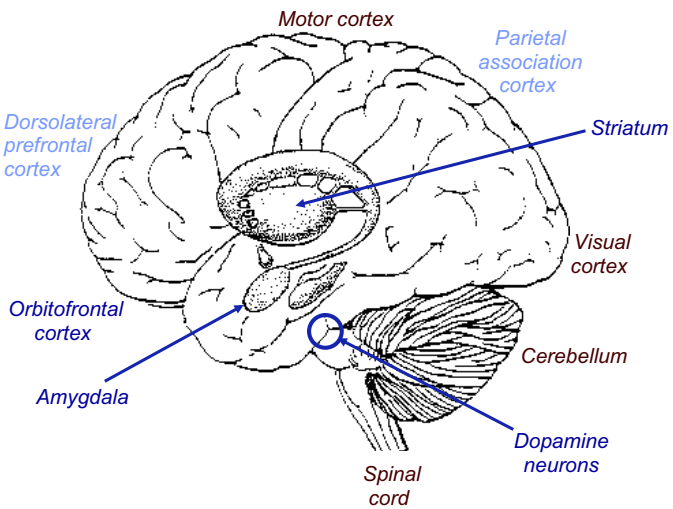


FIGURE 3. Principal brain structures for reward and decision-making. Dark blue: main structures containing various neuronal subpopulations coding reward without sensory stimulus or motor action parameters ("explicit reward signals"). Light blue: structures coding reward in conjunction with sensory stimulus or motor action parameters. Maroon: non-reward structures. Other brain structures with explicit or conjoint reward signals are omitted for clarity.

Despite the absence of specific reward receptors, there are chemical, thermal, and mechanical receptors in the brain, gut, and liver that detect important and characteristic reward ingredients and components, such as glucose, fatty acids, aromatic amino acids, osmolality, oxygen, carbon dioxide, temperature, and intestinal volume, filling, and contractions. In addition to these exteroceptors, hormone receptors are stimulated by feeding and sex (24, 46). These receptors are closest to being reward receptors but nevertheless detect only physical, sensory reward aspects, whereas reward value is still determined by internal brain activity.

The absence of dedicated receptors that by themselves signal reward value may not reflect evolutionary immaturity, as rewards are as old as multicellular organisms. Rather valuation separate from physical receptors may be an efficient way of coping with the great variety of objects that can serve as rewards at one moment or another, and of adapting reward value to changing requirements, including deprivation and satiation. Rather than having a complex, omnipotent, polysensory receptor that is sensitive to all possible primary and conditioned rewards and levels of deprivation, however infrequent they may occur, it might be easier to have a neuronal mechanism that extracts the reward information from the existing sensory receptors. The resulting neuronal reward signal would be able to detect rewarding properties in a maximum number of environmental objects, increase the harvest of even rare rewards, and relate their value to current body states, which all together enhance the chance of survival. Such a signal would be an efficient solution to the existential problem of benefitting from the

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

largest possible variety of rewards with reasonable hardware and energy cost.

III. LEARNING

A. Principles of Reward Learning

1. Advantage of learning

Learning is crucial for evolutionary fitness. It allows biological organisms to obtain a large variety of rewards in a wide range of environments without the burden of maintaining hard-wired mechanisms for every likely and unlikely situation. Organisms that can learn and adapt to their environments can live in more widely varying situations and thus acquire more foods and mating partners. Without learning, behavior for these many situations would need to be preprogrammed which would require larger brains with more energy demands. Thus learning saves brain size and energy and thus enhances evolutionary fitness. These advantages likely prompted the evolutionary selection of the learning function of rewards.

Learning processes lead to the selection of those behaviors that result in reward. A stimulus learned by Pavlovian conditioning elicits existing behavior when the stimulus is followed by reward. Natural behavioral reactions such as salivation or approach that improve reward acquisition become more frequent when a stimulus is followed by a reward. Male fish receiving a Pavlovian conditioned stimulus before a female approaches produce more offspring than unconditioned animals (223), thus demonstrating the evolutionary benefit of conditioning. Operant learning enhances the frequency of existing behavior when this results in reward. Thorndike's cat ran around randomly until it came upon a lever that opened a door to a food source. Then its behavior focused increasingly on the lever for the food. Thus Pavlovian and operant learning commonly lead to selection of behavior that is beneficial for survival.

Learning is instrumental for selecting the most beneficial behaviors that result in the best nutrients and mating partners in the competition for individual and gene survival. In this sense selection through learning is analogous to the evolutionary selection of the fittest genes. For both, the common principle is selection of the most efficient characteristics. The difference is on the order of time and scale. Selection of behavior through learning is based on outcome over minutes and hours, whereas selection of traits through evolution is based on survival of the individual. Evolutionary selection includes the susceptibility to reward and the related learning mechanisms that result in the most efficient acquisition of nutrients and mating partners.

2. Pavlovian learning

In Pavlov's experiment, the dog's salivation following the bell suggests that it anticipates the sausage. A Pavlovian conditioned visual stimulus would have a similar effect. The substances that individuals need to live are packaged in objects or liquids or are contained in animals they eat. We need to recognize a pineapple to get its juice that contains necessary substances like water, sugar, and fibers. Recognition of these packages could be hard wired into brain function, which would require a good number of neurons to detect a reasonable range of rewards. Alternatively, a flexible mechanism could dynamically condition stimuli and events with the necessary rewards and thus involve much less neurons representing the rewards. That mechanism makes individuals learn new packages when the environment changes. It also allows humans to manufacture new packages that were never encountered during evolution. Through Pavlovian conditioning humans learn the labels for hamburgers, baby food, and alcoholic beverages. Thus Pavlovian conditioning touches the essence of reward functions in behavior and allows individuals to detect a wide range of rewards from a large variety of stimuli while preventing run away neuron numbers and brain size. It is the simplest form of learning that increases evolutionary fitness and was thus selected by evolution.

Pavlovian conditioning makes an important conceptual point about reward function. We do not need to act to undergo Pavlovian conditioning. It happens without our own doing. But our behavior reveals that we have learned something, whether we wanted to or not. Pavlov's bell by itself would not make the dog salivate. But the dog salivates when it hears the bell that reliably precedes the sausage. From now on, it will always salivate to a bell, in particular when the bell occurs in the laboratory in which it has received all the sausages. Salivation is an automatic component of appetitive, vegetative approach behavior.

Although initially defined to elicit vegetative reactions, Pavlovian conditioning applies also to other behavioral reactions. The bell announcing the sausage elicits also approach behavior involving eye, limb, and licking movements. Thus Pavlovian learning concerns not only vegetative reactions but also skeletal movements. Furthermore, Pavlovian conditioning occurs in a large variety of behavioral situations. When operant learning increases actions that lead to reward, the more reward inadvertently conditions the involved stimuli in a Pavlovian manner. Thus Pavlovian and operant processes often go together in learning. When we move to get a reward, we also react to the Pavlovian conditioned, reward-predicting stimuli. In choices between different rewards, Pavlovian conditioned stimuli provide crucial information about the rewarded options. Thus Pavlovian and operant behavior constitute two closely linked forms of learning that extend well beyond laboratory experiments. Finally, Pavlovian conditioning concerns not

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

only the motivating components of rewards but also their attentional aspects. It confers motivational salience to arbitrary stimuli that elicits stimulus-driven attention and directs top-down attention to the reward and thus focuses behavior on pursuing and acquiring the reward. Taken together, Pavlovian conditioning constitutes a fundamental mechanism that is crucial for a large range of learning processes.

3. Reward prediction and information

Appetitive Pavlovian conditioning takes the past experience of rewards to form predictions and provide information about rewards. The bell in Pavlov's conditioning experiment has become a sausage predictor for the animal. By licking to the bell the dog demonstrates an expectation of the sausage that was evoked by the bell. Thus Pavlovian conditioning of the bell to the sausage has made the intrinsically neutral bell a reward predictor. The external stimulus or event (the bell) has become a predictor (of the sausage) and evokes an internal expectation (of the sausage) in the animal **(FIGURE 4)**. The same holds for any other reward predictor. The pineapple and the hamburger are predictors of the nutrients they contain.

Predictions tell us what is going to happen. This includes predictors of probabilistic rewards, even if the reward does not occur in every instance. Pavlov's bell predicts a sausage to the dog and at the same time induces salivation, licking, and approach. Thus Pavlovian conditioning confers two components, a predictive and an incentive property. The predictive component indicates what is going to happen now. The incentive property induces action, such as salivation, licking, and approach, which help to obtain and ingest the sausage. The two properties are separated in time in the classic and widely used delayed response tasks in which initial instructive stimuli have reward-predicting properties without eliciting a reward-directed action (but ocular saccades), whereas the final trigger or releasing stimulus induces the behavioral action and thus has both predictive and incentive properties. The predictive and incentive properties are separated spatially with different stimulus and goal (lever) positions and are dissociable in the behavior of particular rat strains. Specially bred sign-tracking rats approach the conditioned predictive stimulus, whereas goal trackers go directly to the reward upon stimulus appearance, indicating separation of predictive and incentive properties in goal trackers (163).

The predictive component can be further distinguished from an informational component. Once a stimulus has been Pavlovian conditioned, it confers information about the reward. The information does not necessarily predict explicitly what is actually going to happen every time, not even probabilistically. Through Pavlovian conditioning I have learned that a particular sign on a building indicates a pub because I have experienced the beer inside. The atmosphere and the beer represent a value to me that is assigned in a Pavlovian manner to the pub sign. I can pass the pub sign without entering the pub, however difficult that may be. Thus the sign is informational but does not truly predict a pint, only its potential, nor does it have the incentive properties at that moment to make me go in and get one. Then I may run into an unknown pub and experience a different beer, and I undergo another round of Pavlovian conditioning to the value of that particular pub sign. When I need to choose between different pubs, I use the information about their values rather than explicit predictions of getting a beer in every one of them. Thus Pavlovian conditioning sets up predictions that contain reward information. The predictions indicate that a reward is going to occur this time, whereas the reward informations do not necessarily result in reward every time.

The distinction between explicit prediction and information is important for theories of competitive decision-making. Value information about a reward is captured in the term of action value in machine learning, which indicates the value resulting from a particular action, without requiring to actually choosing it and obtaining that value (575). In analogy, object value indicates the value of a reward object irrespective of choosing and obtaining it. Individuals make decisions by comparing the values between the different actions or objects available and then selecting only the action or object with the highest value (see below **FIGURE 36*C***). Thus the decision is based on the information about

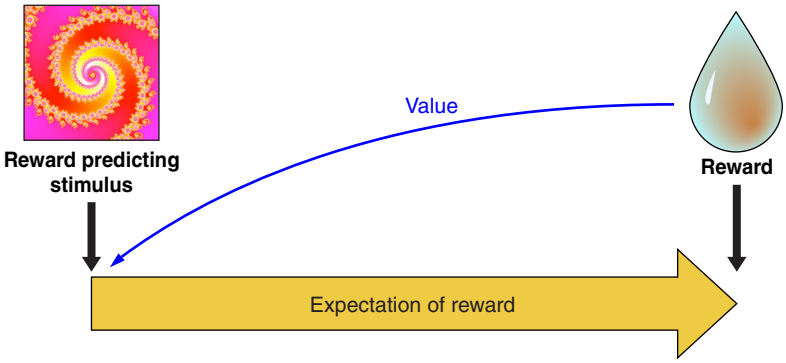


FIGURE 4. Pavlovian reward prediction. With conditioning, an arbitrary stimulus becomes a reward predictor and elicits an internal expectation of reward. Some of the behavioral reactions typical for reward occur also after the stimulus (Pavlovian stimulus substitution), in particular approach behavior, indicating that the stimulus has acquired reward value (blue arrow).

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

each reward value and not on the explicit prediction that every one of these rewards will be actually experienced, as only the chosen reward will occur. In this way, Pavlovian conditioning is a crucial building block for reward information in economic decisions. Separate from the decision mechanism, the acquisition and updating of action values and object values requires actual experience or predictions derived from models of the world, which is captured by model free and model-based reinforcement learning, respectively. Other ways to establish reward predictions and informations involve observational learning, instructions, and deliberate reflections that require more elaborate cognitive processes. All of these forms produce reward information that allows informed, competitive choices between rewards.

4. Pavlovian learning produces higher order rewards

Reward-predicting stimuli established through Pavlovian conditioning become higher order, conditioned rewards. By itself a red apple is an object without any intrinsic meaning. However, after having experienced its nutritious and pleasantly tasting contents, the apple with its shape and color has become a reinforcer in its own right. As a higher order, conditioned reward, the apple serves all the defining functions of rewards, namely, learning, approach behavior, and pleasure. The apple serves as a reinforcer for learning to find the vendor's market stand. When we see the apple, we approach it if we have enough appetite. We even approach the market stand after the apple has done its learning job. And seeing the delicious apple evokes a pleasant feeling. The apple serves these functions irrespective of explicitly predicting the imminent reception of its content or simply informing about its content without being selected. Thus Pavlovian conditioning labels arbitrary stimuli and events as higher order rewards that elicit all reward functions. All this happens without any physical change in the apple. The only change is in the eye of the beholder, which we infer from behavioral reaction.

The notion that Pavlovian conditioning confers higher order reward properties to arbitrary stimuli and events allows us to address a basic question. Where in the body is the original, unconditioned effect of rewarding substances located (517, 640)? The functions of alimentary rewards derive from their effects on homeostatic mechanisms involved in building and maintaining body structure and function. In the case of an apple, the effect might be an increase in blood sugar concentration. As this effect would be difficult to perceive, the attraction of an apple might derive from various stimuli conditioned to the sugar increase, such vision of the apple, taste, or other conditioned stimuli. Mice with knocked out sweet taste receptors can learn, approach, and choose sucrose (118), suggesting that the calories and the resulting blood sugar increase constitute the unconditioned reward effect instead of or in addition to taste. Besides being an unconditioned reward (by evoking pleasure or via innate mechanisms, see above), taste may also be a conditioned reward, similar to the sensory properties of other alimentary rewards, including temperature (a glass of unpleasant, luke warm water predicting reduced plasma osmolality) and viscosity (a boring chocolate drink predicting calories). These sensations guide ingestion that will ultimately lead to the primary reward effect. However, the conditioned properties are only rewarding as long as the original, unconditioned reward actually occurs. Diets lacking just a single essential amino acid lose their reward functions within a few hours or days and food aversion sets in (181, 239, 480). The essential amino acids are detected by chemosensitive neurons in olfactory cortex (314), suggesting a location for the original effects of amino acids with rewarding functions. Thus the better discernible higher order rewards facilitate the function of primary rewards that have much slower and poorly perceptible vegetative effects.

A similar argument can be made for rewards that do not address homeostasis but are based on the pleasure they evoke. The capacity of some unconditioned rewards to evoke pleasure and similar positive emotions is entirely determined by brain physiology. Pleasure as an unconditioned reward can serve to produce higher order, conditioned rewards that are also pleasurable. A good example are sexual stimuli, like body parts, that have unconditioned, innate reward functions and serve to make intrinsically neutral stimuli, like signs for particular stores or bars, predictors of sexual events and activity.

Taken together, the primary, homeostatic, or pleasurable reward functions are innate and determined by the physiology of the body and its brain that emerged from evolutionary selection. Individuals without such brains or with brains not sensing the necessary rewards have conceivably perished in evolution. The primary rewards come in various forms and "packages" depending on the environment in which individuals live. The same glucose molecule is packaged in sugar beets or bananas in different parts of the world. Conditioning through the actual experience within a given environment facilitates the detection of the primary rewards. Thus, in producing higher order rewards, Pavlovian learning allows individuals to acquire and optimally select homeostatic and pleasurable rewards whose primary actions may be distant and often difficult to detect.

5. Operant learning

Getting rewards for free sounds like paradise. But after Adam took Eve's apple, free rewards became more rare, and Pavlovian learning lost its exclusiveness. We often have to do something to get rewards. Learning starts with a reward that seems to come out of nowhere but actually came from an action, and we have to try and figure out what made it appear. The rat presses a lever by chance and receives a drop of sugar solution. The sugar is great, and it presses again, and again. It comes back for more. The frequency of its

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

behavior that results in the sugar increases. This is positive reinforcement, the more reward the animal gets the more it acts, the essence of Thorndike's Law of Effect (589). Crucial for operant learning is that the animal learns about getting the sugar drop only by pressing the lever. Without lever pressing it would not get any sugar, and the behavior would not get reinforced. If the sugar comes also without lever pressing, the sugar does not depend on lever pressing, and the rat would not learn to operate the lever (but it might learn in a Pavlovian manner that the experimental box predicts sugar drops).

Operant reinforcement means that we act more when we get more from the action. However, due to the somewhat voluntary nature of the action involved, the inverse works also. The more we act, the more we get. Our behavior determines how much we get. We get something for an action, which is the bonus or effort function of reward. We can achieve something by acting. We have control over what we get. Knowing this can motivate us. Operant learning allows behavior to become directed at a rewarding goal. We act to obtain a reward. Behavior becomes causal for obtaining reward. Thus operant learning represents a mechanism by which we act on the world to obtain more rewards. Coconuts drop only occasionally from the tree next to us. But we can shake the tree or sample other trees, not just because the coconut reinforced this behavior but also because we have the goal to get more coconuts. In eliciting goal-directed action, operant learning increases our possibilities to obtain the rewards we need and thus enhances our chance for survival and reproduction.

6. Value updating, goal-directed behavior, and habits

Value informations for choices need to be updated when reward conditions change. For example, food consumption increases the specific satiety for the consumed reward and thus decreases its subjective value while it is being consumed. In addition, the general satiety evolving in parallel lowers the reward value of all other foods. Although the values of all rewards ever encountered could be continuously updated, it would take less processing and be more efficient to update reward values only at the time when the rewards are actually encountered and contribute to choices. Reward values that are computed relative to other options should be updated when the value of any of the other option changes.

Operant learning directs our behavior towards known outcomes. In goal-directed behavior, the outcome is represented during the behavior leading to the outcome, and furthermore, the contingency of that outcome on the action is represented (133). In contrast, habits arise through repeated performance of instrumental actions in a stereotyped fashion. Habits are not learned with a new task from its outset but form gradually after an initial declarative, goal-directed learning phase (habit formation rather than habit learning). They are characterized by stimulus-response (S-R) performance that becomes "stamped-in" by reinforcement and leads to inflexible behavior. Despite their automaticity, habits extinguish with reinforcer omission, as do other forms of behavior. In permitting relatively automatic performance of routine tasks, habits free our attention, and thus our brains, for other tasks. They are efficient ways of dealing with repeated situations and thus are evolutionary beneficial.

Value updating differs between goal-directed behavior, Pavlovian predictions, and habits (133). Specific tests include devaluation of outcomes by satiation that reduces subjective reward value (133). In goal-directed behavior, such satiation reduces the operant response the next time the action is performed. Crucially, the effect occurs without pairing the action with the devalued outcome, which would be conventional extinction. The devaluation has affected the representation of the outcome that is being accessed during the action. In contrast, habits continue at unreduced magnitude until the devalued outcome is experienced after the action, at which point conventional extinction occurs. Thus, with habits, values are only retrieved from memory and updated at the time of behavior. Devaluation without action-outcome pairing is a sensitive test for distinguishing goal-directed behavior from habits and is increasingly used in neuroscience (133, 399, 646). Like goal-directed behavior, Pavlovian values can be sensitive to devaluation via representations and update similarly without repairing, convenient for rapid economic choices (however, Pavlovian conditioning is not goal "directed" as it does not require actions). Correspondingly, reward neurons in the ventral pallidum start responding to salt solutions when they become rewarding following salt depletion, even before actually experiencing the salt in the new depletion state (591). In conclusion, whether updating is immediate in goal-directed behavior or gradually with habit experience, values are only computed and updated at the time of behavior.

B. Neuronal Signals for Reward Learning

1. Basic conditions of learning: contiguity and contingency

In both Pavlovian and operant conditioning, an event (stimulus or action) is paired in time, and often in location, with a reinforcer to produce learning. In delay conditioning, the stimulus or action lasts until the reinforcer occurs. In trace conditioning, which is often less effective (47, 423), the stimulus or action terminates well before the reinforcer. Trace conditioning, but not delay conditioning, with aversive outcomes is disrupted by lesions of the hippocampus (47, 356, 553), although presenting the stimulus briefly again with the outcome restores trace conditioning (27). Thus temporal contiguity is important for conditioning.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

What Pavlov did not know was that stimulus-reward pairing was not sufficient for conditioning. Rescorla showed in his "truly random" experiment that the key variable underlying conditioning is the difference in outcome between the presence and absence of a stimulus (or action) (459). Conditioning occurs only when reinforcement is dependent or "contingent" on the stimulus (or an action), and this condition applies also to rewards (126). When the same reward occurs also without the stimulus, the stimulus is still paired with reward but carries no specific reward information, and no conditioning occurs. Thus contingency refers to the intuitive notion that we learn only something that carries specific information or that decreases the uncertainty in the environment (reduction of informational entropy). Contingency is usually defined as difference in reinforcement between stimulus (or action) presence and background (stimulus absence) (**FIGURE 5*A***). A positive difference produces excitatory conditioning (positive learning, left of diagonal), whereas a negative difference produces inhibitory conditioning (negative learning, right of diagonal). Zero contingency produces no conditioning (diagonal line) (but likely perceptual learning). Backward conditioning (reward before stimulus) is ineffective or even produces inhibitory conditioning because the reward occurs without relation to the stimulus. Timing theories of conditioning define contingency as stimulus-to-intertrial ratio of reinforcement rates (175). More frequent reward during the stimulus compared with the no-stimulus intertrial interval makes the reward more contingent on the stimulus. When the same reward occurs later after stimulus onset (temporal discounting), stimulus-to-intertrial reward ratio decreases and thus lowers contingency, although other explanations are also valid (see below). Thus, after contiguity and stimulus-outcome pairing, contingency is the crucial process in conditioning. It determines reward predictions and informations, which are crucial for efficient decisions, and it produces higher order rewards, which drive most of our approach behavior.

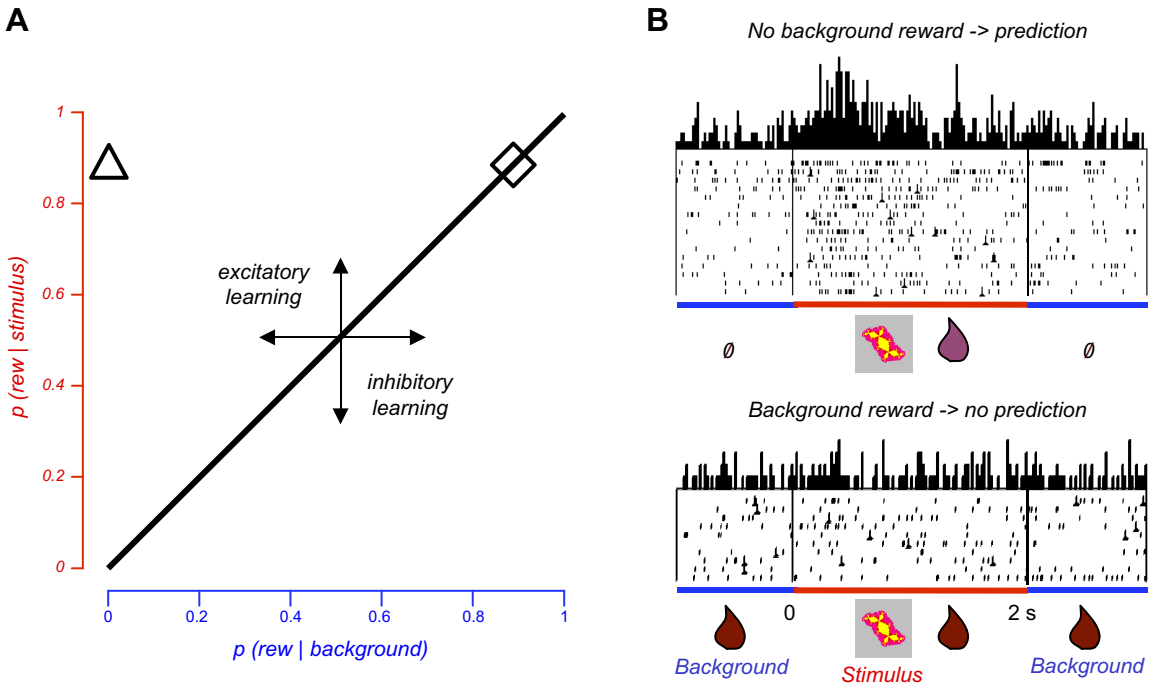


FIGURE 5. Reward contingency. *A:* role of contingency in learning. Contingency is shown as reward difference between stimulus presence and absence (background). Abscissa and ordinate indicate conditional reward probabilities. Higher reward probability in the presence of the stimulus compared with its absence (background) induces positive conditioning (positive contingency, triangle). No learning occurs with equal reward probabilities between stimulus and background (diagonal line, rhombus). Reward contingency applies to Pavlovian conditioning (shown here; reward contingent on stimulus) and operant conditioning (reward contingent on action). [Graph inspired by Dickinson (132).] *B:* contingency-dependent response in single monkey amygdala neuron. *Top*: neuronal response to conditioned stimulus (reward $P = 0.9$; red) set against low background reward probability ($P = 0.0$; blue) (triangle in *A*). *Bottom*: lack of response to same stimulus paired with same reward ($P = 0.9$) when background produces same reward probability ($P = 0.9$) (rhombus in *A*) which sets reward contingency to 0 and renders the stimulus uninformative. Thus the neuronal response to the reward-predicting stimulus depends entirely on the background reward and thus reflects reward contingency rather than stimulus-reward pairing. A similar drop in neuronal responses occurs with comparable variation in reward magnitude instead of probability (41). Perievent time histograms of neuronal impulses are shown above raster displays in which each dot denotes the time of a neuronal impulse relative to a reference event (stimulus onset, time = 0, vertical line at left; line to right indicates stimulus offset). [From Bermudez and Schultz (41).]

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

2. Neuronal contingency tests

Standard neuronal learning and reversal studies test contiguity by differentially pairing stimuli (or actions) with reward. Although they sometimes claim contingency testing, the proper assessment of contingency requires the distinction against contiguity of stimulus-reward pairing. As suggested by the "truly random" procedure (459), the distinction can be achieved by manipulating reward in the absence of the stimulus (or action).

A group of amygdala neurons respond to stimuli predicting more frequent or larger rewards compared with background without stimuli (**FIGURE 5*B***, *top*). These are typical responses to reward-predicting stimuli seen for 30 years in all reward structures. However, many of these neurons lose their response in a contingency test in which the same reward occurs also during the background without stimuli, even though stimulus, reward during stimulus and stimulus-reward pairing are unchanged (41) (**FIGURE 5*B***, *bottom*). Thus the responses do not just reflect stimulus-reward pairing but depend also on positive reward contingency, requiring more reward during stimulus presence than absence (however, amygdala neurons are insensitive to negative contingency). Although the stimulus is still paired with the reward, it loses its specific reward information and prediction; the amygdala responses reflect that lost prediction. Correspondingly, in an operant test, animals with amygdala lesions fail to take rewards without action into account and continue to respond after contingency degradation (399), thus showing an important role of amygdala in coding reward contingency.

Reward contingency affects also dopamine responses to conditioned stimuli. In an experiment designed primarily for blocking (see below **FIGURE 8*A***), a phasic environmental stimulus constitutes the background, and the crucial background test varies the reward with that environmental stimulus. Dopamine neurons are activated by a specific rewarded stimulus in the presence of an unrewarded environmental stimulus (positive contingency) but lose the activation in the presence of a rewarded environmental stimulus (zero contingency) (618). The stimulus has no specific information and prediction when the environmental stimulus already fully predicts the reward. Another experiment shows that these contingency-dependent dopamine responses may be important for learning. Using optogenetic, "rewarding" stimulation of dopamine neurons, behavioral responses (nose pokes) are only learned when the stimulation was contingent on the nose pokes, whereas contingency degradation by unpaired, pseudorandom stimulation during nose pokes and background induces extinction [truly random control with altered p(stimulation|nose poke)] (641).

Taken together, amygdala and dopamine neurons take the reward in the absence of the stimulus into account and thus are sensitive to contingency rather than simple stimulus-reward pairing. They fulfill the crucial requirement for coding effective reward information and prediction and thus may provide crucial inputs to neuronal decision mechanisms. Also, these neuronal signals demonstrate a physical basis for the theoretical concept of contingency and thus strengthen the plausibility of the most basic assumption of animal learning theory.

3. Prediction error learning theory

Contingency, in addition to stimulus-reward pairing (contiguity), is necessary for rewards to induce conditioning. Reward is contingent on a stimulus only when it differs between stimulus presence and absence. When reward is identical, including its time of occurrence, between stimulus presence and absence, it is already predicted by the objects that exist during stimulus presence and absence and therefore not surprising. Hence, the stimulus will not be learned.

The formal treatment of surprise in conditioning employs the concept of prediction error. A reward prediction error PE is the difference between received reward λ and reward prediction V in trial t

$$\mathrm{PE}(t) = \lambda(t) - \mathrm{V}(t) \qquad (1)$$

This formal definition of "error" extends beyond the colloquial meaning of inaccurate behavior. Prediction errors occur whenever *Equation 1* applies, irrespective of whether they simply occur during behavioral learning or are actually being used for learning.

Animal learning theory aims to explain the role of contingency in conditioning by formalizing how prediction errors update reward predictions from previously experienced rewards (460). The new prediction V for trial $t+1$ derives from the current prediction V(t) and the prediction error PE(t) weighted by learning rate α

$$\mathrm{V}(t+1) = \mathrm{V}(t) + \alpha * \mathrm{PE}(t) \qquad (2)$$

Note that V captures the sum of predictions if several stimuli are present. According to this account, conditioning constitutes error driven learning **(FIGURE 6*A*)**. When the outcome differs from the prediction, a prediction error proportional to that difference is generated and added to the current prediction after weighting by learning rate α. When α is set to 1.0, the new prediction is updated in one trial. However, learning rates are usually <1.0, which leads to the typical asymptotic learning curves during which prediction errors decline gradually **(FIGURE 6*B*)**.

Rewards that are better than predicted generate positive prediction errors and lead to response acquisition, whereas worse rewards generate negative prediction errors and lead to extinction of behavior. When the prediction error becomes zero, no further learning occurs and the prediction

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

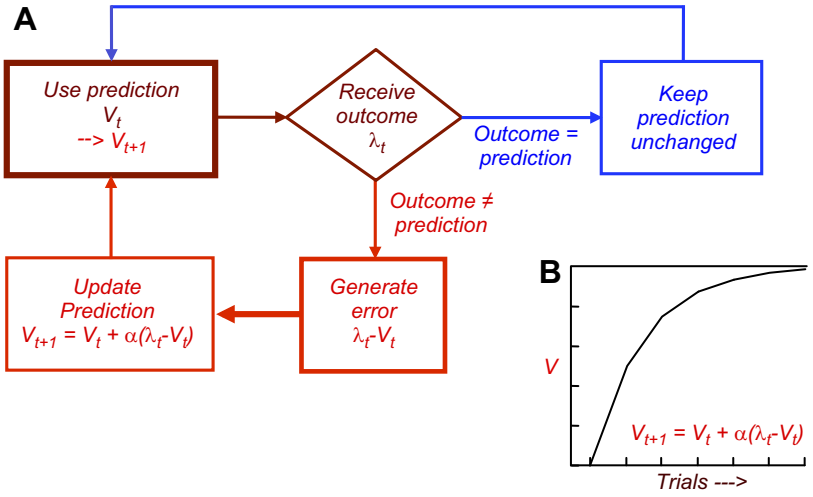


FIGURE 6. Learning with prediction errors. *A:* feedback circuit diagram for prediction updating by error. An error is generated when outcome (reward, punisher) differs from its prediction. In Pavlovian conditioning, a prediction error after an outcome change leads to prediction updating which leads to a behavioral change. In operant conditioning, a prediction error after an outcome change leads to a behavioral change which leads to prediction updating. In contrast, no prediction error is generated when the outcome matches the prediction, and behavior remains unchanged. V is reward prediction, λ is reward, α is learning rate, and t is trial. *B:* typical learning curve, generated by gradually declining prediction errors [$\lambda(t) - V(t)$].

remains stable. Thus learning advances only to the extent to which a reward is unpredicted and slows progressively with increasingly accurate predictions. This formalism views learning intuitively as a change in behavior that occurs when encountering something new or different than predicted, whereas behavior stays the same when everything occurs according to "plan."

Positive prediction errors direct organisms to maximal rewards, whereas negative prediction errors direct them away from poor outcomes. Obviously, individuals appreciate positive prediction errors and hate negative ones. Optimal behavior for maximizing rewards is characterized by many positive prediction errors and few negative ones.

Contingency is based on prediction errors. Contingency is non-zero, and prediction errors are elicited, when reward differs between stimulus presence and absence. This is the crucial requirement for learning. When reward is identical during stimulus presence and absence, there is no contingency, no prediction error, and no learning.

Specific learning situations benefit from extensions of the basic Rescorla-Wagner rule. Attentional learning rules relate the capacity to learn (associability) to the degree of attention evoked by the stimulus, which depends on degree of surprise (prediction error) experienced in the recent past (prediction error in the current trial cannot contribute to attention to the stimulus, as the error occurs only after the stimulus). Learning rate α is modified according to the weighted unsigned (absolute) prediction error in the recent past (425)

$$\alpha = \kappa \left| \lambda(t) - V(t) \right| \qquad (3)$$

with κ as weighting parameter. Thus learning rate increases with large prediction errors, as in the beginning of learning or with large step changes, resulting in strong influences of prediction error and thus fast learning according to *Equation 2*. In contrast, learning slows with small prediction errors from small changes.

When rewards are drawn from different probability distributions, comparable learning can be achieved by scaling the update component α*PE to the standard deviation σ of the respective reward distribution by either (438)

$$\alpha^{\circ} = \alpha / \sigma \qquad (3A)$$

or

$$PE(t)^{\circ} = PE(t) / \sigma \qquad (3B)$$

Behavioral tests with measurable prediction errors demonstrate that scaling learning rate α (*Equation 3A*) by standard deviation may produce constant learning rates with probability distributions differing only in standard deviation (384), which can be formulated as

$$\alpha = \kappa / \sigma \left| \lambda(t) - V(t) \right| \qquad (3C)$$

Thus risk may affect learning in two separate ways, by increasing learning rate α via the attentional effect of the prediction error (*Equation 3*), and by decreasing learning rate or prediction error via division by standard deviation σ (*Equations 3A* and *3B*).

The Rescorla-Wagner learning rule models Pavlovian conditioning: changes in outcome lead to prediction errors that lead to changes in prediction and consequently changes in behavior. The learning rule extends to operant conditioning: changes in outcome produce prediction errors that lead to behavioral changes which then result in changes in prediction.

In realistic situations, rewards may be unstable, which results in frequent prediction errors, continuous updating of predictions, and correspondingly varying behavior. Thus learning of a single, stable reward is a special case of, and

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

formally equivalent to, the general process of updating of reward predictions.

Although the Rescorla-Wagner rule models the acquisition of predictions (V in *Equation 2*), it does not conceptualize predictions as conditioned, higher order reinforcers. However, higher order reward properties are responsible for most reward functions in learning, approach behavior, and emotional reactions, as outlined above (640). The temporal difference (TD) reinforcement model in machine learning captures the learning functions of higher order rewards (574). Its prediction error includes both unconditioned and higher order reinforcers, without distinguishing between them, as

$$\mathrm{TDPE}(t) = [\lambda(t) + \gamma\sum \mathrm{V}(t)] - V(t-1) \quad (4)$$

The t stands now for time step within a given trial, γ is a temporal decay (discounting) factor. A temporal, rather than trial by trial, prediction error is adequate when including higher order reinforcers that derive from reward predictors and necessarily occur at earlier times than unconditioned reinforcers. The term $[\lambda(t) + \gamma\sum \mathrm{V}(t)]$ replaces the $\lambda(t)$ term in *Equation 1* and incorporates at every time step t the unconditioned reward $\lambda(t)$ and the discounted (γ) sum ($\sum$) of future rewards predicted by $\mathrm{V}(t)$ serving as higher order reward. The TDPE includes a higher order prediction error that indicates how much the reward predicted by a conditioned stimulus varies relative to an earlier reward prediction. Thus TDPEs are generated by successive conditioned stimuli to the extent that these stimuli carry different reward predictions than the preceding stimulus.

Substituting PE of *Equation 2* by TDPE of *Equation 4* leads to the TD learning model

$$\mathrm{V}(t+1) = \mathrm{V}(t) + \alpha * \mathrm{TDPE}(t) \quad (5)$$

Thus TD learning advances backwards from an unconditioned reward to the earliest reward predictor. Updating of the earliest reward prediction employs the prediction value of the subsequent stimulus and does not need to wait for the final, unconditioned reward. Thus each of the conditioned, reward-predicting stimuli serves as a higher order reward, which is biologically plausible by reflecting the frequent occurrence of conditioned rewards and addressing the often difficult distinction between the primary or conditioned nature of natural rewards (640). The use of truly unconditioned rewards in TD learning is straightforward (573). TD reinforcement models are efficient and learn complex tasks involving sequential states, like balancing a pole on a moving platform (32) or playing backgammon (586).

4. Dopamine reward prediction error signal

Most midbrain dopamine neurons show rather stereotyped, phasic activations with latencies of <100 ms and durations of <200 ms following unpredicted food or liquid rewards **(FIGURE 7, *A* AND *B*)**. The response codes the prediction error, namely, the quantitative difference between received and predicted reward value in each trial t, which can be generically be expressed as

$$\mathrm{DaResp}(t) \sim \lambda(t) - \mathrm{V}(t) \quad (6)$$

The neuronal response is bidirectional, as a reward that is better than predicted elicits an activation (positive prediction error response) and a reward that is worse than predicted induces a depression (negative error response) in monkeys, rats, and mice **(FIGURE 7*C*)**. A fully predicted reward draws no response. These responses occur with all-or-none reward prediction errors (321, 365, 410, 503, 517, 521, 524), and with graded, quantitative prediction errors (more or less reward than predicted) (33, 102, 144, 161, 376, 598). Thus the dopamine response codes a reward prediction error (*Equation 1*) that may constitute a biological instantiation of the reinforcement term in Rescorla-Wagner learning (365), TD learning (374), and spectral timing (71).

The negative dopamine reward prediction error response consists of a depression of activity against low background activity. Adequate assessment requires measuring impulse rate over the maximal length of the depression (34), thus avoiding an almost ungraded negative response (33). Even with a graded negative response (161), the negative error response has less dynamic range than the positive error response. However, the complete cessation of dopamine activity for varying durations may have stronger impact on postsynaptic neurons than graded increases and thus may compensate for the limited dynamic range. Even before it comes to synaptic transmission, voltammetrically measured dopamine concentration changes resulting from negative reward prediction errors are symmetric to the changes resulting from positive errors (204). Thus the dopamine reward prediction error signal seems to be bidirectional.

Dopamine responses show features beyond the Rescorla-Wagner learning rule. About one-third of dopamine neurons show also slower activations preceding reward that vary with the standard deviation (risk) of reward probability distributions (161). The risk response reflects the unsigned (absolute) prediction error and, in keeping with the attentional learning rules (425), may affect learning rate α (*Equation 3*) and thus serve to adjust learning speed. Furthermore, the prediction error response itself scales with the standard deviation of reward probability distributions (598) according to *Equation 3B*, which may be driven by the slower activation coding standard deviation and serve to achieve comparable learning with different standard deviations. Third, the prediction error response is sensitive to predictions in model-based learning (see below).

The dopamine response at the time of reward fulfils stringent tests for prediction errors conceptualized by animal

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

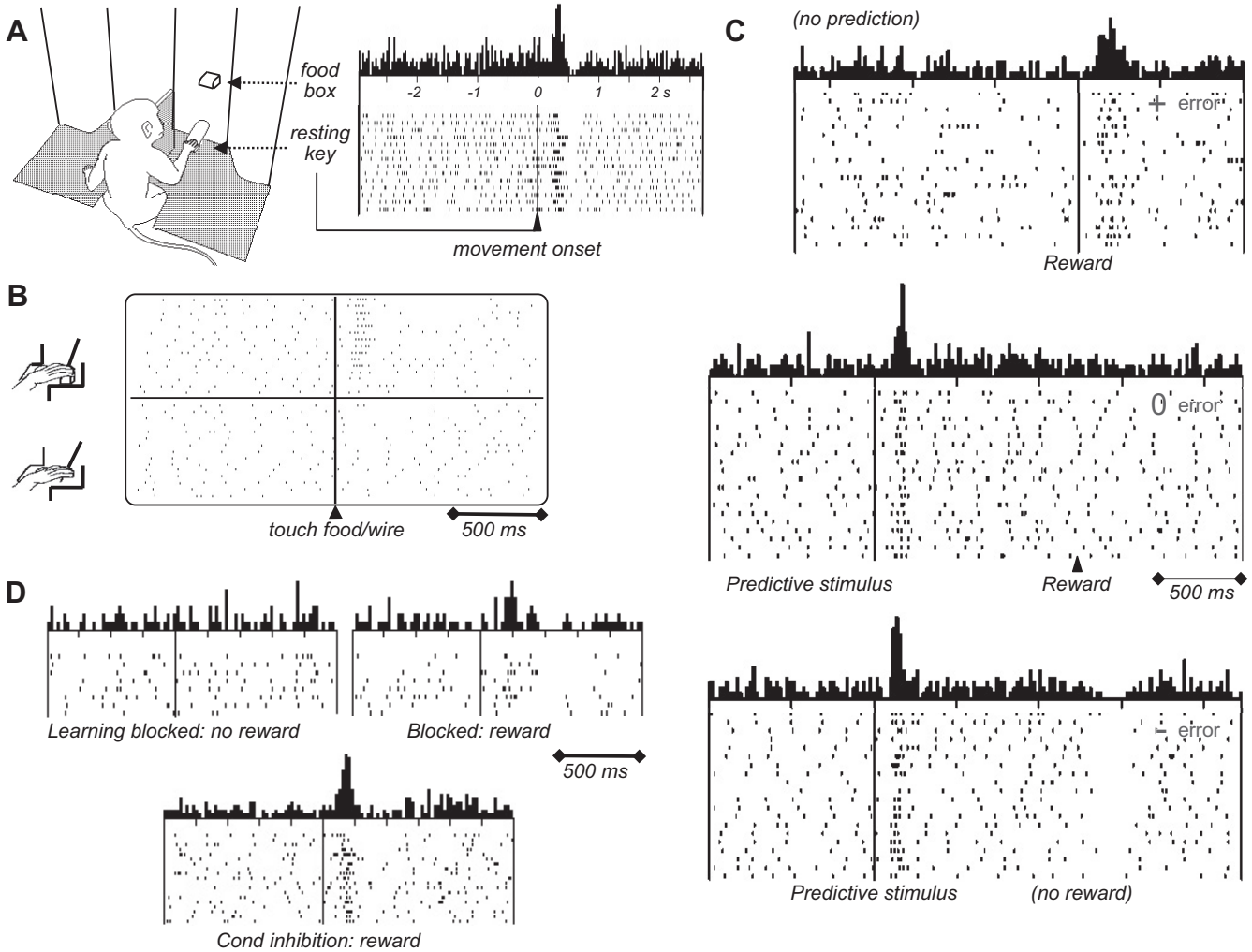


FIGURE 7. Dopamine prediction error responses at the time of reward in monkeys. *A:* dopamine responses to touch of food without any phasic stimuli predicting the reward. The food inside the box is invisible but is touched by the hand underneath the cover. Movement onset is defined as release of resting key. *B:* differential response to touch of a piece of apple (*top*) but not to touch of a bare wire (*bottom*) or known inedible objects. Left graphs show the animal's hand entering the covered food box. Inside the box, touch of a bare wire or a wire holding the food elicits an electric signal for temporal reference (vertical line at right). [*A* and *B* from Romo and Schultz (491), with kind permission from Springer Science and Business Media.] *C:* reward prediction error responses at time of reward (*right*) and reward-predicting visual stimuli (left in 2 bottom 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 (- errror, *bottom*). [From Schultz et al. (524).] *D:* reward prediction error responses at time of reward satisfy stringent prediction error tests. *Top*: blocking test: lack of response to reward absence following the stimulus that was blocked from learning (*left*), but activation by surprising reward after blocked stimulus (*right*). [From Waelti et al. (618).] *Bottom*: conditioned inhibition test. Supranormal activation to reward following an inhibitory stimulus explicitly predicting no reward. [From Tobler et al. (597).]

learning theory **(FIGURE 7*D*)**. In the blocking test, a test stimulus is blocked from acquiring reward prediction when the reward is fully predicted by another stimulus (zero contingency). If reward does occur after the nonpredictive test stimulus, it will produce a positive prediction error, and dopamine neurons are activated by that reward (618). In the conditioned inhibition test, a reward occurring after a stimulus explicitly predicting reward absence elicits a super strong positive reward prediction error (because of a negative prediction being subtracted from the reward, *Equation 1*), and dopamine neurons accordingly show supranormal activation by the surprising reward (*Equation 6*) (598). Thus the phasic dopamine responses follow the formal theoretical requirements for prediction error coding.

Reward-predicting stimuli induce phasic activations in most dopamine neurons (60-75%) **(FIGURE 7*C*)** (322, 503) and increase correlations between dopamine impulses (256). These responses are gradually acquired through

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

learning and thus code the prediction V of Rescorla-Wagner and TD models. Correspondingly, voltammetrically measured striatal dopamine release shifts from reward to reward-predicting stimuli during conditioning (117). Acquisition of dopamine responses to conditioned stimuli requires prediction errors, as the blocking test shows. Compounding a novel stimulus with an established reward predictor without changing the reward fails to elicit a prediction error and prevents the stimulus from being learned **(FIGURE 8*A*)**. A conditioned inhibitor elicits a depressant instead of an activating response **(FIGURE 8*B*)**.

Results from tests with sequential reward-predicting stimuli suggest that dopamine responses signal higher order reward prediction errors compatible with TD learning (374). Thus the dopamine responses are well described by extending *Equation 6* and using t as time steps, which can be generically be expressed as

$$\mathrm{DaResp}(t) \sim [\lambda(t) + \gamma\sum \mathrm{V}(t)] - \mathrm{V}(t-1) \quad (7)$$

The dopamine response jumps sequentially backwards from the "unconditioned" liquid reward via stimuli predicting this reward to the earliest reward-predicting stimulus, while losing the response to the predicted intermediate stimuli and the predicted reward **(FIGURE 8*C*)** (453, 521). Similar to TD learning, the dopamine responses do not distinguish between primary and higher order reward prediction errors. Only their magnitudes, and fractions of responding neurons, decrease due to temporal discounting, which in TD learning is incorporated as γ. With several consecutive conditioned stimuli, dopamine neurons code the TD prediction error exactly as the predicted value changes between the stimuli (398). In such tasks, dopamine responses match closely the temporal profiles of TD errors that increase and decline with sequential variations in the discounted sum of future rewards **(FIGURE 8*D*)** (144). Thus dopamine neurons code reward value at the time of conditioned and unconditioned stimuli relative to the reward prediction at that moment.

The reward-predicting instruction stimuli in delay tasks have little incentive properties and thus provide a distinction between predictive and incentive stimulus properties (see above). These instructions contain information and elicit saccadic eye movements for acquiring the information contained in the stimulus, but the animal is not allowed to react immediately with a movement towards a rewarded target; it must wait for a later, incentive, movement-triggering stimulus. The consistent dopamine responses to these predictive instructions **(FIGURE 8*C*)** (301, 349, 453, 521, 527, 579), and their lack of ocular relationships (527), suggest predominant coding of predictive stimulus properties, without requiring incentive properties. In contrast, in specific sign- and goal-tracking rat strains, striatal voltammetric dopamine responses are stronger to stimuli with incentive compared with reward-predictive properties (163). These differences may be due to different methods (electrophysiological impulses vs. voltammetric dopamine concentrations) and genetic differences in the specifically bred sign versus goal trackers differentially affecting the dopamine response transfer from reward to predictive stimulus.

Hundreds of human neuroimaging studies demonstrate activations induced by reward related stimuli and reward-related actions in the main reward structures (280, 590). A BOLD neuromagnetic (fMRI) signal reflecting reward prediction error is found in the ventral striatum (351, 390) and constitutes probably the most solid reward response in the brain. It likely reflects the dopamine prediction error signal, as it increases with dopamine agonists and decreases with dopamine antagonists (431). It is detected more easily in striatal and frontal dopamine terminal areas than in midbrain cell body regions (111, 431, 615), because it presumably reflects summed synaptic potentials (328), which may be stronger after the massive axon collaterilization in the striatum. The human reward signal allows also to assess neuronal correlates of positive emotional reward functions, which is intrinsically difficult in animals despite recent efforts (44). With the use of raclopride binding with positron emission tomography (PET), amphetamine-induced dopamine concentrations correlate highly with subjective euphoria ratings in ventral (but not dorsal) striatum (140). BOLD responses in ventral striatum correlate with pleasantness ratings for thermal stimuli (along with orbitofrontal and anterior cingulate activations) (486) and with momentary happiness ratings derived from a combination of certain rewards, gambles, and reward prediction errors (along with anterior cingulate and anterior insula activations) (498). These responses in the dopamine receiving striatum suggest a role of dopamine in signaling pleasure and happiness.

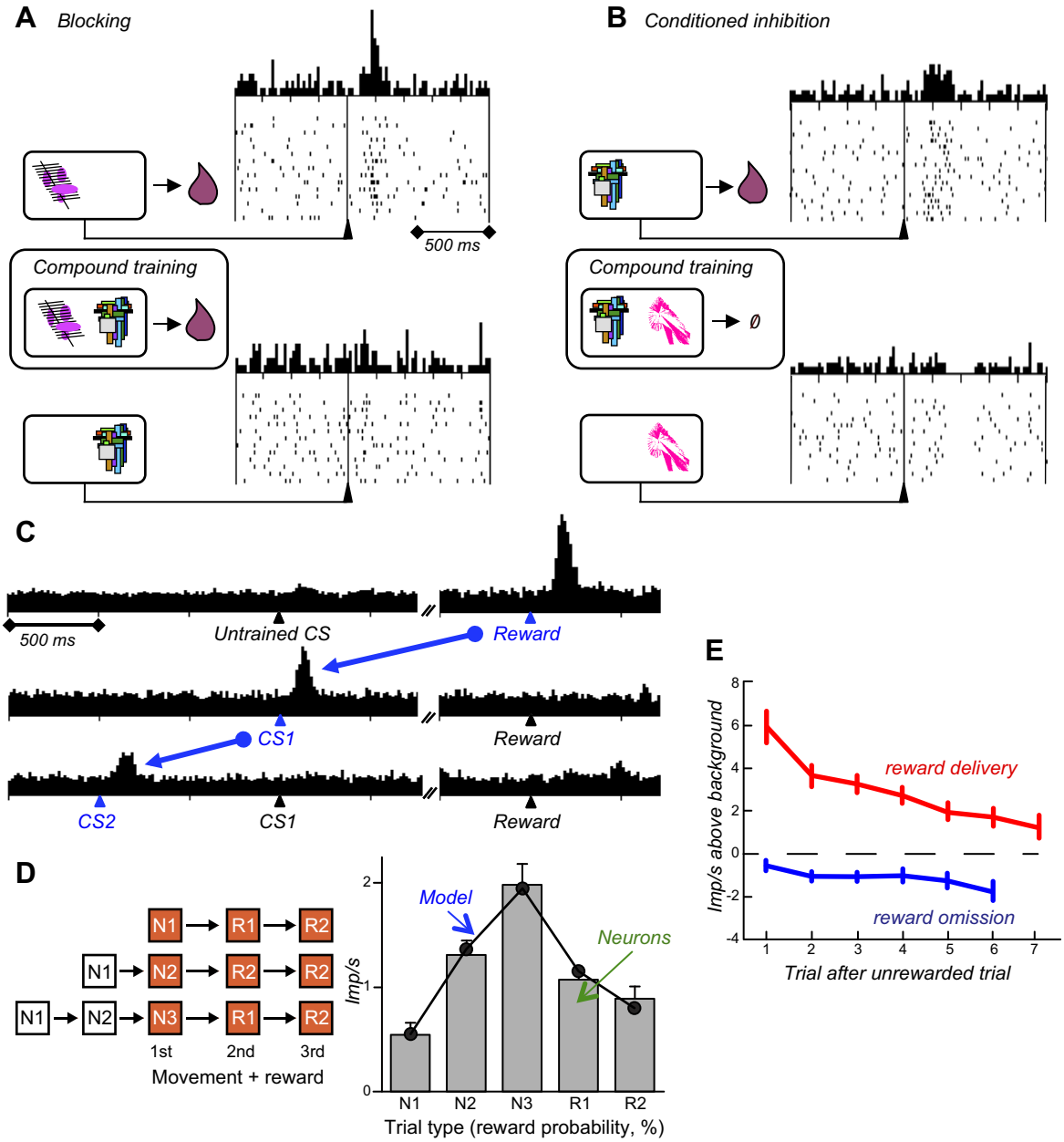
5. Dopamine error signal incorporates model-based predictions

Standard conditioning occurs through prediction errors derived from actually experienced outcomes. However, individuals also learn about contexts, situations, rules, sequences, and conditional probabilities, which may be summarily called models of the world. Knowledge derived from such models can affect the prediction in the neuronal prediction error computation, result in a more appropriate teaching signal, and thus tremendously improve learning.

Model-based learning involves two separate processes, the acquisition and updating of the model, and the influence of the model on reinforcement learning, which likely occurs by influencing predictions of the outcome (139, 575). The acquisition and updating of some of the models likely involves cortical rather than dopamine signals. In contrast, dopamine responses seem to incorporate the predictive information from models once they are established.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

In one such experiment, sequences of unrewarded trials lead to higher conditional reward probability with increasing trials (increasing hazard function). The underlying temporal structure constitutes a model in which reward prediction increases progressively. Due to increasing reward prediction, later reward delivery induces increasingly weaker positive prediction errors, and reward omission elicits stronger negative prediction errors. In contrast, model-free reinforcement learning would only consider the past unrewarded trials and hence generate progressively decreasing reward prediction and an opposite pattern of errors. In this experiment, dopamine prediction error responses decrease with increasing numbers of trials since the last reward and thus reflect the increasing reward prediction **(FIGURE 8*E*)** (382). In another experiment, dopamine reward prediction error responses adapt to previously learned reward probability distributions (598). The responses scale to expected value and standard deviation and show similar magnitudes with 10-fold differences in reward magnitude. In another example of model-based learning, acquisition of a reversal set allows animals to infer reward value of one stimulus after a single reversed trial with the other stimulus. Accordingly, dopamine prediction error responses reflect the inferred reversed reward prediction (68). In each of these examples, the observed dopamine prediction error responses are not fully explained by the immediate experience with reward but incorporate predictions from the task structure (model of the world). Thus dopamine neurons process prediction errors with both model free and model-based reinforcement learning.



Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

6. Dopamine response in cognitive tasks

Bidirectional dopamine prediction error responses occur also in tasks with elaborate cognitive components. Dopamine neurons, and their potential input neurons in lateral habenula, show bidirectional prediction error responses that signal values derived from differential discounting with conditioned reinforcers (66, 67), although concurrent risk influences on subjective value need to be ruled out. More complex tasks require distinction of many more possible confounds. Early studies used spatial delayed, spatial alternation, and delayed matching-to-sample tasks typical for investigating neurons in prefrontal cortex and striatum and report positive and negative prediction error responses to rewards and reward-predicting stimuli compatible with TD learning models (321, 521, 579). Similar dopamine error responses occur in tasks employing sequential movements (503), random dot motion (389), and somatosensory detection (122). Advanced data analyses from these complex tasks reveal inclusion of model-based temporal predictions into the dopamine error response (382) and compatibility with TD models (144). Variations in stimulus coherence (389) and visual search performance (349) result in graded reward probabilities that induce reward prediction errors that are coded by dopamine neurons, although the error coding may not be easily apparent (426). Thus, in all these high-order cognitive tasks, the phasic dopamine response tracks only, and reliably, the reward prediction error.

7. Dopamine event detection response

The dopamine reward prediction error response is preceded by a brief activation that detects any sufficiently strong event before having identified its reward value (**FIGURE 9, *A–C***, blue areas). The early response component occurs with latencies of <100 ms with all sensory modalities, including loud noises, light flashes, and rapidly moving visual stimuli (229, 322, 527, 563) and evokes dopamine release (117). The activation reflects the sensory impact and physical salience of rewards, punishers, and their predictive stimuli (157, 159, 160). It does not fully reflect motivational salience, as it is absent with negative motivating events of low sensory impact (160, 366), negative reward prediction errors (321, 517, 521, 524, 597, 618), and conditioned inhibitors (597). The detection response covaries inversely with the hazard rate (of temporal event occurrence) and thus codes surprise salience (286) or a temporal prediction error (389) as conceptualized in TD models (574). Thus the early dopamine response component reflects the detection of an event before having identified its value.

This initial response component develops into the graded reward prediction error response, which ranges from depression with strong negative prediction errors to strong activation with positive prediction errors (158, 285, 301, 376, 597). The transition from the first to the second response component depends on the time necessary for assessing the value of the stimulus and thus may vary considerably between behavioral tasks. With easy identification, the transfer takes ~50 ms and fuses the two components, for example, when fixed stimulus positions allow rapid stimulus identification without eye saccades (**FIGURE 9, *A* AND *B***). Then the two components are fused and may appear as a single response whose identity has long been unclear and may occasionally be labeled as attentional (355, 356). A possible intermediate sensory identification step is either not discernible or is not engaging dopamine neurons. In contrast, the components separate into two separate responses with slower stimulus identification, as seen with random dot motion stimuli requiring >200 ms behavioral discrimination time (**FIGURE 9*C***) (389). Thus the early dopamine activation detects events rapidly before identifying them properly, whereas the subsequent bidirectional response component values the event and codes a proper reward prediction error.

The accurate value coding by the second response component is reflected in the error response at the time of the reward (398, 618). Despite the initial dopamine activation by an unrewarded stimulus, a subsequent reward is registered as a surprise eliciting a positive prediction error re-

FIGURE 8. Dopamine responses to conditioned stimuli in monkeys. *A*: stimulus responses of a single dopamine neuron during a blocking test. A pretrained stimulus predicts liquid reward and induces a standard dopamine response (*top*). During compound training, a test stimulus is shown together with the pretrained stimulus while keeping the reward unchanged (*middle left*). Thus the reward is fully predicted by the pretrained stimulus, no prediction error occurs, and the test stimulus is blocked from learning a reward prediction. Correspondingly, the test stimulus alone fails to induce a dopamine response (*bottom*). [From Waelti et al. (618).] *B:* stimulus responses of a single dopamine neuron during a conditioned inhibition test. The reward normally occurring with the pretrained stimulus (*top*) fails to occur during compound training with a test stimulus (*middle left*). This procedure makes the test stimulus a predictor of no reward which correspondingly induces a dopamine depression (*bottom*). [From Tobler et al. (597).] *C:* stepwise transfer of dopamine response from reward to first reward-predicting stimulus, corresponding to higher order conditioning conceptualized by temporal difference (TD) model. CS2: instruction, CS1: movement releasing (trigger) stimulus in delayed response task. [From Schultz et al. (521).] *D:* reward prediction error responses closely parallel prediction errors of formal TD model. *Left*: sequential movement-reward task. One or two stimuli (white N1, N2) precede the movements leading to reward (orange, 1st, 2nd, 3rd). *Right*: averaged population responses of 26 dopamine neurons at each sequence step (gray bars, numbers indicate reward probabilities in %) and time course of modeled prediction errors $\{[\lambda(t) + \gamma\sum V(t)] - V(t-1)\}$ (black line). [From Enomoto et al. (144).] *E:* dopamine prediction error responses reflect model-based reward prediction derived from temporal task structure. The model specifies an increase in conditional reward probability P(reward | no reward yet) from initial $P = 0.0625$ to $P = 1.0$ after six unrewarded trials. Correspondingly, positive prediction errors with reward occurrence decrease across successive trials, and negative errors with reward omission increase. Averaged population responses of 32 dopamine neurons show similar temporal profiles (blue and red, as impulses/s above neuronal background activity). [From Nakahara et al. (382), with permission from Elsevier.]

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

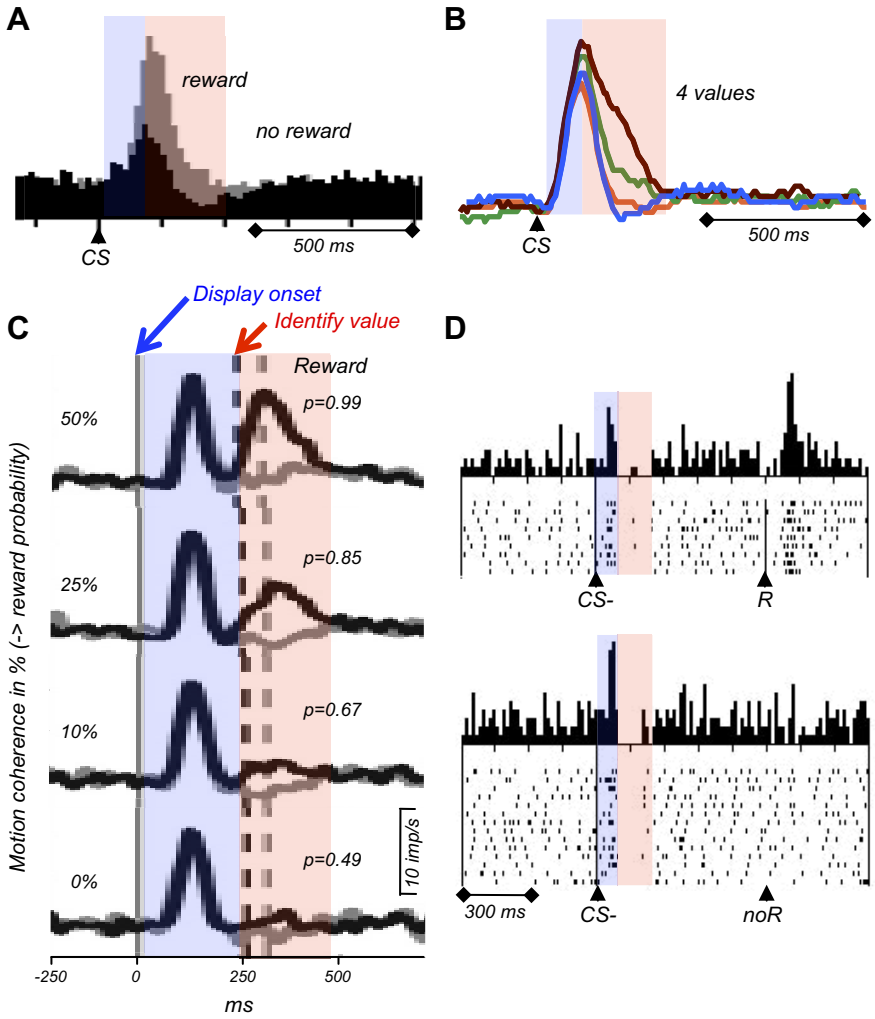


FIGURE 9. Two components of phasic dopamine responses. *A:* averaged population responses of 69 monkey dopamine neurons to conditioned stimuli (CS) predicting reward (gray) and no reward (black). Note the initial indiscriminate detection response component (blue) and the subsequent reward response component distinguishing between reward and no reward prediction (red). [From Tobler et al. (597).] *B:* averaged population responses of 54 monkey dopamine neurons to conditioned stimuli (CS) predicting rewards at different delays (2, 4, 8, and 16 s; brown, green, orange, and blue, respectively). The value reduction due to temporal discounting affects only the second, reward prediction error component (red). [From Kobayashi and Schultz (285).] *C:* differentiation of dopamine response into initial detection response and subsequent prediction error response. Increasing motion coherence (from 0 to 50%) improves binary dot motion discrimination and translates into increasing reward probability (from $P = 0.49$ to $P = 0.99$). The first response component is nondifferentially constant (blue), whereas the second component grows with increasing reward value (derived from probability, *bottom* to *top*, red). [From Nomoto et al. (389).] *D:* accurate value coding at time of reward despite initial indiscriminate stimulus detection response. After the unrewarded conditioned stimulus (CS-), surprising reward (R) elicits a positive prediction error response (*top*), whereas predicted reward absence (noR) fails to elicit a negative error response (*bottom*). [From Waelti et al. (618).]

sponse (**FIGURE 9*D***, *top*); accordingly, reward absence fails to induce a negative prediction error response (**FIGURE 9*D***, *bottom*). Thus, from the second response component on, and before a behavioral reaction can occur, dopamine responses accurately reflect the unrewarded stimulus nature. This constitutes the information postsynaptic neurons are likely to receive, which may explain why animals do not show generalized behavioral responses despite initial neuronal response generalization.

The initial dopamine detection response component varies with four factors. First, it increases with the sensory impact of any physical event, which confers physical salience **(FIGURE 10*A*)** (160). This is most likely what the early studies on dopamine salience responses observed (229, 322, 563). Second, also involving memory processing stages, the dopamine detection component is enhanced by stimulus novelty (conferring novelty salience) (322) or surprise (conferring surprise salience) (255, 286, 345). Although novelty may constitute outright reward (458), dopamine neurons are not activated by novelty per se but require stimuli with sufficient sensory impact. The neurons do not respond to small novel stimuli, even though they show prediction error responses to the same small stimuli after learning (597, 618). The total dopamine response decreases with stimulus repetition, together with the animal's ocular orienting responses, and increases again with reward prediction learning **(FIGURE 10*B*)** (517). Third, the dopamine detection component increases with generalization to rewarded stimuli, even when animals discriminate behaviorally well between them. The neuronal generalization is analogous to generalization in behavioral conditioning when an unconditioned stimulus resembles closely a conditioned stimulus and the "associative strength" spills over (335). The incidence of activations to unrewarded, including aversive, stimuli increases with closer similarity and same sensory modality as rewarded stimuli (527, 597, 618). A change from auditory to visual modality of the rewarded stimulus, while leaving the visual aversive stimulus unchanged, increases the initial activations from <15 to 65% of dopamine neurons **(FIGURE 10*C*)** (366). Generalization likely also induces motivational salience via similar positively valued stimuli. Generalization may explain the more frequent dopamine activations to conditioned compared with primary aversive stimuli (345), which defies basic learning concepts. The initial response component evoked by reward

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

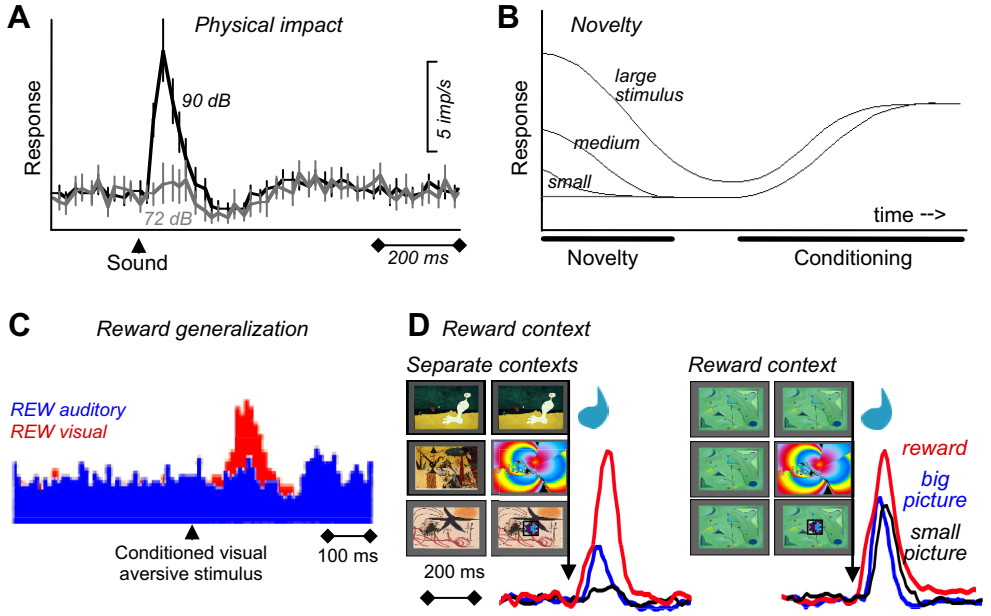


FIGURE 10. Four factors influencing the detection component of phasic dopamine responses. *A:* detection response generated by sensory impact, conferring physical salience. Louder, nonaversive sounds with higher physical salience generate stronger activations (72 and 90 dB, respectively; behavioral choice preferences demonstrated their nonaversive nature). Averaged population responses measured as impulses/s (imp/s) of 14 and 31 monkey dopamine neurons, respectively. [From Fiorillo et al. (160).] *B:* detection response, and possibly also second response component, enhanced by stimulus novelty, conferring novelty or surprise salience. Stimulus novelty itself is not sufficient to induce dopamine activations, as shown by response absence with small stimuli (horizontal line), but enhances detection response when stimuli are physically larger and more salient (vertical axis). Neuronal responses wane with stimulus repetition due to loss of novelty and increase again with conditioning to reward (from *left* to *right*). [Composite scheme from Schultz (517), derived from original data (221, 597, 618).] *C:* detection response enhanced by generalization to rewarded stimuli. Blue: minor population response to conditioned visual aversive stimulus alternating with auditory reward-predicting stimulus (REW auditory) (active avoidance task). Red: substantial activation to identical visual aversive stimulus when the alternate reward-predicting stimulus is also visual (REW visual), a situation more prone to stimulus generalization. As control, both auditory and visual reward-predicting stimuli induce typical dopamine activations (not shown). [From Mirenowicz and Schultz (366).] *D:* detection response enhanced by reward context. *Left* (separate contexts): minor dopamine population activations induced by unrewarded big and small pictures when non-reward context is well separated from reward context by testing in separate trial blocks, using distinct background pictures and removing liquid spout in picture trials. *Right* (common reward context): major activations by same unrewarded pictures without separation between non-reward and reward context. [From Kobayashi and Schultz (286).]

generalization induces dopamine release in nucleus accumbens (117). Fourth, the dopamine detection component increases in rewarding contexts (286), via pseudoconditioning or higher order context conditioning (367, 536, 537), without explicit reward pairing. Accordingly, reducing the separation between unrewarded and rewarded trials, and thus infecting unrewarded trials with reward context, boosts the incidence of dopamine activations to unrewarded stimuli (from 3 to 55% of neurons) **(FIGURE 10*D*)** (286), as does an increase of reward from 25 to 75% of trials (from 1 to 44% of neurons) (597, 618). Thus the initial response component reflects the detection of environmental events related to reward in the widest possible way, which conforms well with the reward function of the phasic dopamine signal.

Other neuronal systems show similar temporal evolution of stimulus responses. During visual search, frontal eye field neurons respond initially indiscriminately to target and distractor and distinguish between them only after 50–80 ms (see below, **FIGURE 38*B***) (588). Neurons in primary visual cortex V1 take 100 ms after the initial visual response to show selective orientation tuning (468) and spatial frequency selectivity (61). Over a time course of 140 ms, V1 neurons initially detect visual motion, then segregate the figure created by dot motion from background motion, and finally distinguish the relevant figure from a distractor (474). V1 and frontal eye field neurons respond initially indiscriminately to targets and distractors and distinguish between them only after 40–130 ms (434). During perceptual decision making, LIP and prefrontal neurons respond initially indiscriminately to stimulus motion and distinguish between motion strength only at 120–200 ms after stimulus onset (see below, **FIGURE 38*A***) (274, 483, 535). Pulvinar neurons show initial, nondifferential detection responses and subsequent differential stimulus ambiguity responses

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

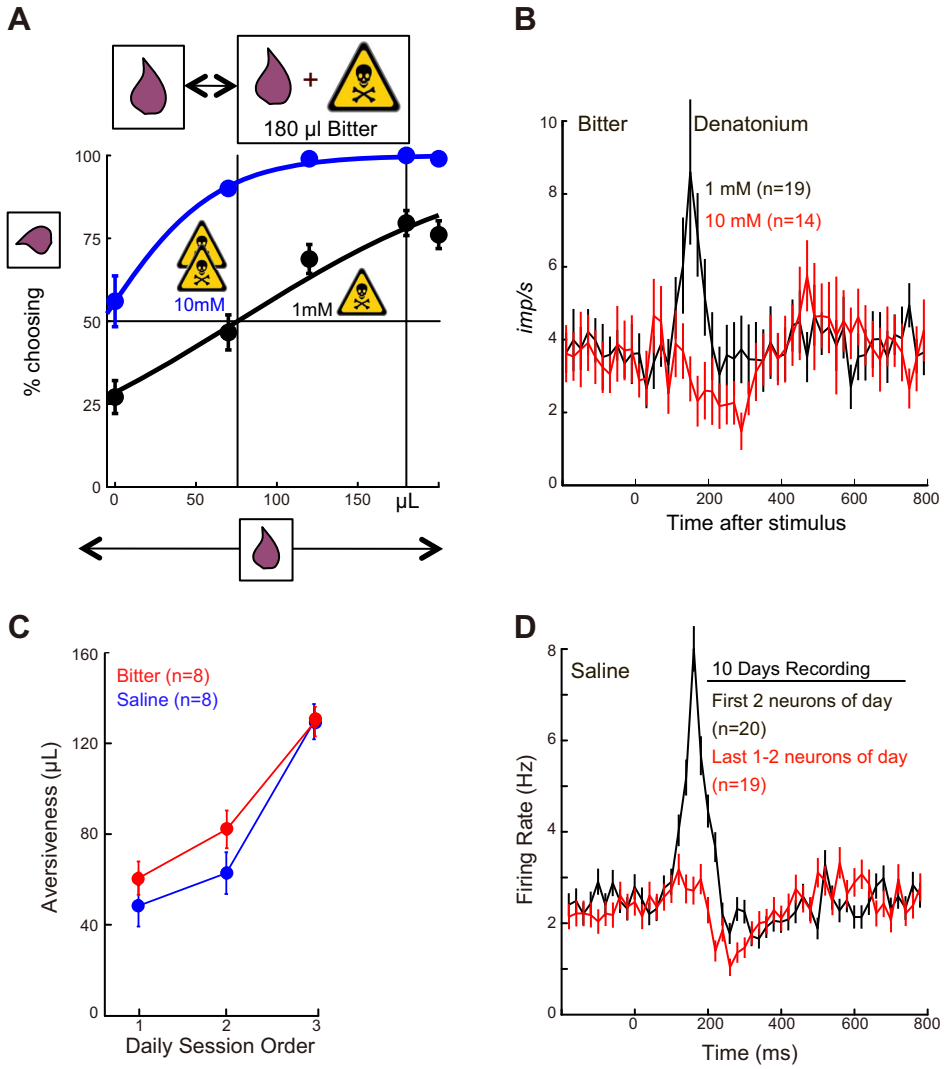


FIGURE 11. No aversive coding in monkey dopamine activations. *A:* psychophysical assessment of aversiveness of bitter solution (denatonium) as prerequisite for investigating quantitative neuronal processing of aversiveness. Monkey chose between single juice reward and juice reward + denatonium (*top*). The blue and black psychophysics curves show that increasing the volume of single juice reward induces more frequent choices of this reward against constant alternative juice + denatonium. Aversiveness of denatonium is expressed as difference between volume of single juice and juice together with denatonium at choice indifference (50% choice). Thus 1 mM denatonium is worth −100 µl of juice (black), and 10 mM denatonium is worth −180 µl (blue). The *x*-axis shows volume of single juice, and *y*-axis shows percent of choices of single juice. *B:* inverse relationship of neuronal activations to psychophysically quantified aversiveness of bitter solutions (behavioral method shown in *A*). *N* = number of neurons. Imp/s indicate firing rate. *C:* development of behavioral aversiveness of solutions within individual test days, as assessed with method shown in *A*. Liquid solutions of salt or bitter molecules lose reward value through gradual satiation and thus become increasingly aversive (thirsty monkeys work for mildly "aversive" saline solutions, and thus find them rewarding, as outcome value is dominated by liquid over salt). *D:* decreasing dopamine activations with increasing aversiveness of saline solution within individual test days (behavioral aversiveness assessment shown in *C*), suggesting subtraction of negative aversive value from gradually declining juice value due to satiation. [*A–D* are from Fiorillo et al. (160).]

during visual categorization (288). Tactile responses in mouse barrel cortex require 200 ms to become differential (300). Similar multicomponent responses are seen with reward coding. Amygdala neurons initially detect a visual stimulus and may code its identity and then transition within 60–300 ms to differential reward value coding (9, 422, 428). V1 and inferotemporal cortex responses show initial visual stimulus selectivity and only 50–90 ms later distinguish reward values (371, 557). Thus there is a sequence in the processing of external events that advances from initial detection via identification to valuation. Whereas sensory processing involves only the first two steps, reward processing requires in addition the third step.

8. No aversive dopamine activation

Aversive stimuli are well known for >30 years to activate 10–50% of dopamine neurons in awake and anesthetized monkeys, rats, and mice (102, 193, 255, 345, 362, 366, 525, 604) and to enhance dopamine concentration in nucleus accumbens (75). Five neurons close to juxtacellularly labeled ventral tegmental area (VTA) dopamine neurons were activated by footshocks (64), similar to the effects of airpuffs and footshocks on non-dopamine neurons in the area (102, 582). The activations may have derived from dopamine neurons or not; the 30–45% of dopamine neurons lacking spontaneous activity (76, 166, 176, 211) may include nonactivated neurons and thus go undetected while taking up the label. In monkeys, some putative dopamine neurons, often located above substantia nigra, are activated by air puffs (345), although they respond more frequently to conditioned stimuli than primary rewards and thus defy the higher aversiveness of primary than conditioned punishers. However, just as rewards, punishers have distinct sensory, attentional, and motivating components, and none of these studies has distinguished between them. A recent psychophysical study that made this distinction by varying these components independently reports that dopamine activations by aversive stimuli reflect the physical, sensory stimulus impact rather than their aversiveness **(FIGURE 11)** (160). Once the sensory impact is controlled for, the aversive nature of stimuli either fails entirely to affect dopamine

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

neurons (157), reduces the total reward value of liquid solutions and their evoked dopamine activations (157, 160), and induces outright depressant dopamine responses (73, 102, 345, 362, 366, 608). In awake animals, the aversive nature of stimuli does not generate dopamine activations, but induces negative reward prediction errors either because of the negative value or by not being rewarding, and correspondingly induces depressant dopamine responses (157). Thus the different fractions of "aversively" activated dopamine neurons seen in earlier studies likely reflect the different intensities of physical stimuli employed (159). The graded activations with punisher probability (345) might reflect salience differences of conditioned stimuli (286), or graded anticipatory relief to punisher termination (75) that is considered as rewarding (552). Salience, rather than negative reinforcer value, might also explain the surprising higher incidence of activations to conditioned stimuli compared with primary punishment (345). The sensory rather than negative value nature of dopamine responses to aversive stimuli would also explain the dopamine activations by aversive stimuli under anesthesia (64, 525, 604) where motivational impact is ruled out.

The "aversive" dopamine activation constitutes a short latency response that is shorter than the reward response, is substantially boosted by generalization from rewarded stimuli, and is often curtailed by a subsequent depression (366). With these characteristics, the aversive dopamine activations seem to constitute the initial dopamine detection response that reflects mostly physical salience. In agreement with this interpretation, the aversive dopamine neurons are also driven by rewards and unrewarded stimuli and thus are highly sensitive to a large range of stimuli (366) compatible with coding physical impact rather than motivational value. Furthermore, dopamine neurons are neither depressed nor activated by negative aversive prediction errors (157, 255, 345), and show only enhanced activations with positive aversive prediction errors possibly reflecting surprise salience. Thus the current evidence suggests that aversive stimuli activate some dopamine neurons through physical impact rather than through their negative, aversive motivational value, although some dopamine neurons with truly aversive activations can never be completely ruled out.

9. Comprehensive account of phasic dopamine response

Of the two phasic dopamine response components **(FIGURE 12)**, the initial detection activation occurs before the neurons have identified the current reward value of the stimulus and arises even with motivationally neutral events and punishers. It increases with sensory impact, novelty, reward generalization, and rewarded contexts. Thus the initial component detects potential rewards and provides an early report before having identified the value of the event properly, which relates closely to the reward coding of dopamine neurons. The second response component arises once the

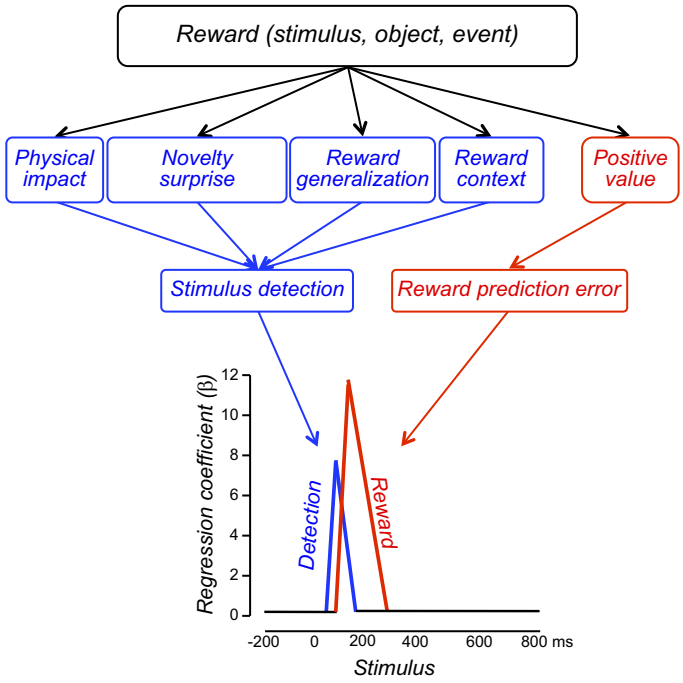


FIGURE 12. Reward components inducing the two phasic dopamine response components. The initial component (blue) detects the event before having identified its value. It increases with sensory impact (physical salience), novelty (novelty/surprise salience), generalization to rewarded stimuli, and reward context. This component is coded as temporal event prediction error (389). The second component (red) codes reward value (as reward prediction error).

neurons have well identified the reward value of the stimulus and codes only reward. As longest and strongest component it constitutes the major phasic dopamine influence on postsynaptic processing. The two components are integral parts of the dopamine reward prediction error response, which seems to be split into an early temporal error component and a subsequent value error component (also time sensitive), thus providing an interesting biological instantiation of a TD reinforcement teaching signal (574).

The phasic dopamine reward signal is homogeneous in latency, duration, and sensitivity across different dopamine neurons and very similar between different rewards, thus contrasting sharply with the widely varying reward processing in other brain structures. Graded variations in dopamine responses derive from the initial detection response component that varies in a continuous manner across the mediolateral or dorsoventral extent of the ventral midbrain. If this graded initial component occurs with aversive stimuli and is interpreted as aversive rather than detection response, some dopamine neurons may seem categorically different from those showing a smaller or no detection component (64, 255), suggesting neuronal response heterogeneity (345). However, these variations, as well as variations in the subsequent reward prediction error response (159), conform to a single-peak probability distribution rather than amounting to categorical differences and distinct dopamine subpopulations. In addition to their continuous phasic response distribution, dopamine neurons are het-

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

erogeneous in most other anatomical, physiological, and neurochemical respects.

The initial activation component may reflect physical, motivational, and surprise salience and thus may enhance the impact of the prediction error response in reward learning compatible with associability learning rules (425, 336). In being highly sensitive to novelty, reward generalization, and rewarded contexts, this component may reflect an early assumption about the nature of an event that is novel, resembles a reward, or occurs within a generally rewarding context. Such events have a chance to be rewards and thus are potential rewards. In responding to such stimuli, dopamine neurons detect a maximal range of potentially rewarding events and objects. Through the very short latency of these responses, dopamine neurons detect these stimuli very rapidly, even before having identified their value, and can rapidly initiate neuronal processes for approach behavior. Once the stimulus has been identified and turns out not to be a reward, it is still time to alter neuronal processing and modify or halt behavioral reactions. If the stimulus is indeed a reward, precious time may be gained and the reward approached before anybody else without such a rapid detection system arrives. By overreacting and processing also potential rewards, the mechanism would prevent premature asymptotes in reward detection and minimize reward misses (avoiding "I do not move unless the reward is for sure"). Such a mechanism would be particularly important when sparse resources present challenges for survival. With early detection facilitating rapid behavioral reactions, the two-component nature of the phasic dopamine response results in better reward acquisition and thus provides competitive evolutionary advantages.

The phasic dopamine reward signal is only one manifestation of dopamine function in the brain. In addition to the distinct and slightly less phasic dopamine risk signal (see below), dopamine exerts a tonic enabling effect on postsynaptic neurons, whose function is considerably more heterogeneous than the rather stereotyped dopamine prediction error response, even when considering its initial detection component (518). Many of these tonic functions depend on the function of the brain structure influenced by tonic extracellular dopamine concentrations, as inferred from behavioral and neuronal alterations arising from inactivations, receptor agonist or antagonist applications, and polymorphisms. These functions extend from motor processes in the striatum evidenced by Parkinsonian movement deficits to prefrontal processes of working memory, attention, reversal learning, categorization, and response control (470). An active involvement of phasic dopamine signals in these vastly different functions is questionable, in particular for functions that are recovered by dopamine receptor agonists without reinstating phasic dopamine signaling. Parkinsonism does not result from deficient dopamine reward prediction error signaling. Thus dopamine function in the brain is vast and heterogeneous and extends well beyond the phasic dopamine reward signal. Serotonin has a similar tonic enabling function, which includes aversive and reward processes (100, 201, 481, 534, 635). The question "what is dopamine doing" reflects the initial idea of "one neurotransmitter-one function" that may originate from "one brain structure-one function." In view of the many different neuronal dopamine processes, a better question might be "what is the function of the specific dopamine process we are looking at." For the phasic dopamine signal, the answer seems to be bidirectional reward prediction error coding, with additional, slightly slower risk coding.

10. Dopamine responses during reward learning

During learning episodes, dopamine responses show gradual changes that reflect the evolution of prediction errors. Dopamine neurons acquire responses to reward-predicting stimuli (322) that transfer back to the first reward-predicting stimulus in longer task schedules (144, 521). Their activations disappear gradually with extinction (598), possibly due to inhibition from neighboring pars reticulata neurons activated during extinction (408). When learning several tasks in sequence, positive reward prediction errors are generated every time a reward occurs during initial trials but subside once rewards become predictable and performance approaches asymptote. Dopamine neurons show corresponding prediction-dependent reward responses during these periods **(FIGURE 13, *A* AND *B*)** (322, 521). A temporal difference model using dopamine-like prediction errors replicates well behavioral learning and performance in these tasks **(FIGURE 13*C*)** (573). During learning set performance with one rewarded and one unrewarded option, reward is initially predicted with $P = 0.5$. Chance selection of each option leads to positive and negative 0.5 prediction errors, respectively, and to corresponding phasic dopamine activations and depressions during initial learning trials (221). The error responses decrease with increasingly correct choices.

The transfer of dopamine activation from "unconditioned" reward to reward-predicting stimuli constitutes conditioning formalized by the Rescorla-Wagner and TD learning rules. Straightforward implementations of the TD model assume gradual TD error backpropagation via imaginary "internal" stimuli (374, 524). However, biological dopamine neurons do not show such gradual backpropagating responses (162) but transfer in single steps from reward to preceding actual stimuli and show simultaneous responses to both reward and stimuli with intermediate learning steps (410). Dopamine concentration transients in ventral striatum show similar transfer **(FIGURE 14*A*)** (117, 598). Biologically plausible TD implementations and spectral timing models achieve the single step response transfer to an earlier stimulus with stimulus and eligibility traces that mark recently active

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

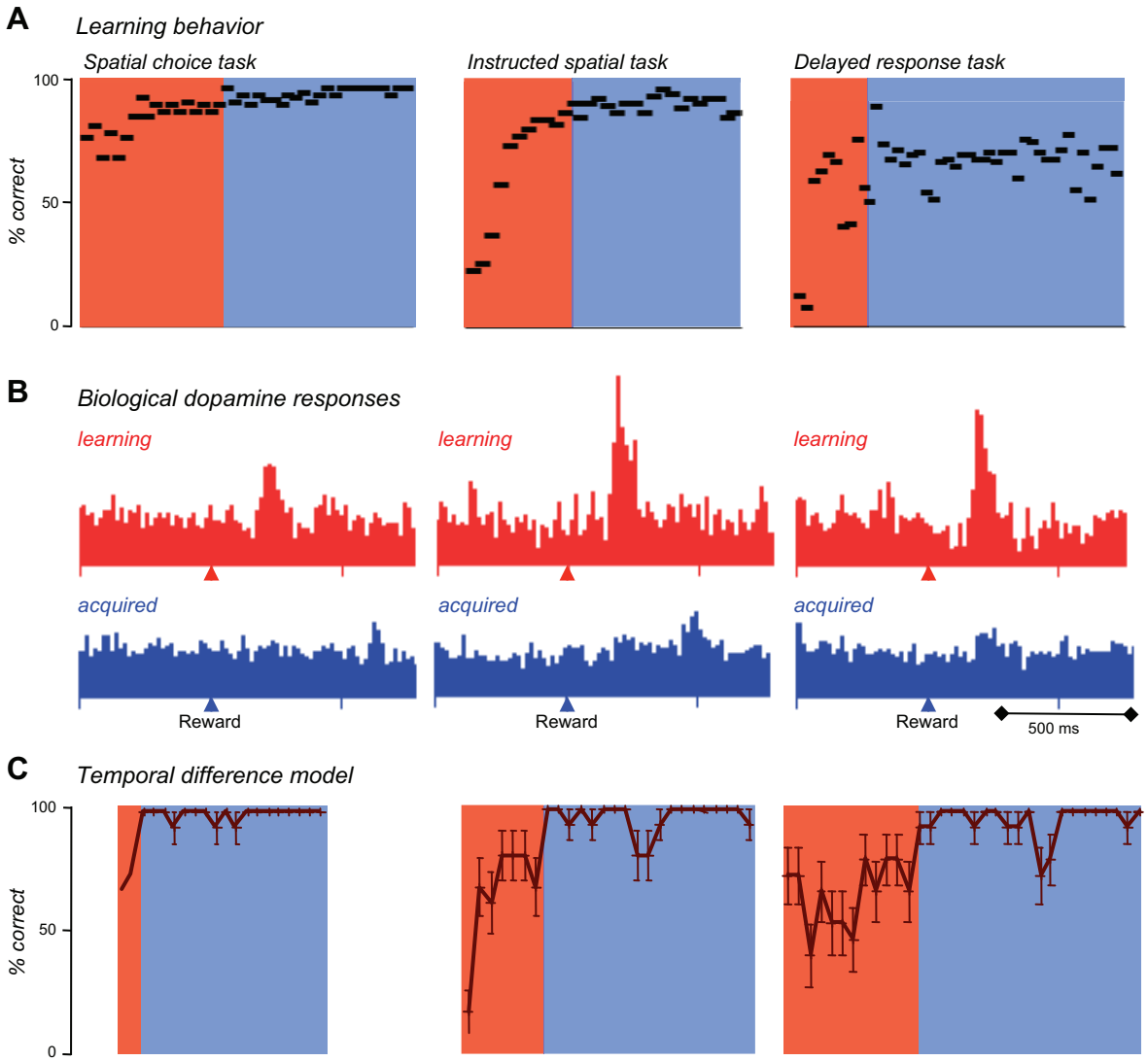


FIGURE 13. Dopamine prediction error responses reflect changing reward prediction during learning. *A:* stepwise learning of spatial delayed response task via intermediate spatial subtasks. Each dot shows mean percentage of correct behavioral performance in 20–30 trials during learning (red) and asymptotic performance (blue). *B:* positive dopamine reward prediction error responses during learning of each subtask, and their disappearance during acquired performance. Each histogram shows averaged responses from 10–35 monkey dopamine neurons recorded during the behavior shown in *A*. [*A* and *B* from Schultz et al. (521).] *C:* behavioral learning and performance of temporal difference model using the dopamine prediction error signals shown in *B*. The slower model learning compared with the animal's learning of the final delayed response task is due to a single step increase of delay from 1 to 3 s in the model, whereas the delay increased gradually in animals (*A*). [From Suri and Schultz (573), with permission from Elsevier.]

postsynaptic neurons for plasticity, as originally suggested **(FIGURE 14, *B–D*)** (71, 410, 573, 574). The eligibility traces may consist of intracellular calcium changes (154, 637), formation of calmodulin-dependent protein kinase II (236), IP_3 increase in dopamine neurons (203), and sustained neuronal activity in striatum and frontal cortex (522) reflecting previous events and choices in prefrontal neurons (30, 531). Thus use of eligibility traces results in good learning without small step backpropagation via imaginary stimuli (410, 573). These models account also better for the temporal sensitivity of prediction error responses (113, 573).

The acquisition of dopamine responses to conditioned stimuli enhances the strength of excitatory synapses onto dopamine neurons (568), and cocaine administration increases spike-time-dependent LTP in dopamine neurons (19, 318). In elucidating the mechanism, burst stimulation of dopamine neurons, mimicking dopamine reward responses, induces NMDA receptor-dependent LTP of synaptic responses in dopamine neurons (203). Compatible with the temporal characteristics of behavioral conditioning (423, 589), the burst needs to follow the synaptic stimulation by at least 0.5 or 1.0 s, whereas postsynaptic burst omission reverses LTP, and inverse timing induces long-term depres-

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

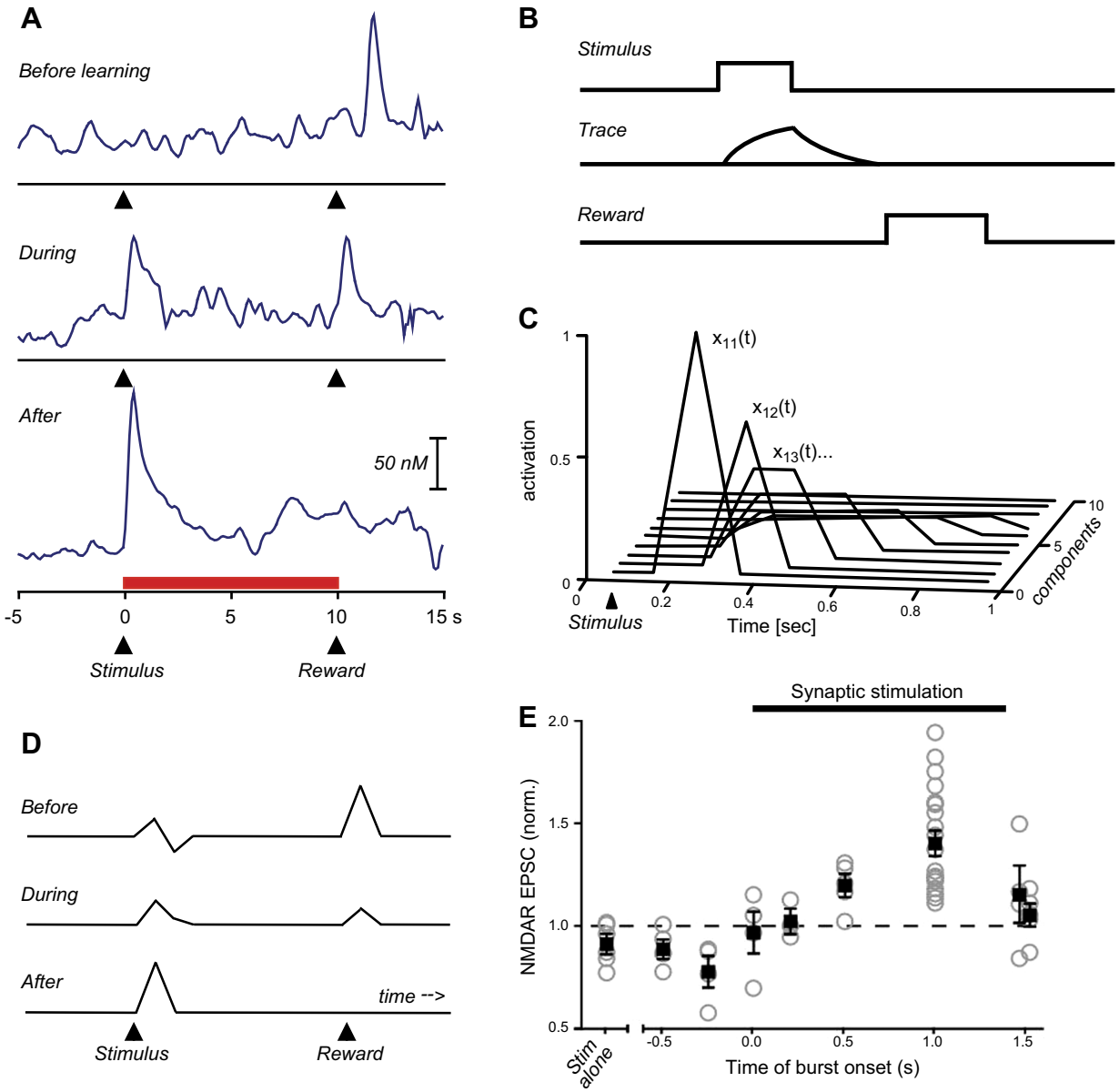


FIGURE 14. Plasticity of dopamine reward prediction error responses. *A:* transfer of dopamine response from primary reward to conditioned, reward-predicting stimulus during learning. The plots show voltammetrically measured dopamine concentration changes in ventral striatum of rats. Note simultaneous occurrence of responses to reward and stimulus during intermediate learning stage. [From Stuber et al. (598). Reprinted with permission from AAAS.] *B:* stimulus eligibility traces linking reward to stimulus during learning in original formulation of temporal difference model (TD) model. [Redrawn from Sutton and Barto (574).] *C:* stimulus eligibility traces used in biologically plausible implementation of TD model of dopamine prediction error responses. The traces allow single step transfer of dopamine prediction error signal from reward to an earlier stimulus. *D:* transfer of modeled dopamine prediction error response from reward to conditioned stimulus using eligibility traces shown in *B* and *C*. [*C* and *D* from Suri and Schultz (573), with permission from Elsevier.] *E:* time-sensitive plasticity of dopamine neurons in rat midbrain slices. LTP induction depends on timing of burst stimulation of dopamine neurons relative to their synaptic excitation. Only bursts occurring 0.5–1.5 s after synaptic stimulation lead to measurable LTP (excitatory postsynaptic currents, EPSC, mediated by NMDA receptors). [From Harnett et al. (203), with permission from Elsevier.]

sion **(FIGURE 14*E*)**. The 0.5- to 1.0-s delay is required to increase intracellular IP_3 as potential stimulus eligibility trace. Thus excitatory synapses onto dopamine neurons seem to mediate the acquisition of dopamine neurons to conditioned stimuli.

11. Non-dopamine prediction error signals

The colloquial meaning of error refers to incorrect motor performance. Climbing fibers of the cerebellum and neurons in superior colliculus and frontal and supplementary eye field

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

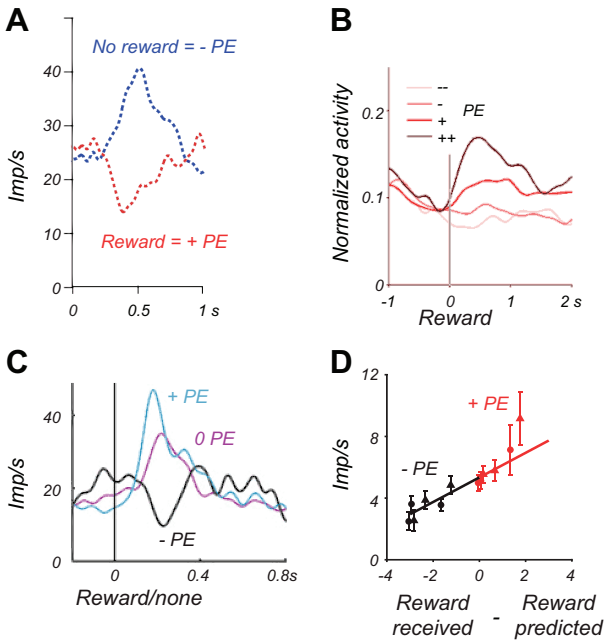


FIGURE 15. Bidirectional non-dopamine reward prediction error signals. *A:* averaged responses from 43 neurons in monkey lateral habenula during first trial of position-reward reversals. Red: positive prediction error; blue: negative prediction error. Note the inverse response polarity compared with dopamine error responses. PE = prediction error. [From Matsumoto and Hikosaka (344). Reprinted with permission from Nature Publishing Group.] *B*: averaged responses from 8 neurons in rat striatum. Subjective reward values (tiny, small, large, huge) are estimated by a Rescorla-Wagner reinforcement model fit to behavioral choices. [From Kim et al. (275).] *C:* response of single neuron in monkey amygdala. [From Belova et al. (36), with permission from Elsevier.] *D*: response of single neuron in monkey supplementary eye field. [From So and Stuphorn (551).]

signal errors of arm or eye movements (182, 266, 278, 292, 395, 569, 607). Cerebellar circuits implement a Rescorla-Wagner type prediction error for aversive conditioning (273, 357). The identification of bidirectional dopamine error signaling extended neuronal error coding to reward.

Bidirectional reward prediction error signals occur in several brain structures. Lateral habenula neurons show bidirectional reward prediction error signals that are sign inverted to dopamine responses and may affect dopamine neurons via inhibitory neurons **(FIGURE 15*A*)** (344). Select groups of phasically and tonically firing neurons in the striatum and globus pallidus code positive and negative reward prediction errors bidirectionally **(FIGURE 15*B*)** (15, 134, 227, 275, 402). Some neurons in the amygdala display separate, bidirectional error coding for reward and punishment **(FIGURE 15*C*)** (36). In the cortex, select neurons in anterior cingulate (261, 531) and supplementary eye field **(FIGURE 15*D*)** (551) code reward prediction errors bidirectionally. The bidirectional reward prediction error responses in these subcortical and cortical neurons are straightforwardly appropriate for affecting plasticity at specific postsynaptic neurons for reinforcement learning.

Positive and negative components of reward prediction errors are also coded in select groups of non-dopamine reward neurons. Some neurons in the pedunculopontine nucleus are activated by both predicted and unpredicted rewards and by conditioned stimuli (136, 396, 410) but fail to show depressions to omitted rewards and thus process reward without producing a clear error signal (281). Pedunculopontine neurons with midbrain projections above substantia nigra show bidirectional error responses to reward versus no-reward-predicting cues, but only positive prediction error responses to the reward (228). Positive responses to unpredicted but not predicted rewards occur in norepinephrine neurons as part of their attentional sensitivity (22), and in nucleus basalis Meynert neurons (466). In the striatum, some phasically or tonically firing neurons are influenced by unpredicted reward delivery or unpredicted reward omission (21, 255, 275, 556). Some of these striatal responses are stronger in Pavlovian than operant tasks (17) or occur only after particular behavioral actions (556), suggesting selectivity for the kind of behavior that resulted in the reward. In the amygdala, some neurons are more activated by unpredicted than predicted rewards (36, 40) or show unidirectional, rectified activations with both unpredicted reward delivery and omission (475). In anterior cingulate and dorsolateral prefrontal cortex, activations induced by reward omission or erroneous task performance are known for >30 years (386). In some orbitofrontal neurons, unpredicted rewards outside of tasks elicit activations, whereas reward omission fails to elicit responses (602). Some neurons in anterior and posterior cingulate and supplementary eye field are activated by unpredicted delivery or omission of reward (11, 21, 246, 261, 354, 551) and by unpredicted positive or negative feedback provided by conditioned visual reinforcers (344). Unpredictedness enhances existing reward responses in anterior cingulate cortex (206). The unidirectional activations to unpredicted delivery or omission of reward may reflect surprise salience or rectified positive and negative reward prediction errors. In reflecting surprise salience, they may code the learning rate parameter α of attentional learning rules (425) according to *Equation 3* and serve to adjust the speed of learning (475).

12. Non-dopamine responses during reward learning

Tonically active striatal interneurons (TANs) acquire discriminant responses to reward-predicting auditory stimuli within 15 min of pairing with liquid, and lose these responses within 10 min of extinction (14). Neurons in rat striatum, orbitofrontal cortex, and amygdala acquire discriminatory responses to movement-instructing and reward-predicting visual, auditory, and olfactory stimuli during learning and reversal (253, 422, 512). The acquired responses are rapidly lost during extinction (29).

Repeated learning of new stimuli leads to learning set performance in which animals acquire novel stimuli within a few trials (202), thus allowing to record from single neurons

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

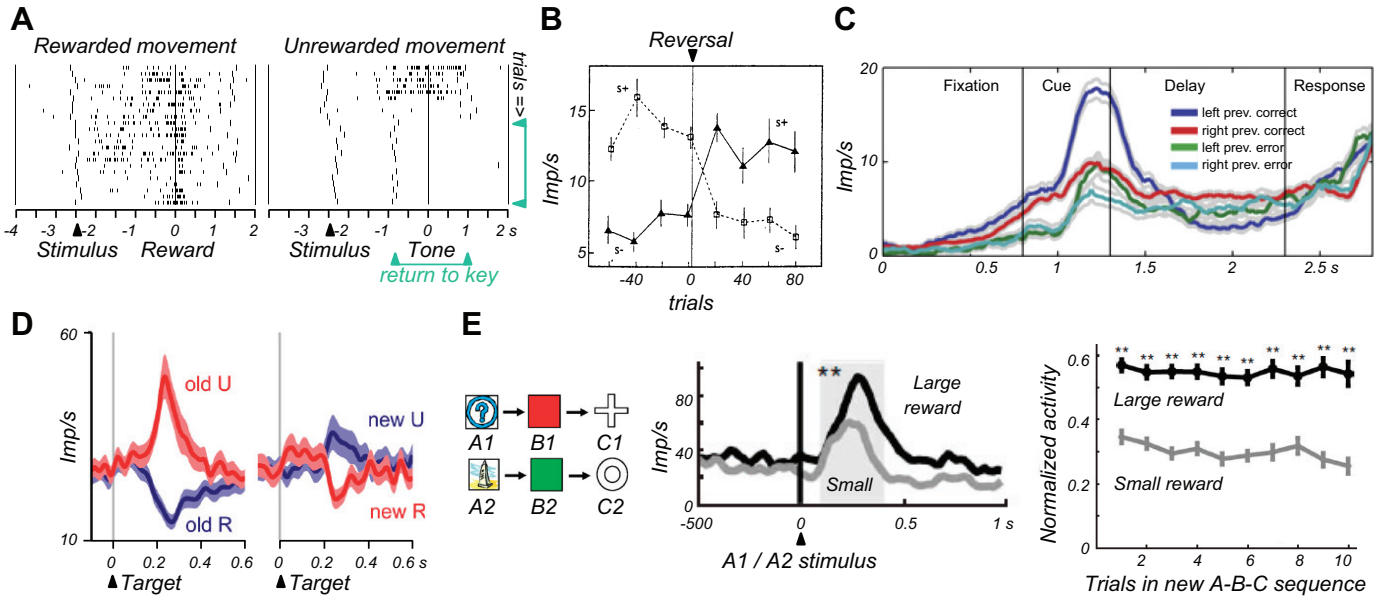


FIGURE 16. Overview of reward learning-related responses in monkey non-dopamine neurons. *A:* adaptation of reward expectation in single ventral striatum neuron during learning. In each learning episode, two new visual stimuli instruct a rewarded and an unrewarded arm movement, respectively, involving the acquisition of differential reward expectations for the same movement. Reward expectation is indicated by the return of the animal's hand to the resting key. This occurs with all rewarded movements after reward delivery (long vertical markers in left rasters, right to reward). With unrewarded movements, which alternate pseudorandomly with rewarded movement trials, the return occurs in initial trials after the tone (*top right*) but subsequently jumps before the tone (green arrows), indicating initial default reward expectation that disappears with learning. The reward expectation-related neuronal activity shows a similar development during learning (from *top* to *bottom*). [From Tremblay et al. (600).] *B*: rapid reversal of stimulus response in orbitofrontal neuron with reversed stimulus-reward association. S+ and S− are two different visual stimuli that are initially rewarded and unrewarded, respectivey. With reversal, they stay physically constant but inverse their reward prediction. [From Rolls et al. (485).] *C:* cue response in dorsolateral prefrontal cortex neuron reflecting correct, as opposed to erroneous, performance in previous trial in a delayed conditional motor task with reversal. This typical prefrontal neuron discriminates between left and right movements. Errors reduce differential left-right movement-related activity (previous trial correct: blue and red vs. error: green and blue). [From Histed et al. (217), with permission from Elsevier.] *D*: inference-related reversal of neuronal population responses in lateral habenula. These neurons are activated by unrewarded targets and depressed by rewarded targets. *Left*: in the first trial after target-reward reversal, before any new outcome occurs, the neurons continue to show activation to the old unrewarded target (red, old U) and depression to the old rewarded target (blue, old R). *Right*: in the second trial after reversal, after having experienced one outcome, the neuronal responses reflect the reward association of the other target. Thus the neurons are activated by the newly unrewarded target (blue, new U) and depressed by the newly rewarded target (red, new R), based entirely on inference from the outcome of only one target. [From Bromberg-Martin et al. (68).] *E:* inference of reward value from differentially rewarded paired associates in prefrontal cortex. Monkeys are first trained with two complete A-B-C sequences (*left*). Then, before the first trial of a new sequence, the two C stimuli are followed by large and small reward, respectively. For testing, the animal is presented with an A stimulus and choses the corresponding B and then C stimuli from the first new sequence trial on, even before sequential A-B-C-reward conditioning could occur. Neurons show corresponding reward differential activity (center, single neuron) from the first trial on (right, population average from 107 neurons), suggesting inference without explicit linking A and B stimuli to reward. [From Pan et al. (411). Reprinted with permission from Nature Publishing Group.]

during full learning episodes. Striatal and orbitofrontal neurons show three forms of reward related activity during learning (626, 600, 603, 639). Some neurons fail initially to respond to novel stimuli and acquire reward discriminatory responses during learning. Apparently they code valid reward predictions. To the opposite, other neurons respond initially indiscriminately to all novel stimuli and differentiate as learning advances, possibly reflecting exploration. A third group of neurons shows reward expectation activity that reflects an initial general expectation and becomes selective for rewarded trials as learning progresses **(FIGURE 16*A*)**. The increases in learning rate with electrical striatal stimulation suggest that some of these striatal neurons may mediate reward learning (639).

In analogy to learning set, repeated reversals of reward predictions of the same stimuli lead to reversal set performance in which animals switch after a single reversed trial

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

to the alternative option rather than requiring multiple learning trials, even without having experienced a reward. Prefrontal, orbitofrontal, and amygdala neurons show rapid response loss to previously learned stimuli and acquisition of the currently rewarded stimuli (626), closely corresponding to behavioral changes **(FIGURE 16*B*)** (422, 485). The changes occur several trials earlier in striatal neurons compared with prefrontal neurons (421), but are similar between orbitofrontal and amygdala neurons (378). Some striatal and cortical responses reflect the correct or incorrect performance of the previous trial in reversal sets, thus bridging information across consecutive trials **(FIGURE 16*C*)** (217).

Both learning sets and reversal sets define task structures and behavioral rules which allow individuals to infer reward values of stimuli without having actually experienced the new values. During reversals, responses in globus pallidus, lateral habenula, and dopamine neurons reflect the new reward value of a target already after the first reversed trial, based solely on inference from the other target **(FIGURE 16*D*)** (68). Neuronal responses in dorsolateral prefrontal cortex reflect reward value based on inference from paired associates **(FIGURE 16*E*)** (411) or transitivity (409). Neuronal responses in the striatum reflect inference by exclusion of alternative stimuli (409). These responses seem to incorporate the acquired rule into their responses, supposedly by accessing model-based learning mechanisms.

C. Dopamine Implementation of Reward Learning

1. Origin of reward prediction error response

The dopamine reward prediction error signal may derive from two basic mechanisms. The dopamine neurons may receive a complete bidirectional reward prediction error signal from presynaptic input. Alternatively, they may compute the signal from heterogeneous components by subtracting the predicted from the experienced reward at the time of the reward and, in the spirit of TD models, at the time of reward predictors. The computation would require separate inputs for positive and negative errors, both at the time of the final reward and at the time of cues predicting all future rewards.

Dopamine neurons receive main reward information from striatum, amygdala, subthalamic nucleus, pedunculopontine nucleus, rostromedial reticular nucleus, and GABAergic neurons of pars reticulata of substantia nigra (625). They receive homeostatic signals from lateral hypothalamus (464). However, dorsal, medial, and orbital prefrontal cortical areas in monkeys provide probably less direct inputs to dopamine neurons than often assumed (625) but could influence them transsynaptically via the striatum.

Bidirectional reward prediction error response components may arise from lateral habenula signals **(FIGURE 15*A*)** (344) which themselves derive from globus pallidus (227) but are sign inverted compared with dopamine responses. The habenula exerts strong and exclusively inhibitory influences on dopamine neurons (95, 250, 344) via GABAergic neurons in the rostromedial reticular nucleus (226, 249). The same habenula neurons show bidirectional punishment prediction error signals opposite to their reward prediction error responses. They respond to positive reward prediction errors in a very similar way as to negative punishment prediction errors (by depression), and to negative reward prediction errors similarly as to positive punishment prediction errors (by activation), thus coding value monotonically across rewards and punishers (346). However, it is unclear why the aversive habenula input would not impact on dopamine neurons (157, 160). Bidirectional reward prediction error inputs may also arise from select neurons in a number of other structures, including the striatum **(FIGURE 15*B*)** (15, 134, 275), amygdala **(FIGURE 15C)** (36), anterior cingulate (261, 531), and supplementary eye field **(FIGURE 15*D*)** (551). To make these neurons input candidates for the bidirectional dopamine prediction error signal would require demonstration of their projection to dopamine neurons, which may become technically possible by using selective optogenetic stimulation of error processing neurons that project to dopamine neurons.

Positive dopamine prediction error response components may arise from direct excitatory inputs from pedunculopontine neurons (131) responding to sensory stimuli, reward-predicting cues, and rewards (136, 228, 396, 410). In support of this possibility, inactivation of pedunculopontine neurons reduces dopamine stimulus responses (410). Norepinephrine neurons activated by attentional reward components (22) and nucleus basalis Meynert neurons activated by rewards (466) may also project to dopamine neurons. Different groups of striatal neurons exert activating influences on dopamine neurons via double inhibition (407) and respond to reward-predicting stimuli irrespective of the stimuli being themselves predicted (222) or incorporate reward predictions (17, 21, 255, 275, 402, 556). Positive response components may also derive from amygdala neurons that are activated by rewards irrespective of prediction (422), by unpredicted rewards (40), or by unpredicted reward delivery and omission (475). Reward responses in monkey frontal cortex (11, 21, 206, 246, 261, 347, 354) may reach dopamine neurons via the striatum.

Negative dopamine error responses may arise from GABAergic inputs from VTA, substantia nigra pars reticulata, or striatum (625). Some of these inputs are activated by aversive stimuli (10, 102, 582). Optogenetic activation of pars reticulata GABAergic neurons reduces dopamine impulse activity (613). Inhibition from GABAergic pars re-

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

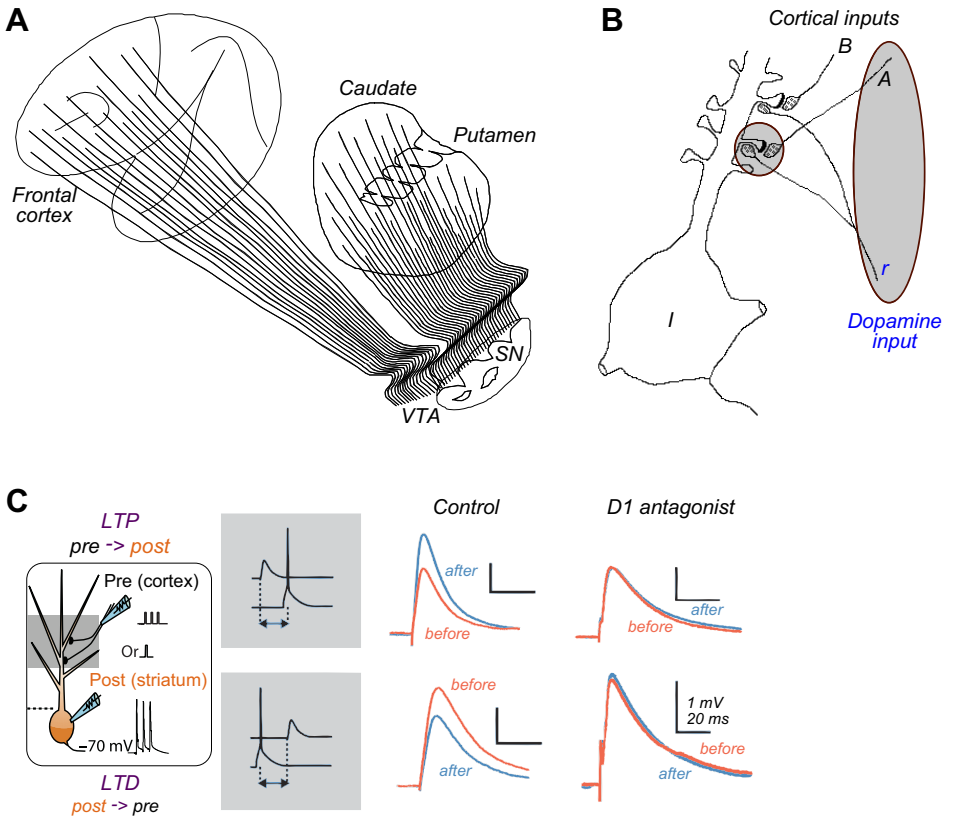


FIGURE 17. Anatomical and cellular dopamine influences on neuronal plasticity. *A:* global dopamine signal advancing to striatum and cortex. The population response of the majority of substantia nigra (SN) pars compacta and ventral tegmental area (VTA) dopamine neurons can be schematized as a synchronous, parallel volley of activity advancing at a velocity of 1–2 m/s (525) along the diverging projections from the midbrain to large populations of striatum (caudate and putamen) and cortex. [From Schultz (517).] *B:* differential influence of global dopamine signal on selectively active corticostriatal neurotransmission. The dopamine reinforcement signal (r) modifies conjointly active Hebbian synapses from active input (*A*) at striatal neuron (I) but leaves inactive synapses from inactive input (*B*) unchanged. Gray circle and ellipse indicate simultaneously active elements. There are ~10,000 cortical terminals and 1,000 dopamine varicosities on each striatal neuron (138, 192). Drawing based on data from Freund et al. (171) and Smith and Bolam (548). [From Schultz (520).] *C:* dopamine-dependent neuronal plasticity in striatal neurons. *Left*: experimental in vitro arrangement of cortical input stimulation and striatal neuron recording using a spike time dependent plasticity (STDP) protocol in rats. Control: positive STDP timing (striatal EPSP preceding action potential, Δt = 20 ms) results in long-term potentiation (LTP) (*top*), whereas negative STDP timing (striatal EPSP following action potential, Δt = −30 ms) results in long-term depression (LTD) (1, orange, and 2, blue, refer to before and after stimulation). Dopamine D1 recepor antagonist SCH23390 (10 μM) blocks both forms of plasticity, whereas D2 antagonist sulpiride (10 μM) has less clear effects and affects only plasticity onset times (not shown). [From Shen et al. (540). Reprinted with permission from AAAS. From Pawlak and Kerr (424).]

ticulata neurons showing sustained activations during reward prediction may also result in cancellation of reward activations in dopamine neurons by predicted rewards (102).

The computation of reward prediction error requires a reward prediction to be present at the time of the primary or higher order reward. This prediction may be mediated by the well-known sustained activations preceding rewards and reward-predicting stimuli in neurons of structures projecting mono- or polysynaptically to dopamine neurons, including orbitofrontal cortex (544, 602, 628), dorsal and ventral striatum (18, 205, 215, 523), and amygdala (40) (see below, **FIGURE 37, *D* AND *E***).

2. Postsynaptic influences

The phasic, time specific, reward prediction error reporting dopamine signal is propagated from specific dopamine populations in VTA and substantia nigra pars compacta (SNpc) to nucleus accumbens, dorsal striatum, frontal cortex, amygdala, and other forebrain structures. The widespread and diverging anatomical dopamine projections to postsynaptic structures show limited topography **(FIGURE 17*A*)** (177, 251, 342). Dopamine release and postsynaptic influence vary within small regions and are locally finely regulated (92). The tonic local dopamine concentration has separable, powerful enabling effects on a large variety of motor, cognitive, and motivational processes that are chal-

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

lenged by impaired dopamine function induced by Parkinson's disease, lesions, and psychopharmacological interventions (518). Distinction of the tonic function against the phasic signal is crucial for adequate conceptualization of dopamine's role in behavior. Importantly, interference of dopamine neurotransmission does not allow distinction between the two functions without special procedures. The deficits derive often from impaired tonic dopamine function and thus do not match the normal functions conveyed by phasic signals.

Dopamine synapses on dendritic spines of postsynaptic neurons in striatum and frontal cortex show a triad arrangement with cortical afferents that constitutes an architecture for three-factor Hebbian plasticity and allows the global dopamine reinforcement signal to exert differential influences on selectively active corticostriatal neurotransmission **(FIGURE 17*B*)** (171, 186, 461, 517). Indeed, electrical midbrain stimulation induces dopamine D1 receptor dependent long-term potentiation in striatal neurons in vivo (461) and expansion of cortical auditory fields for costimulated frequencies (28). Dopamine plays crucial roles in long-term potentiation (LTP) and depression (LTD) in striatum (82, 265, 294, 583), frontal cortex (194, 400), hippocampus (401), and amygdala (495). Dopamine turns tetanus stimulation-induced cortical LTD into LTP when paired with NMDA receptor activation (343). In spike-time-dependent plasticity (STDP) protocols, LTP occurs differentially when presynaptic stimulation precedes postsynaptic stimulation by a few tens of milliseconds, whereas LTD occurs with reverse sequence (48). Slice STDP protocols induce anatomical enlargement of striatal dendritic spines by burst-stimulating dopamine axons 0.3–2.0 s after depolarization of medium spiny striatal neurons, involving dopamine D1, but not D2, receptors and striatal NMDA receptors (642). The specific sequence of striatal excitation followed, but not preceded, by dopamine stimulation corresponds well to the temporal requirements for behavioral conditioning (423, 589). Correspondingly, dopamine D1 receptors are necessary for both LTP and LTD in striatal neurons, whereas D2 receptors may be more involved in the time course of plasticity **(FIGURE 17*C*)** (424). The effects may differ between striatal neuron types (540), as LTP in indirect pathway neurons does not depend on dopamine (294). In cultured hippocampus neurons, stimulation of dopamine D1 receptors enhances STDP LTP and turns LTD into LTP (648). These in vitro STDP data suggest a necessary role of dopamine in plasticity and should be carried to in vivo preparations for better functional assessment of their role in natural learning situations. Formal modeling demonstrates the suitability and power of dopamine-like reward prediction error signals in STDP learning (170, 247, 446), including prevention of run-away synaptic weights (375, 457).

3. Necessary dopamine involvement in learning (inactivation)

Hundreds of lesioning and psychopharmacological studies over the past 30 years demonstrate behavioral learning deficits with impaired dopamine function in a large variety of tasks. Following the three-factor Hebbian learning scheme (461, 517), intra-accumbens or systemically applied D1 receptor blockers or knockout of NMDA receptors on D1 receptor expressing striatal neurons impairs simple types of stimulus-reward learning **(FIGURE 18*A*)** (130, 163, 413). Similarly, intracortical D1 antagonist application impairs acquisition of differential neuronal responses in monkey prefrontal cortex and parallel behavioral learning in a delayed conditional motor task (440). Learning is somewhat less impaired by systemic D1 receptor blockade in tasks involving less direct reactions to stimuli and engaging less phasic dopamine responses (sign trackers versus goal trackers, **FIGURE 18*B***) (45, 163, 415). The learning deficits occur irrespective of performance deficits (163, 440, 624, 649, 650). Thus prediction error responses of dopamine neurons are not indiscriminately involved in all learning forms and

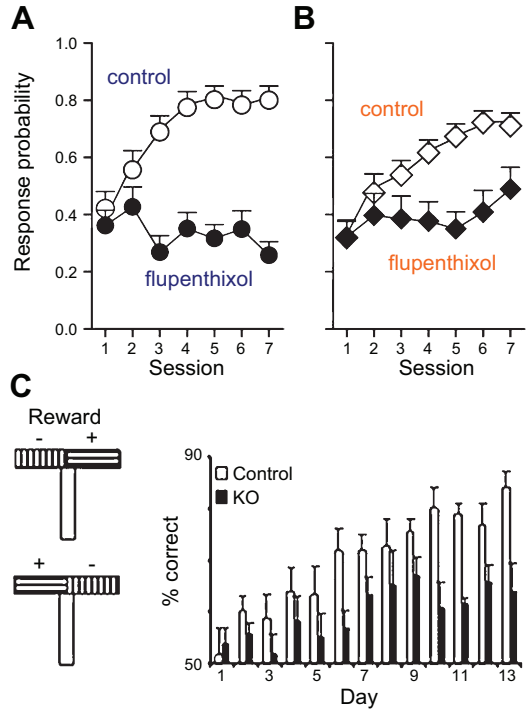


FIGURE 18. Differential learning impairment by dopamine receptor blockade. *A:* dopamine receptor blockade induces severe learning deficits in sign-tracking rats (learning under flupenthixol, not shown; test drug free, shown here). Sign trackers contact a conditioned stimulus before approaching the goal. *B:* dopamine receptor blockade by flupenthixol induces less severe learning deficits in goal-tracking rats. Goal trackers bypass the stimulus and directly approach the goal. [*A* and *B* from Flagel et al. (163). Reprinted with permission from Nature Publishing Group.] *C:* knockout (KO) of NMDA receptors on mouse midbrain dopamine neurons reduces phasic impulse bursts (not shown) and induces learning deficits in pseudorandomly alternating T-maze reward arms. [From Zweifel et al. (650).]

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

play more important roles when prediction errors have more direct effects.

Genetic NMDA receptor knockout on mice dopamine neurons reduces neuronal burst responses to reward-predicting stimuli and, in parallel, induce deficits in a wide range of learning situations including conditioned place preference, maze learning, and operant conditioning **(FIGURE 18*C*)** (624, 649, 650). Correspondingly, $GABA_A$ receptor knockout on dopamine neurons enhances dopamine release and T-maze and lever press learning (414). Without permitting adaptation during ontogenetic development, inactivation of dopamine neurons by local muscimol impairs learning from unexpected reward increase or omission (478). Transient direct or transsynaptic optogenetic inhibition of dopamine neurons induces place dispreference learning in choices between two compartments (see **FIGURE 19*E*** below) (244, 582).

Taken together, learning depends on intact dopamine function in simple reward contiguity situations with explicit, easily identifiable rewards and conditioned stimuli that engage phasic dopamine responses. Learning may not depend on dopamine neurons in situations in which they do not show phasic responses. Learning may also occur despite dopamine impairments when other learning systems compensate within the employed time frames, which may consist of switching to other struc-

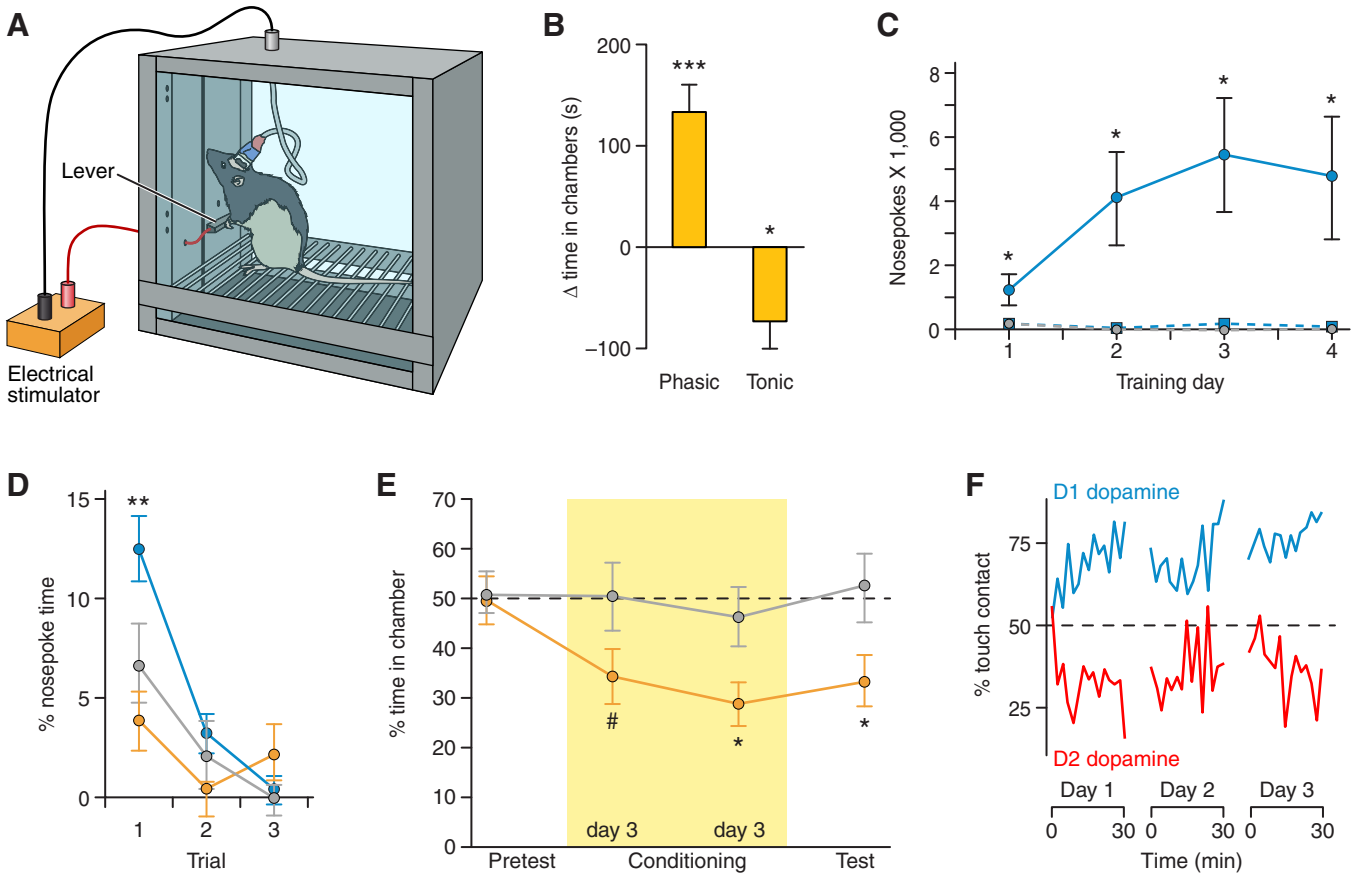


FIGURE 19. Effects of dopamine stimulation on behavioral learning. *A:* operant self-stimulation chamber. Pressing the lever elicits an electrical or optogenetic stimulus delivered to a specific brain structure through an implanted electrode/optrode. [From Rentato M. E. Sabatini, The history of electrical stimulation of the brain, available at http://www.cerebromente.org.br/n18/history/stimulation_i.htm.] *B:* place preference conditioning by phasic but not tonic optogenetic activation of dopamine neurons in mice. [From Tsai et al. (605). Reprinted with permission from AAAS.] *C:* operant nosepoke learning induced by optogenetic activation of dopamine neurons at stimulated target (blue) but not at inactive target (black) in rats. [From Witten et al. (641), with permission from Elsevier.] *D:* optogenetic activation of dopamine neurons unblocks learning of visual stimulus for nosepoke in rats, as shown by stronger response to stimulus paired with optogenetic stimulation in a blocking procedure (blue) compared with unpaired stimulus (orange) and stimulation in wild-type animals (gray). Response decrements are typical for unreinforced tests ("in extinction"). [From Steinberg et al. (562). Reprinted with permission from Nature Publishing Group.] *E:* place dispreference conditioning by direct optogenetic inhibition of dopamine neurons in mice (yellow; optical stimulation alone, gray). [From Tan et al. (582), with permission from Elsevier.] *F:* operant conditioning of approach (blue) and avoidance (red) behavior by optogenetic activation of D1 and D2 dopamine receptor expressing striatal neurons, respectively, in mice. [From Kravitz et al. (293). Reprinted by permission of Nature Publishing Group.]

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

tures after receptor blockade or lesions or altering ontogenetic development in gene knockouts. Fine-grained behavioral analysis may nevertheless reveal subtle remaining learning deficits. Given the crucial importance of reward for survival, multiple, flexible systems that sustain basic forms of learning despite partial impairments are biologically plausible and would enhance evolutionary fitness.

4. Sufficient dopamine involvement in learning (stimulation)

Intracranial electrical self-stimulation (398) induces operant learning of lever pressing **(FIGURE 19*A*)**. The effect is partly based on activation of VTA and SNpc dopamine neurons (103, 155) and involves dopamine release onto D1 receptors in nucleus accumbens (117). Optogenetic activation of midbrain dopamine neurons in rodents elicits learning of place preference, nose poking and lever pressing, and restores learning in a blocking procedure **(FIGURE 19, *B–D*)** (2, 244, 271, 562, 605, 641). Optogenetically induced nose poke learning occurs also with activation of dopamine axons in rat nucleus accumbens and is attenuated by intra-accumbens infusion of dopamine D1 and D2 receptor antagonists, suggesting involvement of dopamine projections to nucleus accumbens (562). The stimulations are similar in frequency to natural dopamine activations (20–50 Hz) but often exceed their natural durations of 150–250 ms by 2–7 times (500-1,000 ms), although stimulation durations of 200 ms are also effective (271, 641). The longer-than-natural activations induce more dopamine release in vitro (641) but may be compromised by the limited capacity of dopamine neurons for prolonged discharges due to pronounced tendency for depolarization block (54, 463). In contrast to 50-Hz phasic optogenetic dopamine activation, tonic (1 Hz) stimulation does not induce place preference learning (605), emphasizing the efficacy of phasic activations. In monkeys, electrical 200 Hz/200 ms VTA microstimulation induces choice preferences and neuromagnetic striatal activations, presumably reflecting dopamine activations (20). Opposite to excitations of dopamine neurons, direct optogenetic inhibition of dopamine neurons, or their indirect inhibition via activation of local, presynaptic GABA neurons, leads to place dispreference learning **(FIGURE 19*E*)** (244, 582). Conceivably the optogenetic excitation and inhibition of dopamine neurons mimic positive and negative dopamine prediction error signals and affect learning accordingly.

Learning induced by stimulation of dopamine neurons and their postsynaptic targets demonstrates the acquisition of associations with stimulation reward rather than simply reflecting arousal, response bias, direct effects on behavior, or performance enhancement, as the learned behavior is retained over several days, declines only gradually during extinction (271, 293, 582, 605) and is sensitive to devaluation by contingency degradation (641). The learning may involve striatal dopamine receptors, as optogenetic stimulation of D1 receptor expressing striatal neurons induces approach learning and stimulation of D2 receptor containing striatal neurons induces avoidance learning **(FIGURE 19*F*)** (293). The dichotomous striatal functions go back to learning models showing choice preference mediated by D1, and dispreference by D2, receptor containing striatal neurons and their respectively associated direct and indirect pathways (169). Taken together, electrical and optogenetic activation of dopamine neurons seem to mimic the natural dopamine prediction error responses to rewards and elicit learning behavior. The wider question is, to which extent would dopamine responses to a natural reward bias Pavlovian associations with that reward or operantly condition actions leading to the reward?

IV. APPROACH AND CHOICE

Rewards induce approach behavior and serve as arguments for economic choices. These functions involve a fundamental requirement, the prediction of reward that elicits reward expectation. We can approach an object because we expect it to be a reward. Without that expectation, we would approach an unknown object only when we are exploring the world. The same holds for choices. Final outcomes are often not apparent at the time of choices. Without such information, choices would be guesses and thus be very inefficient. One simply cannot make informed decisions without knowing what to expect from each option. Explorations are not informed choices but are helpful for detecting potentially better rewards. Thusgoal-directed approach behavior and informed choices are based on predictions of reward.

Predictions are acquired in the most basic form by Pavlovian conditioning. The sound of the bell predicts the sausage to Pavlov's dog. Also, the stimuli occurring during operant conditioning become Pavlovian conditioned reward predictors as the improving behavior results in more rewards. Each stimulus may become a higher order reward through conditioning and learning and thus has distinct economic value. Money does not have homeostatic or reproductive functions on its own but allows individuals to acquire nutritional and mating rewards and thus is a non-primary reward. More complex forms of prediction learning involve observation of other individuals, inference of events based on acquired knowledge, as in Bayesian updating in which the "posterior" probability of event occurrence arises from the "prior" probability, and conditional and reflective reasoning ("if I do this I can expect to obtain that reward"). In mediating Pavlovian and operant conditioning, in addition to stimulus-reward and action-reward pairing, the three-term contingency under which a discriminative stimulus signals an effective action-reward relationship constitutes the key factor for goal-directed approach behavior and informed choices. Contingency requires prediction errors, which serve to acquire and update the crucial predictions. Reward neurons are sensitive to contingency **(FIGURE 5)** (41) and code reward prediction errors **(FIGURE 7)** (517).

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

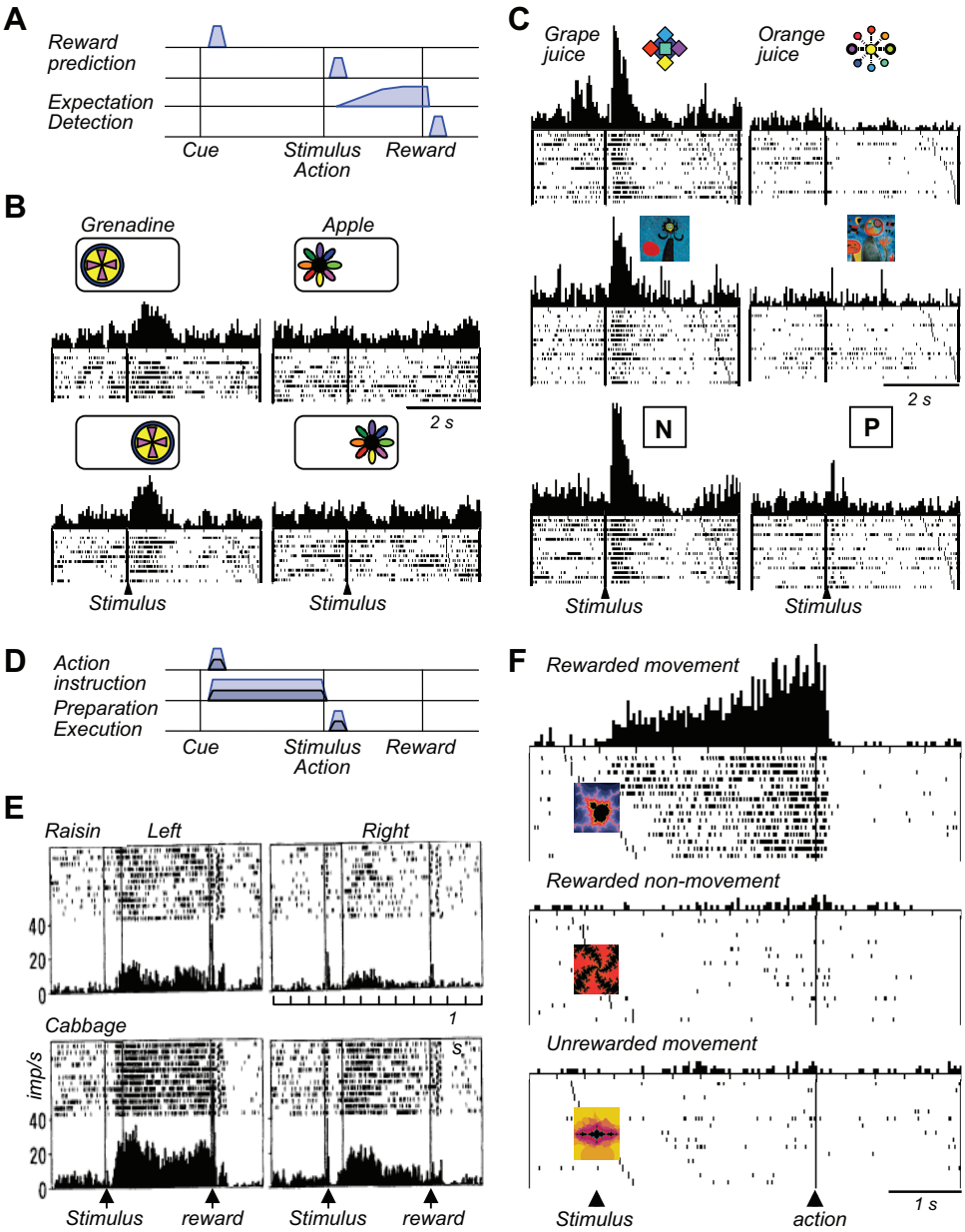


FIGURE 20. Explicit and conjoint neuronal reward signals in monkeys. *A:* scheme of explicit reward processing. Reward-predicting responses occur to an initial cue or to an action-inducing stimulus. Anticipatory activity occurs during the expectation of reward elicited by an external stimulus or action. Reward detection responses occur to the final reward. These different activities occur in separate neurons, except for dopamine neurons which all show similar responses (and no sustained reward expectation activations). *B:* neuronal reward processing irrespective of spatial position in orbitofrontal cortex. Differential activation by conditioned stimulus predicting grenadine juice but not apple juice, irrespective of spatial stimulus position and required movement (spatial delayed response task). *C:* neuronal reward processing irrespective of visual stimulus features in orbitofrontal cortex. Differential activations by conditioned stimuli predicting grape juice but not orange juice, irrespective of visual stimulus features. [*A* and *B* from Tremblay and Schultz (601).] *D:* scheme of conjoint reward-action processing. Predicted reward affects neuronal activity differentiating between different movement parameters during the instruction, preparation, and execution of action (e.g., spatial or go-nogo, blue vs. gray). *E:* conjoint processing of reward type (raisin vs. cabbage) and differentiating between spatial target positions in dorsolateral prefrontal cortex (spatial delayed response task). [From Watanabe (627). Reprinted with permission from Nature Publishing Group.] *F:* conjoint processing of reward (vs. no reward) and movement (vs. no movement) in caudate nucleus (delayed go-nogo task). The neuronal activities in *E* and *F* reflect the specific future reward together with the specific action required to obtain that reward. [From Hollerman et al. (222).]

A. Basic Reward Processing

1. Events eliciting reward responses

Explicit neuronal reward signals code only reward information. They occur as phasic responses to the delivery of "unconditioned" rewards, as phasic responses to conditioned, reward-predicting stimuli, and as sustained activity during the expectation of reward **(FIGURE 20A)**. They reflect only reward properties and do not vary with sensory or motor aspects **(FIGURE 20, *B* AND *C*)**.

"Unconditioned" liquid or food rewards elicit explicit reward signals in all main components of the brain's reward system (see **FIGURES 7**; **8, *A–C***; **13*B***; **AND 27*A***) (16, 18, 42, 60, 205, 215, 322, 383, 387, 405, 422, 451, 492, 541, 592). Reward responses are found also in premotor, prefrontal, cingulate, insular, and perirhinal cortex (8, 246, 369, 569). Some prefrontal neurons code the reward of the preceding trial (30, 179).

Conditioned, reward-predicting stimuli induce phasic explicit reward signals in all main reward structures of the brain, including orbitofrontal cortex, striatum, amygdala, and dopamine neurons (VTA and SNpc) (**FIGURE 20, *B* AND *C***; see also **FIGURES 5*B***; **7*C***; **8*C***; **9, *A*, *B*, AND *D***; **27, *B* AND *D***) (e.g., Refs. 41, 205, 276, 319, 371, 387, 405, 422, 516, 542, 601). These responses occur irrespective of the physical properties of the conditioned stimuli. Similar reward-related stimulus responses occur also in dorsolateral prefrontal cortex and anterior insula (338, 369, 411).

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

During reward expectations evoked by reward-predicting stimuli, select neurons in orbitofrontal cortex, striatum, and amygdala show slower, sustained explicit reward signals (see **FIGURES 16*A* AND 38, *D* AND *E***) (18, 37, 215, 601) which may also reflect reward timing (528) and temporal reward structure (40).

In addition to inducing explicit reward signals, rewards affect existing sensory-specific and action-specific activity. Neurons coding reward together with differential sensory information are found in orbitofrontal, dorsolateral prefrontal, perirhinal, and inferotemporal cortex (231, 319, 371, 394, 411). Reward affects neuronal activity differentiating between different movement parameters during the instruction, preparation and execution of action in prefrontal and premotor cortex **(FIGURE 20, *D* AND *E*)** (135, 282, 313, 348, 476, 606, 627), anterior and posterior cingulate cortex (354, 542), parietal cortex (381, 427, 433), striatum **(FIGURE 20*F*)** (107, 205, 222, 260, 308), globus pallidus (178), substantia nigra pars reticulata (502), superior colliculus (243), and amygdala (428). By processing information about the forthcoming reward during the preparation or execution of action, these activities may reflect a representation of the reward before and during the movement toward the reward, which fulfills a crucial requirement for goal-directed behavior (133). However, in motor structures, increased movement-related activity with larger rewards may reflect the more energized movements rather than representing a true conjoint reward-action signal (476). After reward delivery, responses in dorsolateral prefrontal neurons differentiate between movement directions (606).

Taken together, reward neurons show passive responses to reward-predicting stimuli and rewards, and sustained activities in advance of predicted rewards. Explicit reward signals in a limited number of brain structures reflect reward information but code neither sensory nor motor information. In contrast, sensory and motor neurons in several brain structures process reward information conjointly with sensory responses and action activity **(FIGURE 21)**. Reward coding involves the large majority of dopamine neurons, which all respond in a very similar manner to the effective events. In contrast, only specific and much smaller fractions of neurons in the other reward structures show reward activity, and their responses can be quite selective for one or the other of several task events.

2. Subjective reward perception

Neuronal reward responses may depend on the subjective perception of reward-predicting stimuli. In a signal detection task, monkeys correctly or incorrectly report the presence of stimuli (hit and false alarm, respectively), and correctly or incorrectly report stimulus absence (correct rejection and miss) (122). Dopamine neurons are activated by an initial task stimulus only when the animal correctly reports its detection, whereas they are not activated by the same physical stimulus when the animal reports its absence (miss). Thus subjective perception of the stimulus is necessary to activate dopamine neurons. However, subjective perception is not sufficient, as dopamine neurons do not respond when the physically absent stimulus is incorrectly reported as present (false alarm) **(FIGURE 22)**. The prediction error responses to a subsequent cue and to the final reward reflect also the subjective perception of the initial

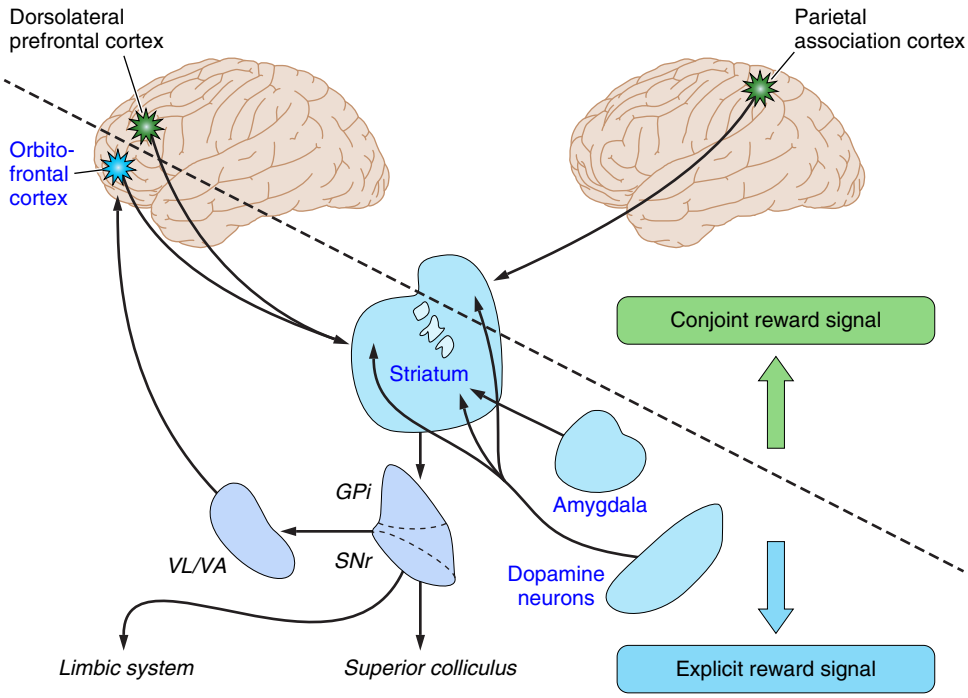


FIGURE 21. Simplified scheme of brain structures and connections involved in explicit and conjoint reward processing. The diagonal line schematically separates brain structures with neurons processing explicit reward information from structures whose neurons process reward together with sensory or action information. The striatum contains both classes.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

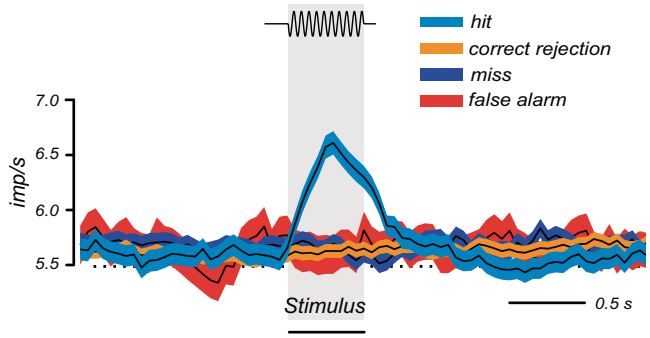


FIGURE 22. Dopamine reward response requires correct perception of reward-predicting stimulus. The reward prediction error response occurs only when the stimulus is present and is correctly reported as being present (hit; vibratory stimulus, *top*) but not when the stimulus is missed (miss) or is absent but incorrectly reported as present (false alarm). [From de Lafuente and Romo (122).]

stimulus. The responses are lower, and thus incorporate the prediction from the initial stimulus, when the animal had accurately reported stimulus presence compared with inaccurately reporting stimulus absence, even though the stimulus was physically the same. Thus dopamine responses reflect the animal's subjective perception and valuation of reward-predicting stimuli beyond their purely physical properties.

3. Distinction to unrewarding events

Reward neurons differentiate reward against neutral and aversive events. Dopamine neurons code reward but are not activated by aversive stimulus components **(FIGURE 11)** (157, 159, 160), even though their initial event detection response covaries with physical salience, novelty salience, stimulus generalization, and reward context **(FIGURE 10)**. In contrast, neurons in other reward structures respond selectively to aversive stimuli. Non-dopamine VTA neurons are activated by airpuffs and footshocks (102, 582). Some amygdala neurons are activated by airpuffs and conditioned stimuli predicting airpuffs, separately from reward neurons (422). Neurons in nucleus accumbens are differentially activated by sucrose or quinine, including their predictive stimuli (484). Neurons in orbitofrontal and anterior cingulate cortex, and TANs in striatum, respond differentially to aversive airpuffs, reward liquids, and their predictive stimuli and follow reversal between punishment and reward (10, 255, 452, 592). In dorsolateral prefrontal cortex, more neurons respond to reward than airpuff (283) and are differentially activated by gains or losses of reward tokens (532). In medial prefrontal cortex, dorsal neurons respond more frequently to air puffs, whereas ventral neurons are more sensitive to liquid reward (372). Thus punishment responses are often well separated from reward processing.

Some reward neurons respond also to punishers and thus process the motivational salience common to both reinforcers rather than reward or punishment separately. These reinforcer neurons are seen in amygdala (387), dorsolateral prefrontal cortex (283), and parietal cortex (309). In contrast to separate monotonic variations with reward and punishment magnitude, salience responses peak with high reward and high punishment and thus follow a U function (309). Thus restricted populations of reward neurons fail to discriminate against punishers and thus code motivational salience.

However, there is a caveat when interpreting the results of studies using aversive stimuli for probing negative motivational value. As indicated for dopamine responses above, it is crucial to distinguish between the physical impact of the stimulus and its motivational aversiveness. However, most studies apply presumably aversive stimuli without assessing their aversiveness to the individual animal. A good method to do so in neurophysiological studies is to estimate the negative value quantitatively on a common currency basis, using psychophysical choice procedures against rewards **(FIGURE 11*A*)** (160). Even substantial airpuffs are surprisingly not reported by the animals as having negative reward value (160).

B. Reward Value

1. Central role of subjective reward value

The ultimate function of reward is evolutionary fitness, which requires us to survive and reproduce. Thus rewards have survival value that is specific for us, and thus subjective. "Subjective" does not refer to conscious awareness but simply to the own person. Our brain detects the rewards that are necessary for this function. It is not really important what the objective, physical amount of a reward is, and often not even possible to know, as long as the reward provides us with the best chance for survival and reproduction. If we have plenty of food, a bit more does not matter much, but if we are starving, even a small amount is important. We do not necessarily go for the largest reward if it does not mean anything for us. We go for the reward that most satisfies our needs at that moment. A rat normally readily does not drink a salt solution but will do so when it is deprived of salt (472). Striving for the highest subjective value is built into our behavior by evolutionary selection. If we do not choose the highest value, we are at a disadvantage and over time fall behind and ultimately lose out against competitors that get better values. The striving for subjective value might also explain why we sometimes make choices that others do not understand. They are good for us at the moment, because we make them based on our own, private, subjective values, irrespective of whether we are consciously aware of them or not. Subjective valuation is also mechanistically reasonable. Reward function in individuals is defined by behavior driven by the brain, and the brain has evolved for our individual survival and reproduction rather than for an impersonal cause. We can identify all

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

substances in a wine and add up their prices from a catalog, but as a chemist and not as a wine connoisseur who is trying to maximize the personal pleasure based on subjective taste preferences. It is not the chemists's impersonal cause directed at the physical molecules but the wine connoisseur's subjective pleasure that contributes to fitness (by providing relaxation and encouraging taste discrimination training).

2. Assessing subjective value from behavior

Although subjective value is the key variable underlying approach and choices, it is a theoretical construct and not an objective measure. Utilitarianism aimed to achieve the greatest happiness for the greatest number of people (39). This might be assessed by an external person with honorable intentions (the benevolent dictator), by a hypothetical hedonimeter that measures happiness (142), or by subjective ratings (for example as "experienced utility") (259). However, benevolent dictators are implausible, hedonimeters are technically difficult, and ratings are subjective. Furthermore, there is no way of comparing subjective values between individuals because we cannot compare their feelings or, as a famous economist put it, "You cannot feel how much my tooth aches." In contrast, behavioral reactions and choices are measurable. They are often described in terms of preferences. However, preferences are internal private states and therefore prone to misjudgement and misreport. To operationalize preferences and remove their metaphysical and emotional connotations of purpose, desire, and pleasure, Samuelson (501) coined the term *revealed preferences* that are elicited by choices and follow specific axioms, notably consistency and transitivity. Thus choices elicit private and unobservable preferences hidden deep inside us and make them measurable as utility in an objective way. Other, more loosely defined, terms for value include survival value (survival of individuals or genes, referring to the reward function in evolutionary fitness), motivational or incentive value (eliciting approach behavior), and affective value (linking reward to emotion). There will be a more specific definition of economic utility further down with neuronal value signals.

There are several mechanisms for inferring subjective value from approach behavior and choices. In approach behavior, reward value is revealed by the speed (e.g., reaction time), duration, intensity, frequency, accuracy (e.g., error rate), and persistence of behavioral reactions. The reward eliciting the strongest approach behavior presumably has the highest "value." In choices, reward value is revealed by the probability of choosing one option over all other alternatives. The reward being consistently selected is considered to have the highest subjective value (assuming "rational" choices). In animals, psychophysical assessment of choices provides good measures for value. In humans, auction mechanisms allow to assess the subjective value of goods (willingness to pay). Thus approach and choices reveal the unobservable variable of value. When I see someone move faster towards an object or choose the object over another one, I get the intuitive notion that the object must have a higher value for the person. Value becomes simply a shortcut term denoting the impact of a reward on measurable behavioral choices. To avoid the circular definition when using value for explaining behavior we say that individuals behave "as if" they are maximizing value. Deducing the variable of value from behavior allows us to search for neuronal value signals for approach and choices. Such neuronal value signals would provide a physical basis for the theoretical notion of value and thus inform and validate economic choice theory.

3. Relative choices and common currency value

Behavioral choices reflect the values of choice options relative to each other. They will remain stable irrespective of value variations as long as the relative values of the options remain unchanged. For example, doubling of all option values should not affect choices. Thus value can only be expressed relative to other options within a given set of options. We can define a common reference reward, compare behavioral choices between that reference reward and any other reward, and thus establish a common value ranking or numeric scale defined by the reference reward which then serves as common currency. Money is a good example. It assesses the value of any reward in common monetary units. Any reward can be used as common currency for rewards differing in type, magnitude, probability, risk, delay, and other attributes. Just as different physical objects are measured on common scales of length or weight, different rewards can be measured on a common scale of value. Note that the common scale is only valid for an individual decision maker and does not provide a common measure to all decision makers. One dollar, or any other common currency, may constitute a different value for you than for me.

How can we choose between apples and oranges? We can compare the sensory properties of different objects using standardized scales of physical and chemical measures. As reward value is not a physical but a behavioral measure, this does not help. Rather, we should assess the values of such incommensurable rewards relative to each other. The idea of common currency allows us to quantify value relative to a unique reference that quantifies the value of rewards with vastly different properties. The activity of orbitofrontal and dopamine neurons reflects the subjective, common currency value integrated from different liquid rewards in close correspondence to behavioral choices (405) and extends to food rewards (301) **(FIGURE 23)**, suggesting a physical basis for the theoretical notion of common currency. Common currency is a necessary and sufficient feature of value coding, as it reflects independence from sensory properties and follows the idea of a reward retina as earliest neuronal detector of value (519).

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

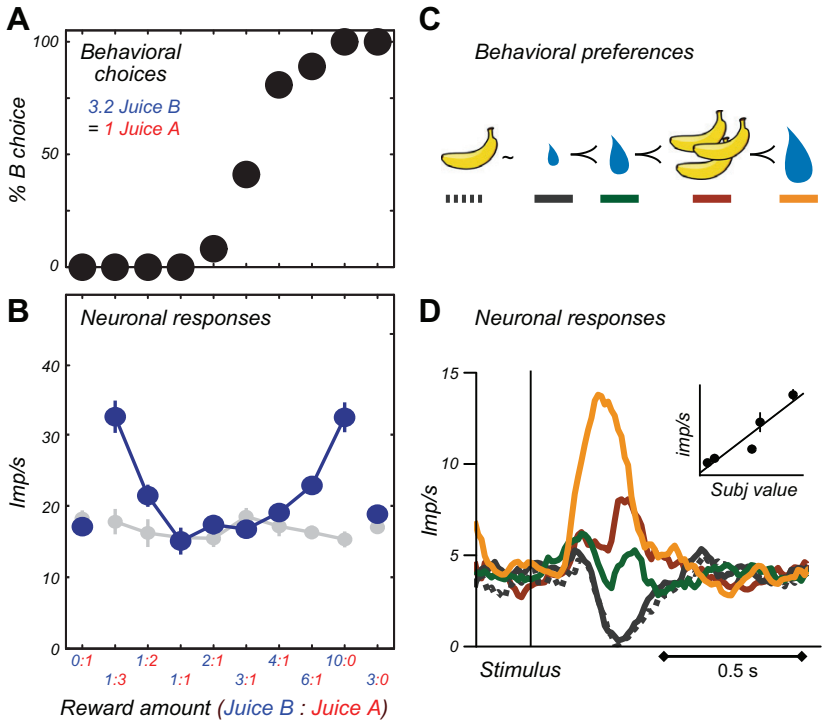


FIGURE 23. Common currency coding of subjective value derived from different rewards in monkeys. *A:* behavioral choices between two juice rewards of varying amounts reveal common scale subjective economic value. Amounts at choice indifference (50% choice of either reward) reveal 3.2 times lower value of reward B than reward A (which serves as common scale reference). *B:* common currency coding of subjective value derived from different juice rewards in single orbitofrontal neuron. Response to reward-predicting stimulus increases with amount of either juice, irrespective of juice. This activity codes the decision variable of chosen value. [*A* and *B* from Padoa-Schioppa and Assad (405). Reprinted by permission from Nature Publishing Group.] *C:* rank ordered, ordinal behavioral preferences for different liquid and food rewards, as assessed in behavioral choices between blackcurrant juice (blue drop) and mashed mixture of banana, chocolate, and hazelnut food (yellow bananas). ~, indifferent; <, preferred. *D:* common currency coding of subjective value in dopamine neurons (averages from 20 neurons). Colors refer to the rewards shown in *C*. [*C* and *D* from Lak et al. (301).]

Relative values do not always correlate with choices. Decision makers have a tendency to select the highest option unproportionally more often than its relative value against other options. Such maximizing choices favor the highest valued option irrespective of value variations as long as the ordinal value ranking among the options remains unchanged. Maximizing is reduced during exploratory behavior, when individuals occasionally check an inferior option for potential improvement, but exploration declines with increasingly stable values. These mechanisms are captured by the "temperature" parameter of the softmax function that models choices (333, 575). Exploration, modeled by higher temperature, may lead to effective decorrelation between choice probability and value and reveal serious constraints on the computation of value from choices. However, specific test methods control for these possible misreadings, including assessment of indifference points (subjective equivalence between choice options) in psychophysical procedures.

4. Value in natural rewards

Natural rewards, for which reward processing has evolved, contain nutrients and other substances of survival value. These rewards, like foods and drinks, have sensory properties, like shape, color, texture, viscosity, taste, and smell that help to detect, identify, and distinguish the reward objects. However, the sensory properties do not have nutrient value. Approach behavior and choices maximize reward value, not sensory properties. The transfer from sensory properties to value involves experiencing the value of the object, associating that value with the object's sensory properties using Pavlovian and higher forms of learning, and then using the predicted value for approach and choices. Thus sensory discrimination serves to identify the object via sensory receptors rather than determining its value. The sensory-value distinction lies at the core of reward processing; the sensory properties belong to the external object, whereas the reward properties derive from internal brain function attributing value to external objects.

Reward neurons distinguish between the sensory and value attributes of rewards. Amygdala and dopamine neurons show faster sensory and slower value responses (160, 422). Orbitofrontal and amygdala neurons follow reversed value associations of physically unchanged stimuli (422, 485). Orbitofrontal and striatal reward neurons discriminate between different food and liquid rewards but are insensitive to their visual and spatial attributes (205, 405, 601). Satiation on particular rewards reduces reward value responses in orbitofrontal and ventromedial prefrontal cortex (58, 105) but leaves sensory responses in primary gustatory cortex unaffected (487). Salt depletion induces reward responses in ventral pallidum to physically unchanged salt solutions (591). Thus the intuitive distinction between sensory and value attributes of rewards has a neurophysiological basis.

5. Non-value factors in choices

Besides being based on positive and negative economic value (gain and loss and their corresponding emotions) and exploration, real life economic choices are also determined by other factors. The factors include strategy, heuristics,

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

personal history, peer example, social pressure, conventions, traditions, prejudice, idiosyncrasies, morals, ethics, culture, religion, superstition, nationalism, chauvinism, daily irrationalities, and many others. Some originate from experiences that maximized gains and minimized losses and then led to automatisms that are no longer questioned for economic value. (Note that such automatisms are distinct from habits which may develop from goal-directed behavior and are based on value.) Other choice factors have evolved without explicit understanding of distant gains. Just as birds do not know the partial differential equations describing the coordination of their movements, individuals engage in behaviors without understanding why they do them, how they do them, and what the benefits are. These automatisms have evolved through evolution and provide guidance of behavior to varying but often large extents. Their advantages are transmission of beneficial behavioral traits, but their possible disadvantages are inconsistencies and domination by higher values during choices. Decisions based on automatisms would involve brain processes related to schemata and models of behavior, rules, emotions, social memory, theory of mind, and other cognitive functions.

The impact of automatisms should be considered in relation to value as key variable for economic choices. The true character of most automatisms differs fundamentally from that of gains and losses, and decision makers may not compute economic values when following habitual strategies or conventions. Then deriving value from economic choices would be very difficult and needs to be restricted to specific situations in which the automatism component can be estimated. This is often difficult and may easily lead to inconsistent choices when different situations engage different automatisms. What is truly missing is a definition of value independent of choice behavior. A neuronal value signal might provide such a measure and allow interpretations of economic choices. For example, a neuronal value signal identified independently of a particular choice situation might help to assess the contribution of value in choices in which automatisms also play a role.

C. Construction of Neuronal Value Signals

1. Value functions

Intuitively, the value of a reward increases monotonically with its physical magnitude. A larger drop of juice is worth more than a smaller drop, except when satiation results in asymptotic and then decreasing value and thus breaks monotonicity (see below). The value of a reward depends also on the frequency of its occurrence which is modeled as probability. Stimuli that predict more likely rewards are valued higher. Given their occurrence with specific magnitudes and probabilities, rewards constitute probability distributions of individual reward magnitudes, and choices between two rewards can be viewed as choices between probability distributions (417). The first statistical moment in probability distributions is the expected value (EV), which constitutes the sum of probability-weighted magnitudes (first moment of probability distributions)

$$EV = \sum_i [m_i * p_i(m_i)]; \text{ over all } i \qquad (8)$$

with m as magnitude (objective property of good), p as objective probability, and i as individual reward occurrences. The mean of increasingly large samples approaches the EV.

Probability is a theoretical construct that is derived from the measured frequency of event occurrence. Probability theory, as developed by Pascal and Fermat, allowed transferring the measure of experienced event frequency into the predictive measure of probability. For biological agents and their neurons, this transfer requires learning and memory, as the frequency of events impinging on sensory receptors needs to be counted in relation to other events and expressed as probability. The fact that there are no sensory receptors for probability strengthens the notion of a theoretical construct. Thus probability assessments mediate the transition from reacting to individual events to predicting future event occurrence. The past experience of reward frequency is registered by neurons and, true to the nature of predictions, processed to predict rewards and guide future behavior. By presenting events with a given frequency to animals, we assume that this transfer takes place, and we can search for neuronal activity that reflects reward probabilities. As described above with Pavlovian conditioning, the activity is necessarily predictive of the overall frequency of reward occurrence and thus reflects the reward value of the predictive stimulus. Thus the brain reacts to and assesses the frequency of individual rewards and transfers this measure into predictive activity reflecting the probability of future rewards.

The intuition of subjective value underlying economic choices combines with the failure to explain human choices by EV (43) and led to the axiomatic formulation of Expected Utility Theory (504, 617), which defines expected utility as a measure of subjective value in analogy to *Equation 8* as

$$EU = \sum_i [u(m_i) * p_i(m_i)]; \text{ over all } i \qquad (9)$$

with $u(m)$ as (subjective) utility. The main utility functions are power, logarithmic, exponential, and quadratic as well as their combinations (295), which can model human choices adequately (225). Logarithmic utility may combine the logarithmic Weber function that describes the sensory impact of reward objects with a logarithmic valuation function of objective, physical reward magnitude and probability.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

Utility is a hypothetical variable reflecting the subjective preferences elicited by behavioral choices. Whereas objective, physical value is measurable directly in milliliters or grams, utility is not measurable directly but inferred from behavioral choices (and maximized by decision makers). Thus the only physical measure from which utility can be inferred is behavioral choices. Choices are the tool to assess utility, and choices are the key method that economists accept for this assessment. There is a crucial difference between subjective value and utility. Although both subjective value and utility are estimated from measured behavioral choices, utility is a mathematical function of objective value [u(m)] that allows to predict the subjective value even for choices that have not been measured. Such mathematical functions allow to determine whole distributions of subjective values and establish useful terms such as EU for comparison between choice options. Utility is a universal measure of subjective value that does not require immediate behavioral assessment every time and thus constitutes the fundamental variable of economic decision theory.

The curvature of utility functions reflects marginal utility. Marginal utility is defined as the increment in utility gained from one additional unit of consumption, or the decrement from one unit less consumption, which is mathematically the first derivative of the utility function. Progressively decreasing marginal utility leads to concave functions (downward concave, as viewed from below, decreasing first derivative), which models the decreasing welfare one derives from ever more reward **(FIGURE 24A)**. A "well-behaved" economic utility function is concave, continuous, monotonically increasing, and nonsaturating (295, 341). It is based on money of which apparently one cannot get enough. Utility functions vary between different rewards, different risk levels (631), and different individuals. Utility functions for other rewards, like foods and drinks, do saturate and may even decrease, as too much food or liquid may become aversive and thus have negative marginal utility, effectively destroying monotonicity of the utility function **(FIGURE 24B)**. The "bliss point" is the utility producing maximal satisfaction. In contrast, increasing marginal utility leads to convex functions **(FIGURE 24C)**, suggesting that wins in high ranges are considered more important than the same gains in lower ranges ("small change"). In some cases, gains are valued increasingly more but ultimately become less important, leading to an initially convex and then concave utility function (173, 339) **(FIGURE 24D)**. Effectively, all combinations of curvatures of utility functions are possible, but the typical, "well-behaved" concave utility function is the standard model in economics.

The subjective weighting of value extends to reward probability. In particular, low and high probabilities are often distorted. Prospect Theory refines *Equation 9* to

$$\mathrm{EU} = \sum_i \{\mathrm{u}(m_i) * \pi[p_i(m_i)]\}; \text{ over all } i \qquad (10)$$

with π as probability weighting function (and P replacing EU in popular notation) (258) (Prospect Theory in addition incorporates references, different curvatures, and different slopes for losses than gains). The probability weighting function $\pi(p)$ is most nonlinear at $0.0 < P < \sim 0.2$ and $\sim 0.5 < P < 1.0$ and can be modeled with a one- or two-parameter function (α) (435) with an inflection point $\pi(p) = p$ at about $P = 0.37$, or a linear-in-log-odds function (187) in which gamma reflects curvature (low = inverted S, high = regular S) and α reflects elevation. Low probabilities can be distorted by a factor of 10,000. Well-trained monkeys show similar nonlinear, inverted S-shape probability weighting as humans with similar inflections points around $P = 0.37$ **(FIGURE 25)** (559).

Given that rewards affect behavior primarily through their subjective value, and that utility u(m) constitutes the central subjective reward function in economics, we can use *Equations 9* and *10* to state reinforcement learning *Equations 1*, *2*, *4*, and *5* in terms of economic utility

$$\mathrm{UPE}(t) = \lambda(t) - \mathrm{EU}(t) \qquad (10A)$$

$$\mathrm{EU}(t+1) = \mathrm{EU}(t) + \alpha * \mathrm{UPE}(t) \qquad (10B)$$

$$\mathrm{UTDPE}(t) = \left[\lambda(t) + \gamma \sum \mathrm{EU}(t)\right] - \mathrm{EU}(t-1) \qquad (10C)$$

$$\mathrm{EU}(t+1) = \mathrm{EU}(t) + \alpha * \mathrm{UTDPE}(t) \qquad (10D)$$

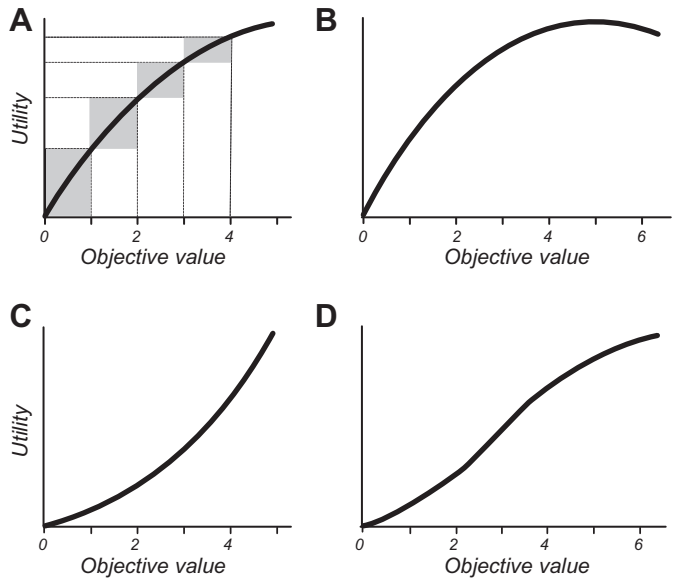


FIGURE 24. Utility functions. *A:* a "well behaved," gradually flattening ("concave"), continuous, monotonically increasing, nonsaturating utility function, typically for money. Gray rectangles indicate decreasing marginal utility (*y*-axis) with increasing wealth despite identical changes in objective value (*x*-axis). *B:* a saturating, nonmonotonic utility function typical for many biological rewards. Marginal utility becomes negative when an additional unit of consumption reduces utility (right end of function). *C:* a convex utility function, with gradually increasing marginal utility. *D:* a convex-concave utility function typical for progressively increasing marginal utility in lower ranges and progressively decreasing marginal utility in higher ranges.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

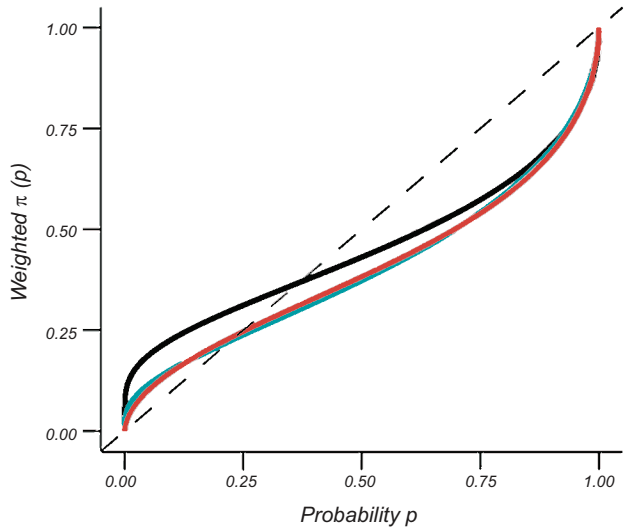


FIGURE 25. Nonlinear reward probability weighting functions estimated from behavioral choices in monkey. Fitting by one-parameter Prelec function (435) (black), two-parameter Prelec function (green), and linear-in-log-odds function (red) (187, 305). [From Stauffer et al. (565).]

Note that λ refers to the utility u(m) of the actual single reward received, whereas EU derives from the statistical utility distribution. Thus the utility prediction errors UPE(t) and UTDPE(t) constitute the key reinforcement terms for decision variables defined by economic utility.

Economists distinguish three principal forms of utility. First, scalar utility is subjective value as a mathematical function of physical, objective value. Utility can be derived from marginal utility under risk (to result in cardinal utility) or from the marginal rate of substitution without risk (amounting to ordinal utility). Cardinal utility functions established by choices under risk (617) provide a quantitative, numeric, mathematical characterization of the preference structure (257). Cardinal utility is valid up to positive affine transformations ($y = ax + b$) and can be psychophysically estimated from certainty equivalents at choice indifference between certain and risky options, starting at the distribution ends and advancing centrally by iteration (fractile or chaining procedure) (87). Cardinal utility can be expressed as u(m) (*Equation 9*) and is a requirement for establishing neuronal utility functions as neuronal signals have cardinal characteristics. In contrast, ordinal utility provides only a rank order of utilities and indicates whether a good is better or worse than another, without indicating on a numeric scale how much better or worse it is. It is only of limited value for establishing neuronal correlates of economic utility (higher versus lower neuronal responses for better or worse goods without numeric scaling). The marginal rate of substitution indicates the amount of good A a consumer is willing to give up to obtain one unit of good B while maintaining the same utility. The "distance" between curves of constant utility (indifference curves) indicates at least ordinal utility. Second, mean-variance utility is derived from objective value and risk via Taylor series expansion (see below, *Equations 21* and *21A*). Third, prospect theory extends utility by incorporating probability distortion (**FIGURE 25 AND** *Equation 10*), different loss than gain slopes (see below **FIGURE 29*E***) and reference dependency (see below **FIGURE 32** and *Equation 22A*).

The term of utility derives from the original utilitarian philosophers (39, 363). In economic decision theory, utility is defined by formal axioms (504, 617) that postulate how individuals behave in a meaningful manner "as if" they were maximizing utility. The term of utility is often used for convenience without requiring formal axiomatic tests in every instance (150, 324). This article follows this tradition, although that usage should not discourage axiomatic testing when identifying neuronal utility signals. Thus economic utility is a specific form of subjective value, which is derived from objective behavioral measures in individuals (but not across individuals). Utility has stricter requirements than subjective value by assuming an underlying, usually nonlinear, mathematical function that derives utility from objective value [u(m), as used in *Equation 9*]. If assessed from risky choices, utility is unique up to a positive affine transformation, which defines a cardinal function.

2. Neuronal value coding

The definition of cardinal, quantitative, numeric utility is important for investigating neuronal utility signals. The prime neuronal signal that is communicated to other neurons is the action potential. With sensory information, the signal arises from dedicated receptors whose stimulation induces numeric neuronal signals. Its strength as a signal is quantitatively expressed as firing rate (impulses/s) (3) and, to some extent, as pattern of action potentials. Firing rate is quasi-continuous between 0 and 1,000 impulses/s within the usual periods of 10–60 min of data sampling and usually increases or decreases monotonically with the phenomenon it codes. The same neuronal processing principles could be used for coding reward utility. For such neuronal signals to approach a meaningful mathematical function of utility, utility needs to be cardinal. It would be incorrect to derive a cardinal neuronal function from ordinal utility. This requirement would not apply if one would only look for ordinal higher firing rate for a higher (or lower) ranked, ordinal reward. With these characteristics, neuronal utility signals are physical (hardware) manifestations of a hypothetical variable that can only be inferred from behavioral choices. Although choices are the key method for assessing utility from behavior, utility neurons should also reflect utility at the time of reward reception, as this is the most important event for the survival of the organism.

Reward magnitude exerts a physical impact on sensory receptors. Reward magnitudes induce monotonically graded responses in dopamine neurons and in neurons in the stria-

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

tum, amygdala, and orbitofrontal, ventrolateral prefrontal, dorsolateral prefrontal, cingulate and posterior parietal cortex (12, 42, 107, 205, 262, 313, 354, 381, 405, 433, 598, 620). For rewards that do not vary in physical impact, like money, magnitude coding needs to be explained by other mechanisms including Pavlovian and cognitive processes.

Stimuli predicting specific reward probabilities induce graded responses in dopamine, striatum, and posterior parietal neurons and also in orbital, ventromedial, ventrolateral, dorsolateral, and cingulate prefrontal neurons, along with graded behavioral reactions and choices (12, 15, 17, 161, 261, 262, 354, 372, 376, 381, 402, 418, 433, 500). The transfer of dopamine responses from rewards to reward-predicting stimuli **(FIGURES 7*C* AND 14, *A* AND *D*)** may constitute a suitable correlate for the transition from reporting experienced reward frequency to coding reward probability. There are no neurophysiological studies on probability distortions. Human neuroimaging has shown probability distortions of values analogous to *Equation 10* (241, 595) but no behavioral distortions with valueless probabilities (110). Thus reward neurons transform the theoretical construct of probability into another theoretical construct called value (both of which are measurable from behavioral choices).

Increasing EVs induce graded neuronal responses in dopamine, striatum, orbitofrontal, ventromedial prefrontal, ventrolateral prefrontal, dorsolateral prefrontal, cingulate, and posterior parietal cortex neurons **(FIGURE 26)** (12, 261, 262, 354, 381, 402, 433, 418, 500, 598). In addition to characterizing reward coding, these data suggest a general biological basis for the mathematical constructs of EV, for which there are no sensory receptors. However, it is unlikely that these neurons code objective EV specifically rather than the subjective value that is monotonically related to objective value. Nevertheless, neuroscience applauds to the genius of the earlier mathematicians postulating biologically implemented theoretical terms.

Appropriate behavioral tests reveal that reward neurons typically code subjective value. Reward responses in orbitofrontal cortex, ventromedial prefrontal cortex, parietal cortex, striatum, globus pallidus, amygdala, lateral habenula and dopamine neurons closely reflect subjective behavioral choices **(FIGURE 23)** (301, 405, 433). Subjective value coding is also evident in specific situations, including satiation or deprivation without physical reward changes (58, 105, 591), differently delayed conditioned reinforcers (66, 67), temporal discounting of reward value (see below **FIGURE 28*D***) (79, 158, 285, 477, 479), and adaptation to identical reward magnitudes (see below **FIGURE 33, *A* AND *B***) (42, 284, 332, 403, 601). Details will be presented in the following sections.

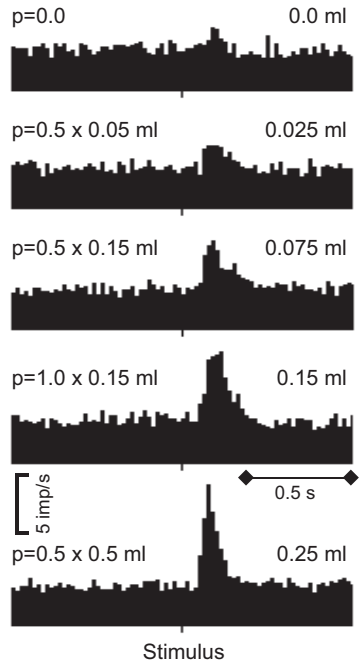


FIGURE 26. Neuronal value coding by monkey dopamine neurons. Monotonically increasing responses to increasing expected value (EV), irrespective of individual probability-magnitude combinations. EVs (*right*) are from 5 binary probability distributions with different probabilities and magnitudes of juice reward indicated at *left*. Population average from 55 dopamine neurons. Note that these responses do not suggest specific coding of EV as opposed to expected utility (EU), as at the time this distinction was not made experimentally. [From Tobler et al. (598).]

Although subjective value coding confirms the close behavioral-neuronal relationships of reward processing, it is agnostic about the relationship to physical reward value. In contrast, economic utility defines this relationship mathematically. Neuronal correlates of utility are found in dopamine neurons (560). Utility functions in monkeys may transition from convex via linear to concave with increasing reward volumes **(FIGURES 24*D* AND 27*A*)**, which reflects nonmonotonic marginal utility (first derivative of utility) and results in nonmonotonic utility prediction errors. This utility profile is closely paralleled by nonmonotonically changing dopamine prediction error responses with gambles placed at different parts of the utility function **(FIGURE 27*B*)**. With unpredicted rewards eliciting positive prediction errors, dopamine responses increase monotonically **(FIGURE 27*C*)** and thus follow the monotonic increase in utility prediction error (Δu) but not the nonmonotonic variation in marginal utility ($\Delta u/\Delta x$ or du/dx). Thus the dopamine reward prediction error response constitutes a utility prediction error signal that reflects marginal utilty but does not explicitly code marginal utility. Marginal utility relates well to prediction error, as both terms conceptualize deviations from a reference (prediction and current wealth, respectively).

Given its coding of utility, the dopamine signal can be stated as

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

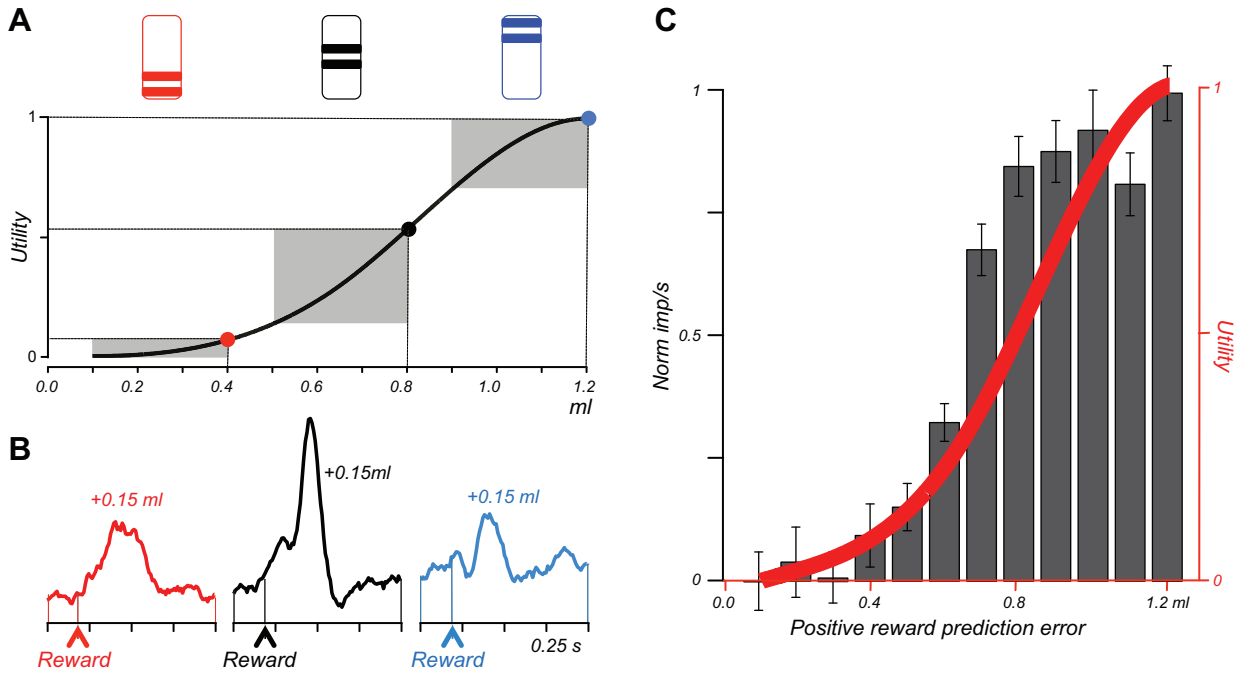


FIGURE 27. Utility prediction error signal in monkey dopamine neurons. *A*, *top*: gambles used for testing (0.1–0.4; 0.5–0.8; 0.9–1.2 ml juice; $P = 0.5$ each outcome). Height of each bar indicates juice volume. *Bottom*: behavioral utility function in monkey. Delivery of higher reward in each gamble generates identical positive physical prediction errors across gambles (0.15 ml, red, black, and blue dots). Due to different positions on the convex-concave utility function, the same physical prediction errors vary nonmonotonically in utility. Shaded areas indicate physical volumes (horizontal) and utilities (vertical) of tested gambles. *B:* positive neuronal utility prediction error responses (averaged from 52 dopamine neurons) to higher gamble outcomes in same animal (colored dots on utility function in *A*). The nonmonotonically varying dopamine responses reflect the nonmonotonically varying first derivative of the utility function (marginal utility). *C:* positive utility prediction error responses to unpredicted juice rewards. Red: utility function. Black: corresponding, nonlinear increase of population response ($n = 14$ dopamine neurons) in same animal. [*A–C* from Stauffer et al. (560).]

$$\text{DaResp}(t) \sim \text{UPE}(t) \qquad (10E)$$

and formulated for temporal difference (TD) utility error

$$\text{DaResp}(t) \sim \text{UTDPE}(t) \qquad (10F)$$

and can be fully expressed by using *Equations 6* and *7*, replacing V by EU and applying *Equations 10A* and *10E*

$$\text{DaResp}(t) \sim \lambda(t) - \text{EU}(t) \qquad (10G)$$

$$\text{DaResp}(t) \sim \left[\lambda(t) + \gamma\sum\text{EU}(t)\right] - \text{EU}(t-1) \qquad (10H)$$

As some neurons in other reward structures code also reward prediction errors, DaResp in *Equations 10E* through *10H* may be replaced by error signals from such neurons if their utility coding were established in future experiments.

As utility is not simply a measure of subjective value but derives mathematically from objective reward measures, utility coding extends the dopamine response well beyond subjective reward value coding seen with multiple rewards, effort cost, and risk (116, 156, 301, 570). Utility as higher specification of the dopamine response likely applies also to multiple rewards, effort cost, and risk whose neuronal coding is so far characterized only as subjective value (risk is already incorporated into the utility response in **FIGURE 27*B***).

3. Motivational state

Based on the intuition that the primary function of nutrient rewards derives from the need to acquire necessary substances for survival, reward value is closely related to the state of the animal. Deprivation on a particular substance enhances its subjective value, whereas satiation reduces its value. These states can be general or specific for a substance (e.g., "general" versus "sensory specific" satiety).

Satiation is captured by decreasing marginal utility in gradually flattening (concave) utility functions **(FIGURE 24*A*)**. Higher satiation leads to complete saturation and subsequent repulsion characterized by zero and then negative marginal utility **(FIGURE 24*B*)**, although such satiety does not typically occur with money. In contrast, reward deprivation is captured by the steeper part of the utility function with higher marginal utility.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

Rather than directly affecting marginal utility and EU, an alternative account may conceptualize motivational states as likely nonlinear and nonmonotonic utility (EUstate) which adds to EU and results in overall EUnet, similar to other influences described below

$$\text{EUnet} = \text{EU} + \text{EUstate} \quad (11)$$

With deprivation, EUstate is positive and increases EUnet, whereas satiation turns EUstate negative and decreases EUnet. We can distinguish general from sensory specific effects on EUnet by using separate state utilities

$$\text{EUnet} = \text{EU} + \alpha * \text{EUgenstate} + \beta * \text{EUspecstate} \quad (11A)$$

with weighting coefficients α and β. General deprivation and satiety affect all rewards to some extent (EUgenstate), whereas sensory specific satiety affects primarily the particular reward (EUspecstate). The effects of deprivation and satiety on EUnet can be accommodated by the reinforcement framework by substituting EU in *Equations 10A* through *10D* by EUnet.

Satiety on individual rewards reduces behavioral reactions and neuronal responses in monkey orbitofrontal cortex to the smell, taste, and sight of rewards (105) and in ventromedial prefrontal cortex to water (58). The response decrease is specific for the satiating reward, including blackcurrant, cream, vanilline, limonène, or monosodium glutamate liquids (105). Some responses to the nonsatiating rewards remain uninfluenced, suggesting sensory specific rather than general satiety. General satiety over the daily course of the experiment reduced most responses of dopamine neurons to auditory cues for food pellets in mice (496). Opposite to satiation, salt depletion makes salt rewarding. Rats show behavioral approach to NaCl solutions after, but not before, salt depletion by furosemide while reward neurons in the ventral pallidum begin to respond to salt (472, 591). These results provide neuronal correlates for the influence of motivational states on subjective value compatible with *Equation 11*.

4. Context

Rewards occur always in an environmental context, and these contexts exert considerable influence on subjective reward value. Imagine yourself in a gourmet restaurant and hearing a loud "wow." Then imagine watching a tennis match and hearing the same "wow." Obviously the behavioral reaction to the "wow" differs depending on the context, suggesting that information from the context (restaurant versus tennis) affects the interpretation of the explicit event ("wow"). Adding a nonlinear subjective value function EUcxt captures the value gained by rewarded contexts

$$\text{EUnet} = \text{EU} + \text{EUcxt} \quad (12)$$

Animal learning theory conceptualizes the influences of context on behavioral reactions in several forms. In general, Pavlovian-instrumental transfer, instrumental (operant) learning is enhanced in the presence of a previously Pavlovian conditioned stimulus compared with an unconditioned stimulus (145). In renewal, previously extinguished conditioned responding may return depending on the contextual environment (59). In pseudoconditioning, a primary reinforcer endows the context and all its cues with reward value through conditioning. Pavlov noted "a conditioned reflex to the environment" (423) and Konorski mentioned "conditioned to situational cues" (289). Pseudoconditioning elicits approach behavior to unpaired events in rewarded contexts (367, 536, 537), as if reward properties "spill over" from the context to the explicit event. For example, in an experiment on rabbit jaw movement, the value of the unpaired stimulus was close to 0 (analogous to EU = 0), but amounted to 70% of the value of the rewarded stimulus after that stimulus had been conditioned in the same context (EUcxt = 70% of rewarded stimulus EU) (537). In all these forms, otherwise ineffective stimuli occurring in the same context gain motivational value and induce behavioral responses (191, 335), even though they have never been explicitly paired with the primary reinforcer.

Influences of contextual information are well known from everyday experiences. Labeling an odor as cheese rather than body odor, or describing flavors as rich and delicious rather than vegetable water, increases pleasantness ratings and responses in human orbitofrontal cortex (118, 189). Stating the brand name of soft drinks, like Coke and Pepsi, strongly increases preference ratings for the drink indicated and enhances prefrontal reward responses (352). In wine tasting, stated prices influence pleasantness ratings and value responses in orbitofrontal cortex (432). In risky decisions with only partly known outcomes, additional information of outcomes after the decision may reduce or enhance the value of a received outcome (329). Thus contexts add EUcxt to the EU of odors, flavors, soft drinks, and wine and gambles to modify EUnet.

5. Economic cost

Natural rewards do not only provide nutrients and calories, their acquisition usually requires calorie expenditure that reduces the gains provided by them. We need to work for them, approach them, collect them, and consume them. This intuition applies to all rewards and is conceptualized as economic cost. Every good has its cost, and we usually calculate the net benefit before handing over the money. We need to ensure that the cost does not wipe out the benefit. The value of Eve's apple is reduced by the loss of paradise.

Net benefit utility can be calculated generically as a subtraction

$$\text{EUnet} = \text{EU} - \text{EUcost} \quad (13)$$

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

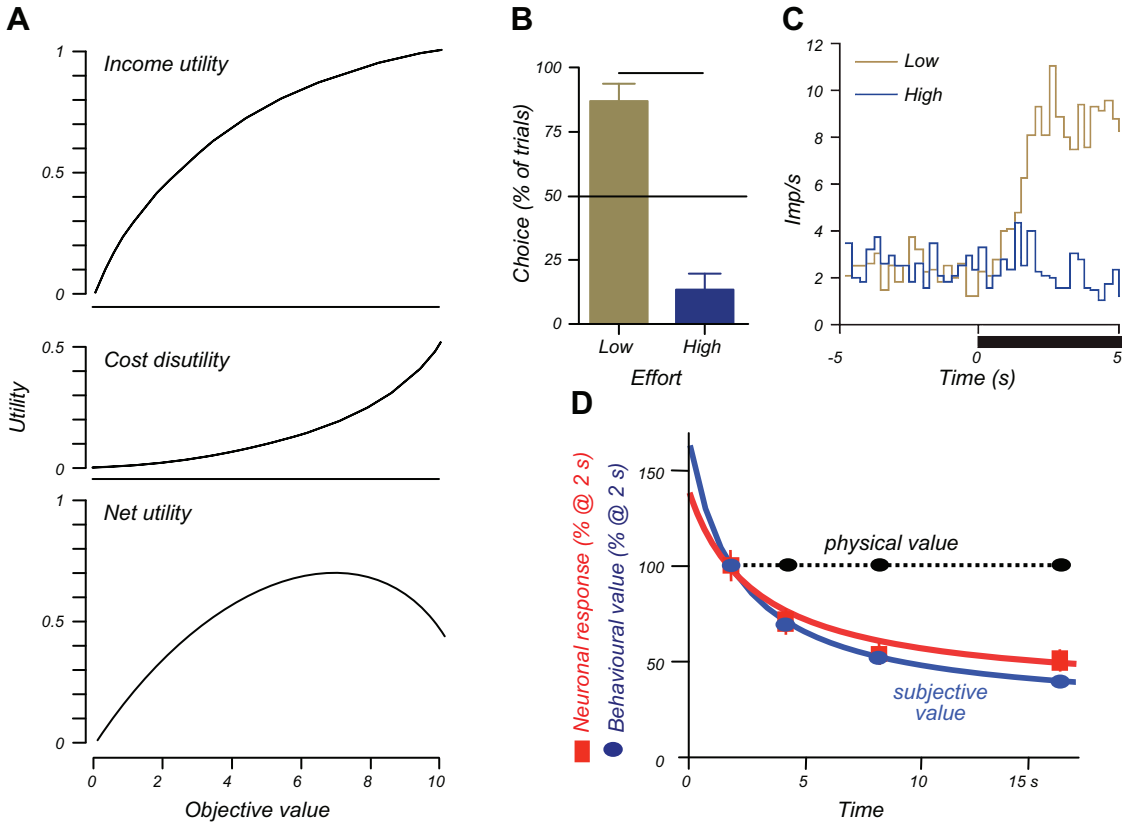


FIGURE 28. Neuronal value coding incorporates reward cost and delay. *A:* construction of net utility from reward cost. Concave income utility (*top*) minus convex cost disutility (*middle*) results in nonmonotonic net utility (*bottom*). *B:* behavioral choice preferences reveal stronger reduction of subjective reward value by high compared with low effort cost (number of lever presses) in rats (income is identical). *C:* stronger reduction of cue response of single rat nucleus accumbens neuron by high (blue) compared with low (gray) effort cost. [*B* and *C* from Day et al. (116). Copyright 2010 John Wiley and Sons.] *D:* temporal discounting in monkey choice behavior (blue) and corresponding responses of dopamine neurons (red) across reward delays of 2–16 s (hyperbolic fittings; n = 33 neurons). [From Kobayashi and Schultz (285).]

Equation 13 is based on physical net benefit, physical income, and physical cost (such as calories)

$$\text{Net benefit} = \text{income} - \text{cost} \qquad (14)$$

and becomes net utility by applying different utility functions u(income) and d(cost)

$$\text{Unet} = \text{u(income)} - \text{d(cost)} \qquad (15)$$

Income utility u(income) is often concave (gradually flattening due to decreasing marginal utility) (**FIGURE 28A**, *top*). Physical effort, as the energy expenditure required to obtaining a reward, is an economic cost, as long as it is associated with the pursuit and acquisition of that reward. As higher efforts become increasingly harder, cost is often assumed to be convex (**FIGURE 28A**, *middle*). With convex cost, the disutility of high effort may outweigh the increase of income utility, resulting in decreasing Unet and thus loss of monotonicity (**FIGURE 28A**, *bottom*). The effort cost in *Equations 13–15* needs to be distinguished from temporal discounting, which may be partly due to another form of cost, opportunity cost, and reduces reward value through a separate mechanism (see *Equations 16-18A*).

A gross simplification may neglect the different curvatures possible with *Equation 15* and apply a common logarithmic utility and disutility function

$$\text{Unet} = \log(\text{income}) - \log(\text{cost}) \qquad (15A)$$

which is equivalent to the benefit-cost ratio

$$\text{Unet} = \log(\text{income/cost}) \qquad (15B)$$

Appropriate behavioral tests would involve separate assessment of income-utility and cost-disutility functions in a common reward value currency, using choices between reward alone [to obtain u (reward)] and reward combined with cost [to obtain d(cost)], similar to composite outcomes involving punishers **(FIGURE 11A)**.

In analogy to temporal discounting (see *Equations 17* and *18* below), alternative effort models used in some neurobiological studies (419, 439) incorporate effort cost into net benefit by exponential discounting

$$\text{Net benefit} = \text{income} * e^{-k*\text{cost}} \qquad (15C)$$

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

to result in a utility measure

$$\text{Unet} = \text{u(income)} * e^{-k*\text{cost}} \quad (15D)$$

or by hyperbolic discounting

$$\text{Net benefit} = \text{income}/(1 + k * \text{cost}) \quad (15E)$$

$$\text{Unet} = \text{u(income)}/(1 + k * \text{cost}) \quad (15F)$$

where k is discounting factor, reflecting subjective cost sensitivity.

The pursuit of any reward reduces the chance of obtaining another reward. This opportunity cost arises with mutually exclusive rewards and is defined as the value of the highest foregone alternative reward. Thus, in principle, *Equations 13-15D* apply to every reward obtained in the presence of a viable, mutually exclusive alternative. Opportunity cost applies also when rewards are delayed and other beneficial options cannot be pursued, which is conceptualized as temporal discounting (see below).

Movements and effort may not only constitute a cost but also have intrinsic reward value (31, 358, 546, 547). A behavior may be undertaken for the pure pleasure of it and without resulting in any other reward. A kid is having fun building an airplane without getting monetary or social benefits. The intrinsic reward value can be considered as income, the required effort (and missed opportunity for homework) is the cost, and *Equation 15* becomes

$$\text{Unet} = \text{u'(intrinsic income)} - \text{d(cost)} \quad (15G)$$

with u' as utility function distinct from u. Intrinsically rewarding behavior may also lead to extrinsic reward. A mechanic loves to rebuild historic car engines and also sells them well, a typical scenario for people that love their jobs. Here, income utility derives from both intrinsic and extrinsic reward value, and *Equation 15* becomes

$$\text{Unet} = \text{u(extrinsic income)} + \text{u'(intrinsic income)} - \text{d(cost)} \quad (15H)$$

Importantly, the added value from intrinsic rewards may outweigh high effort costs. Even if u(extrinsic income) in *Equation 15H* is below d(cost), Unet remains positive, and the behavior viable, if the utility gained from intrinsic reward [u'(intrinsic income)] is sufficiently high. People that love their jobs may be content with unspectacular pay. Alternative formulations may consider intrinsically rewarding behavior as reward in its own right and apply *Equation 25* for multiple rewards (see below). Taken together, effortful behavior may both enhance and reduce net benefit. As the intrinsic gains and costs of behavior are subjective and not deducible from physical movement measures alone, it is particularly important to assess the involved utilities in choices against a reference utility using quantitative psychophysical methods. Simply assuming that movements represent only an economic cost, and estimating its utility from physical rather than behavioral measures, may be misleading.

Behavioral studies on cost follow primarily *Equation 14*. Experiments involve rats crossing barriers in T mazes (499) or working for intracranial self-stimulation (62) and monkeys performing sequential movements (60) or exchanging tokens for reward (69). Tests on humans sampling different numbers of targets comply with *Equation 15B* (108). Binary choices increase with reward value and decrease with effort cost measured as number of lever presses in monkeys (263) and rats (115, 116, 621) and lever movement against resistance in monkeys (233, 419) **(FIGURE 28*B*)**. These choice procedures allow comparison of subjective net benefit between the options analogous to Unet, but without explicitly assessing u(income) and d(cost). Behavioral choices are fit better by the physical cost subtraction model of *Equation 14* than by exponential or hyperbolic cost discounting of *Equations 15C* and *15E* (233). Reaction times increase with increasing effort, well fit by exponential or hyperbolic cost discounting of *Equations 15C* and *15E* (419).

Neurons in anterior cingulate cortex show activity increases with income and decreases with effort cost (233, 263), thus following the components of *Equation 14*. In contrast, orbitofrontal and dorsolateral prefrontal cortex neurons code either income or cost but not both (233). Neurophysiological responses and dopamine release evoked by reward-predicting cues in nucleus accumbens decrease with higher fixed ratio effort (115, 116) **(FIGURE 28*C*)**. Exponential or hyperbolic cost discounting (*Equations 15C and 15E*) affects some dopamine neurons in substantia nigra (419). Thus neuronal responses in reward structures typically reflect the influence of effort cost on reward value.

The dopamine sensitivity to effort is at odds with the failure of finding reductions in voltammetric dopamine reward responses in nucleus accumbens with effort (224, 621). However, the effort in these studies was intercorrelated with temporal discounting, which is well established to reduce reward value and corresponding electrophysiological and voltammetric dopamine responses (116, 158, 285) **(FIGURE 28*D*)**. Thus it is unclear why not at least temporal discounting reduced the dopamine responses. The effort-insensitive voltammetric dopamine changes were slow (lasting >8 s) and one order of magnitude lower than the typical phasic voltammetric dopamine changes [0.5–15 nM (224, 621) versus 30–70 nM (116)], which raises additional, methodological issues. The slower responses resemble dopamine changes of similar durations and magnitudes that have no correlate in dopamine impulse activity (237). Thus the slower, lower, effort insensitive dopamine changes likely do not reflect subjective value (224) but unlikley bear on the

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

role of phasic dopamine signals in economic value coding. Thus the slower, lower, effort insensitive dopamine changes may not reflect subjective value (224) which, however, does not bear on the phasic dopamine coding of all subjective value components tested so far, including reward amount (598), reward type (301), risk (156, 301, 560, 570), delay (116, 158, 285), and effort cost (116) all the way to formal economic utility (560).

6. Temporal discounting

Typical primary nutrient rewards, like fruits or meat, mature and decay over time. Once consumable, they risk destruction from many causes including competition by other reward seekers. Rewards often need to be consumed at particular times due of specific energy demands, and later rewards are less tangible in general. Thus specific rewards in specific situations should be harvested at specific moments. This is why we cannot wait for the pay slip, use refrigerators, drink wine at its best age, and mate only after maturation. These temporal constraints favor temporal sensitivity in the evolution of the brain's reward system and lead to personal, potentially beneficial traits of impatience, impulsivity, and risk taking. The characteristics extend to all rewards and explain the variation of subjective reward value over time. Although temporal discounting likely originates in the perishable nature of goods, it occurs even when the reward remains physically unchanged and overall reward rate is held constant, thus dissociating subjective from objective reward value.

Temporal discounting studies focus mostly on monotonic, forward decreases of value (4, 462, 473). However, value may evolve nonmonotonically in time as birds store food (443), often increasing initially and then decreasing, although the inverse is also possible. The change of reward value over time occurs between a reward-predicting stimulus and reward delivery when no operant action is required (Pavlovian conditioning) or between an action and the resulting reward. In animal learning theory, increasing temporal delays decrease the proximity (contiguity) between stimulus or action and reinforcer. Delays position the reward away from the stimulus or action and more into the background, thus reducing the subjectively perceived reward contingency (dependency of reward on stimulus or action) (175, 220). Thus both subjective devaluation and subjective contingency reduction reduce the power of reinforcers for learning.

Temporal discounting constitutes an opportunity cost when waiting prevents individuals from other activity leading to reward or from using the reward for beneficial investments. Thus temporal discounting can be stated generically as an economic cost, in analogy to *Equation 13*

$$\mathrm{EUnet} = \mathrm{EU} - \mathrm{EUtemp} \qquad (16)$$

Obtaining net benefit Unet requires two steps, assessment of u(m) from objective reward magnitudes and then its temporal discounting (296). Temporal discounting can be implemented by exponential functions, resulting in constant discount rate

$$\mathrm{Unet} = \mathrm{u}(m) * e^{-k*\mathrm{D}} \qquad (17)$$

where k is discount factor. D is subjective delay, as assessed from maximal responses at the subjectively perceived delay (peak interval procedure, Ref. 469). Alternatively, hyperbolic functions show initially higher and then gradually decreasing discounting (5, 218)

$$\mathrm{Unet} = \mathrm{u}(m)/(1 + k * \mathrm{D}) \qquad (18)$$

Hyperbolic discounting is particularly prominent against immediate rewards. When all rewards have longer delays, exponential discounting may describe the process better. The generalized hyperbola combines the two models by adding a free parameter a (325)

$$\mathrm{Unet} = \mathrm{u}(m)/(1 + k * \mathrm{D})^{a/k} \qquad (18A)$$

Unet can be assessed psychophysically with intertemporal choices between a variable early and a constant delayed reward (adjusting-amount procedure). Unet of the late reward is inferred from the utility of the early reward that produces choice indifference. Unet in monkeys decreases monotonically across delays of 2, 4, 8, and 16 s by ~25, 50, and 75%, respectively (**FIGURE 28*D***, blue) (285), with slightly better fits by hyperbolic than exponential discount functions. Temporal discounting decreases from rodents to monkeys to humans.

Some theories of temporal discounting distinguish between a prefrontal system controlling immediate, impulsive consumption and a separate valuation system (352, 538), whereas more straightforward accounts assume that a single reward value system directly discounts value according to *Equations 16-18A*, although the truth might lie in between. Some reward value coding neurons in all reward systems show decreased responses with increasing delays. Responses of monkey dopamine neurons to reward-predicting stimuli decrease monotonically across reward delays of 2–16 s (**FIGURE 28*D***, red), despite constant physical reward magnitude, thus matching closely the behavioral discounting (compare blue and red) (158, 285). Corresponding to human behavior (277), lower reward magnitudes are associated with steeper neuronal discounting (285). The same dopamine neurons code reward magnitude. Temporal discounting may reflect reduced value at the time of the predictive stimulus (input stage) or during the decision process (output stage). The reduced dopamine response occurs in imperative tasks without choices, suggesting value alterations in the input stage. The reduced neuronal response leads to reduced dopamine release in nucleus accumbens (116). Inversely, dopamine prediction error responses to reward delivery increase with longer delays. Both responses are fit better by hyperbolic than exponential func-

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

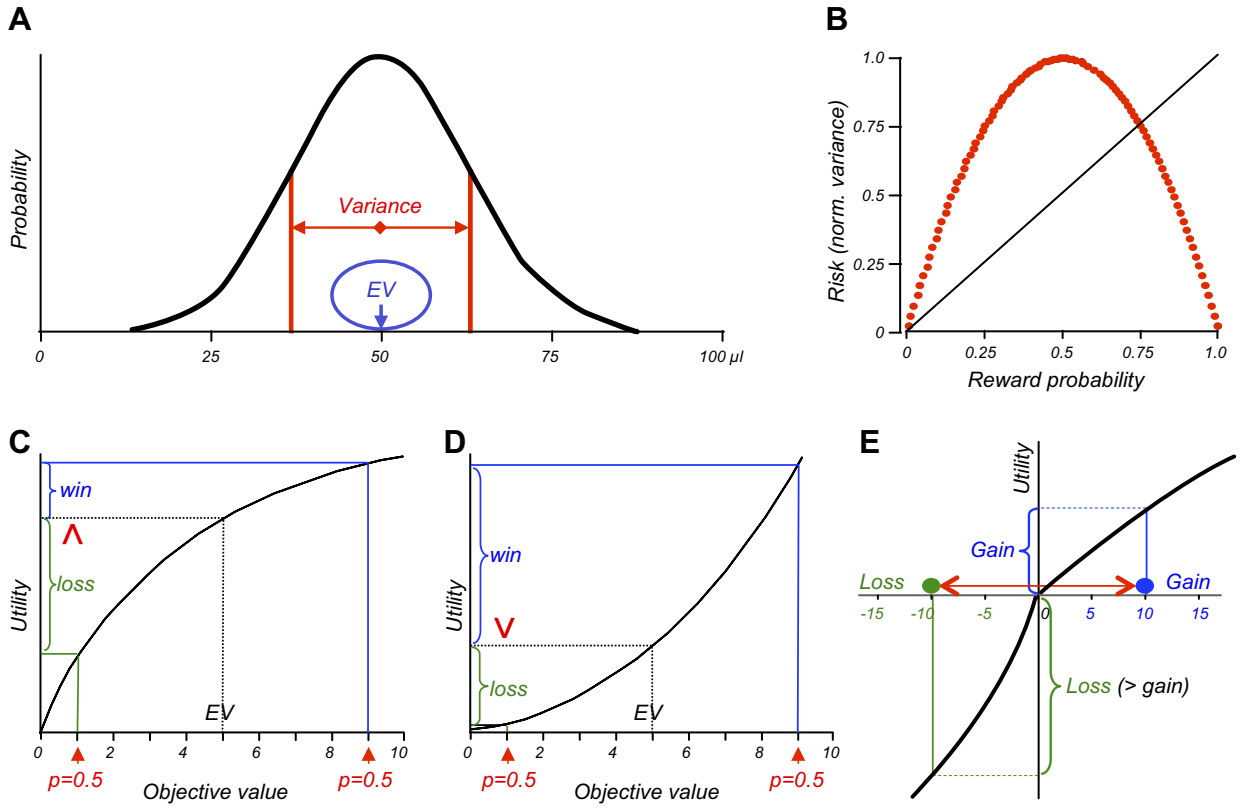


FIGURE 29. Risk and utility functions. *A:* two principal measures in probability distributions (density function). Expected value (EV, first raw moment) denotes value. Variance (second central moment) denotes spread of values and is a good measure for symmetric risk. Red arrows show ± 1 units of standard deviation (SD; square root of variance). Not included are other important risk measures, including informational entropy and higher statistical moments such as skewness and kurtosis. *B:* variance risk (normalized) as a nonmonotonic function of probability. For contrast, dotted line shows monotonic increase of value with probability. *C:* concave utility function associated with risk avoidance. The risky gamble (red, 1 vs. 9, each at $P = 0.5$) induces stronger utility loss when losing the gamble (green) than the gain when winning the gamble (blue) relative to the utility of the expected value of the certain outcome (EV = 5). *D:* convex utility function associated with risk seeking. Gain from gamble win is stronger than loss from gamble losing. *E:* explaining subjective risk notions by gain-loss functions. Due to the steeper loss slope ("loss aversion"), the loss from the low outcome of the binary gamble looms larger than the gain from the high outcome. Hence, risk is often associated with the notion of loss. [Value function from Kahneman and Tversky (258), copyright Econometric Society.]

tions. The almost perfect overlap between behavioral and neuronal discounting **(FIGURE 28*D*)** suggests that temporal delays affect dopamine responses via changes in subjective reward value. Dopamine neurons process delayed rewards as having less value.

Many reward value coding neurons in other reward structures show temporal discounting, including orbitofrontal cortex (478), prefrontal cortex (272, 475), anterior cingulate (233), premotor cortex (475), parietal cortex (477), and dorsal and ventral striatum (79, 115). Reversal of cue-delay associations leads to reversed neuronal responses, suggesting an influence of delay rather than visual properties (475). Often the same neurons code also reward magnitude (476), suggesting common subjective value coding across different rewards. Taken together, neuronal population in all major reward structures code subjective reward value during temporal discounting.

7. Risk

Rewards are, as all environmental events, inherently uncertain. The uncertainty can be formally approached via the description of rewards as probability distributions of values. In contrast to the first statistical moment of probability distributions that defines value (EV, *Equation 8*), the higher moments reflect risk, including variance, skewness, and kurtosis **(FIGURE 29*A*)**. Risk refers to the degree of uncertainty in known probability distributions, whereas ambiguity denotes uncertainty when probability distributions are incompletely known. Risk, and also ambiguity, has at least three functions in reward processing. First, individuals should perceive risk appropriately, irrespective of making a decision at a given moment. Second, as the terms of risk avoidance and risk seeking indicate, risk affects the subjective value of rewards. These two functions, and their processing by reward neurons, will be described in this section.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

Third, risk informs about the spread of reward probability distributions which neurons may take into account when processing reward value within an existing distribution. This function will be described subsequently as adaptive processing.

Popular risk measures are variance

$$\text{var} = \sum_i [p_i(m_i) * (m_i - \text{EV})^2]; \text{ over all } i \quad (19)$$

where P is probability and m is magnitude. Standard deviation (SD), the square root of variance, is also a good and frequently used risk measure. EV and SD fully define Gaussian and equiprobable binary probability distributions. However, the coefficient of variation, CV = SD/EV, is often a better measure for risk in economic choices (633). Other risk measures are Shannon informational entropy (analogous to thermodynamic entropy) that seems appropriate for neuronal information processing systems (477).

Although probabilistic rewards are intrinsically risky, probability is not a monotonic measure for risk but for value (*Equation 8*). Whereas reward value increases monotonically with probability, variance as a measure for risk follows probability as an inverted U function **(FIGURE 29*B*)**. It peaks at $P = 0.5$ where the chance to gain or miss a reward is equal and maximal. Variance risk is lower at all other probabilities with more certain gains and misses.

Risk is perceived subjectively through genuine risk avoiding/seeking tendencies ("I hate/love the risk"), risk distortions ("I have no idea/am confident about the outcome"), the value at which loss/win occurs ("no gambling at high values/peanuts"), the domain, situation, and familiarity of the risky event, and hormones (86, 87, 630, 632). Some of these factors may play a role when defining subjective risk in analogy to *Equation 19* as variance of utility

$$\text{varU} = \sum_i \{\pi[p_i(m_i)] * [\text{u}(m_i) - \text{EU}]^2\}; \text{ over all } i \quad (20)$$

A fuller account would include more of the mentioned subjective factors.

It is important for the understanding of risk attitude to distinguish between objective and subjective value. The famous example is the hungry bird with challenged energy balance (87). The bird requires 100 calories to survive the night and has the choice between a certain option of 90 calories and an equiprobable gamble of 0 and 110 calories ($P = 0.5$ each). The bird would hopefully choose the gamble, as this provides it at least with a 50% chance of survival. However, from the perspective of objective value, the usually risk-avoiding bird seems to be surprisingly risk seeking, as it prefers a risky option with an EV of 55 calories over a certain 90 calories. However, the bird's behavior follows first-order stochastic dominance when we consider survival values (0 vs. 0/1 survival), which tests for meaningful and rational choices (341).

Risk avoidance and risk seeking suggest that risk affects economic choices, which can be visualized via the curvature of the utility function (*Equation 9*). With concave utility **(FIGURE 29*C*)**, the initially steeper slope indicates higher marginal utility in lower ranges. Losses move utility into the lower, steeper range and thus appear subjectively more important than gains which move utility into the upper, flatter range, resulting in risk avoidance. In contrast, with a convex utility function **(FIGURE 29*D*)**, the steeper slope in higher ranges makes gains appear subjectively more important than losses, thus encouraging risk seeking. In some cases, the curvature of utility functions can change, often from risk seeking with low values to risk avoidance with high values. Risk seeking at low values may suggest that the attraction of gambling outweighs potential losses ("peanuts effect") (152, 436, 630).

The subjective value of risk combines with riskless utility to result in the utility of risky choices, as expressed generically

$$\text{EUnet} = \text{EUrl} - \text{EUrisk} \quad (21)$$

with EUrl as riskless utility, and EUrisk as disutility due to risk. This approach recognizes the difference between riskless and risky utility functions (631). Positive EUrisk reduces EUnet and characterizes risk avoidance, whereas negative EUrisk enhances utility and characterizes risk seeking. This approach follows in kind the construction of utility from objective measures of risky outcomes (EV and variance) by the mean-variance approach in financial economics based on Taylor series expansion (238, 315)

$$\text{EU} = \text{EV} - b * \text{variance} \quad (21A)$$

with b as subjective risk weighting coefficient. *Equation 21A* applies, strictly spoken, only to quadratic utility functions.

In addition to subjective risk perceptions, the subjective effect of risk is determined by the curvature of utility functions that are typically concave **(FIGURE 29*C*)** and asymmetric for gains and losses (steeper gain functions, or steeper loss functions inducing "loss aversion") (258). Subjective pessimistic/optimistic probability distortions are further subjective influences ("today is my bad/lucky day"). These subjective factors transform symmetric variance risk into the common asymmetric notion of risk as the danger to lose (higher loss than gain utility) **(FIGURE 29*E*)**.

The most simple and controlled risky outcome is provided by a binary, equiprobable gamble ($P = 0.5$ each outcome) in which risk can be varied without changing the EV (mean-preserving spread) and which has negligible, constant probability distortion and is not skewed (497) **(FIGURE 30*A*)**. Risk attitude can be estimated by eliciting the certainty

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

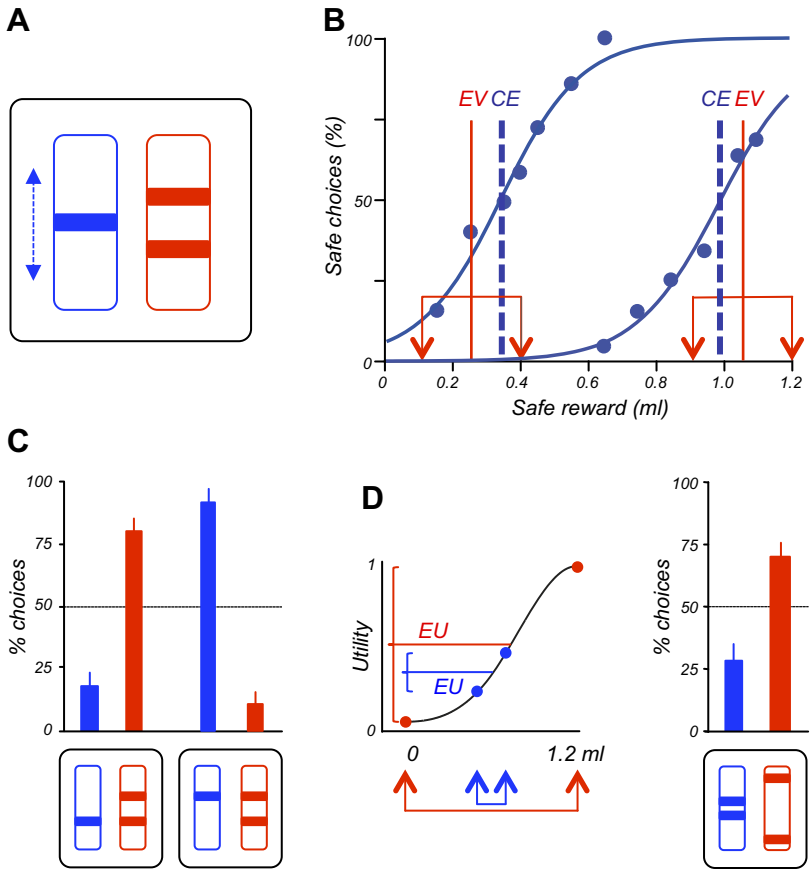


FIGURE 30. Behavioral risk measures in monkeys. *A:* stimuli indicating an adjustable certain (riskless, safe) outcome (blue) and a minimal risky binary gamble (red). Heights of horizontal bars indicate reward magnitude (higher is more). Typically, each gamble outcome occurs with probability $P = 0.5$. *B:* psychophysical assessment of subjective value of binary risky gambles by eliciting the certainty equivalent (CE). The CE is the value of the adjustable certain (riskless, safe) outcome at which choice indifference against the risky gamble is obtained (oculomotor choices in monkeys). A CE exceeding expected value (EV) indicates risk seeking (gamble at *left*), CE < EV indicates risk avoidance (*right*). Red arrows indicate the two outcomes of each gamble. *C:* choices for better options satisfy first-order stochastic dominance in monkey, indicating meaningful choice behavior. In choices against an equiprobable gamble ($P = 0.5$ each outcome), monkeys avoid the low certain reward (set at low gamble outcome, *left*) and prefer the high certain reward (set at high gamble outcome, *right*). Averaged data from binary choices involving four gambles: 0.1–0.4 ml, 0.5–0.8 ml, 0.9–1.2 ml, 0.1–1.2 ml. *D:* choices for riskier options with mean preserving spread satisfy second-order stochastic dominance for risk seeking in monkey, indicating meaningful incorporation of risk into choices. When presented with gambles in the risk-seeking domain of the utility function (left; expected utility, EU, higher for riskier gamble, red), monkeys prefer the riskier over the less risky gamble. [*B–D* from Stauffer et al. (560).]

equivalent (CE) in behavioral choices between an adjustable certain (riskless) outcome and a risky gamble. The CE is defined as the value of the certain option at choice indifference against the risky option. It is lower than EV in risk avoiders and higher in risk seekers. The risk premium reflects the value reduced or added by risk (CE < EV or CE > EV, respectively). The risk premium is usually defined as EV − CE (>0 for risk avoidance, <0 for risk seeking, 0 for risk neutrality), or sometimes as EV/CE (>1 for risk avoidance, <1 for risk seeking, 1 for risk neutrality) (558). Rhesus monkeys are often risk seeking with standard small liquid volumes (CE > EV) and risk avoiding with volumes above 0.5–0.8 ml (CE < EV) **(FIGURE 30*B*)** (355, 560, 644). Their utility function is initially convex, then linear and ultimately concave **(FIGURES 24*D* AND 27*A*)**, thus showing an inflection as conceptualized and found in human utility functions (152, 339, 436, 630).

Formal tests of stochastic dominance (341) confirm that monkeys understand the risky stimuli (560). First-order stochastic dominance helps to define meaningful choices without requiring assessment of a utility function. However, higher EV does not necessarily satisfy first-order stochastic dominance, due to nonlinear utility, nonlinear probability weighting, and risk attitude. In statewise dominance, a case of first-order stochastic dominance, a gamble dominates another gamble if all its outcomes are equal except at least one outcome that is better. Monkeys' behavior satisfies this criterion in choices between a certain reward and an equiprobable, binary gamble whose low or high outcome equals the certain reward. They avoid the low certain reward (**FIGURE 30*C***, *left*) and prefer the high certain reward to the gamble (*right*). Second-order stochastic dominance tests whether risk is incorporated into subjective reward value in a consistent fashion. Monkeys' choices satisfy also this test. With mean preserving spreads, they prefer more risky over less risky gambles in the lower, risk-seeking range of reward volumes **(FIGURE 30*D*)** and show stronger risk preference against certain outcomes (391). Also, monkeys prefer risky over ambiguous options (207). Thus monkeys make meaningful choices under risk consistent with expected utility theory.

The orbitofrontal cortex plays a key role in risky choices (317). Human patients and animals with lesions in orbitofrontal cortex or nucleus accumbens show altered risk and ambiguity sensitivity distinct from value (35, 88, 174, 240, 268, 364, 370, 445). Correspondingly, specific orbitofrontal neurons in monkeys show responses to stimuli predicting binary, equiprobable gambles **(FIGURE 31*A*)** (391). These risk responses occur irrespective of visual stimulus properties and only very rarely reflect reward value, which is coded in separate orbitofrontal neurons. The absence of value coding argues also against the coding of salience, which is associated with both risk and value. Some monkey orbitofrontal neurons are activated by unsigned (absolute)

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

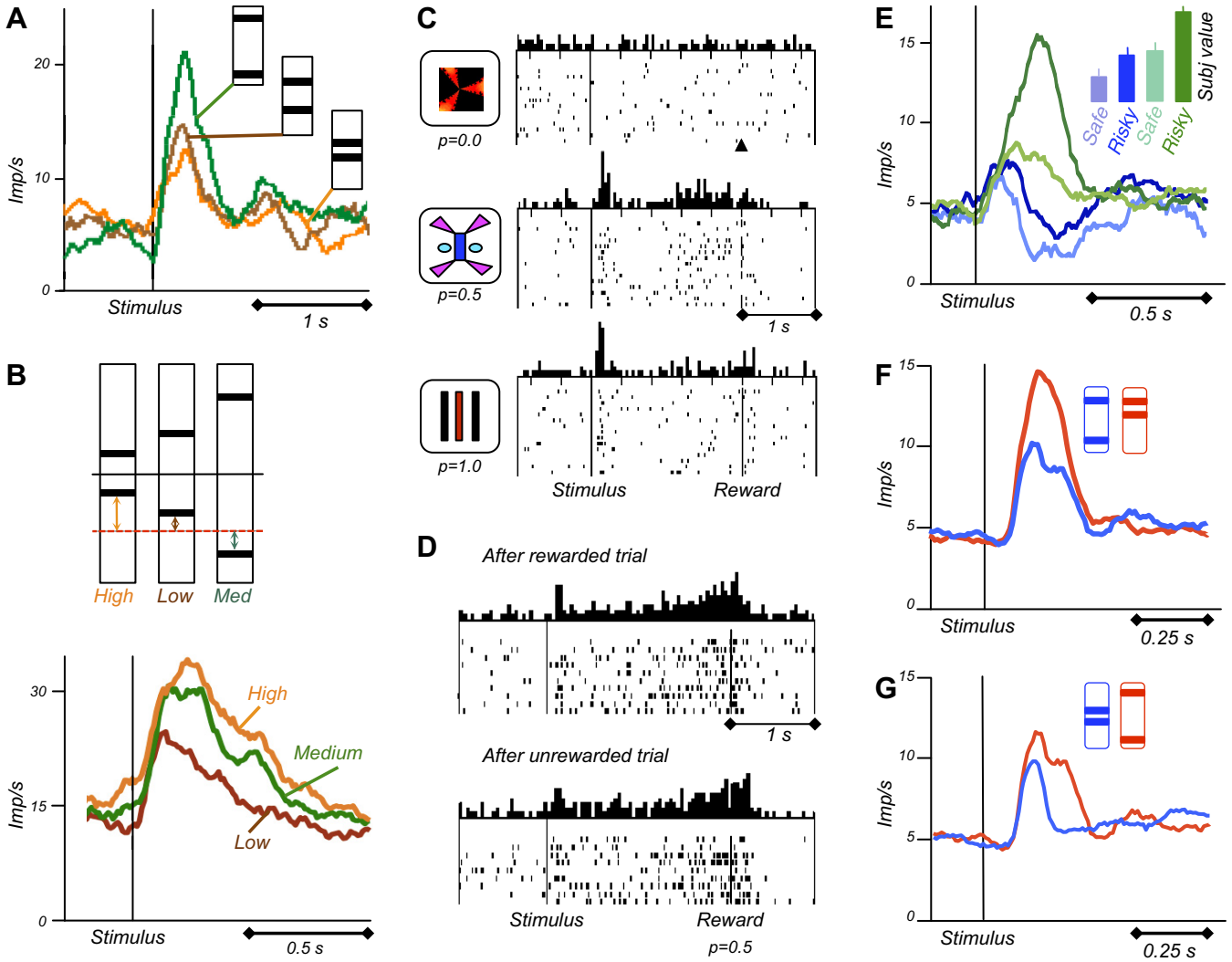


FIGURE 31. Neuronal risk processing in monkeys. *A:* coding of risk in single neuron of orbitofrontal cortex. Height of bars indicates liquid reward volume, two bars within a rectangle indicate an equiprobable gamble ($P = 0.5$ each). The three gambles have same mean but different variance (9, 36, and 144 ml $\times$ 10^{-4}) (mean-preserving spread). This neuron, as most other orbitofrontal risk neurons, failed to code reward value (not shown). [From O'Neill and Schultz (391), with permission from Elsevier.] *B:* coding of risk prediction error in orbitofrontal cortex. *Top*: risk prediction error (colored double arrows), defined as difference between current risk (vertical distance between bars in each gamble indicates standard deviation) and predicted risk (common dotted red line, mean of standard deviations of the three gambles). Colored double vertical arrows indicate unsigned (absolute) prediction errors in variance risk. *Bottom*: averaged population responses from 15 orbitofrontal neurons to unsigned risk error. Such signals may serve for updating risk information. [From O'Neill and Schultz (392).] *C:* risk coding in single dopamine neuron during the late stimulus-reward interval. The activation is maximal with reward probability of $P = 0.5$, thus following the inverted U function of variance risk shown in **FIGURE 29*B***. Trials are off line sorted according to reward probability. [From Fiorillo et al. (161).] *D:* independence of risk activation in single dopamine neuron from reward outcome in preceding trial. Activation is not stronger after positive prediction errors (*top*) than after negative prediction errors (*bottom*). Thus prediction error responses do not backpropagate in small steps from reward to stimulus. Both rewarded and unrewarded current trials are shown in each graph. [From Fiorillo et al. (162).] *E:* influence of risk on subjective value coding in dopamine neurons. The population response is stronger with subjectively higher valued certain (safe) juice rewards (green: blackcurrant; blue: orange; both 0.45 ml), demonstrating value coding. This value response is enhanced with risky juice volumes (equiprobable gamble of 0.3 and 0.6 ml, same mean of 0.45 ml), corresponding to the animal's risk seeking attitude (higher subjective value for risky compared with certain rewards, *inset*). [From Lak et al. (301).] *F:* dopamine responses satisfy first-order stochastic dominance, demonstrating meaningful processing of stimuli and reward value under risk (averages from 52 neurons). In the two gambles (blue, red), the higher outcomes are equal, but the lower red outcome is higher than the lower blue outcome, defining gamble dominance. *G:* dopamine responses satisfy second-order stochastic dominance, demonstrating meaningful incorporation of risk into neuronal utility signal (averages from 52 neurons). The red gamble has greater mean-preserving spread than the blue gamble, defining dominance with risk seeking. [*F* and *G* from Stauffer et al. (560).]

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

risk prediction errors, defined as unsigned difference between current and predicted risk, which may serve for updating risk information **(FIGURE 31*B*)** (392). In a categorical discrimination task, responses in orbitofrontal neurons reflect the riskiness of olfactory stimuli but in addition differentiate between correct and erroneous choices in correlation with decision confidence (264). The orbitofrontal risk signal is distinct from movement-related activity (153), some of which correlates with particular, asymmetric forms of salience (393). In human prefrontal cortex, BOLD signals show direct risk coding, distinct from value (599), which reflects risk perception irrespective of choice. Outside of orbitofrontal cortex, supplementary eye field neurons differentiate categorically between certain and risky trials and code certain and risky reward value differentially (551), and septum neurons show slow risk responses closely resembling the risk ramp of dopamine neurons (373).

Dopamine neurons code reward risk during the stimulus-reward interval after the value prediction error response **(FIGURE 31*C*)** (161). The activation ramps up to the reward and varies with probability in an inverted U function, analogous to risk measures such as variance, standard deviation, and entropy. It increases with variance and standard deviation, even when entropy is kept constant. Thus the signal reflects risk rather than value **(FIGURE 29*B*)**. An analogous signal is also seen in the human ventral striatum, which is distinct from value and most likely reflects synaptic activity induced by the electrophysiologically measured dopamine risk signal (437). Although its latency is too long for immediate decision processes, the dopamine risk signal might affect the subsequent prediction error signal at the time of the reward. In coding risk, the signal varies with surprise salience and thus may affect the learning constant of attentional (associability) learning theories (425, 336). Also, the slower risk signal might contribute to the risk sensitivity of the faster, phasic dopamine signal (see below) (156, 301, 560, 570).

Straightforward implementations of TD models may produce similar ramps by backpropagation of prediction error responses from reward to the preceding conditioned stimulus in small steps via imaginary stimuli (374, 524), thus resembling the dopamine risk response (388). As the backpropagating activation would be small and dispersed, it would appear as ramp only when averaging across multiple trials (183, 388). However, the dopamine risk signal occurs in single trials without averaging and does not backpropagate via imaginary stimuli, as suggested by its absence after prediction errors **(FIGURE 31*D*)** (161, 162), but jump in single steps from reward back to the preceding actual stimulus **(FIGURE 14*A*)** (117, 410, 598). The ramp is not a necessary feature of TD models but occurs only in specific implementations (374, 388, 524). Biologically more plausible TD implementations and spectral timing models use stimulus eligibility traces, as originally suggested **(FIGURE 14*B*)** (574), and model well the single step backward transition of dopamine prediction error responses from reward to the next preceding conditioned stimulus without error backpropagation and thus without a ramp during the stimulus-reward interval **(FIGURE 14, *C* AND *D*)** (71, 410, 573). In particular, models using a biologically consistent striatal and midbrain circuit replicate well the dopamine risk ramp (581). Thus the dopamine ramp represents a genuine risk response and not the necessary byproduct of TD models.

In addition to the late and slow dopamine risk signal, risk affects the phasic dopamine value response. The subjective reward value signal of dopamine neurons is enhanced when monkeys prefer risky over certain outcomes of low or intermediate reward magnitudes **(FIGURE 31*E*)** (156, 301), suggesting incorporation of risk into the final construction of utility in possible compliance with *Equation 21A*. Correspondingly, dopamine concentration changes following risky cues are reduced in risk avoiding rats and enhanced in risk seekers (570). In formal tests informed by economic theory (341, 497), dopamine responses follow first- and second-order stochastic dominance **(FIGURE 31, *F* AND *G*)**, suggesting meaningful coding of economic value and risk and incorporation of risk into neuronal substrates of subjective value. Thus three types of experiment demonstrate that dopamine neurons incorporate risk appropriately into neuronal substrates of subjective value and formal utility, namely, value responses modified by risk in correspondence with risk attitude **(FIGURE 31*E*)**, second-order stochastic dominance between different risky outcomes **(FIGURE 31*G*)**, and processing of utility assessed via CEs across risk seeking and risk avoiding domains **(FIGURE 27*B*)** (560).

Risk influences also subjective value coding in orbitofrontal cortex (444, 482) and modifies saccadic activity in posterior cingulate cortex (355). In human prefrontal cortex, risk affects value responses in close correspondence to individual risk attitudes in analogy to *Equation 21A* (596). Direct risk coding for risk perception and incorporation of risk into subjective value constitute two different processes that allow multiple uses of risk information for behavior.

Taken together, the distinct neuronal coding of value and risk suggest that Pavlovian reward predictions concern probability distributions defined by value and risk. In contrast to reward magnitudes that impact directly on sensory receptors, the coding of EV, probability, and risk of reward is not explained by peripheral sensory stimulation but requires central neuronal mechanisms including memory. Irrespective of these implementation details, the separation of reward signals according to the statistical moments of probability distributions, and the possible correspondence to Taylor series expansion, provides a biological basis for the mathematical treatment of rewards as probability distributions, demonstrates the biological plausibility and imple-

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

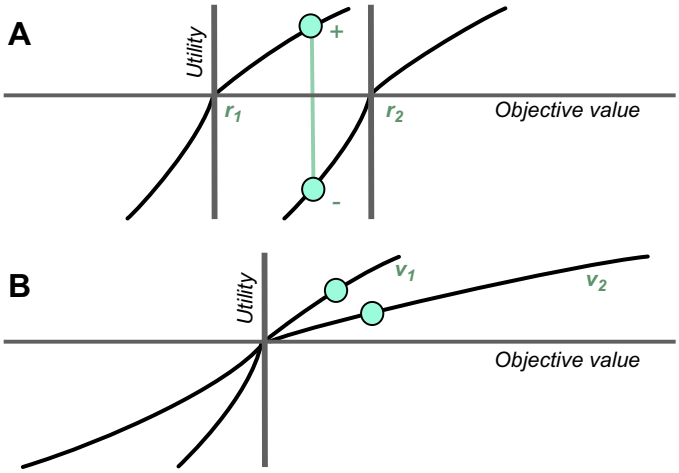


FIGURE 32. Scheme of adaptive reward processing. *A:* adaptation to expected value (r_1, r_2): shift of reference point changes utility gain to loss for identical objective values. *B:* adaptation to variance (v_1, v_2): slope adaptation to variance results in larger utility for lower objective value. [Value function from Kahneman and Tversky (258), copyright Econometric Society.]

mentation of these mathematical concepts, and incorporates them into the neuronal basis of learning theory.

8. Adaptive processing

Whereas the number of neurons and their maximal firing rate limits the processing capacity of the brain, the number of possible rewards is huge. Dedicating equal processing capacity to all possible rewards would lead to low slopes of reward-response functions and thus poor reward discrimination. However, reward availability varies considerably between situations. An efficient discrimination mechanism would restrict neuronal processing to the rewards that are currently available in a given environment, similar to sensory adaptation to probability distributions of ambient visual intensities (148, 235, 307).

For optimal response slopes and thus best use of the limited neuronal coding range, reward processing may adapt to all statistical moments of reward distributions, including EV and variance **(FIGURE 29*A*)**. When schematically separating these two moments, changes in EV lead to processing shifts with maintained response slopes **(FIGURE 32*A*)**, whereas variance affects response slopes, which become steeper with narrower distributions and flatter with wider distributions **(FIGURE 32*B*)**. These adaptations lead to optimal reward discrimination with steepest possible response slopes while maintaining as wide reward coverage as possible.

The consequences of adaptive processing are changes in net utility, which may be expressed generically by adding a term (EUadapt) that summarily covers adaption to all statistical moments of the current probability distribution of objective values

$$\mathrm{EUnet} = \mathrm{EU} - \mathrm{EUadapt} \qquad (22)$$

This approach includes reference-dependent valuation in prospect theory derived from *Equation 10*

$$\mathrm{EU} = \sum_i \{u(m_i - r) * \pi[p_i(m_i)]\};\ \text{over all } i \qquad (22A)$$

with r as reference value (258). The term r is replaced by EVr if reference r derives from a probability distribution rather than being a single value

$$\mathrm{EU} = \sum_i \{u(m_i - \mathrm{EV}r) * \pi[p_i(m_i)]\};\ \text{over all } i \qquad (22B)$$

Equation 22B describes adaptation to the first moment of the probability distribution of objective values. The adaptation to EV can lead to vastly different utilities for identical objective values **(FIGURE 32*A*)**.

A milder approach to reference-dependent valuation combines nonadapted and adapted utility (290). This approach would refine the generic *Equation 22* to

$$\mathrm{EUnet} = \alpha\mathrm{EU} + \beta(\mathrm{EU} - \mathrm{EUadapt});\ \alpha + \beta = 1 \qquad (22C)$$

α and β are additional weighting coefficients. In the terminology of Köszegi and Rabin (290), EU represents consumption utility, and (EU-EUadapt) is reference dependent gain-loss utility. Thus EUnet consists of weighted absolute and relative utilities. Considering nonadapted value recognizes the fact that we know cognitively the absolute reward value, even though we value it relative to references or discount it over time. We know that \$100 is still the same \$100 irrespective of having received \$50 or \$500 the last time around, even though we may feel more elated receiving them as a positive than negative surprise.

Adaptive processing may not be limited to a single reference value and should be extended to full distributions. Inclusion of other statistical moments such as variance adds slope changes. We extend *Equations 10* and *22B* by introducing ω for slope adaptation to variance (or standard deviation)

$$\mathrm{EU} = \sum_i \{u[\varpi * (m_i - \mathrm{EV}r)] * \pi[p_i(m_i)]\};\ \text{over all } i \qquad (22D)$$

A convenient term for ω is 1/standard deviation (384, 438). The adaptation to variance can lead to larger utilities for identical or even lower objective values **(FIGURE 32*B*)**.

Adaptations to reward probability distributions underlie several interesting behavioral phenomena. The negative contrast effect derives from a downshift of reward and reduces behavioral reactions and preferences compared with the same low reward occurring without preceding higher values (51, 104, 593), even when the lower outcome adds value (513). A positive but less prominent contrast effect exists also. The opponent process theory of motivation ac-

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

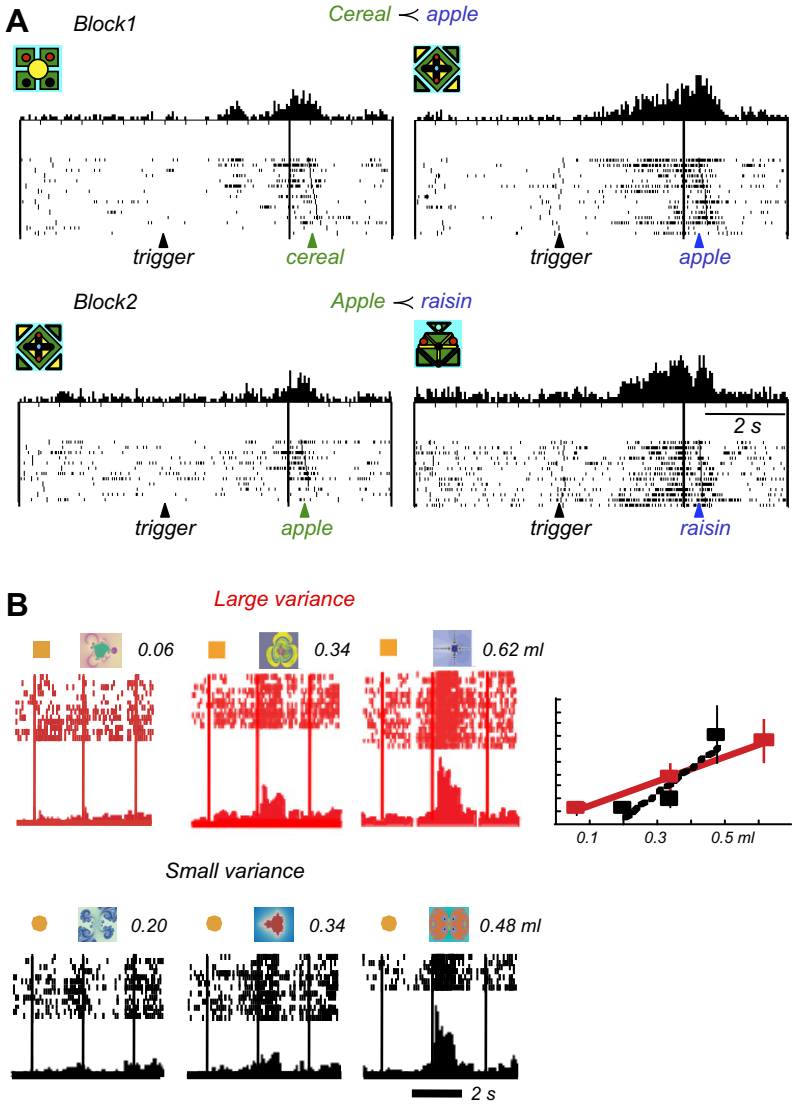


FIGURE 33. Adaptive neuronal reward processing in monkey orbitofrontal cortex. *A:* response adaptation to approximate expected subjective value in single neuron. Trial block 1 offers pieces of cereal or apple, separate block 2 offers same apple (indicated by same stimulus) or raisin. Behavioral choices ranked cereal < apple < raisin. Visual instructions predict type of reward, trigger stimuli elicit arm movement leading to reward. [From Tremblay and Schultz (601).] *B:* response adaptation to variance of reward amount (ml) in single neuron. *Inset* shows change of reward-response regression slope in same neuron. [From Kobayashi et al. (284).]

counts for observations indicating that cessation of reward is aversive, whereas end of punishment is rewarding (552). This effect may well derive from short-term adaptation to the initial reward or punishment, respectively. Adaptations to distributions may also explain the dichotomy in humans between "experienced utility" reflecting utilitarian, hedonic value (39) and "decision utility" measured in overt choices, although overweighting of recent and maximal value is also important (259). This effect is also seen in monkeys (52). The distinction may explain irrational, suboptimal, dominated choices. Monkeys favor [one food morsel plus a 50% chance of a second morsel] over [two food morsels with 50% chance of losing one], although EVs are both 1.5 morsels, although endowment effect and loss aversion may also play a role (91).

Reward responses in dopamine neurons, striatum, amygdala, parietal cortex and orbital, dorsolateral, and cingulate prefrontal cortex adapt to reward magnitude and reward probability distributions as suggested by *Equation 22* (137, 628). More formally, the adaptations may occur to changes in EV **(FIGURE 33*A*)** (42, 106, 232, 601), variance **(FIGURE 33*B*)** (284, 420, 598), or both combined (80, 403). In subtracting reward value from the mean (prediction error) and dividing it by standard deviation, dopamine neurons code a z-score of reward value (598). Divisive normalization from sensory physiology describes well the adaptations to changes in individual values of choice options in reward neurons of parietal cortex (332). The ventral striatum is importantly involved, as lesions here reduce reward contrast in rats (326). Taken together, the adaptations would be driven by neuronal EV and risk signals and result in a match between the probability distribution of neuronal responses and the probability distribution of current reward values. The result would be dynamically varying, optimal reward response slopes and thus efficient reward processing for choices. The loss of the actual values would be immaterial in choices comparing the options relative to each other. Thus, in adapting to the two key moments of reward probability distributions, reward neurons make ef-

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

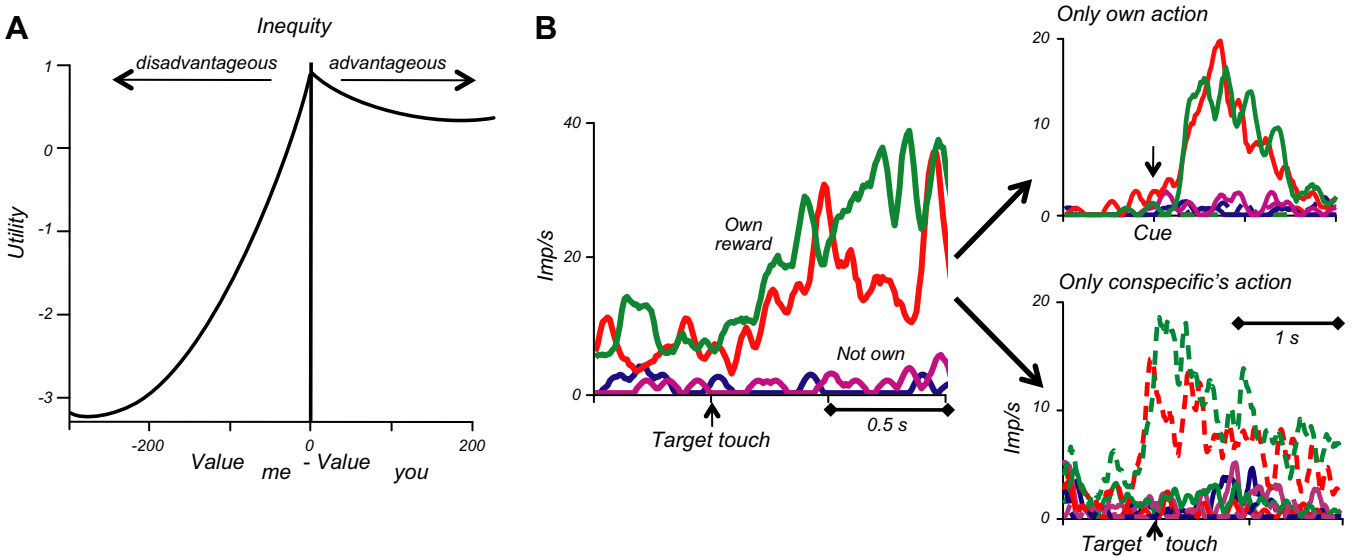


FIGURE 34. Social reward processing. *A:* social reward inequity aversion. Difference between own and other's reward reduces my utility. Usually, disadvantageous inequity (getting less than others, negative difference, *left*) has stronger and more consistent effects than advantageous inequity (getting more than others, *right*). [From Loewenstein et al. (324).] *B:* action-dependent coding of social reward in single striatal neurons. In the imperative reward giving task (modified dictator game), two monkeys sit opposite each other across a touch-sensitive computer monitor and give reward to itself and the conspecific. *Left*: neuronal activation when receiving own reward, either only to itself (red) or together with conspecific (green), but no activation with reward only to conspecific (violet) or to nobody (blue). *Right*: activation with own reward occurs only with own action (*top*, continuous lines) or only with conspecific's action (*bottom*, dotted lines, different neurons between *top* and *bottom*). [From Báez-Mendoza et al. (25).]

ficient use of processing resources for optimal reward discrimination and choice.

The EU shift between environments and the slope adaptation may result in lower utility of higher absolute reward values **(FIGURE 32*B*)**. Thus adaptation may lead to violation of rank order, menu invariance, independence from irrelevant alternatives, and transitivity, which are crucial for meaningful economic choices. However, at least two processes may serve to maintain transitivity in the presence of adaptive valuation. First, the mentioned predictive EU and risk signals in dopamine neurons (598) and orbitofrontal cortex (391) convey information about the two key moments of the reward probability distribution of each environment. These signals may combine with the adapted value signal to generate an accurate, absolute, nonadapted value signal and thus maintain transitivity at the neuronal level. Second, adaptations require time and may only concern subpopulations of reward neurons. Indeed, only a fraction of orbitofrontal reward neurons show adaptation, whereas the remainder codes reward magnitude irrespective of other rewards (403, 406, 284). Although longer test periods might have increased the number of adapting neurons, some orbitofrontal reward neurons seem to code nonadapted, absolute value necessary for maintaining transitivity. Adaptation and absolute value coding by different neuronal populations would combine efficacy with transitivity.

9. Social rewards

Observation of social partners and comparison of rewards between them is crucial for competition and cooperation that improve performance and give individuals access to otherwise unobtainable resources and ultimately enhance fitness. Social factors and contexts have at least two distinct reward functions.

First, viewing and encountering others is a reward on its own. Rhesus monkeys sacrifice quantities of liquid reward for viewing body parts of conspecifics, demonstrating that social images have reward value (124). Neurons in monkey parietal cortex code the reward value of conspecific's faces and hindquarters (279), although orbitofrontal and striatal neurons code these social stimuli often separately from liquid rewards (629).

Second, reward to others affects own reward processing. With benevolent exceptions ("Mother Teresa" and many other mothers), individuals hate to receive less reward than others when all other factors are equal. Disadvantageous inequity aversion typically induces a sharp, gradually declining, drop in the utility of own reward **(FIGURE 34*A*)** (324). Emotional correlates are envy and distaste for unequal or unfair outcomes. Individuals often also dislike, usually to a lesser extent, receiving more reward than others. This advantageous inequity aversion may elicit a similar

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

but weaker utility drop. Associated emotions are compassion, sense of fairness, guilt, dislike of inequity, positive reciprocity, group welfare, "warm glow," and maintaining personal social image (83, 149). The aversion may turn into advantageous inequity seeking (getting more than others) with competition and other advantage seeking situations associated with pride, sense of achievement, superiority feelings, and retaliation. Good behavioral tools for assessing inequity are the ultimatum game in which rejection of unfair offers assesses disadvantageous inequity aversion, and the dictator game in which the reward fraction handed to the opponent reflects advantageous inequity aversion (83, 149).

Final own expected utility in social settings EUnet can be stated generically as

$$\mathrm{EUnet} = \mathrm{EU} - \mathrm{EUsocial} \quad (23)$$

EU stands not only for income utility (*Equation 10*) but also for any utility EUnet derived from reward components (*Equations 11, 12, 13, 18, 21*, and *22*). EUsocial is derived from the difference between own value x and value x_o received by the opponent(s), weighted by disadvantageous (α) and advantageous (β) inequity coefficients in analogy to a popular model (150), resulting in

$$\mathrm{EUnet} = \mathrm{EU} - \alpha * \max\left[(x_0 - x), 0\right] - \beta * \max\left[(x - x_0), 0\right] \quad (23A)$$

where α and β are disadvantageous and advantageous inequity coefficients, respectively. Alternative accounts focus on reciprocal "fairness weights" which enhance own reward value when being nice to a nice opponent and reduce own value when being mean to a mean opponent, thus stabilizing cooperation and disfavoring unsocial behavior (442). Both models suggest that reward to others has appreciable value for the observer. Although condensing social reward value into a common value variable is tempting and would help to conceptualize decisions, nonvalue factors such as strategy may importantly influence decisions (50, 150, 151).

Chimpanzees and rhesus, cebus, and marmoset monkeys show prosocial tendencies and actively observe rewards in conspecifics (77, 123, 359). Cebus show disadvantageous but no advantageous inequity aversion (69). Rhesus prefer receiving own reward alone over both animals receiving the same reward, suggesting advantageous inequity seeking (negative β of *Equation 23*), and they prefer seeing conspecifics receive reward over nobody receiving reward, suggesting disadvantageous inequity seeking (negative α) in this setting (23, 89). Disadvantageous inequity reduces the effort cebus are willing to spend (612). Many of these social effects may depend on the relative social positions of the monkeys within their group and would merit further investigation. These social tendencies seem simplistic compared with those of humans but may constitute basic building blocks for social encounters and exchanges.

Reward neurons in monkey orbitofrontal cortex and striatum process primarily own rewards in social settings, whereas neurons in anterior cingulate cortex distinguish between own and other's reward or sense primarily conspecific's reward (**FIGURE 34*B***, *left*) (23, 25, 89). Striatal reward neurons distinguish between the social agents whose action leads to own reward (**FIGURE 34*B***, *right*) (25). These neurons are only active when either own action or the other's action results in own reward, and many of them do not show this activity with a computer instead of a monkey opponent. Neurons in medial frontal cortex show activations during observation of behavioral errors of conspecifics, sometimes together with observation of reward omission (647). Some premotor mirror neurons distinguish between reward and no reward irrespective of who is receiving it (78). These neuronal signals mediate the distinction between own and others' rewards and actions and thus convey basic components of social reward function. When testing social competition, reward neurons in prefrontal cortex differentiate between competitive and noncompetitive video game performance (234). In humans, disadvantageous inequity reduces, and advantageous inequity activates, ventral striatal reward responses (165), reflecting α and β of *Equation 23A* (the coding direction may reflect different signs of the two coefficients or different neuronal coding slopes). Taken together, social reward neurons code fundamental components of competition and cooperation.

10. Multiple reward attributes

Single rewards come seldom with a single attribute. Attributes include income utility, cost, delay, risk, and social interaction quantified by *Equations 9-23A*. Their aggregate subjective reward value can be condensed into a single variable called total utility. This has advantages. First, extracting the variable once from the multiple reward attributes and then using it for choices would save the decision maker computing time in crucial moments and thus enhance evolutionary fitness, even if occasional attribute changes require partial recalculations. Second, the brain could transform the condensed neuronal value responses into signals for specific decision variables for competitive decision mechanisms such as winner-take-all (see below).

Total utility EUtotal of a single reward with multiple attributes can be derived from a combination of attributes, each with their own expected utility derived from likely nonlinear and possibly nonmonotonic utility functions

$$\mathrm{EUtotal} = \mathrm{EU} + f_i * \mathrm{EUstate} + f_2 * \mathrm{EUcxt} - f_3 * \mathrm{EUcost} - f_4 * \mathrm{EUdiscount} - f_5 * \mathrm{EUrisk} - f_6 * \mathrm{EUref} - f_7 * \mathrm{EUsocial} + \mathrm{INTER} \quad (24)$$

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

where f_1 to f_7 are weighting coefficients reflecting the different contributions of each reward attribute. In some cases, the attributes can be so closely linked to individual rewards that the total utility u(total) of a reward can be directly derived from adding or subtracting their utility u(att) to the income utility u(m)

$$\mathrm{u(total)} = \mathrm{u}(m) - \mathrm{u(att)} \quad (24A)$$

Establishing a full probability distribution from u(total) of all rewards would allow computation of EU with *Equations 9* or *10* that amounts to the same EUtotal as stated in *Equation 24*.

Take the example of a hamburger in a pub and apply *Equations 10, 11, 12, 13, 18, 21, 22*, and *23*. Its expected utility EU is defined by the probability distribution of the utilities of the slightly varying hamburger sizes u(burger). My hunger enhances EU by EUstate, and the familiar pub context associated with food and drink adds to the burger's attraction with EUcxt. However, the burger does not come for free (EUcost) and takes a moment to cook because of a busy kitchen, subjecting it to temporal discounting (EUdiscount). Burger quality varies, to which I am averse, thus reducing its EU by EUrisk, although it is still the best food in the pub, which puts its EU above a mean reference and adds EUref (in a gourmet restaurant it would have lower rank and thus lose EUref). When the burger finally arrives, it is smaller than my neighbor's and thus elicits disadvantageous inequity aversion, but better than my dieting other neighbor's salad, inducing mild advantageous inequity aversion, which both reduce the burger's EU by EUsocial. Some of these attributes are known to interact which is captured summarily by INTER. For example, smaller rewards are discounted steeper (lower EUdiscount with smaller EU) (277) and larger stakes increase risk aversion (lower EUrisk with larger EU) (152, 630).

11. Multiple rewards

Most rewards are composed of several individual reward objects whose values add up and may interact. Our hamburger meal is composed of the burger with its calories and taste, a tomato salad with taste and appealing visual aspects, and a pint of ale with water, calories, taste, and alcohol, thus totaling eight individual rewards. However, ale is sometimes poorly pulled (not in my pub), which would constitute an aversive component that would need to be subtracted from the positive components to obtain a summed final reward. If utilities of several rewards are assessed separately on a common scale, they can be weighted and added up to a summed utility, with possible interaction (INTER)

$$\mathrm{EUsum} = \sum_i (f_i * \mathrm{EUtotal}_i) + \mathrm{INTER}; \quad i = 1,\ n \text{ rewards} \quad (25)$$

with f_i as weighting coefficients. Each of the rewards is likely to have its own saturation dynamics. INTER models multisensory integration as interaction that can make aversive taste or odor in some, usually small quantities add to reward value, rather than subtract from it, like pepper on meat, spicy mustard on sausage, or quinine in sugary fizzy drinks.

Equation 25 can be used for assessing the value of an otherwise unquantifiable reward on a scale defined by a common reference reward. For example, in choices between [juice plus picture] versus [juice alone], the value of the picture is reflected by the sacrificed amount of its associated juice to obtain choice indifference (124). Punishers can be included into *Equation 25* where they reduce EUsum. They constitute negative value but not economic cost, as they do not reflect energetic, temporal, or monetary expenditures. Behavioral choices of monkeys reveal the value summed from rewards and punishers compatible with *Equation 25* **(FIGURE 11*A*)** (160).

With the specification of final utility EUsum, the reinforcement *Equations 10A–D* can be restated as

$$\mathrm{UPEsum}(t) = \lambda(t) - \mathrm{EUsum}(t) \quad (26)$$

$$\mathrm{EUsum}(t+1) = \mathrm{EUsum}(t) + \alpha * \mathrm{UPEsum}(t) \quad (26A)$$

$$\mathrm{UTDPEsum}(t) = \left[\lambda(t) + \gamma \sum \mathrm{EUsum}(t)\right] - \mathrm{EU}(t-1) \quad (26B)$$

$$\mathrm{EUsum}(t+1) = \mathrm{EUsum}(t) + \alpha * \mathrm{UTDPEsum}(t) \quad (26C)$$

Analogous definitions can be made for utility prediction errors of multiple attributes of single rewards by replacing EUsum by EUtotal of *Equation 24*. With multiple rewards and reward attributes, summed, single prediction errors of *Equations 26* and *26B* would not require separate computations of utility prediction error for every reward and attribute. Such prediction errors can be used by economic decision mechanisms for straightforward EU updating by standard reinforcement rules.

Alternatively, prediction errors may be computed from separate, possibly multi-attribute, rewards that are added up in a weighted manner, in analogy to *Equation 25*

$$\mathrm{UPEsum}(t) = \sum_i \left[f_i * \mathrm{UPEtotal}_i(t)\right] + \mathrm{INTER}; \ i = 1,\ n \text{ rewards} \quad (26D)$$

$$\mathrm{UTDPEsum}(t) = \sum_i \left[f_i * \mathrm{UTDPEtotal}_i(t)\right] + \mathrm{INTER}; \ i = 1,\ n \text{ rewards} \quad (26E)$$

A similar calculation is used for incorporating fictive outcomes into gamble values (329). This computation takes the prediction error of each reward into account. An analogous breakdown of prediction error computation, and hence updating, can be made for individual reward components listed in *Equation 24*.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

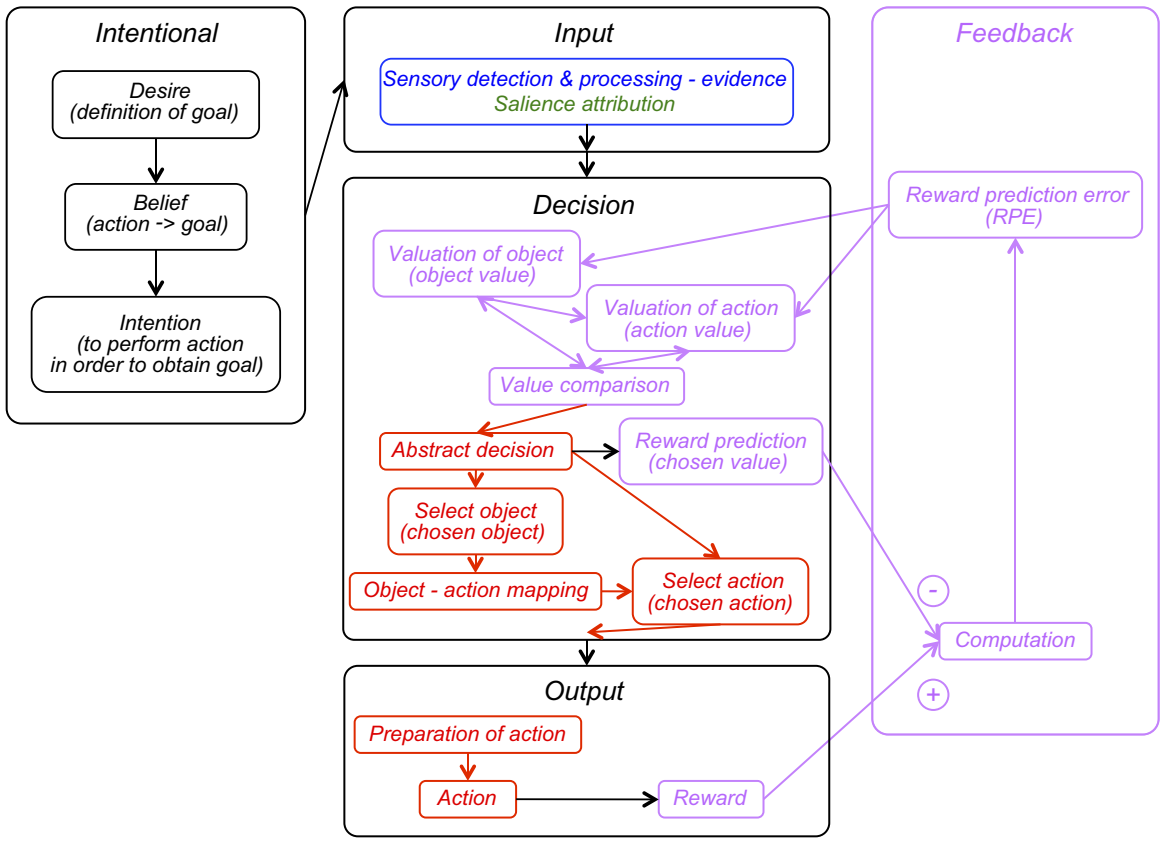


FIGURE 35. Steps and components of voluntary economic decisions. Five component processes occur in sequence, and to some extent in parallel depending on the characteristics of the decision. Decision variables are indicated in central box below their respective mediating processes (object value, action value, chosen value, chosen object, chosen action).

Given that the dopamine prediction error signal codes utility (560), the dopamine responses (*Equations 10E* and *10H*) can be restated for multiple rewards by replacing EU by EUsum

$$\text{DaResp}(t) \sim \lambda(t) - \text{EUsum}(t) \qquad (26F)$$

and for the temporal difference error

$$\text{DaResp}(t) \sim [\lambda(t) + \gamma \sum \text{EUsum}(t)] - \text{EUsum}(t-1) \qquad (26G)$$

DaResp in *Equations 26F* and *26G* would apply also to utility error signals in other brain structures.

Common scale value coding is a prerequisite for applying *Equation 25* to neuronal signals and is seen in orbitofrontal cortex with different liquid rewards (405) and images of body parts (629), in parietal cortex with body part images (279), and in dopamine neurons with different liquid and food rewards (301) **(FIGURE 23, *B* AND *D*)**. Neuronal signals reflect the sum of positive and negative values from liquid or food rewards and aversive liquids or air puff punisher in monkey dopamine neurons **(FIGURE 11*D*)** (157, 160) and anterior cingulate cortex (10), and from odor rewards in human orbitofrontal cortex (190). Compatible with coding negative value as component for *Equation 25*, some frontal cortical neurons are depressed by losses of gained rewards (532). Ideally, neuronal investigations would test EUsum with all constituent components, although this is impractical and experimenters usually investigate only a few of them at a time.

D. Economic Decision Mechanisms

1. Component processes of decisions

In a highly simplified scheme, economic decisions may involve five component processes **(FIGURE 35)**. The initial intentional component comprises the desire of obtaining a reward (*left*). This emotion is based on the knowledge, or at least a hunch, that there is a reward that is worthwhile to pursue. With a belief that an action can be attributed to the reward, an intention can be formed to perform an action to get the reward. Thus the reward becomes the goal of the intentional action. In the practicalities of laboratory experiments, these processes can be driven by explicit, temporally specific stimuli indicating the availability of rewards or arise spontaneously by processes internal to the decision maker based on environmental information. The desire to get a

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

beer may be driven externally by the sight of a pub or internally by the heat of a summer day.

Once the intentional stage has been completed, the inputs to the decision mechanism are processed (**FIGURE 35,** *top*). Perceptual decisions require stimulus detection and then evidence accumulation for identifying the sensory reward component, tracking the object and assessing its associated salience components (**FIGURE 1,** *left* and *middle*). Correct identification of each reward option is the basis for choosing the best one. For example, perceptual decisions identify motion direction from the degree of dot motion coherence; the percentage of correct identification translates into reward probability (134, 389).

The perceptual stage is followed by valuation of the individual reward objects (**FIGURE 1,** *right*) or, depending on the characteristics of the decision, by valuation of the actions required to obtain each reward object. Valuation comprises the immediate identification or gradual accumulation of value evidence, similar to the accumulation of sensory evidence in perceptual decisions. The process may either parallel the gradual identification of the object or action or start only after they have been completely identified. The decision stage (**FIGURE 35,** *middle*) begins with object and action valuation and comprises the selection of the best reward by comparing instantaneously or gradually the values of the available objects or actions. Valuation and value comparison may evolve partly in parallel (double pink arrows). Purely sensory and motor aspects no longer play a role, unless they affect value, such as effort cost (*Equation 13*). Following valuation and value comparison, the abstract decision specifies the result of the process. It concerns only the decision itself and leads to three selections: the winning object, the winning action, and the prediction of the reward value extracted from the winning object values and action values. If the decision primarily specifies the winning object, subsequent object-action mapping determines the action required to obtain the reward object.

Obtaining the selected object and performing the selected action requires motor decisions for preparing and executing the necessary action (**FIGURE 35,** *bottom*), as every decision will ultimately lead to an action for acquiring the chosen reward. The quality of motor decisions determines the efficacy of the economic decision process. Decision makers want to get the action right to get the best reward. In cases in which different possible actions have the same cost and lead to the same reward, motor decisions may become independent of economic decisions. In these cases, the choice of the particular action may depend on noise or on motor exploration. Apart from equivaluable actions, it is unclear how much motor decisions can be independent of economic consequences. If actions are chosen for their own sake, they have intrinsic reward value and thus are economic decisions. After experiencing the selected reward object, its value is compared with the value predicted during the decision process (**FIGURE 35,** *right*). The resulting prediction error is conveyed to the valuation process in the input stage where it serves to update object values and action values.

The scheme of **FIGURE 35** should not suggest strictly sequential processing of decision variables that transitions from sensory identification via value assessment to decision-making. Real life situations are often characterized by noisy evidence, unstable values, moving objects, and changing actions and thus are likely to engage several of these processes in parallel. Furthermore, dopamine, striatum, amygdala, frontal, and parietal neurons show not only hugely different but also considerably overlapping reward valuation and decision activity. Several arguments suggest nonsequential processing of evidence, value, and decisions. First, dorsolateral frontal and posterior parietal neurons are involved in spatial movement processes but also show considerable reward value coding **(FIGURES 20*E* AND 21)** (313, 381, 433, 627). Second, ramping activity for evidence accumulation is the hallmark of cortical perceptual decisions but occurs also in subcortical structures during reward expectation and valuation (18, 40, 215, 221, 260, 523). Third, abstract decision coding and chosen action coding overlaps with reward valuation in subcortical structures, including amygdala (188) and striatum (79, 275, 306). Although the different processes may involve different neuronal pools within the same brain structures, the data nevertheless question a strictly sequential processing scheme and suggest more interactive and overlapping processing of evidence, value, and decision-making within the different structures and through subcortical loops.

In addition to many nonvalue factors that determine economic choices (see above), even purely value-based decisions may not involve the complete or correct consideration, computation, or comparison of values. A first important deviation from proper value processing is due to noise. Noise can be derived from uncertain sensory evidence (experimentally modeled, for example, by partly coherent random dot motion), varying salience attribution, imprecise estimation of economic value (economic noise, for example, when estimating probabilities from changing frequencies, or when deriving value from these probabilities), and noisy learning constants inducing unstable value predictions. Except for sensory evidence, which is external to the brain, the noise results from neuronal noise in sensory receptors and subsequently engaged neurons. Ways to deal with noisy information are averaging out the noise and accumulating the evidence over time. A second deviation from proper value processing occurs with exploration. Options that were previously suboptimal might have improved in the meantime and should better be recognized to prevent missing the best options and ensure utility maximization in the long run. Exploration is captured by ε-greedy and softmax functions (575). A single parameter models the proportion

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

of exploration and dominated choices (higher ε and softmax "temperature" for more exploration). Thus utility maximization involves a healthy mix of exploitation (choosing known best options) and exploration (choice of dominated but potentially better options). The proportion of exploration should be fine tuned to the dynamics of the environment to result in maximal reward. It should be higher in volatile (higher ε or temperature) compared with more stable situations (lower ε or temperature). A third deviation from proper value processing occurs with satisfying when individuals fail to consider all available options and choose only from a subset of many options that seem to satisfy their immediate desires (85, 545). Individuals may stop searching for higher value and thus make suboptimal choices. However, in some cases, the cost of considering all options may be high and reduce substantially the net benefit (*Equations 13–15*), in which cases satisfying may constitute the optimal behavior. A fourth deviation from proper value processing may be due to adaptive value processing that includes inaccessible and thus irrelevant options. In these cases, irrelevant alternatives do affect option values and may make less valuable options appear more valuable and direct choices towards them. The common result of the various forms of incomplete value consideration may be selection of suboptimal, dominated options whose values are lower than those of the best option, a behavior that compromises utility maximization.

2. Decision variables

A decision variable is what the decision maker tries to control and use to achieve the goal. It is the crucial independent variable driving the decision process. The basic decision variable in economic choices is subjective value and, more specifically, formal economic utility. The basic assumption in all economic decision theories is that decision makers aim to maximize utility. Utility is derived from several attributes including reward magnitude, probability, motivational state, context, cost, delay, risk, reference, and social interaction, as stated in *Equations 9–25* above. This notion of decision variable differs from the more technical definition that includes all sources of inputs to decision rules and mechanisms (185).

An additional definition of decision variables derives from the specific decision mechanism that serves to control them. Decision variables should be appropriate for this mechanism. Economic decision variables constitute the inputs to the decision mechanism and reflect the value of each option irrespective of the actual choice (**FIGURE 35,** *top*). The value is expressed as formal economic utility. It concerns specifically the value of a reward object (object value) and the value obtained by an action (action value). Although the decision process cannot change what is on offer, it maximizes the value by selecting the best object or action from the existing inputs (**FIGURE 35,** *middle*). This process requires sensory evidence for identification of each option as prerequisite for subsequent valuation of the options. Sensory evidence itself is not a decision variable in economic decisions, as it is imposed by the environment, and controlling it would not necessarily lead to the best reward. Furthermore, decision variables are closely related to the outputs of the decision process. The outputs controlled by the decision maker concern the reward value predicted from the decision (chosen value), the object that conveys that value (chosen object), and the selection of the valuable action (chosen action). Preparation, initiation, and execution of the action for obtaining that value are downstream processes (**FIGURE 35,** *middle* and *bottom*).

The value in the decision variables is rarely innate but acquired and updated using various forms of learning, in particular Pavlovian processes, operant habit learning, and operant goal-directed behavior, as described above. Through these three forms, objects and actions acquire formal economic utility, using model-free reinforcement learning as most basic mechanism, as stated in *Equations 10B, 10E, 26A*, and *26C*. More complex learning and updating of utility may use various forms of model-based reinforcement learning.

Perceptual decision models, which are crucial for the sensory identification of reward objects (**FIGURES 1,** *left*, and **35**, *top*), assume the gradual accumulation of evidence that needs to be extracted from noisy stimuli. Two main types of perceptual decision models vary in the way the inputs are compared with each other. Random walk, drift, or diffusion models have a single integrator with separate decision thresholds specific for each option **(FIGURE 36*A*)**. Evidences are continuously compared between options through mutual inhibition at the input level (254, 302, 447, 454, 549, 565). The resulting difference signal grows and diverges over time towards one of the decision thresholds. The decision is executed for the option whose threshold is reached first. In contrast, race models have distinct accumulators for each option, each one with its own decision threshold. The evidences grow separately until the first option reaches a specific threshold, which then generates the corresponding behavioral choice **(FIGURE 36*B*)** (327, 614, 622). Race models are appropriate for choices involving several options that are tracked separately until a final comparison can be made. Intermediate forms between diffusion and race models differ in architecture and parameter setting (55). The valuation of each option (**FIGURES 1,** ***right*****, and 35,** ***top***) may occur as soon as the perceptual decisions have been made and each reward object has been identified. In other cases, the value may not be immediately apparent and requires some time to evolve into a stable representation. Only then can an economic decision variable arise. Thus a general formulation of economic decision mechanisms should include accumulation processes for sensory and value evidence.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

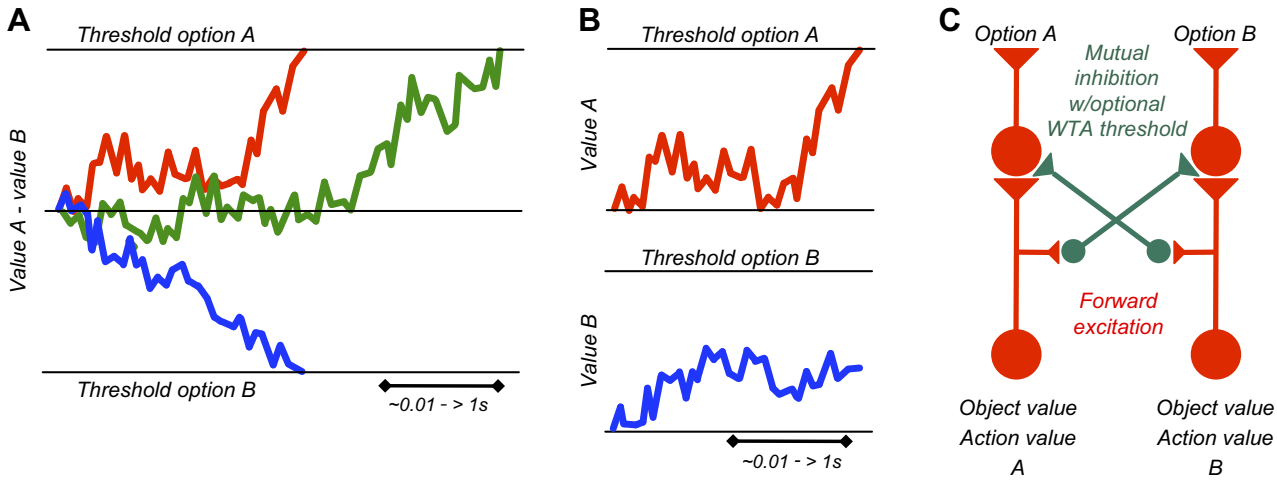


FIGURE 36. Schematics of decision mechanisms. *A:* diffusion model of perceptual decisions. Difference signals for sensory evidence or value increase in a single integrator towards a specific threshold for one or the opposite choice option (red, blue). Later threshold acquisition results in later decision (green). Time basis varies according to the nature of the decision and the difficulty in acquiring sensory evidence and valuing the options. *B:* race model of perceptual decisions. Separate signals for each choice option increase in separate integrators towards specific thresholds. The option whose value signal reaches the threshold first (red) will be pursued, whereas the lower signal at this point loses (blue). *C:* basic competition mechanism. Separate inputs for each choice option (*A*, *B*) compete with each other through lateral inhibition. Feedforward excitation (red) enhances stronger options, and mutual inhibition (green) reduces weaker options even more, thus enhancing the contrast between options. This competition mechanism defines object value and action value as input decision variables. For a winner-take-all (WTA) version, a threshold cuts off the weaker of the resulting signals and makes the stronger option the only survivor.

The basic mechanism of the decision process is competition between independent options and underlies a large variety of decision models (55, 56, 441, 585, 623). It would mediate the on-going comparison and divergence of evidence in the diffusion models **(FIGURE 36*A*)**, achieve the final comparison between options in the race models **(FIGURE 36*B*)**, or constitute a decision mechanism on its own when evidence is immediate and accumulation negligible. The minimal model comprises two inputs that map onto two outputs while inhibiting each other's influence **(FIGURE 36*C*)**. Through forward excitation coupled with mutual lateral inhibition, the stronger neuronal signal becomes even more dominant by not only inhibiting more the weaker signal but also by being less inhibited by the weaker input. The mechanism amplifies the graded differences between the inputs, analogous to lateral inhibition of sensory systems, and, with various alterations, forms the basis for competition in a wide variety of diffusion and race models for perceptual and motor decisions (55, 56, 441, 585, 622, 623) and may also apply to economic decisions. Recurrent excitation would generate gradually increasing evidence accumulation in components of the network (622, 623). An additional threshold produces a winner-take-all (WTA) mechanism by removing the weaker signal and turning the graded difference into an all-or-none output signal that reflects only the value of the strongest option or consists of an all-or-none decision signal for the winning option. This comparative mechanism provides definitions for the major decision variables.

Neuronal decision signals should reflect the decision variables defined by the decision mechanisms. The art is to identify formal, biologically plausible decision models that employ the mechanisms defining the decision variables and are implemented with identifiable neuronal activities during various trial epochs **(FIGURE 20, *A* AND *D*)**. The activities concern the initial accumulation of evidence, implemented by gradually increasing ramping activity in perceptual decisions (**FIGURES 35, *top*, and 36, *A* AND *B***), and the subsequent valuation of the options. The central competitive process and its outputs (**FIGURES 35,** *middle* and *bottom*, and 36*C*) are reflected in distinct neuronal activities coding the input decision variables of object value (405) and action value (500), the abstract decision output (188), and the chosen value (405), chosen object (38), and chosen action (543). Updating of decision variables via prediction errors occurs globally via dopamine neurons and locally through specific groups of non-dopamine reward neurons (**FIGURE 35,** *right*). Neuronal decision signals are typically investigated in choice tasks, although basic processing aspects may be studied in imperative, forced-choice trials, which allow to assign neuronal coding to individual choice options and to predictions from a single stimulus. The following sections are devoted to these descriptions.

3. Evidence and valuation

The initial stage of the decision process concerns the rapid or gradual acquisition of evidence and the valuation of the

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

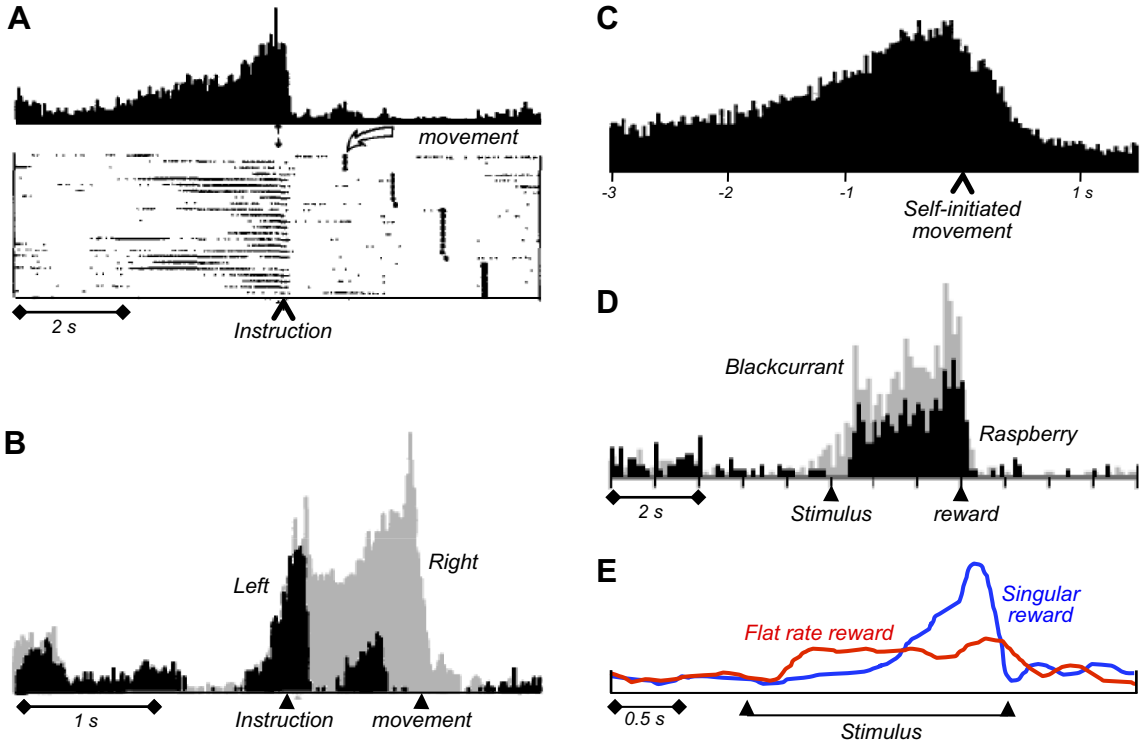


FIGURE 37. Neuronal ramping activity preceding stimuli, action, and reward. *A:* gradual activity increase during the expectation of an initial visual instruction stimulus in premotor cortex neuron. Open arrow points to movement triggering stimulus. [From Mauritz and Wise (350), with kind permission from Springer Science and Business Media.] *B:* gradual increase of left-right differential activity during movement preparation in premotor cortex neuron. Activity terminates just after the movement triggering stimulus and before the movement. Left (black) and right (gray) refer to target positions. [From Kurata and Wise (298).] *C:* ramping activity preceding self-initiated movements in striatum neuron. Activity begins gradually without external imperative stimuli and terminates with movement onset. [From Schultz and Romo (528), with kind permission from Springer Science and Business Media.] *D:* differential reward expectation activity in striatum neuron. Activity is lower in anticipation of raspberry juice (black) than blackcurrant juice (gray). [From Hassani et al. (205).] *E:* reward expectation activity in amygdala neuron reflecting instantaneous reward probability. Activity ramps up steeply before singular reward occurring at predictable time after stimulus onset (increasing reward expectation), but shows lesser increase before reward occurring pseudorandomly during stimulus (flat reward expectation with flat occurrence rate). [From Bermudez et al. (40).]

identified options (**FIGURE 35**, *top*). Acquisition of evidence takes time, in particular when the evidence itself is noisy. Neurons in many brain structures have the capability to bridge time gaps and accumulate information through persistent activity that may ramp up to future events. Ramping activity is one of the major neuronal mechanisms underlying decisions. This form of neuronal activity concerns sensory information in perceptual decisions, motor processes in action decisions, and economic utility with all its components in economic decisions (*Equations 9–13, 16*, and *21–25*).

Ramping activity reflects a general propensity for persistent, gradually increasing activity in many cortical and subcortical neurons. It occurs without choice during several processes. During the expectation of sensory events, activity increases gradually toward predictable visual stimuli in premotor cortex (**FIGURE 37*A***) (350), prefrontal cortex (65), parietal cortex (248), and striatum (6, 18, 215, 526, 528). During the instructed preparation of eye and arm movements, activity ramps up and differentiates between movements in primary motor cortex (467), premotor cortex (**FIGURE 37*B***) (298, 491, 493, 636), frontal eye field (74, 506) [and its projection neurons to superior colliculus (147)], supplementary eye field (101, 506), supplementary motor area (491, 493), parietal cortex (109, 184), superior colliculus (379), and striatum (**FIGURE 20*F***) (526, 528). Ramping activity to a threshold occurs also several seconds in advance of internally chosen, self-initiated movements in neurons of premotor cortex, supplementary motor area, parietal cortex, and striatum; these neurons are typically not activated with externally triggered movements (**FIGURE 37*C***) (299, 311, 337, 380, 397, 491, 493, 526, 528). Similar human encephalographic ramping activity resembling the readiness potential is the basis for an accumulator decision model for movement initiation (515). During reward expectation, ramping activity occurs in orbitofrontal cortex (216, 544, 602), dorsal and ventral striatum (**FIGURES 32*C* AND 37*D***) (18, 205, 215, 523), and amygdala (**FIGURE 37*E***) (40). The anticipatory ramping activity seen with sensory events,

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

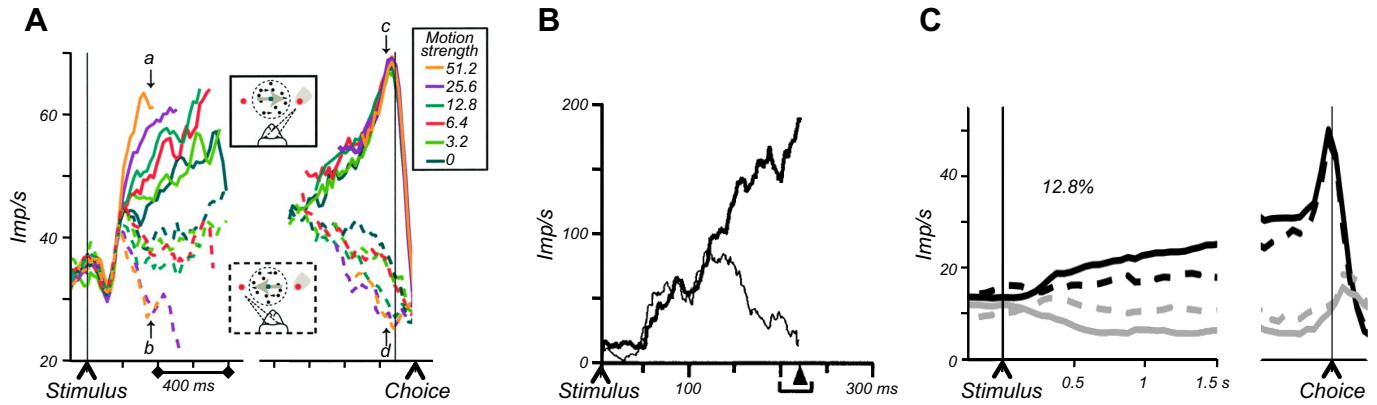


FIGURE 38. Neuronal ramping activity during perceptual decisions. *A:* ramping activity during dot motion discrimination in monkey lateral intraparietal cortex (average from 54 neurons). Continuous and dotted lines indicate saccades into and out of neuronal response field, respectively. Motion strength indicates the percentage of coherently moving dots, and thus the facility to discriminate between their direction (out of two possible directions). The buildup is steeper with higher motion coherence but ends at same height at the time of choice. Stimulus marks onset of search array (target and distractors); choice marks saccade onset. [From Roitman and Shadlen (483).] *B:* ramping activity during visual search in a monkey frontal eye field neuron. Heavy and thin lines indicate target position inside and out of neuronal response field, respectively. Stimulus marks onset of search array (target and distractors); filled arrow marks saccade onset. [From Thompson et al. (588).] *C:* average ramping activity from 104 neurons in monkey lateral intraparietal cortex reflects subjective perception of dot motion direction rather than objective motion direction (12.8% coherently moving dots). Correct trials (full line) and error trials (dashed) differentiate between saccades into and out of neuronal response field (black and gray), irrespective of actual dot motion direction. [From Shadlen and Newsome (535).]

actions, and reward may constitute a basic neuronal mechanism for perceptual decisions, economic decisions, and internal movement generation.

During perceptual decisions, primary sensory systems code evidence about physical stimulus parameters irrespective of the subject's perception and decision (120), whereas activity in secondary sensory cortex and premotor cortex reflects the animal's subjective perception (120, 121). Ramping activity occurs over several seconds in neurons in parietal cortex, frontal cortex, and superior colliculus during the gradual accumulation of noisy evidence and reaches a criterion threshold at the moment of choice. Ramping activity is absent when sensory evidence is immediate and thus may not simply reflect neuronal processing (334). Ramping activity occurs during dot motion discrimination in prefrontal cortex, lateral intraparietal cortex (LIP), and superior colliculus **(FIGURE 38*A*)** (230, 274, 483, 535); length and brightness discrimination in superior colliculus (448, 450); and visual search and visual stimulus identification in frontal eye field neurons **(FIGURE 38*B*)** (49, 587, 588). Although typically tested with two choice options, ramping activity in LIP occurs also with four options (97). The slope of the ramp buildup to criterion threshold varies in LIP and superior colliculus with the speed and difficulty of the decision derived from the coherence of moving dots **(FIGURE 38*A*)** (483), differences in reaction time (97, 270, 448, 483), and differently rewarded options (494). The height of activity at choice threshold is the same irrespective of sensory coherence and motor response time in LIP **(FIGURE 38*A*)** (483). The statistical variance of neuronal activity in LIP increases gradually and declines earlier for one option compared with the other option immediately before the action (96). The probability of accurately predicting the animal's behavioral choice from neuronal activity in LIP increases during the ramp (535), suggesting that intraparietal activity is suitable for affecting subsequent neurons involved in expressing the decision through action. Accordingly, ramping activity in prefrontal neurons changes its direction while monkeys change their decision (269). Choice accuracy depends on the starting point of the neuronal ramp but not its threshold in LIP (199). The ramping activity in perceptual choice tasks does not correlate with objective sensory stimulus parameters but reflects the subjective perception expressed by the animal's choice, as shown with errors in direction judgement in LIP neurons **(FIGURE 38*C*)** (270, 535). Ramping activity may not always reflect motor processes, as it occurs in the frontal eye field even without action (587, 588) and in LIP correlates with motion strength evidence but not action (535).

During action decisions, even when straightforward sensory signals do not require lengthy accumulation of information, ramping activity in frontal eye fields builds up to a threshold for movement initiation before saccadic choices (198) and before correctly countermanded saccades **(FIGURE 39*A*)** (507). The rate of neuronal build-up determines saccadic reaction time **(FIGURE 39*B*)** (198, 508). Neuronal activity in cingulate motor area, frontal eye fields, and LIP ramps up before monkeys change movements to restore declining reward **(FIGURE 39, *C* AND *D*)** (101, 543). During delayed action instruction, neurons in premotor cortex

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

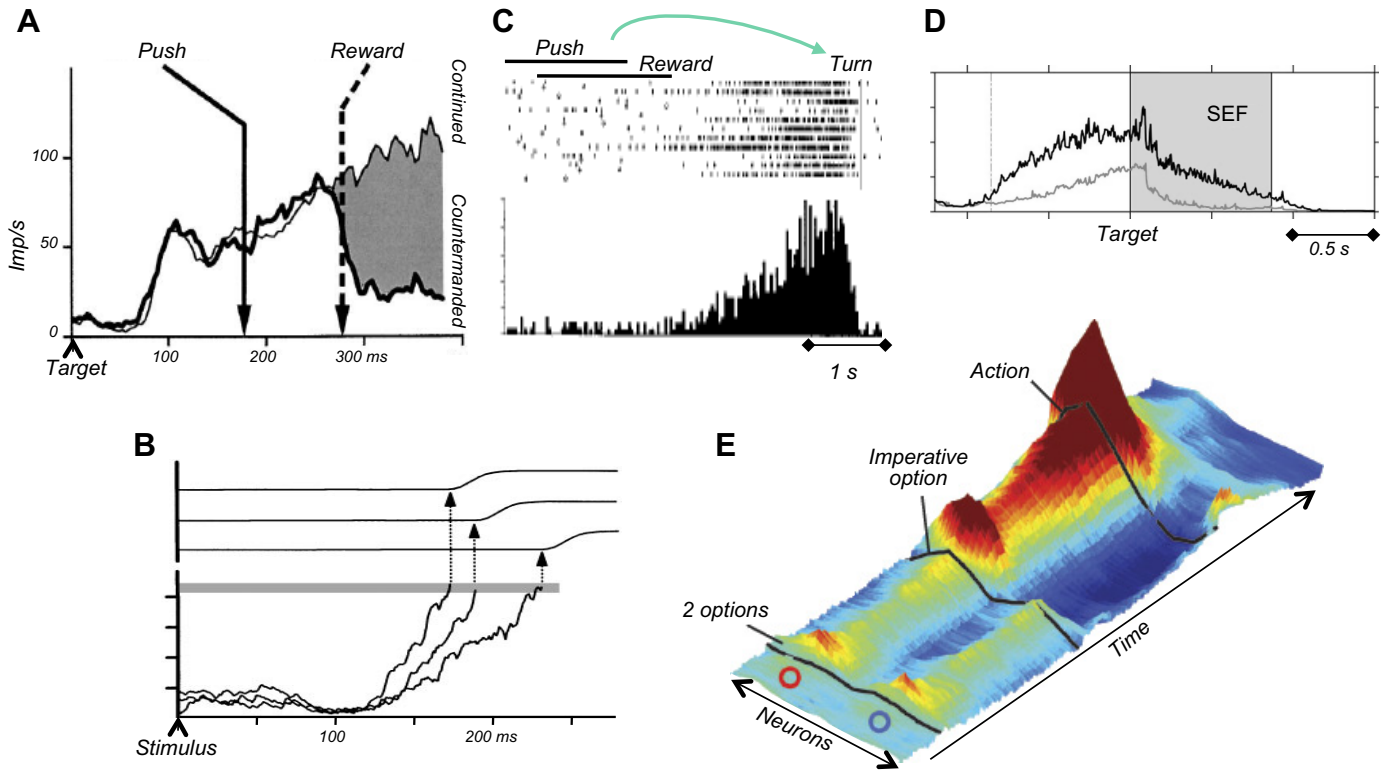


FIGURE 39. Neuronal ramping activity during action decisions. *A:* ramping activity during eye movement countermanding in a monkey frontal eye field neuron. Thin and heavy lines indicate continued and correctly stopped (countermanded) saccades, respectively. [From Schall et al. (507), with permission from Elsevier.] *B:* steeper ramping activity associated with earlier saccade choice in a monkey frontal eye field neuron. Neuronal activity has reached the same height at time of choice irrespective of slope. [From Schall and Thompson (508). Copyright 1999, Annual Reviews.] *C:* ramping activity during action decisions in monkey cingulate motor area. Upon decrease of reward amount, the monkey selects an alternate arm movement (turn) after the previous movement (push), thus restoring full reward amount. [From Shima and Tanji (543). Reprinted with permission from AAAS.] *D:* ramping activity preceding choice target reflects future choice in monkey supplementary eye field neuron. Black, free choice into neuronal response field; gray, opposite choice. [From Coe et al. (101).] *E:* differentiation of action coding in monkey premotor cortex. Neuronal population activity is initially segregated according to each action (red and blue circles) but becomes selective for final action after an imperative cue instructs the action. [From Cisek and Kalaska (99). Copyright 2010, Annual Reviews.]

show initial nondifferential activity that becomes stronger, selective, and ramps up for the action once the target is revealed **(FIGURE 39*E*)** (98). As these activities occur in motor areas, reaching the threshold may reflect the completed decision process or the initiation of action.

With economic decisions between differently rewarded options, neurons in LIP show ramping activity with noisy reward options (494). This activity likely reflects the accumulation of sensory evidence before the reward valuation rather than constituting the valuation process itself. However, LIP neurons show ramping activity also with immediately evident, non-noisy information about the reward options, such as differently colored lights at specific spatial positions **(FIGURE 40)** (571). The neuronal activity scales with the economic decision variable of fractional income and predicts the chosen action. Neurons in anterior cingulate cortex show ramping build-up towards a single threshold when animals leave a foraging patch with decaying reward value (209). Steeper ramp slopes reflect shorter reaction times for leaving the patch, as the overt choice is initiated when the common threshold is reached. These ramping activities cannot reflect noisy evidence but may constitute components of the neuronal decision process in the transition from sensory processing to reward valuation. They might also reflect associated processes such as stimulus or reward expectation or movement preparation. Nevertheless, ramping activity is not a necessary feature for economic decision activity in frontal cortex neurons (404), although its absence in a studied neuronal pool would not contradict ramping activity in other, unrecorded neurons, as cortical neurons show vastly inhomogeneous task relationships. In a probabilistic choice task with unambiguous visual symbols, parietal cortex activity reflects the economic decision variable of log likelihood ratio of probabilistic rewards (643). A potential ramp is due to averaging of asynchronous, nonramping activity. Without outright decisions, responses of amygdala and dopamine neurons reflect the

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

transition from initial sensory processing to reward valuation **(FIGURE 9)** (159, 422), which is considerably prolonged when the sensory evidence is noisy (389). Neurophysiological dopamine responses show a ramp with reward risk **(FIGURE 31, *C* AND *D*)** (161) but not with reward value. Striatal dopamine release ramps up before reward (237) which likely reflects presynaptic influences from reward expectation related cortical or amygdala inputs to striatum (40, 216, 544, 602) or from striatal neurons (18, 205, 215, 523) (Figures 32C and **37, *D* and *E***), but not a dopamine impulse ramp which does not occur with riskless rewards.

The morphology of ramping activity corresponds to the number of thresholds per neuron in the different decision models (56). The two-threshold drift-diffusion model assumes neuronal activity that ramps up differentially towards upper and lower thresholds for respective options **(FIGURE 36*A*)** (254, 302, 447, 565). Such activity is seen during perceptual decisions **(FIGURE 38, *A* AND *C*)** (97, 449, 483, 535) and economic decisions **(FIGURE 40)** (571). The observed gradual divergence of ramping activity in the same neuron may represent the output of an ongoing graded comparison with mutual inhibition in competing presynaptic neurons **(FIGURE 36*C*)**. The single-threshold type of decision model conceptualizes a gradual divergence of ramping activity in the same neuron. The ramp either reaches the threshold for a specific option or fades back to baseline for the alternative option **(FIGURES 36*B*)**. Such activity would also form the basis for race models in which activities from different pools of differentially active neurons race toward their individual thresholds (609, 614). Differential ramping activity is seen during the expectation of sensory events **(FIGURE 37*A*)** (6, 18, 65, 215, 248, 350, 526, 528), motor preparation **(FIGURE 37*B*)** (74, 98, 101, 109, 147, 184, 298, 299, 311, 337, 379, 380, 397, 467, 491, 493, 506, 526, 528, 636), reward expectation **(FIGURE 37, *D* AND *E*)** (18, 40, 205, 215, 216, 523, 544, 602), perceptual decisions **(FIGURE 38, *A* AND *B*)** (274, 483, 588), motor decisions **(FIGURE 39*A*)** (198, 507, 508, 587), and economic decisions **(FIGURE 40)** (571). The final step to the choice in these models may involve a graded or WTA competition between different neurons **(FIGURE 36*C*)**.

Taken together, neuronal ramping activities constitute key mechanisms in the identification of choice options from sensory evidence and may partly underlie their valuation. They reflect both the evidence that the decision maker cannot influence and the decision variables controlled by the decision maker. Ramping activities resemble statistical decision processes for accumulating stochastic information about sensory evidence (185) and rewards and follow major assumptions of decision models **(FIGURE 36, *A* AND *B*)**. The activities are widespread in the brain and reflect internal processes related to expectation of events, decisions between options, and preparation of action. As they occur also without choices, they may represent a general feature of neuronal physiology that is used for decision-making. Gradually increasing activity provides a focusing influence on downstream neurons through greater signal strength and higher signal-to-noise ratio, allowing better selection of behavioral actions **(FIGURE 36*C*)**.

4. Object value

Object value denotes the value of a specific reward object. It derives intuitively from nutrient or other survival values of rewards that are sought by humans and animals for survival. In human societies, object value is fundamental for economic exchanges of goods in which a quantity of one good with a specific value is traded against another good with a specific value. Object value is the basis for the marginal rate of substitution in economics which is the amount of *good A* a consumer is willing to give up to obtain one unit of *good B*, utility remaining constant across the exchange (341). To choose optimally, we assess the utility of each object separately, compare between them, and then select the best one. The value updating after each experienced outcome involves reinforcement learning (*Equations 10B, 10E, 26A*, and *26C*) with Pavlovian, habit, and goal-directed processes. Thus object value is defined as an input variable for competitive decision mechanisms (**FIGURES 35,** *top*, and **36*C***), in analogy to action value conceptualized in machine learning (575).

Object value has three crucial characteristics. First, it varies with the values of a reward object and, in principle, should be quantified as formal economic utility (*Equations 9–12*). Second, the value refers to one specific choice object. When choosing between several reward objects, such as a piece of meat and a loaf of bread, each object has its own and inde-

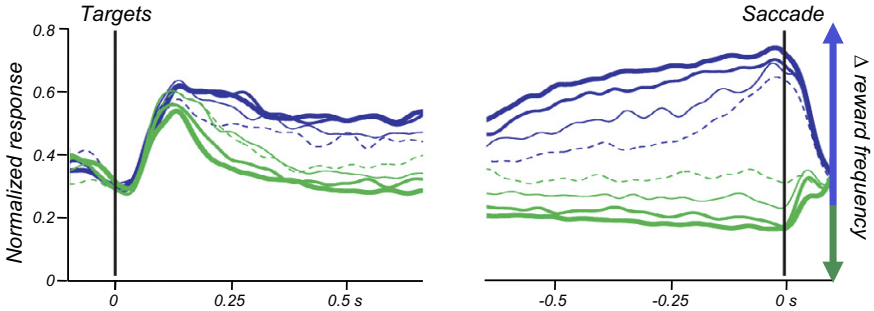


FIGURE 40. Neuronal ramping activity during economic decisions. Ramping activity in monkey lateral intraparietal cortex neurons during oculomotor matching choice between more and less frequently rewarded options (average from 43 neurons). Activity reflects the "fractional income" (ratio of chosen reward value to summed value of all options). Blue, saccade into neuronal response field; green, opposite movement. Fractional income varies between 0 (dotted lines) and 1.0 (think lines). [From Sugrue et al. (571). Reprinted with permission from AAAS.]

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

pendent value. "Objects" are solid or liquid rewards indicated by their appearance (color of fruit or fluid) and specific, intrinsically arbitrary stimuli associated with rewards. Different containers of the same kind of milk are examples of stimulus objects. Without being outright physical objects, events, situations and activities endowed with value have functions analogous to objects for economic decision mechanisms. The objects gain value through the animal's actual experience with the reward or from inferences based on non-Bayesian or Bayesian models of the world. Third, object value is independent of the actual choice of the object. It indicates the value of a good irrespective of whether we will imminently receive or consume it, thus complying with the notion of reward information, rather than explicit reward prediction. Economic exchanges are based on value comparisons irrespective of immediate consumption. A pound of meat has an object value of \$15 irrespective of whether I buy it now or not. Without this property, object values cannot serve as independent inputs to comparative decision mechanisms. Thus minimal tests for object value require two different reward objects, two values separately for each object, and independence from object choice.

Object value neurons process the value of one particular object irrespective of the animal's action. There would be grape juice neurons tracking the value of grape juice irrespective of the animal's choice, or banana neurons indicating how much the monkey will get if it were to climb the tree. The separate neurons tracking value for the specific objects constitute inputs for competitive decision models. In a binary decision, each input neuron, or pool of neurons, coding the value of its specific object competes with the input neurons coding the value of the other object. Hybrid subforms may show more pronounced object value coding if one particular object or action is chosen (combination with chosen object or chosen action coding defined below). To make meaningful comparisons, their signals need to vary on a common scale, both between the rewards ("common currency") and between the neurons. Neither sensory differences between rewards nor different coding ranges or slopes between neurons should affect object value signals. The model can be easily extended to three and more options as long as each neuron codes only the value of a single object. Thus object value coding on a common scale appears to constitute a suitable neuronal decision signal for competitive decision mechanisms.

Neurons tracking the value of specific reward objects irrespective of reward choice or reception are found in orbitofrontal cortex. They code selectively the value of a specific fruit juice irrespective of the monkey's choice and the required action [named "offer" value (405)] **(FIGURE 41, *A* AND *B*)** and thus do not predict the choice (404). These activities comply with the competitive formalism that requires independent coding of object value at the input **(FIGURE 36*C*)**. Orbitofrontal object value coding remains selective with three rewards (406), conforming coding specificity for a particular object. Impaired object-reward associations during choices in orbitofrontal lesioned humans and monkeys underline an orbitofrontal role in object value coding (84, 108).

In contrast, most conventional reward responses reflect explicit reward prediction or reception and thus do not have the necessary property of coding object value irrespective of the actual choice, without which they could not serve as independent inputs to competitive decision mechanisms. They reflect the actual, free or forced, behavioral choice and probably code forms of chosen value and reward prediction (see below).

5. Action value

All choices lead to movements. Whether they are ocular saccades, licking, reaching, locomotion, or speech, they all consist of muscle contractions performed to get the best value for the effort. Even verbally expressed choices, and speaking itself, involve muscle contractions (of the larynx, mouth, and hands). Machine learning and reinforcement learning, which are concerned with artificial agents moving about the world, use action value to model the agent's behavior for maximizing reward (575), like robots deciding which car part to weld first. Animals deciding between foraging patches are likely more interested in the value the action can bring them rather than considering each patch as an object. Each action leads to a reward value if it were chosen, thus the shorthand term of action value. Thus, for maximizing utility, an agent selects the action by which they achieve the highest estimated value. Pavlovian, habit, and goal-directed processes provide mechanisms for setting up and updating action value which, in principle, should be quantified as formal economic utility (*Equations 9–25*).

Actions are not only means for obtaining reward but may be themselves pleasurable and rewarding. The intrinsic reward may be added to the action value that usually refers only to the reward obtained from the action. The subjective nature of intrinsic action value emphasizes the need for quantitative behavioral assessments of often nonlinear and even nonmonotonic subjective value functions rather than inferring value from objective reward and action properties.

Like object value, action value is defined as input variable for competitive decision mechanisms (**FIGURES 35,** *top*, and **36*C***). Similar to object value, it requires three conditions. First, action value reflects reward value (amount, probability, EV, EU; *Equations 9–13*, *16*, and *21–25*). Second, value coding is specific for a particular action. The value is attached to an action and varies only for that action, such as an arm movement to the right or the left (spatial differential), an arm or an eye movement (effector differential), or going to work or vacation. Action value may arise from operant conditioning following the animal's actual experi-

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

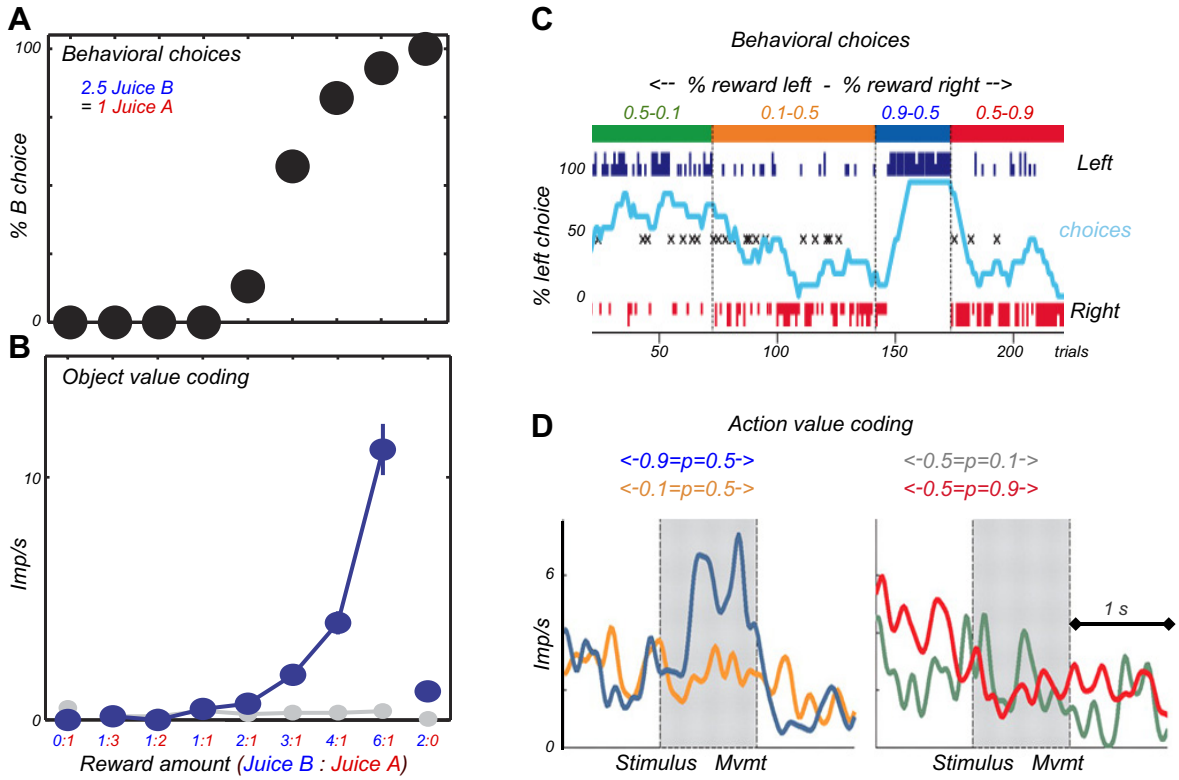


FIGURE 41. Neuronal coding of economic input decision variables for competitive mechanisms. *A:* behavioral choices between two juices that vary in magnitude. Increasing the amount of juice B (blue) increases the frequency of the monkey choosing that juice. The estimated indifference point (50% B choice) reveals that juice B is worth 0.4 units (1/2.5) of juice A. *B:* object value coding of single neuron in monkey orbitofrontal cortex during behavioral choices assessed in *A*. Activity increases with the amount of only juice B, not juice A, suggesting object value coding of reward B. [*A* and *B* from Padoa-Schioppa and Assad (405). Reprinted with permission from Nature Publishing Group.] *C:* behavioral choices between two actions that vary in reward probability. Blue and red ticks show actual choices of left and right targets, respectively. Light blue line shows running % left choices. *D:* action value coding of single neuron in monkey striatum. Activity increases with value (probability) for left action (*left panel*: blue vs. orange), but remains unchanged with value changes for right action (*right panel*), thus coding left action value. [*C* and *D* from Samejima et al. (500). Reprinted with permission from AAAS.]

ence with the reward or from inferences based on models of the world, e.g., Bayesian. Third, action value is independent of the actual action choice and the value obtained by that choice. Without this property, action values cannot serve as independent inputs to competitive decision mechanisms. Thus minimal tests for action value require two different actions, two different values for each action, and independence from choice.

Action value neurons process the value for one particular action irrespective of the animal's action. There would be left action value neurons tracking the value of a movement to the left irrespective of the animal's actual movement on that trial, and there would be separate right action value neurons. Or different neurons would respectively track the value of arm and eye movements irrespective of the animal's actual movement selection. Such separate action value tracking neurons are suitable inputs to competitive decision mechanisms that underlie utility maximization **(FIGURE 36*C*)**. Subforms may show preferential action value coding for particular chosen actions or objects. Action value signals need to vary on a common scale, both between the actions and between the neurons, to constitute comparable neuronal decision signals. Sensory and motor parameters of the action should not affect action value signals.

Neurons coding action values according to these definitions are found in monkey and rat dorsal and ventral striatum (slowly and tonically firing neurons), globus pallidus, dorsolateral prefrontal cortex, and anterior cingulate of monkeys and rats. These neurons code selectively the value of the left or right arm or eye movement irrespective of the animal's choice (**FIGURE 41, *C* AND D**) (245, 267, 275, 306, 500, 530). Action value neurons are more frequent in monkey striatum than dorsolateral prefrontal cortex (530). Action values in these studies are subjective and derived from behavioral models fitted to the choices (245, 267, 306, 500, 530) and from logistic regressions on the animal's choice frequencies (275). However, the required common scale coding is unknown. The anterior cingulate cortex is involved in action-reward mapping during choices, as lesions

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

there reduce preferences for more rewarded movements in humans and monkeys (84, 108).

The substantial action value coding in the striatum may reflect the pronounced motor function of this structure. Indeed, optogenetic stimulation in mice of dopamine D1 receptor expressing striatal neurons (direct pathway) induces within a few minutes behavioral preferences toward an operant touch or nose poke target associated with the stimulation, whereas stimulation of D2 receptor expressing striatal neurons (indirect pathway) induce behavioral dispreference (293). These data are compatible with the notion of striatal action value signals constituting inputs to a competitive decision mechanism located in the striatum or downstream structures.

In contrast, some reward signals reflect the explicitly predicted reward for specific actions while the actions are being planned or executed. These signals do not comply with action value coding, as they don't show the required independence from free or forced choice and action and thus are unsuitable inputs to competitive decision mechanisms (see below under "chosen value"). They code the chosen value during choices and reward prediction in choices or imperative trials. If the signals code in addition specifics of the actions they may reflect goal-directed mechanisms (see above and **FIGURE 20, *E* AND *F***).

6. Object value versus action value

The two decision variables derive from the intuition of objects having reward value that can be traded and require actions for obtaining them. The two variables may play different roles depending on the choice situation. Object value computations are likely to play a major role when values of objects change frequently and need to be tracked before choosing the required action. Once the object has been selected, its value can be directly mapped onto the action, without computing and comparing all action values. When I like to select a fruit at a new lunch place, I check the values of all available fruits (object value), and then I choose the highest valued fruit. Then I will know the action that gets me the chosen fruit (map the object to the action), without need to value each action and choose between their values. I can enjoy the feeling to soon get my beloved fruit, but that comes after the decision has been made and thus does not enter the competitive decision process. Indeed, monkey orbitofrontal neurons code the chosen value and map it onto action activity in dorsolateral prefrontal cortex to initiate the action ("good-to-action transformation," Ref. 81).

In contrast, when object values are rather stable but actions change frequently, action value would be the more appropriate decision variable. With frequent action changes, effort cost may change and should to be tracked continuously and subtracted from income value (*Equation 13*). At the time of choice, the different object values are mapped onto action values while subtracting effort cost of the currently required actions. In realistic choice situations, the progression from object value to action value may not be so stereotyped, nor would object value have primacy over action value or vice versa. Rather, object values, action values, object-action mapping, and motor plans would evolve with partial overlap depending on the choice situation. Neuronal processing of these variables and mechanisms, and the competition mechanisms at each stage, may involve separate neuronal populations in different brain structures. Indeed, orbitofrontal and anterior cingulate lesions induce differential deficits in object-reward and action-reward associations (84, 108).

7. Abstract decision

Once the comparison process has resulted in selection of the best reward object, the output of the selection process may be specified by an abstract decision signal that concerns only the decision and as such may precede the specification of the choice. Abstract decision signals code the choice in a binary, categorical manner and irrespective of value, object, or action.

Abstract decision signals are seen during perceptual decisions. They become apparent when decoupling the decision temporally or spatially from the subsequent reporting action. The signals occur without, separately from, or stronger than object or action signals in prefrontal, premotor, secondary somatosensory, and anterior cingulate cortex with vibrotactile stimulus comparisons **(FIGURE 42*A*)** (212, 312, 334, 340, 489, 490) and visual stimulus detection (360, 361). They are also seen as buildup activity in supplementary eye fields and superior colliculus during visual search (587) and random dot motion discrimination (230).

Abstract decision signals occur also in specific neurons in the amygdala that show graded reward value coding early during the trial and transition to later binary abstract decision coding **(FIGURE 42*B*)** (188). The decision signal from 1–50 pooled neurons predicts the upcoming choice correctly in 75–90% of trials, respectively, and remained predictive even with identical reward between the two options. These neurons are not activated during imperative forced-choice trials, nor do they code visual stimuli indicating the options or the subsequently chosen oculomotor actions. Abstract decision signals during economic choices in the amygdala extends its recently recognized reward function (422) to economic decision-making and challenge the view of a structure primarily known for fear responses.

8. Chosen value

Chosen value reflects the value of the specific object or action that the decision maker obtains from the choice. Its

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

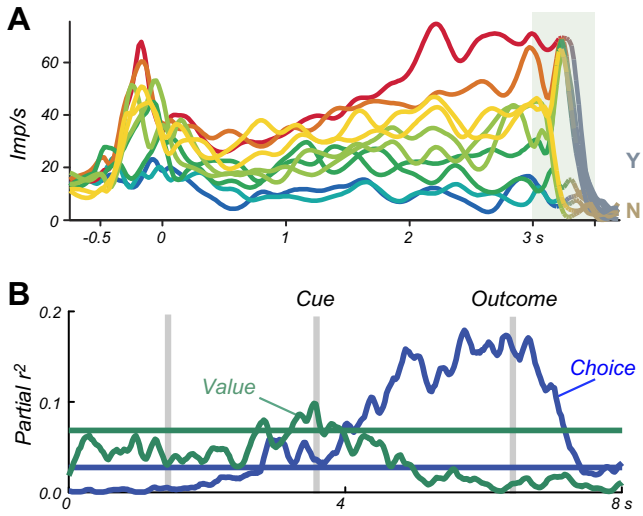


FIGURE 42. Abstract neuronal decision signals in monkeys. *A:* abstract perceptual decision coding in prefrontal cortex neuron. Initial graded frequency coding of vibrotactile stimulus (left peak) transitions to all-or-none coding of difference between two successive vibration frequencies. First vibratory stimulus begins at −0.5 s, second (comparison) vibratory stimulus begins at 3.0 s. Y and N refer to yes/no decision about the first stimulus frequency being higher than the second one. [From Machens et al. (334). Reprinted with permission from AAAS.] *B:* abstract economic decision coding during save-spend choices in single amygdala neuron. In each trial, the animal chose between saving reward with an interest or spending the saved reward. The neuron codes reward value early in the trial (derived from choice preferences, green) and later codes the actual save-spend choice (blue). Shown are time courses of the two most significant partial regression coefficients (r^2) from a multiple regression model that includes abstract save-spend choice, value (reward magnitude), spatial cue position, left-right action, and reaction time. [From Grabenhorst et al. (188).]

function is defined by competitive decision mechanisms (**FIGURES 35,** *middle*, and **36*C***). Whereas object value and action value are attached to an object or action irrespective of the choice, chosen value reflects the explicitly predicted or received reward value following the choice of that object or action. Thus chosen value reflects the output of the decision mechanism. For this important difference, chosen value should not be called object value or action value, even when it is attributed to a specific object or action. Chosen value is a decision variable, as it is controlled by the decision maker, who usually tries to maximize it (with the exceptions mentioned above). Chosen value, in principle, should be quantified as formal economic utility (*Equations 9–25*). Reinforcement learning employing Pavlovian, habit, and goal-directed processes provide mechanisms for setting up and updating chosen value.

Chosen value comes in two forms. Chosen object value reflects the value of a chosen reward object, like apple or orange, or the value of a specific stimulus predicting a particular reward. In contrast, chosen action value reflects the value of a chosen action. This distinction can be exemplified with particular reward magnitudes and probabilities. *Rewards A* and *B* are 1 ml and 2 ml of juice, respectively. *Reward A* occurs at the left in 20% of trials and at the right in 80%, whereas *reward B* occurs at the left in 60% and the right in 40% of trials. Then chosen object values are 1 ml and 2 ml for *rewards A* and *B*, respectively, whereas chosen action values are 1.4 ml for left (1 ml × 0.2 + 2 ml × 0.6) and 1.6 ml for right action (1 ml × 0.8 + 2 ml × 0.4). Most experiments do not specifically address this distinction by not using such discriminating values.

Neuronal chosen value signals reflect the value of an object or action irrespective of nonvalue attributes. These signals do not code the kind of reward (sensory aspects), sensory properties of predictive stimuli, spatial reward occurrence, stimulus positions, and required actions. In free choices between objects, specific neurons in orbitofrontal cortex code the chosen value of two or three different reward juices **(FIGURE 23, *A* AND *B*)** (405, 406). These activities reflect the estimated or experienced subjective economic value of the chosen juice, as assessed psychophysically by behavioral choices, and thus predict the choice (404). The signals scale in the same way across different rewards, thus coding value in common currency. Dopamine neurons code chosen value irrespective of action. Their value signal varies with the reward probability associated with the chosen stimulus (377). Stronger dopamine activations following conditioned stimuli predict the choice of the higher valued stimulus. Chosen value signals in the striatum show temporal discounting in choices between identical reward liquids delivered after different delays (79). Chosen value signals in anterior cingulate cortex code reward magnitude, probability, and effort, often in the same neurons (261). Chosen value signals in the supplementary eye field reflect the subjective value of objects, as estimated from certainty equivalents of behavioral choices between certain and risky outcomes (551). In imperative forced-choice trials, neurons in striatum and substantia nigra pars reticulata distinguish between stimuli predicting different reward magnitudes irrespective of sensory stimulus properties (107, 654).

In free choices between actions, some striatal neurons code chosen value without coding the action itself. The value signals vary with the reward probability associated with the chosen action and predict choice of the better action with higher activity during several task epochs (275, 306). Some striatal reward magnitude coding neurons do not distinguish between actions in imperative trials but reflect the reward value predicted from the choice (107). Chosen value signals in the striatum often occur after the overt behavioral choice, whereas action value signals typically appear before the choice (275, 306), suggesting that these chosen value signals represent the result of the decision process.

Unchosen value refers to the value of an unexperienced outcome. Whereas chosen value reflects the actual reward and is appropriate for experience-based decisions and reinforcement learning, unchosen value reflects some knowl-

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

edge about the world and might be more appropriate for model-based behavior. Animals benefit from the unchosen value information to improve reward gains (1, 208). Some neurons in dorsal and ventral striatum code unchosen value during intertemporal choices in monkeys (79) and during locomotor choices in rats (275). These neurons are distinct from the more frequent chosen value coding neurons in these structures. Some neurons code both chosen value and unchosen value with similar response slopes in anterior cingulate (208) and in orbitofrontal and dorsolateral prefrontal cortex (1). Thus some chosen value neurons code value irrespective of actual outcome.

Even before choosing, subjects may inspect the different options and process the anticipated value of each option. Humans show such inspectional eye movements whose frequencies and durations comply with evidence accumulation in diffusion decision models (291). It would be interesting to see whether neuronal value processing may initially reflect the value of the object or action being considered and subsequently shift to coding the chosen value. During action decisions, neurons in premotor cortex show initial activity for both actions but become selective for the final action after an imperative instruction (98, 99).

Hybrid forms of chosen value include chosen objects or actions. Corresponding neuronal signals reflect the value of the chosen reward together with sensory or motor information about the chosen object or action. If these activities occur well before the overt behavioral choice and increase towards a decision threshold, they may reflect the evidence accumulating in favor of the chosen option and thus play a role in the decision process. If the activities occur after the decision, they would constitute the outputs of the decision process and not drive the decision itself. They are occasionally referred to as object value or action value coding. However, according to the definition by competitive decision mechanisms **(FIGURE 36*C*)**, they constitute chosen value coding, as they reflect the reward value derived from the chosen objects or actions. Neuronal activity combining expected reward value conjointly with object information is only reported from imperative trials in frontal and temporal lobe neurons, as mentioned above (231, 319, 371, 394, 411), but cannot be labeled as "chosen" value due to the nonchoice task nature. Neurons coding chosen value conjointly with action are found with free choices in dorsal and ventral striatum (275, 306, 418, 479); globus pallidus (418); premotor, prefrontal, and anterior cingulate cortex (210, 532, 550, 620); and parietal cortex (433). The actions are eye or hand movements towards different spatial targets in monkeys (210, 306, 418, 433, 494, 550, 571, 620, 643) and nose pokes and whole left versus right body movements in rats (275, 479). In striatum, hybrid chosen value and action coding neurons are distinct from neurons coding only chosen value (306) or action value (306, 572). Conjoint coding exists also between unchosen value and action in orbitofrontal and dorsolateral prefrontal cortex (1). Comparable conjoint coding of value and action occurs also in nonchoice tasks (107, 135, 178, 205, 221, 243, 260, 282, 308, 313, 348, 354, 381, 427, 476, 502, 542, 606, 627), without qualifying for chosen value coding.

9. Relative value

The competitive decision mechanism for maximizing utility **(FIGURE 36*C*)** does not depend on actual values but involves comparison of the options relative to each other and is thus based on the difference or ratio of their values. Thus relative value is an important decision variable which, in principle, should be quantified as utility. Random walk-diffusion models capture the importance of relative value by assuming an early comparison with subsequent evidence growing toward specific thresholds. The mechanism for coding relative value may involve mutual inhibition between inputs that generates a graded difference between the inputs, amplified by the inhibition exerted by stronger inputs onto weaker inputs.

For binary decisions between *options 1* and *2*, the relative value can be stated using the final utility defined in *Equation 25*

$$\mathrm{REUsum} = \mathrm{EUsum1} - \mathrm{EUsum2} \qquad (27)$$

Instead of EUsum, any of the constituent utility functions defined by *Equations 9–13*, *16*, and *21–24* can be used for *Equation 27*. Most simply, REUsum may be derived from objective magnitudes $m1$ and $m2$

$$\mathrm{R}m = m1 - m2 \qquad (27A)$$

to obtain relative utility

$$\mathrm{Ru} = \mathrm{u1} - \mathrm{u2} \qquad (27B)$$

By applying a logarithmic utility function u = log (m), we obtain

$$\mathrm{Ru} = \log(m1) - \log(m2) \qquad (27C)$$

which is equivalent to

$$\mathrm{Ru} = \log(m1/m2) \qquad (27D)$$

As *Equations 8* and *9* show, reward value derives not only from reward magnitudes but also from reward probabilities. We replace $m1$ and $m2$ by probabilities $p1$ and $p2$

$$\mathrm{Ru} = \log(p1/p2) \qquad (27E)$$

Ru is a log odds ratio, also called log likelihood ratio (643). General simplifying formulations would replace objective m and p by their summed product of EV.

Another good measure of relative value is the ratio of objective values (magnitudes or probabilities)

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

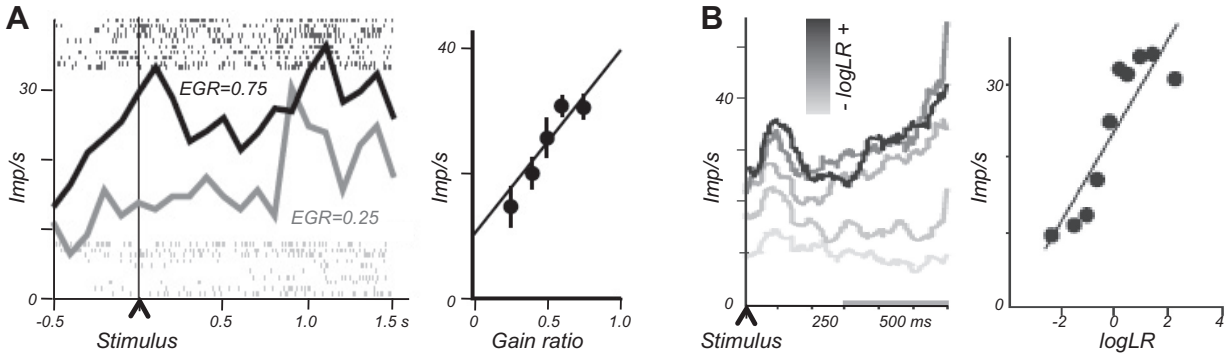


FIGURE 43. Relative value signals in monkey lateral intraparietal cortex. *A:* single neuron activity reflecting relative value during free oculomotor choice. *Left*: higher activity with higher "expected gain ratio" (EGR), lower activity with lower EGR. *Right*: linear regression on EGR. EGR denotes liquid reward volume of chosen option divided by summed magnitude of all options. [From Platt and Glimcher (433). Reprinted with permission of Nature Publishing Group.] *B:* single neuron responses increasing with the log likelihood ratio (logLR) of reward probabilities between chosen and unchosen options (choices into neuronal response field). [From Yang and Shadlen (643). Reprinted with permission from Nature Publishing Group.]

$$\mathrm{R}v = v1/v2 \qquad (27F)$$

The simple value ratio is often extended to include the summed value of all options in the denominator, called expected gain ratio (433) or fractional value (571)

$$\mathrm{RS}v = v1/\sum v_i;\ \text{for } i = 1, N;\ n = \text{number of options} \qquad (27G)$$

Reference value defined by *Equations 27-27G* applies to all forms of economic value, including object value, action value, and chosen value, and to all value functions defined by *Equations 8–13, 16*, and *21–24*. In monkeys, binary behavioral choices follow relative subjective value defined by the log odds ratio (*Equation 27E*) (643), reward magnitude ratio (*Equation 27F*) (306), or the ratio of objective chosen reward magnitude or probability over all options (*Equation 27G*) (433, 571). These measures are good predictors of choice preferences.

The difference value of *Equation 27* is analogous to reference-dependent utility (*Equation 22*), as both terms indicate relative value. Thus difference value and value adaptation serve the same function. They allow comparisons between existing options rather than requiring separate computations of the values of all options. Value ratios of the chosen option over the alternative (*Equation 27F*) or all options (*Equation 27G*) add adaptation to the spread of the reward probability distribution and correspond to divisive normalization in adaptive coding (332).

Some neurons in monkey and rat dorsal and ventral striatum, globus pallidus, and anterior cingulate cortex code difference subjective action value during free choices with eye or body movements, conforming to *Equation 27A* (79, 245, 306, 531, 572). Subjective value is derived from reinforcement models fitted to the choices or logistic regressions on choices. During temporal discounting, neurons code differences in subjective action value in dorsal striatum, and to some extent in ventral striatum (79). The discounted, subjective value is assessed from logistic regressions on intertemporal choices. Due to the relative nature of difference coding, neuronal activity varies positively with the value of one action or negatively with the value of the alternative action. True to the definition of action value, these relative value signals do not depend on actual choice. Neuronal difference signals exist also for chosen value in dorsal and ventral striatum, which take both chosen and unchosen values into account (79, 306). These activities increase with the value of one option and decrease with the alternative's value and, different from action value, depend on the actual choice. Neuronal activity in parietal association cortex reflects the reward ratio of the chosen option over all possible options (*Equation 27G*) **(FIGURE 43*A*)** (137, 433) and reaches a threshold at the time of choice **(FIGURE 40)** (494, 571). Similar parietal activity follows the increasing, logarithmic reward probability ratio between two options (*Equation 27E*) in the statistical sequential ratio test **(FIGURE 43*B*)** (643). Finally, difference signals for hybrid chosen value and oculomotor action exist in dorsal and ventral striatum (79, 306). Taken together, there are groups of distinct reward neurons coding basically every form of value decision variable in a relative way.

10. Chosen object and action

Although economic decision variables center around value, once the decision has been specified internally, the process may transfer to the overt behavioral choice without any longer involving value. Some neuronal signals in principal reward structures code the chosen object or the chosen action (frontal cortex, parietal cortex, and striatum) (**FIGURE 35,** *middle*).

Chosen object signals code the specific reward selected during the decision process or the specific visual stimulus pre-

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

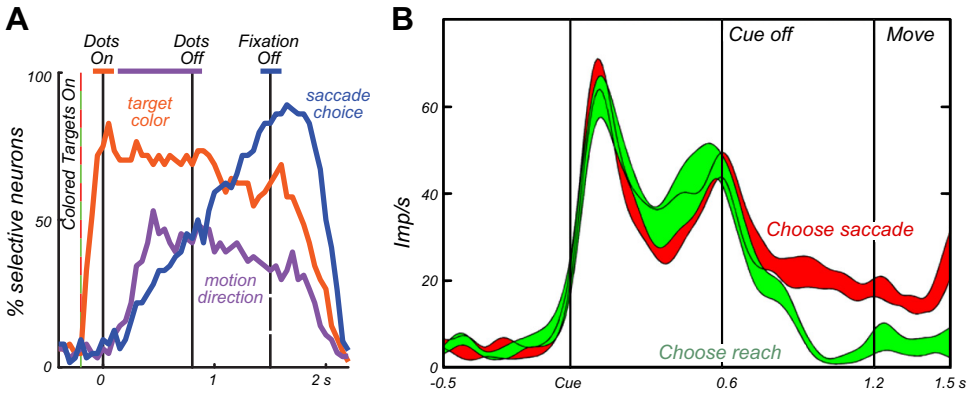


FIGURE 44. Chosen object and action signals in monkey lateral intraparietal cortex. *A:* coding of chosen object (orange color) or chosen action (blue) during perceptual decisions in population (averages from 23 neurons). In the conditional motor task, the animal saccades to a red or green dot (chosen object) depending on the direction of random dot motion. The dots are shown at different spatial positions and thus require different saccades (chosen action). Ordinate is % of significantly responding neurons for chosen object and chosen action (and dot motion direction, magenta). [From Bennur and Gold (38).] *B:* chosen action coding in single neuron. After a common onset, activity differentiates between effectors ~400 ms before action (saccade vs. reach). [From Cui and Andersen (109), with permission from Elsevier.]

dicting a particular reward. They serve to identify the chosen reward object while it is being acquired by an action. The signal does not code value and is not a simple visual or somatosensory response, as it may precede the appearance of the object. With economic decisions, chosen object signals in orbitofrontal cortex reflect different juices (405). With perceptual decisions, LIP neurons code the target color chosen for reporting the earlier, temporally distinct random dot motion direction (**FIGURE 44*A***, orange) (38). The signal is distinct from abstract decision and chosen action coded by other parietal neurons.

Chosen action signals code the specific action for acquiring the reward and may serve to prepare, initiate, and execute the necessary motor processes. These signals reflect the future action rather than reward value. They occur in cingulate motor area, anterior cingulate cortex, frontal and supplementary eye fields, and parietal cortex when monkeys change their actions after the reward for the current action declined (543), even before contacting a switch-over key (209) or before a chosen target occurs **(FIGURE 39*D*)** (101). The signals are absent or less pronounced in imperative forced-choice trials (101, 543). Neurons in LIP code the chosen saccadic action during probabilistic reward choice (433, 571), intertemporal choice (477), and dot motion discrimination (494), both during free choices and imperative forced-choice trials. In monkey and rat striatum, different neurons code chosen action from action value or chosen value during probability matching (306), locomotor choice (275), and intertemporal choice (79). Neurons in ventral striatum show much less chosen action signals than in dorsal caudate nucleus (79). Chosen action signals seem to begin earlier in striatal compared with parietal neurons and in both structures follow the subjective value signals by a few hundred milliseconds (79, 306, 477, 494). Some chosen action signals in parietal cortex are in addition modulated by chosen value (571). Thus chosen action signals occur in reward neurons close to behavioral output.

With perceptual decisions, parietal cortex neurons code the saccadic action for reporting random dot motion direction, distinct from abstract decision coding (**FIGURE 44*A***, blue) (38). Neurons in superior colliculus show chosen action signals superimposed on abstract decision signals (230).

With motor decisions, parietal cortex neurons show differential activity preceding freely chosen eye or arm movements **(FIGURE 44*B*)** (109, 509). During instructed action selection, separate neurons in premotor cortex initially code all options, but their activity differentiates and ramps up specifically for the arm movement imposed by a cue **(FIGURE 39*E*)** (98, 99). Although not reflecting free choice, the activity shows a nice transition from separate option coding to specific option coding as a general template for the progression of neuronal coding of economic decision variables.

11. Updating economic decision variables

Whereas perceptual decisions are based on sensory evidence, economic decisions require additional information about the value obtained by the choice. For meaningful decisions, this information needs to be predictive of the outcome. Choices without predictions amount to guessing. The value prediction does not derive from sensory organs but is learned and updated by error driven reinforcement learning. Its crucial error term is computed as the difference between the experienced reward and its prediction (*Equations 1, 2, 4, 5, 8A*–D, and *26-26C*).

The prediction entering the error computation derives from different sources (575). In model free reinforcement learn-

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

ing (460), the prediction derives only from direct experience with the values of the chosen objects or chosen actions in the past several trials. Other decision variables providing predictions, such as object value and action value, are only partly suitable for computing prediction errors because they are independent of the object or action being chosen and thus independent of the actually experienced reward. Their predictions are only useful for computing prediction errors when the objects or actions are chosen and their outcomes experienced, whereas unchosen rewards cannot enter the error computation. Thus the prediction in standard, model-free reinforcement learning derives from chosen value, which is a decision variable. In model-based reinforcement learning, the prediction incorporates information from models of the world and does not only reflect the recent experience with the outcome. The value prediction can be accurate even for objects and actions that have not recently been chosen and experienced. For example, given a model that assumes a constant sum of left and right action values, the left action value prediction can be accurate after only experiencing a right action and its outcome. Model-based reinforcement learning incorporates information beyond the experienced chosen value into the prediction and thus affects the error computation, whereas the reward is experienced as a matter of fact irrespective of model or not.

The type of prediction entering the error computation affects the form of the reward prediction error and thus the efficacy of value updating in specific learning scenarios (575). Three models are noteworthy in this respect. First, the Sarsa model applies to a range of laboratory tests in which the agent can experience all options during learning. Its prediction error incorporates the prediction derived from chosen value, which reflects the choice, but prediction from object value and action value is only valid when it derives from the actual choice. Thus chosen value neurons in striatum, orbitofrontal cortex, supplementary eye field, and anterior cingulate cortex (79, 208, 245, 261, 275, 306, 377, 405, 551) may serve as prediction inputs for the Sarsa model, and dopamine neurons code prediction errors in chosen value according to the Sarsa model (377). Second, Q learning directly approximates the optimal reward value within a given choice set. Its prediction error incorporates the prediction of the maximal attainable reward value, without taking the choice into account. Neurons in monkey anterior cingulate cortex code the prediction for the maximal probabilistic reward value that the animal usually but not always chooses (called task value) (12). Outright prediction error signals coding maximal value have not been described. Third, actor-critic models incorporate the sum of reward predictions from all options. Neurons in striatum, globus pallidus, and anterior cingulate cortex code the predicted sum (or mean with steeper slope) of reward values (79, 208, 245, 531). Taken together, the Sarsa reinforcement model has a correlate in the prediction error signals of dopamine neurons and other reward neurons that reflect predictions derived from chosen value. Future experiments might search for prediction error signals of the other reinforcement models in a wider range of learning situations.

Error signals suitable for updating decision variables are found in several reward structures. The bidirectional dopamine reward prediction error response (517) occurs with visual, auditory, and somatosensory stimuli (221, 365, 492) irrespective of action (521, 597, 618); integrates reward delay, reward risk, and different liquid and food rewards into a common subjective value signal (285, 301); incorporates model-based predictions (68, 382, 598); and reflects chosen value (377). Subgroups of non-dopamine neurons show bidirectional prediction error signals in lateral habenula, striatum, globus pallidus, amygdala, anterior cingulate cortex, and supplementary eye field (15, 36, 134, 227, 261, 275, 344, 531, 551). Some error signals in striatum code additional sensory and motor parameters (556). Neurons in cingulate cortex and supplementary eye field code bidirectional prediction errors from object value and action value to the extent of the object or action being chosen and leading to actual reward experience (chosen object value and chosen action value) (531, 550). Possible prediction error signals are also the separate, unidirectional positive and negative prediction error activations in anterior cingulate and supplementary eye field (246, 261, 551).

The updating function of the dopamine prediction error signal can be considered within the generic competitive model of decision-making **(FIGURE 45)**. Irrespective of the differences between formal decision models (55, 56, 441, 585, 623), the dopamine response would primarily update the inputs to the comparison process, namely, object value and action value that are coded during various trial epochs **(FIGURE 20, *A* AND *D*)**. The experienced and the predicted reward for the dopamine error computation likely derive from chosen value. The dopamine signal is rather similar among individual neurons and has heterogeneous effects on heterogeneous postsynaptic neurons. However, object value and action value are processed irrespective of the decision. They do not provide the necessary postsynaptic specificity to make the dopamine influence selective, which would indiscriminately update all object values and action values, including those not even experienced. However, as primary and higher order rewards occur only after an event, neuronal updating involves stimulus traces that maintain a label of the engaged synapses beyond the event and make them eligible for modification (574). Thus each object value and action value signal may leave a distinct eligibility trace that decays rapidly unless being stabilized by neuronal inputs **(FIGURE 14, *B* AND *C*)**. Neuronal activity from the specific chosen object/chosen action signal may provide selective inputs to object value/action value neurons for the same object/action for stabilizing their eligibility traces (**FIGURE 45,** *right*), whereas the lack of activity from the unchosen option would not prevent the decay (*left*). Thus the

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

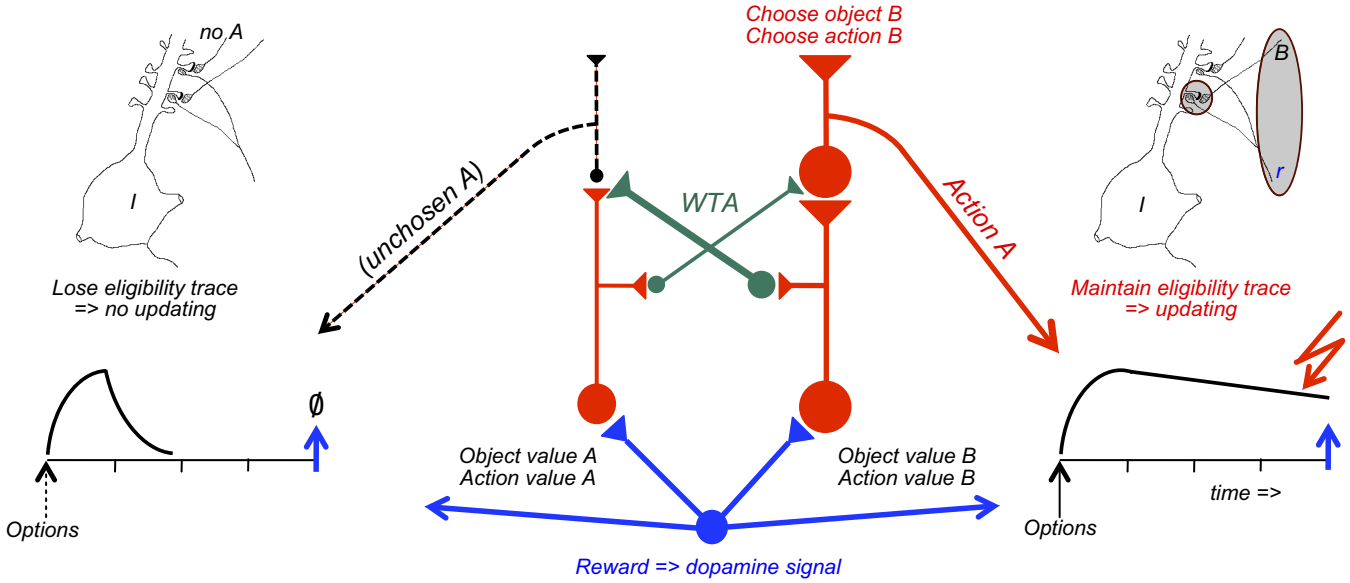


FIGURE 45. Hypothetical mechanism of dopamine updating of decision variables. Within a competitive decision mechanism, the global dopamine reward prediction error signal would act indiscriminately on postsynaptic neurons and changes their synaptic input efficacy by influencing stimulus eligibility traces. It affects only neurons coding object value or action value of the option chosen, as only their eligibility traces are being stabilized and maintained by input from neurons activated by the chosen object or chosen action (*right*), but not affect neurons whose initial eligibility traces are lost due to lack of stabilizing input from neurons not being activated by the unchosen object or unchosen action (*left*). This selective effect requires specific connections from chosen object/chosen action neurons to object value/action value neurons for the same object/action. The prediction error conveyed by the dopamine neurons derives from chosen value (experienced minus predicted reward). Gray zones at *top right* indicate common activations. The weight of dots and lines in the circuit model indicates level of neuronal activity. Dotted lines indicate inputs from unchanged neuronal activities. Crossed green connections are inhibitory; WTA, winner-take-all selection. Scheme developed together with Fabian Grabenhorst.

maintained eligibility trace reflects the output of the decision process and provides the necessary specific label for dopamine influences to act upon. As a result, the indiscriminate dopamine signal would selectively update the object value and/or action value of the chosen option, whereas object value and/or action value of the unchosen option remain unaltered and need to wait for their next selection before being updated (except in model-based reinforcement learning, see above). Thus the influences of chosen object and chosen action on the eligibility trace provide the necessary selectivity for dopamine action. As an example **(FIGURE 45)**, choice of a movement to the right and experience of the corresponding reward would lead to updating of right action value by a prediction error signal. If the right choice resulted in a higher than predicted value and thus a positive prediction error, right action value would be increased; however, if the value is lower than predicted and elicits a negative prediction error, right action value would be decreased. If the right action is not chosen, its value cannot be experienced and thus will not be updated (or might decay depending on the algorithm implemented). Taken together, within a simple comparator model of decision-making, the global influence of the dopamine error signal derived from chosen value may serve to update input decision variables for specific options via eligibility traces selectively stabilized by activity from the chosen option.

Neuronal correlates for this hypothetized dopamine influence consist of the global projection of this reward signal to most or all neurons in striatum and frontal cortex and to many neurons in amygdala, thalamus, and other structures **(FIGURE 17*A*)**. The released dopamine influences the plasticity between conjointly active presynaptic and postsynaptic neurons in a three factor Hebbian learning scheme **(FIGURE 17, *B* AND *C*)** involving striatum, frontal cortex, and other structures in which appropriate dopamine dependent plasticity is found (82, 194, 265, 294, 400, 401, 424, 461, 495, 540, 583, 642, 648). In this way, the dopamine signal could affect synaptic transmission onto striatal and cortical neurons coding object value and action value. These effects may be differential within striatal regions, as action value neurons are more abundant in dorsal compared with ventral striatum (245), whereas ventral striatum neurons are more involved in reward monitoring unrelated to action coding (245, 523). The dopamine plasticity effects may also be differential for striatal neuron types, as optogenetic stimulation of neurons expressing D1 or D2 dopamine receptors induces learning of behavioral preferences and dispreferences, respectively, possibly by updating action values **(FIGURE 19*F*)** (293). These neuronal data support the scheme of dopamine updating of decision variables **(FIGURE 45)**.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

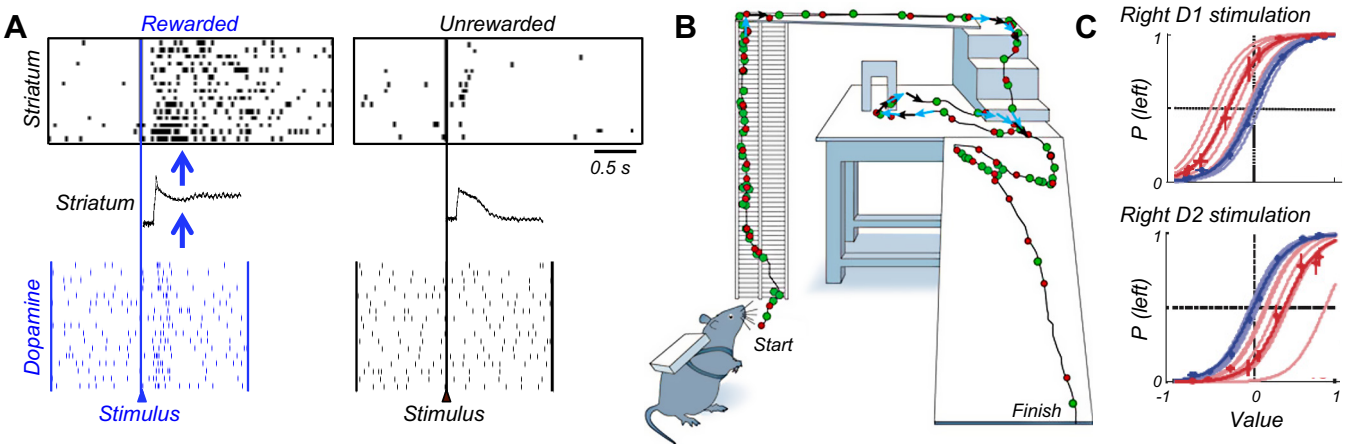


FIGURE 46. Immediate dopamine influences. *A:* hypothetical mechanism of influence of dopamine reward signal on striatal reward response. *Bottom*: a rewarded, but not unrewarded, stimulus differentially activates a dopamine neuron. [From Tobler et al. (598).] *Middle*: possible dopamine influence on striatal synaptic potentials: dopamine D1 receptor agonist prolongs the depolarization of a striatal neuron. [From Hernández-Lopez et al. (213).] *Top*: as a result of the dopamine influence on striatal depolarization, the striatal neuron responds stronger to a rewarded compared with an unrewarded stimulus. [From Hollerman et al. (222).] *B:* influence of dopamine reward signal on behavioral navigation. Electrical stimulation to the left and right somatosensory cortex of rats provides cues for turning; electrical stimulation of the medial forebrain bundle containing dopamine axons induces forward locomotion. The combined stimulation is able to guide a rat through a three-dimensional obstacle course, including an unnatural open field ramp. Colored dots indicate stimulation of forebrain bundle and somatosensory cortex. [From Talwar et al. (580). Reprinted with permission from Nature Publishing Group.] *C:* influence of dopamine reward signal on behavioral choices. Unilateral optogenetic stimulation of mouse striatal neurons expressing dopamine D1 or D2 receptors induces immediate differential choice biases (left bias for right D1 stimulation, *top*; and right bias for right D2 stimulation, *bottom*). [From Tai et al. (576). Reprinted with permission from Nature Publishing Group.]

Updating with non-dopamine error signals may be less global and more selective. These neurons likely exert focused influences on specific postsynaptic neurons through selective anatomical projections. Their error signals are more specific for the type of information they provide and for the type of postsynaptic neuron they influence. Thus they are likely to be more specific for the objects and actions and the type of neuronal value signals they update.

12. Immediate dopamine influences

In addition to affecting synaptic plasticity and learning, phasic dopamine signals exert immediate influences on postsynaptic processing and behavioral performance and choices. The anatomical basis may be the triad arrangement of dopamine synapses on glutamatergic inputs to dendritic spines of striatal and cortical neurons **(FIGURE 17*B*)** (171, 186) and convergence in striatum and globus pallidus (416, 430). The long debated focusing effect assumes that dopamine signals suppress the effects of weaker inputs to the striatum while allowing stronger ones to pass through to the basal ganglia outputs (70, 368, 594). Thus phasic dopamine signals may affect the function of existing anatomical convergence and modulate the use of neuronal connections. At the cellular level, the phasic dopamine signal has excitatory action on D1 receptor expressing striatal neurons of the direct pathway by prolonging transitions to membrane up states (depolarization) (213). In contrast, dopamine action on D2 receptor expressing striatal neurons of the indirect pathway reduces membrane up states and prolongs membrane down states (hyperpolarization) (214). In a hypothetical model of such action, the dopamine enhancement of cortical input efficacy may induce reward related activity in striatal neurons **(FIGURE 46*A*)**. This influence may affect reward processing in all forms of striatal task related activity **(FIGURE 20, *A* AND *D*)**. Impairments of dopamine transmission reveal the immediate dopamine influences. Striatal dopamine depletion reduces learned neuronal responses in striatum (13), alterations of D1 receptor stimulation affect mnemonic behavior and neuronal activity in frontal cortex (505, 616), and reduction of dopamine bursting activity via NMDA receptor knockout prolongs reaction time (650). These deficits relate well to the immediate behavioral effects of phasic dopamine stimulation. Optogenetic activation of mouse dopamine neurons, in addition to providing reinforcement, elicits immediate behavioral actions, including contralateral rotation and locomotion (271). Pairing electrical stimulation of the medial forebrain bundle containing dopamine axons with lateralized stimulation of somatosensory cortex induces forward locomotion in rats and directs the animals in sophisticated and even unnatural spatial navigational tasks **(FIGURE 46*B*)** (580). The forebrain bundle stimulation likely serves as immediate reward, whereas the cortical stimulation provides the necessary spatial information. Corresponding to the action-inducing effects of dopamine stimulation, indirect inhibition of VTA dopamine

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

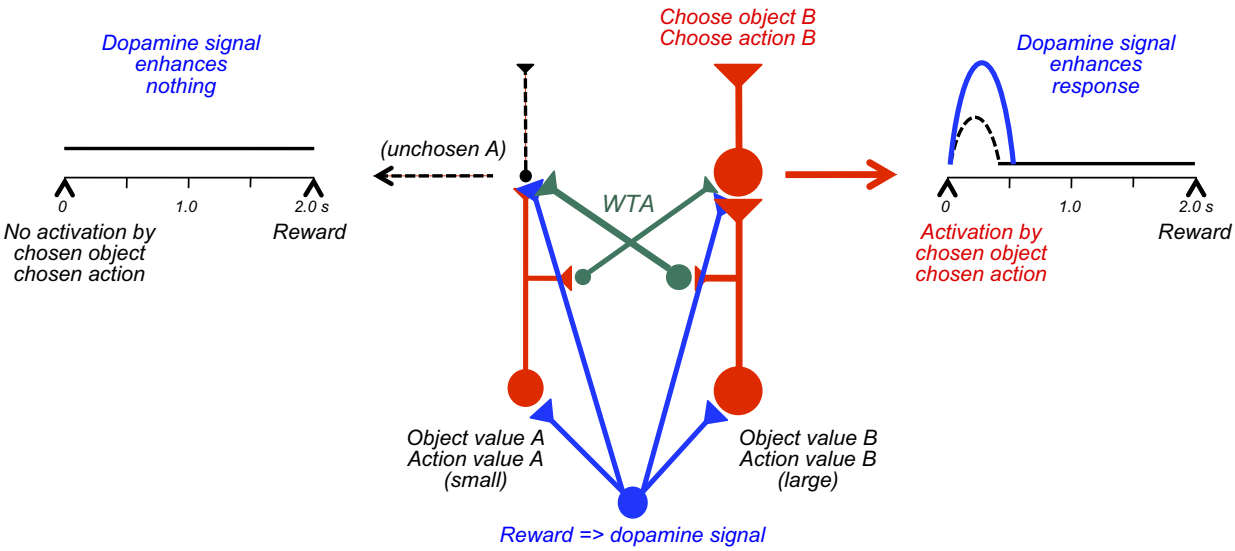


FIGURE 47. Hypothetical mechanism of immediate dopamine influence on neuronal coding of decision variables. For the basic competitive decision mechanism, the dopamine signal arises from a temporal difference (TD) prediction error in chosen value at the time of the decision. The dopamine signal boosts differences in activity between the two options reflecting object value or action value (*bottom*). Alternatively, and more effectively, the signal may enhance activities representing chosen object or chosen action selected by the winner-take-all (WTA) mechanism (*top right*), while leaving nonexisting activity unchanged (*top left*). The weight of dots and lines in the circuit model indicates level of neuronal activity. Dotted lines indicate inputs from unchanged neuronal activities. Crossed green connections are inhibitory.

neurons via optogenetic activation of local, presynaptic GABA neurons reduces reward consumption (613). At the postsynaptic level, optogenetic stimulation of dopamine D1 receptor expressing striatal neurons increases behavioral choices toward contralateral nose poke targets, whereas stimulation of D2 receptor expressing striatal neurons increases ipsilateral choices (or contralateral dispreferences) **(FIGURE 46*C*)** (576), suggesting an immediate, differential effect on reward value for the two actions.

Whereas these effects are straightforward, stimulating a couple synapses upstream of dopamine neurons induces surprisingly complex effects. Optogenetic activation of tegmental inputs to dopamine neurons projecting to nucleus accumbens elicits place preference, whereas stimulation of habenula inputs to dopamine neurons projecting to frontal cortex induces place dispreference (303). Whereas place preference signals reward, dispreference reflects either aversion or reduced reward. The dispreference might be due to monosynaptic excitation of aversive coding dopamine neurons, as suggested by FOS gene expression and excitatory currents (EPSCs) in dopamine neurons and by blunting of the dispreference with prefrontal D1 antagonists (303). However, dopamine neurons are not activated by aversive stimulus components (157, 160), FOS activation is slow and may derive from rebound activation following dopamine depression (158, 566), EPSCs do not necessarily induce action potentials required for dopamine release in target areas, and the optogenetic stimulation of presumed dopamine neurons might have included non-dopamine GABA neurons in used TH:Cre mice (304). Alternatively, dispreference might reflect disynaptic inhibition of dopamine neurons via the inhibitory rostromedial reticular nucleus (613). The monosynaptic habenula-dopamine projection is weak (625), and electrophysiology reports depression-activation sequence of VTA and nigral dopamine impulse activity following habenula stimulation (95, 250, 344, 566) that likely involves the inhibitory rostromedial reticular nucleus (226, 249). Furthermore, habenula glutamate receptor blockade increases striatal and prefrontal dopamine concentrations, probably reflecting blockade of the inhibitory habenula-dopamine projection (310). The prefrontal D1 antagonists likely blunt all prefrontal dopamine effects irrespective of excitation or inhibition. Thus the place dispreference with habenula stimulation unlikely reflects monosynaptic activation of aversive coding dopamine neurons and rather results from inhibiting dopamine neurons via the reticular nucleus. Indeed, a recent study reports inhibition of dopamine neurons and behavioral dispreference by electrical habenula stimulation. Similar place dispreference is elicited by activating midbrain GABA neurons that inhibit dopamine neurons (613) or by directly inhibiting dopamine neurons (244, 582). Thus the approach and dispreference elicited by differential input stimulation are likely the results of transsynaptically induced increased and reduced dopamine reward signals, respectively.

The immediate behavioral dopamine effects may be captured by a simplifying account that incorporates the essentials of immediate dopamine influences on behavioral choices into a generic competitive decision model **(FIGURE 47)**. To have an immediate effect, the dopamine signal

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

should reflect the prediction error at the time of the decision preceding the final reward. This error is the TD reward prediction error between the current prediction reflecting the chosen value minus the preceding reward prediction. The signal may influence neuronal activity for two kinds of decision variables. A dopamine-focusing effect (70, 368, 594) on predecisional object value or action value signals would enhance the value difference between the two options in a graded manner. The indiscriminate dopamine signal would affect all active synapses on neurons coding by definition object value and action value irrespective of the decision. Such dopamine influences on action value coding were assumed to explain the differential effects of optogenetic stimulation of D1 and D2 containing striatal neurons on behavioral choices (576). A more divisive dopamine influence would be achieved by affecting neurons that code the postdecisional variables of chosen object or chosen action before the overt behavioral choice (**FIGURE 47**, *right*). Activity in these neurons is differential due to selection by the WTA decision mechanism. The nonexisting activity from the unchosen option would not be enhanced by the dopamine signal (left). Thus the WTA mechanism would provide selectivity for the indiscriminate dopamine influence. Taken together, the basic comparator model of decision-making provides an architecture possibly underlying the immediate influence of the phasic dopamine signal on decision processes. This hypothesis is intentionally crude and provides only a general account, without addressing details of implementation by different types of dopamine innervated neurons in striatum, frontal cortex, and amygdala.

13. Formal neuronal decision models

The remarkable correspondence of neuronal signals to the key features of decision models encourages the development of biologically inspired neuronal models of decision-making. In many of such models, the economic decision variables of individual choice options are processed in separate pools of spiking neurons. The ramping activity is generated from transient inputs by recurrent synaptic excitation involving AMPA or NMDA receptors to induce self-sustained attractor states and integrate sensory or value information. The competition between choice options is mediated by GABA neurons exerting mutual, lateral, feedforward, or feedback inhibition between the neuronal pools for the choice options. The WTA mechanism combines the inhibition with a thresholding mechanism that leads to acceleration of ramping activity or to all-or-nothing selection between signals. The winning suprathreshold activity then propagates to pools of output neurons that mediate the behavioral action.

For neuronal decision models involving the cerebral cortex, key features are the discovery of explicit neuronal decision signals and the local synaptic connectivity. A generic model for perceptual discrimination of random dot motion direction in LIP neurons uses a difference signal derived from pools of specific direction-sensitive input neurons from middle temporal (MT) cortex (535). For example, larger activity in rightward compared with leftward motion coding MT neurons results in an LIP difference signal for rightward motion. The activity begins with appearance of the moving dots and accumulates over time with the continuing incoming noisy evidence about dot motion direction until it reaches one of the two decision thresholds. This model conforms to a diffusion model with continuous comparison between options and activity diverging towards one of several thresholds (55) and fits well the experimental data (535). Through its generality, such a model would also fit economic decision processes during which LIP neurons code the log likelihood ratio decision variable in a sequential probability ratio test (643), which constitutes an efficient decision mechanism (619).

A cortex model of two-stage somatosensory frequency discrimination uses two mutually inhibitory pools of 250 noisy spiking neurons that represent the two choice options (334). The model neurons show sustained activity that scales with the frequency of the first, transient vibratory stimulus until the comparison stimulus occurs. The subsequent activity conforms to an abstract decision signal that does not vary with individual stimulus frequencies. The sustained activity and its transition from mnemonic coding to decision signal resemble closely the activity of prefrontal and premotor cortex neurons of monkeys in this task (65, 212, 488, 489). The model achieves categorical decisions by discriminating between vibratory stimuli irrespective of their actual frequencies and thus supports the function of the experimentally observed neuronal signals in perceptual decisions.

More elaborate cortical neuronal decision models employ architectures comprising frontal and parietal cortex, basal ganglia, and superior colliculus **(FIGURE 48*A*)** (323, 622, 623). Transient inputs induce persistent activity with slow temporal integration via recurrent cortical NMDA and AMPA mediated excitation. Mutual lateral inhibition between pools of 240 irregularly spiking neurons mediates the competition between the stochastic choice options. By excitatory attractor dynamics and WTA competition throughout decision formation, the ramping activities in the neuronal pools representing the choice options become increasingly more separated in their race toward specific decision thresholds. The ramping activities reflect evidence accumulation or decision variables such as event likelihood ratio for perceptual decisions or object value, action value, and derivatives for economic choices. The cortical activity exceeding the competition threshold is detected by the superior colliculus via direct or double inhibitory influences through caudate nucleus and pars reticulata of substantia nigra and then propagated to the oculomotor system. The modeled neuronal activities replicate well the experimentally observed

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

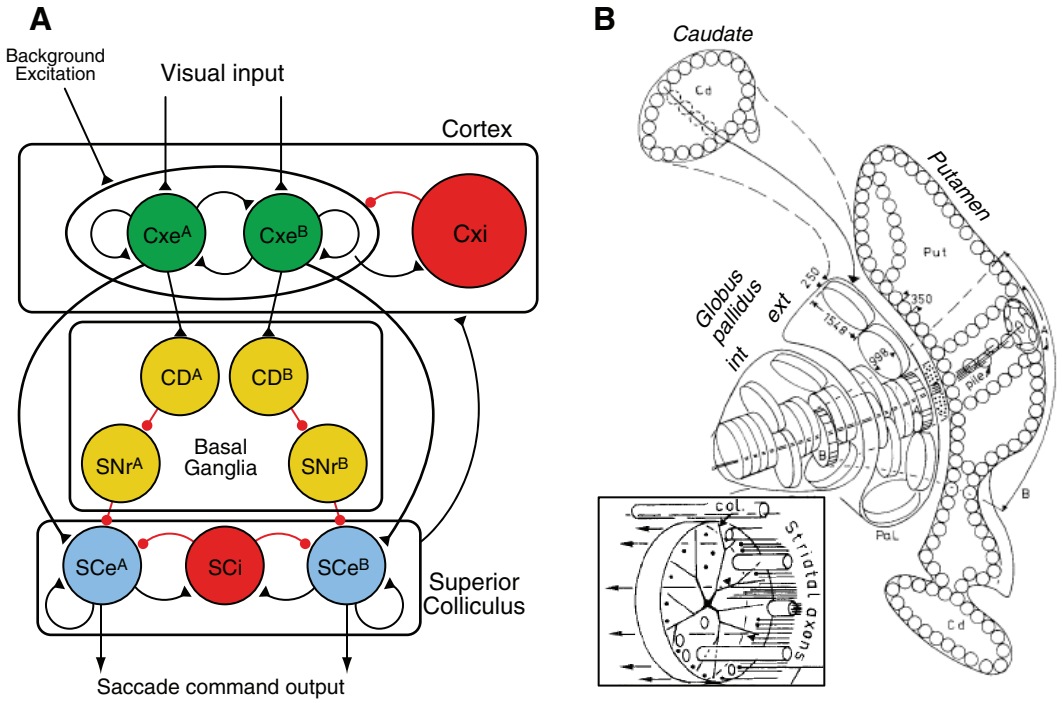


FIGURE 48. An example economic decision model and key anatomical foundations. *A:* architecture of a sensory or value decision model involving the cerebral cortex and basal ganglia. Recurrent excitations induce ramping activity in cortex (input to basal ganglia) and superior colliculus (output of basal ganglia), whereas lateral inhibition between neuronal pools representing the two choice options mediate competitive winner-take-all (WTA) selection of the better option at the input (cortex) and induce a corresponding ocular saccade at the output (superior colliculus). [From Wang (623), with permission from Elsevier.] *B:* anatomical convergence of neurons from striatum (caudate and putamen) onto dendritic disks of neurons in globus pallidus (internal and external segments) in monkey. The dendritic disks represent an abstraction of the wide dendritic arborizations that are oriented orthogonally to axons traversing the globus pallidus from the striatum. *Inset* shows several axons from striatal neurons traversing and contacting the pallidal dendritic disks. These parallel striatal axons induce substantial anatomical striatal-pallidal convergence by contacting the dendrites in the disks. [From Percheron et al. (430). Copyright 1984 John Wiley and Sons.]

ramping in frontal and parietal neurons described above. Dopamine-mediated three-factor Hebbian reinforcement learning may set the decision threshold in caudate nucleus or cortex (554). Thus cortical WTA determines the ramping, superior colliculus detects threshold crossing, and dopamine neurons set the threshold.

For neuronal decision models involving the basal ganglia, key features are the well-organized corticostriatal projections, internal connections, and parallel loops with cortex (7, 141, 164, 196, 197, 252); the synaptic convergence through the component structures **(FIGURE 48*B*)** (416, 430); the separate striatal outputs with sequential inhibitory neurons (93, 127); and the dopamine-mediated plasticity of corticostriatal connections (424, 517, 540). The basal ganglia may be involved in two distinguishable aspects of decision-making, namely, the direct selection of options and the updating of option values through reinforcement learning (57).

Option selection in the basal ganglia employs the dopamine focusing effect (70, 368, 594) and anatomical convergence in striatum and globus pallidus (416, 430) **(FIGURE 47)**. The models employ physiological activities and known input-output relationships of basal ganglia nuclei to demonstrate the feasibility of movement selection (72). By implementing the efficient sequential probability ratio test, the scaled race model shows that the subthalamic nucleus modulates the decision process proportionally to the conflict between alternatives (57). It outperforms unscaled race models. Particularly good fits are obtained with a diffusion model of movement selection (449). The basal ganglia decision models are supported by experimental data not used for their construction and also elucidate economic decision mechanisms, including the differential effects of optogenetic stimulation of dopamine D1 and D2 receptor expressing striatal neurons **(FIGURE 46C)** (576). The opposing functions of the direct and indirect striatal output pathways are modeled by a race between movement and countermanding signals that reflect the relative timing of the two signals as they race towards a threshold (511). The signals fit experimentally observed activities in rat substantia nigra neurons. These studies provide key support for important basal ganglia contributions to motor selection and economic decisions.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

Updating of object value, action value, and their derivatives by the basal ganglia employs dopamine prediction error signals (517) **(FIGURE 45)**. Data analyses routinely employ the modeled subjective values of choice options as regressors for neuronal value coding and action value updating in the striatum (245, 267, 275, 306, 500, 530). Reinforcement models distinguishing between direct and indirect striatal pathways are based on in vitro work (424, 540) and assume that dopamine prediction error signals have differential plasticity effects on D1 versus D2 receptor expressing striatal neurons (72) and induce the learning of respective choice preferences and dispreferences (169). These results from formal models closely reflecting physiological data confirm the capacity of dopamine reinforcement signals for learning and updating of economic decision variables.

E. Reward and Free Will

Current notions about free will range from extreme determinism, in which every event of every individual at every moment is predetermined in advance by the mechanics of the world, to the capricious, unrestrained free will that puts intention and responsibility at each individuals' complete disposition and initiative. The truth lies probably somewhere in between. There are likely neuronal mechanisms involving stochasticity that are constrained by external influences and generate choice options within a limited range. Deliberate actions might originate from these neuronal mechanisms and operate within these constraints. Constraints are provided by reward value and all factors influencing it, including effort cost, delay, risk, and references, that affect brain processes underlying decision-making. Without trying to provide a coherent view, I will sketch a few thoughts along these lines.

1. Origin of free will (the impossible question)

Assuming that free will exists at all, we simply do not know how it may originate, and we have only faint ideas how brain processes might generate it. We argue even whether free will originates in the brain. For Kant, free will lies above all activity, irrespective of physical, biological, psychological, and social constraints. For him, only God has free will, as s/he has created everything. But admitting that we do not know anything is the easy answer. We can try to reduce this impossible question to tractable issues by trying to understand the mechanisms determining free will.

Free will has several necessary components. First, free will depends on the ability to make choices, to do what we want to do (219). Second, free will implies a voluntary action that is self-directed and arises on self-command. Free will does not apply to vegetative reactions that I cannot control, like heartbeat, intestinal contractions, or hormone release, nor to probably most Pavlovian reactions and many habits. Third, the word *will* indicates a goal that we want to achieve. Thus free will would be instrumental in generating voluntary, goal-directed behavior. Fourth, free will involves conscious awareness about the goal directedness. We think we have free will because we know explicitly that we are working towards something we want. And we know that we want this something, however diffuse this may be in exploratory behavior. We feel the freedom to choose when to act and which of several options to select. This feeling is subjective, private, and unobservable by others. It is difficult to measure objectively, and it may not even be real but an illusion (634), nobody knows for sure.

The fact that free will has emerged in evolution suggests a beneficial function for the competitive survival and reproduction of individuals in their function as gene carriers. As much as the idea of a conscious illusion of free will may be disappointing, evolutionary biology teaches us that a structure or function may enhance the chance for survival without necessarily evolving to perfection. Thus it may not matter whether conscious free will is an illusion or real, as long as it is helpful for keeping the genes in the pool. Thus free will is a psychological tool for gene propagation.

What we do know is that we need rewards for survival (essential substances contained in food and drinks) and reproduction (sexual rewards). It would be advantageous to seek these hidden rewards actively rather than waiting for them to appear out of the blue. That is why we can move. Intentions and free will would be a great help to seek and explore rewards actively and deliberately, rather than running after them using deterministic programs. Without the true or false belief of free will, we may have only limited initiative to find hidden rewards. The main advantage of free will is the subjective experience of free will, which makes the search for hidden rewards more deliberate and focused, and thus saves time and is more efficient in many situations. Dennett (128) argues that consciously sensing free will is a crucial characteristic of free will. The evolutionary benefits driving the appearance of free will would be the boost in initiative, resulting from the conscious belief that we can deliberately decide our own behavior. This belief is also the basis for moral responsibility and thus helps social coherence. And adding theory of mind would make me understand others' free will and voluntary behavior and further facilitate social processes.

One view on conscious free will assumes that it is generated from scratch and makes us completely free to do what we want. There are only two possibilities to generate this sort of free will. First, brain activity generates the conscious awareness as necessary hallmark of free will. This brain activity necessarily precedes the conscious awareness and thus cannot be directed by free will. In this case, free will would not be free but an illusion of freedom that follows the uncontrolled generation of free will

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

by the brain. Alternatively, it is not brain activity but something nonmaterial that leads to conscious awareness of free will, and then brain activity carries out what the nonmaterial event has directed it to do and signals free will. Both scenarios seem unresolvable and therefore depressing and ultimately unconstructive. However, the problem may be overcome by taking a step back to consider the neuronal mechanisms by which free will might arise in more realistic situations in which it is seriously constrained by a number of factors. Even in this restricted form, free will would still provide an evolutionary beneficial, conscious control over our actions.

Everyday neurophysiological experiments tell us that many, maybe most, neurons in the brain are spontaneously active without any observable stimuli or actions. Thus brains generate their own activity. We also know that we have many mental experiences that seem to arise spontaneously out of nowhere, even in fully awake and nonpathological states, like thoughts, imagination, and old memories. How could I write this text if I could not sit back (with the proverbial cup of tea) and listen to the ideas that are popping up in my head (most of them rubbish, but occasionally a useful one), and then find the words and put them into grammatically correct order? Neuronal activity in cortex and striatum ramps up over several seconds before self-initiated movements but not before externally instructed actions (299, 311, 337, 380, 397, 491, 493, 526, 528). The key to explaining such brain activity may be the tendency to stochastic activity in many neurons and networks (129, 167, 180, 195). Once initiated by random molecular fluctuations (94, 287, 561, 567), the noise may result at some point in organized activity (168, 200, 297, 330, 465, 564, 611) and influence or settle in attractor states (63, 125, 510) that result in mental events. The hardware of the individual's brain and the individual's experience would constrain the range of attractor states producing mental events. A person who has never done or watched or heard about underwater diving may never experience a spontaneous imagination of the color of corals, nor the free will to go diving to see this. Thus the stochastic brain activity would arise spontaneously, relate to past experiences, and not be initiated from scratch by free will (nor do we need to assume nonmaterial origins).

The spontaneously occurring attractor states arising from initially stochastic neuronal activity would not need to be conscious. Several such elementary unconscious attractor states may develop at the same time, each of them representing a different future option for intention. The possibility of developing attractor states from stochastic activity would not be infinite but depend on brain activity restricted by hardware, current environment, and memory from earlier experience. Thus the attractor states would be restricted to viable options within boundaries. Impulses that never come up cannot be selected. Attractor states related to the intention of exploring the color of corals may only occur if my brain can process color and if I have ever heard about colorful corals. Then a selection mechanism would be set in motion in which viable options emerge with a certain degree of variation within bounds defined by hardware and personal experience. The process may resemble the selection of a reasonable option from a rotating initiator of options. There would be no deliberate, free initiative involved in this selection, as not only the attractor states but also the selection process and its boundaries would be unconsciously mechanical. Only the result of the selection would pop up into consciousness and be interpreted as free will. This conscious free will is not an illusion but the result of unconscious processing. From this point on, it would be possible to decide consciously whether to follow the selected impulse for action or to resist, to select consciously between the emerging mental images and intentions (e.g., separating rubbish from nonrubbish in writing), or to listen further to solutions from those popping up. Thus the unconscious processes produce a scratchpad that pops into consciousness and from which conscious free will can choose freely or veto an intention. This might be what is usually associated with the term of free will and the individual's responsibility for their own intentional actions.

Of course, there may be completely different, currently unknown processes besides stochasticity that may generate brain states underlying free will. The idea would require the discovery of new phenomena in physics (429). However, some of the described steps in the emergence of free will may still hold, in particular the dilemma in the initiation of free will between the brain's hardware and the mind.

2. Constraints on free will

Although free will is associated with being able to make choices (219), I can only choose between the options that are available. We are just not free to choose between anything imaginable, but our options are restricted by the physical state of the world (we cannot fly to Mars), our body's hardware (we cannot fly like birds), and our personal experiences and education (most language scholars do not do partial differential equations). Even among the available options, our choices are influenced by factors not considered as economic value such as emotions driven by external events, social pressure and conventions, altruism, moral rules, and irrational beliefs such as nationalism, chauvinism, prejudices, and superstitions. These external and internal constraints reduce free will in a graded manner, and up to its complete loss when constraints become too serious. Of course, survival is the strongest and most natural constraint on free will. Only within these bounds is free will free.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

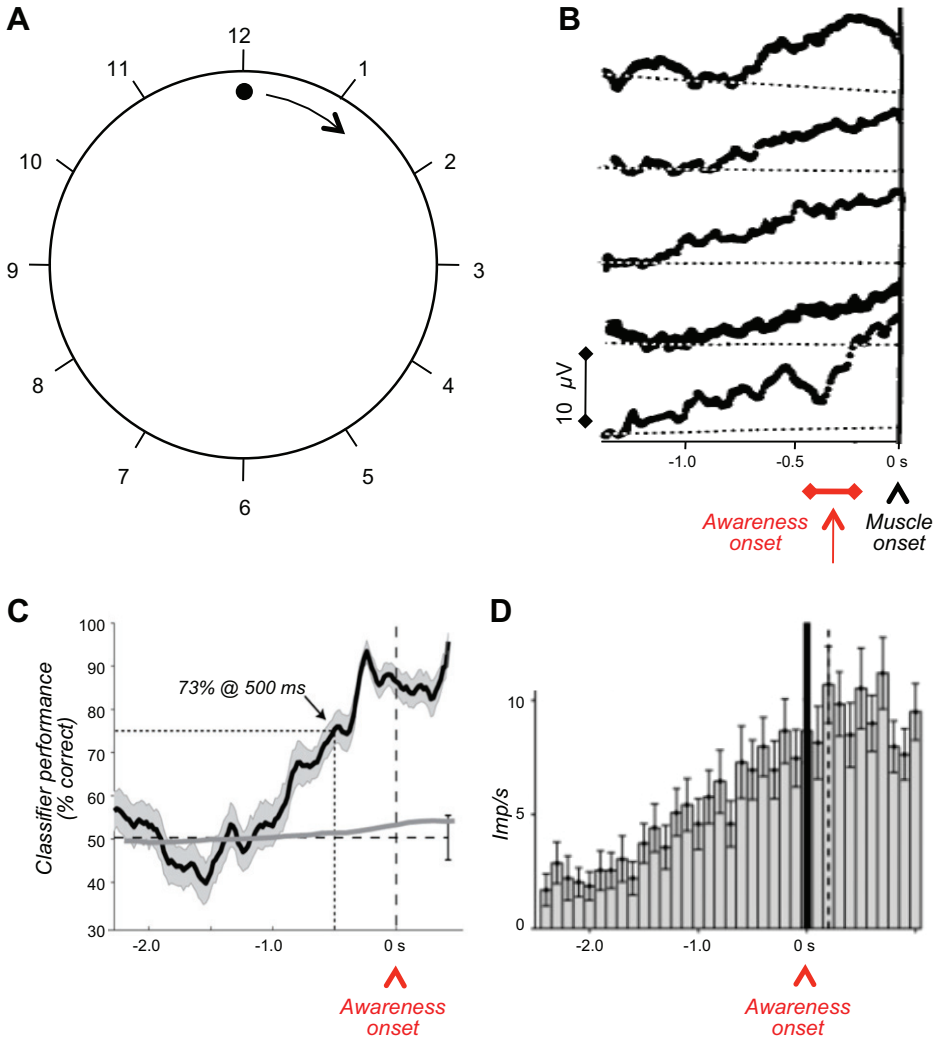


FIGURE 49. Unconscious onset of conscious awareness to act. *A:* design of Libet's classic experiment. A dot moves clockwise around a circle. Subjects remember the dot position at which they became consciously aware of the urge to do a self-initiated finger or wrist movement and report the dot position after being prompted. *B:* electroencephalographic readiness potential recorded at the vertex (top of skull) of one human participant (average of 40 trials). Onset of aware urge to move is indicated in red (bar shows range from eight averages of 40 trials each). Muscle onset denotes onset of electromyographic activity in involved forearm muscle. [From Libet et al. (316), by permission of Oxford University Press.] *C:* detection of time and occurrence of deviation from baseline activity in 37 single and multiple neurons before awareness to act, as determined by a support vector machine classifier. The neurons or multineuron clusters are located in human supplementary motor area (n = 22), anterior cingulate cortex (n = 8), and medial temporal lobe (n = 7). *D:* ramping activity preceding the conscious awareness to act in activity averaged from 59 neurons in human supplementary motor area. [*C* and *D* from Fried et al. (172), with permission from Elsevier.]

Free will can be constrained and focused. A good example is intention in action (529). I choose deliberately a course of action that allows me to form short-lived intentions on the fly while I am moving. I choose to do a walk through town and find a shop window that attracts my intention and makes me buy a pair of boots for the winter. My free will is partly restricted by the decision to take the walk, and the shoe purchase reflects that constraint. My purchase may simply be triggered by the view of the shoes, their price, and their relationship to my purchasing habits and personal preferences. The intention to enter the shop comes instantaneously, almost as a reaction to the sensory input. Very little free will is involved in the final purchasing action after I have initiated the walk.

In a similar situation of very restricted free will, humans indicate the moment at which they consciously perceive the awareness to initiate a simple finger movement **(FIGURE 49*A*)** (316). Indeed, the participants become aware of the urge to move only several hundred milliseconds after the onset of electroencephalographic brain activity for the movement (readiness potential, **FIGURE 49*B***). Sensory and motor confounds are excluded. The only conscious component consists of the possibility to veto the arising urge to move before the actual movement starts. Action-specific preparatory increases of BOLD signals in frontopolar and parieto-cingulate cortex precede the conscious awareness to move by 7–10 s (555). The activity of single neurons in humans preceding self-initiated movements predicts the awareness to move by 500 ms with 70–90% precision **(FIGURE 49*C*)**. These potentials may originate in nonintentional, spontaneous fluctuations reflecting a state of neuronal propensity toward action initiation, true to the meaning of the word "readiness potential" (515). At the neuronal level, such spontaneous fluctuations might occur below action potential threshold but are not conspicuous when recording action potentials during self-initiated movements in the supplementary motor area and associated cortical and subcortical structures in monkeys not expressing awareness to move (299, 311, 337, 380, 397, 491, 493, 526, 528) **(FIGURE 37*C*)** and in the supplementary motor area and anterior cingulate cortex in humans 1–2 s before the awareness to move **(FIGURE 49*D*)** (172). Thus it appears

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

as if the brain's hardware initiated the movement without the participant's own initiative, suggesting that conscious free will does not exist in these experiments. It is tempting to deny the general existence of free will on the basis of these results. However, the voluntary submission to the experiment may put so many constraints on the liberty of the participant that indeed there is no free will left and the brain is able to initiate the movement without conscious free will, just like the shoe buying example. Thus the experiments employ the role of constraints on free will in an extreme manner and test intention in action rather than true free will. In doing so they may reveal the existence of the elementary process outlined above by which stochastic brain activity settles spontaneously into an attractor state without conscious awareness and only afterwards leads to a conscious mental state. The consciously aware urge to move is then simply a retrograde interpretation of an unconsciously initiated action following the conscious selection to undergo the experiment.

3. Rewards exert constraints on free will

Rewards are necessary for survival. Without rewards we would die of thirst within days and of hunger within weeks, depending on our adipose reserves. With that function, rewards influence and thus constrain behavior and restrict free will in several ways. The passive experience of rewards attracts behavioral reactions away from other objects and focuses behavior onto the available rewards. Entering active decisions between rewards is not very free either because we need rewards at some point for survival. Once we are able to decide, we choose the best option, called utility maximization in economics, to ensure the highest chance for survival and gene propagation. Choices in such restricted situations may often reflect intentions in action, similar to the Libet experiments on movements, and involve serious constraints on free will. The choices are easy to do when the options are limited and have clearly different values, in which case there is little scope for free will, but the choices are more difficult when more, and more similar, options remove some constraints on free will. Habits for acquiring rewards impose particularly strong constraints on free will, and addictions to natural and artificial rewards eliminate free will even more. The initial administration of an addictive substance, like nicotine, is done with free will, often constrained by peer pressure, but following the habit once serious addiction has set in is almost deterministic.

All additional factors affecting reward value add to the constraints on free will. Motivational states like hunger and thirst make rewards more valuable and prioritize our behavior, as easily seen before lunch and dinner. Rewarding contexts make us long for rewards. Who would work during an office party when colleagues drink wine and tell interesting stories? Work required to obtain rewards seriously focuses our actions. Many animals spend most of their day searching for food. Temporal discounting of future rewards hinders our planning and favors myopic actions. Reference-dependent valuation makes us slaves of our environment. Nobody wants to lose what they have achieved, and keeping up with the Joneses is hard work, both of which require good focus and constitute constraints.

Risk exerts huge constrains on our behavior. Risk avoidance makes us shy away from risky situations even if they might be beneficial in the long run. The inability to completely eliminate risk induces anxiety, prejudice, and superstition which further direct behavior in serious ways, up to generating wars without reason. Risk management and the desire for risk reduction are prevalent in all cultures. Many activities of daily life are aimed at reducing risk, from paying car and health insurance to ritual sacrifices. The first reaction of the Minoan people to the devastating Santorini volcano eruption was human sacrifices to appease the angry gods. Risk reduction becomes a part of the culture, influences the whole life history of individuals, and thus takes some free will away.

The constraints exerted by rewards on free will extend also to the learning about rewards. Pavlovian reward predictions arise automatically following repeated exposure without the subject's own doing, which constrains free will substantially. Operant conditioning involves own action and allows goal-directed behavior, which constrains free will a bit less than Pavlovian predictions. Conditioning and higher forms of learning are evolutionary beneficial, as they increase reward acquisition for survival and gene propagation, but in doing so they constrain free will considerably.

Genetic dispositions, education, and long-term reward experience are instrumental for establishing personal reward preferences that influence the whole life of individuals. Different individuals prefer different rewards and choose their profession accordingly (among many other factors). They become mechanics because they appreciate precision, bankers for the money, musicians for the beautiful sounds, doctors because of altruism, and scientists because they value curiosity and knowledge highly. The impact of these preferences accumulates over the whole life span and shapes the personality. All of these influences constrain free will. Even seemingly free changes in personal reward preferences due to physiological factors and personal experience do not much enhance free will.

Punishers constitute the second major class of motivators of behavior besides rewards. Punishers derive from external sources over which we have little control, and they need to be avoided to escape their damaging impact. Thus avoidance is a reaction to events that occurs without one's own doing. There may be different options for escape, and thus limited scope for free will, but overall avoidance involves little free will, allows little deliberation, and produces primarily reactive rather than proactive behavior. The choices

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

between punishers seem to allow some degree of liberty that is intrinsically low when being forced into the choice. According to the *New York Times* of April 23, 2010, prisoners sentenced to capital punishment in Utah may choose between hanging and a firing squad. "I would like the firing squad, please" reportedly said Ronnie Lee Gardner. Overall, some liberty exists when being able to proactively search for best solutions for efficient avoidance. We can think of known dangers and generate preventive measures, but it is hard to imagine totally unknown dangers and be creative in planning their avoidance. Thus punishers may exert even more constraints on free will than rewards.

4. Neuronal reward processing constrains free will

Electrical and optogenetic stimulation of dopamine neurons are very efficient for eliciting immediate behavioral reactions and behavioral learning **(FIGURE 19, *A–D*)** (2, 103, 155, 562, 605, 641). The strength of dopamine stimulation affects choices, reaction times, and reward acquisition in a graded manner. The animals are simply driven towards goals associated with the dopamine activation, both during the stimulation (immediate effect) and in the next test sessions (reinforcement learning). Thus, in focusing individuals on rewards at the expense of other pursuits, dopamine stimulation takes parts of free will away from the individual's behavior.

Dopamine stimulations are not experimental artifacts but mimic the natural phasic dopamine responses to reward prediction errors. The dopamine responses are likely to have similar behavioral effects as the stimulation. Even though stimulation parameters differ somewhat from the natural responses, the important feature is the brief, phasic dopamine activation, whereas slow, tonic dopamine stimulation is ineffective (e.g., for place preference learning; Ref. 605). The dopamine effects likely involve synaptic plasticity and focusing in striatum and frontal cortex (70, 368, 424, 540, 594). Indeed, optogenetic stimulation of striatal neurons postsynaptic to dopamine neurons induces differential approach and avoidance learning and behavioral bias depending on the dopamine receptor subtype the stimulated neurons carry **(FIGURES 19*F* AND 46*C*)** (293, 576). Thus the combined evidence from dopamine stimulation and dopamine reward responses indicates that eliciting a strong reward signal in dopamine neurons directs the organism towards reward and thus constrains its free will. This is the mechanism through which primary (unpredicted) rewards and (Pavlovian) reward predictors influence the behavior and constrain free will.

A similar argument may hold for neuronal risk processing. Dopamine and orbitofrontal reward neurons code risk separately from value **(FIGURE 31*C*)** (161, 162, 391, 392). Furthermore, neuronal risk signals, and the influence of risk on neuronal value signals, vary with individual risk attitudes (301). If these risk signals could affect risk components in behavior analogous to value signals, they would provide mechanistic explanations for the constraining influence of risk on free will. To demonstrate such an effect, one would need to demonstrate that selective stimulation of risk neurons, such as those found in orbitofrontal cortex, affects risk attitudes during choices.

The strength of the dopamine response seems to convey the influence of rewards on behavior. Rewards with stronger effects on dopamine neurons are likely to have more impact on learning and choices. Thus the propensity of individual rewards to activate dopamine neurons would determine the influence of reward on behavioral choices. This function may hold also for other reward centers in the brain, although their less compact organization may not be conducive to such dramatic stimulation effects. However, it is unlikely that a response of dopamine neurons or other reward neurons to an automatically conditioned, reward-predicting stimulus induces approach behavior entirely without the subject's own doing. There are other brain mechanisms that would limit the automaticity of such effects, but any neuronal reward signal is nevertheless likely to have a biasing effect towards rewarded stimuli and actions. The neuronal responsiveness to different rewards is likely to vary between individuals. Interindividual differences in reward processing may affect daily preferences but also determine long-term behavior, including professional choices and other important decisions in life. Thus the activity of reward neurons shapes behavior, constrains voluntary decisions, and thus restricts free will.

ACKNOWLEDGMENTS

I thank Anthony Dickinson for 20 years of discussion and collaboration on animal learning theory, Peter Bossaerts and Christopher Harris for 10 years of discussion and collaboration on experimental economics, and Peter Bossaerts, Anthony Dickinson, Fabian Grabenhorst, Armin Lak, Johannes Schultz, William R. Stauffer and two anonymous referees for comments. I am indebted to Colin F. Camerer, David M. Grether, John O. Ledyard, Charles R. Plott, and Antonio Rangel for discussing experimental economics at the Division of Humanities and Social Sciences, California Institute of Technology, Pasadena.

Address for reprint requests and other correspondence: W. Schultz, Dept. of Physiology, Development and Neuroscience, Univ. of Cambridge, Cambridge CB2 3DY, UK (e-mail: ws234@cam.ac.uk).

GRANTS

This work is supported by Wellcome Trust Principal Research Fellowship, Programme and Project Grants 058365, 093270, and 095495; European Research Council Advanced Grant 293549; and National Institutes of Health Caltech Conte Center Grant P50MH094258. Previous sup-

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

port came from the Human Frontiers Science Program, the Wellcome-MRC-funded Behavioural and Clinical Neuroscience Institute Cambridge, and the Swiss National Science Foundation.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author.

REFERENCES

1. Abe H, Lee D. Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron* 70: 731–741, 2011.
2. Adamantidis AR, Tsai HC, Boutrel B, Zhang F, Stuber GD, Budygin EA, Touriño C, Bonci A, Deisseroth K, de Lecea L. Optogenetic interrogation of dopaminergic modulation of the multiple phases of reward-seeking behavior. *J Neurosci* 31: 10829–10835, 2011.
3. Adrian ED, Zotterman Y. The impulses produced by sensory nerve endings. Part 3. Impulses set up by Touch and Pressure. *J Physiol* 61: 465–483, 1926.
4. Ainslie GW. Impulse control in pigeons. *J Exp Anal Behav* 21: 485–489, 1974.
5. Ainslie GW. Specious rewards: a behavioral theory of impulsiveness and impulse control. *Psych Bull* 82: 463–496, 1975.
6. Alexander GE, Crutcher MD. Neural representations of the target (goal) of visually guided arm movements in three motor areas of the monkey. *J Neurophysiol* 64: 164–178, 1990.
7. Alexander GE, DeLong MR, Strick PL. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu Rev Neurosci* 9: 357–381, 1986.
8. Amador N, Schlag-Rey M, Schlag J. Reward-predicting and reward-detecting neuronal activity in the primate supplementary eye field. *J Neurophysiol* 84: 2166–2170, 2000.
9. Ambroggi F, Ishikawa A, Fields HL, Nicola SM. Basolateral amygdala neurons facilitate reward-seeking behavior by exciting nucleus accumbens neurons. *Neuron* 59: 648–661, 2008.
10. Amemori KI, Graybiel AM. Localized microstimulation of primate pregenual cingulate cortex induces negative decision-making. *Nat Neurosci* 15: 776–785, 2012.
11. Amiez C, Joseph JP, Procyk E. Anterior cingulate error-related activity is modulated by predicted reward. *Eur J Neurosci* 21: 3447–3452, 2005.
12. Amiez C, Joseph JP, Procyk E. Reward encoding in the monkey anterior cingulate cortex. *Cereb Cortex* 16: 1040–1055, 2006.
13. Aosaki T, Graybiel AM, Kimura M. Effect of the nigrostriatal dopamine system on acquired neural responses in the striatum of behaving monkeys. *Science* 265: 412–415, 1994.
14. Aosaki T, Tsubokawa H, Ishida A, Watanabe K, Graybiel AM, Kimura M. Responses of tonically active neurons in the primate's striatum undergo systematic changes during behavioral sensorimotor conditioning. *J Neurosci* 14: 3969–3984, 1994.
15. Apicella P, Deffains M, Ravel S, Legallet E. Tonically active neurons in the striatum differentiate between delivery and omission of expected reward in a probabilistic task context. *Eur J Neurosci* 30: 515–526, 2009.
16. Apicella P, Ljungberg T, Scarnati E, Schultz W. Responses to reward in monkey dorsal and ventral striatum. *Exp Brain Res* 85: 491–500, 1991.
17. Apicella P, Ravel S, Deffains M, Legallet E. The role of striatal tonically active neurons in reward prediction error signaling during instrumental task performance. *J Neurosci* 31: 1507–1515, 2011.
18. Apicella P, Scarnati E, Ljungberg T, Schultz W. Neuronal activity in monkey striatum related to the expectation of predictable environmental events. *J Neurophysiol* 68: 945–960, 1992.
19. Argilli E, Sibley DR, Malenka RC, England PM, Bonci A. Mechanism and time course of cocaine-induced long-term potentiation in the ventral tegmental area. *J Neurosci* 28: 9092–9100, 2008.
20. Arsenault JT, Rima S, Stemmann H, Vanduffel W. Role of the primate ventral tegmental area in reinforcement and motivation. *Curr Biol* 24: 1347–1353, 2014.
21. Asaad WF, Eskandar EN. Encoding of both positive and negative reward prediction errors by neurons of the primate lateral prefrontal cortex and caudate nucleus. *J Neurosci* 31: 17772–17787, 2011.
22. Aston-Jones G, Rajkowski J, Kubiak P, Alexinsky T. Locus coeruleus neurons in monkey are selectively activated by attended cues in a vigilance task. *J Neurosci* 14: 4467–4480, 1994.
23. Azzi JCB, Sirigu A, Duhamel JR. Modulation of value representation by social context in the primate orbitofrontal cortex. *Proc Natl Acad Sci USA* 109: 2126–2131, 2012.
24. Badman MK, Flier JS. The gut and energy balance: visceral allies in the obesity wars. *Science* 307: 1909–1914, 2005.
25. Báez-Mendoza R, Harris C, Schultz W. Activity of striatal neurons reflects social action and own reward. *Proc Natl Acad Sci USA* 110: 16634–16639, 2013.
26. Balleine B, Dickinsion A. Goal-directed instrumental action: contingency, and incentive learning and their cortical substrates. *Neuropharmacology* 37: 407–419, 1998.
27. Bangasser DA, Waxler DE, Santollo J, Shors TJ. Trace conditioning and the hippocampus: the importance of contiguity. *J Neurosci* 26: 8702–8706, 2006.
28. Bao S, Chan VT, Merzenich MM. Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature* 412: 79–83, 2001.
29. Barnes TD, Kubota Y, Hu D, Jin DZ, Graybiel AM. Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature* 437: 1158–1161, 2005.
30. Barraclough D, Conroy ML, Lee DJ. Prefrontal cortex and decision making in a mixed-strategy game. *Nat Neurosci* 7: 405–410, 2004.
31. Barto AG, Singh S, Chentanez N. Intrinsically motivated learning of hierarchical collections of skills. In: *Advances in Neural Information Processing Systems 17: Proceedings of the 2004 Conference*. Cambridge, MA: MIT Press, 2005.
32. Barto AG, Sutton RS, Anderson CW. Neuronlike adaptive elements that can solve difficult learning problems. *IEEE Trans Syst Man Cybernet* 13: 834–846, 1983.
33. Bayer HM, Glimcher PW. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* 47: 129–141, 2005.
34. Bayer HM, Lau B, Glimcher PW. Statistics of dopamine neuron spike trains in the awake primate. *J Neurophysiol* 98: 1428–1439, 2007.
35. Bechara A, Damasio AR, Damasio H, Anderson SW. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* 50: 7–15, 1994.
36. Belova MA, Paton JJ, Morrison SE, Salzman CD. Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. *Neuron* 55, 970–984, 2007.
37. Belova MA, Paton JJ, Salzman CD. Moment-to-moment tracking of state value in the amygdala. *J Neurosci* 28: 10023–10030, 2008.
38. Bennur S, Gold JI. Distinct representations of a perceptual decision and the associated oculomotor plan in the monkey lateral intraparietal area. *J Neurosci* 31: 913–921, 2011.
39. Bentham J. *An Introduction to the Principle of Morals and Legislations*, *1789* . Reprinted Oxford, UK: Blackwell, 1948.
40. Bermudez MA, Göbel C, Schultz W. Sensitivity to temporal reward structure in amygdala neurons. *Curr Biol* 22: 1839–1844, 2012.
41. Bermudez MA, Schultz W. Responses of amygdala neurons to positive reward predicting stimuli depend on background reward (contingency) rather than stimulus-reward pairing (contiguity). *J Neurophysiol* 103: 1158–1170, 2010.
42. Bermudez MA, Schultz W. Reward magnitude coding in primate amygdala neurons. *J Neurophysiol* 104: 3424–2432, 2010.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

43. Bernoulli D. Specimen theoriae novae de mensura sortis. *Comentarii Academiae Scientiarum Imperialis Petropolitanae (Papers Imp Acad Sci St Petersburg)* 5: 175–192, 1738 (translated as Exposition of a new theory on the measurement of risk. *Econometrica* 22: 23–36, 1954).

44. Berridge KC. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology* 191: 391–431, 2007.

45. Berridge KC, Robinson TE. What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res Rev* 28: 309–369, 1998.

46. Berthoud HR, Morrison C. The brain, appetite, obesity. *Annu Rev Psychol* 59: 55–92, 2008.

47. Beylin AV, Gandhi CC, Wood GE, Talk AC, Matzel LD, Shors TJ. The role of the hippocampus in trace conditioning: temporal discontinuity or task difficulty? *Neurobiol Learn Mem* 76: 447–461, 2001.

48. Bi GQ, Poo MM. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J Neurosci* 18: 10464–10472, 1998.

49. Bichot NP, Schall JD. Effects of similarity and history on neural mechanisms of visual selection. *Nat Neurosci* 2: 549–554, 1999.

50. Binmore K, Shaked A. Experimental economics: where next? *J Econ Behav Organiz* 73: 87–100, 2010.

51. Black RW. Shifts in magnitude of reward and contrast effects in instrumental and selective learning: a reinterpretation. *Psychol Rev* 75: 114–126, 1968.

52. Blanchard TC, Wolfe LS, Vlaev I, Winston JS, Hayden BY. Biases in preferences for sequences of outcomes in monkeys. *Cognition* 130: 289–299, 2014.

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

54. Blythe SN, Wokosin D, Atherton JF, Bevan MD. Cellular mechanisms underlying burst firing in substantia nigra dopamine neurons. *J Neurosci* 29: 15531–15541, 2009.

55. Bogacz R. Optimal decision-making theories: linking neurobiology with behaviour. *Trends Cog Sci* 11: 118–125, 2007.

56. Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol Rev* 113: 700–765, 2006.

57. Bogacz R, Gurney K. The basal ganglia and cortex implement optimal decision making between alternative actions. *Neur Comput* 19: 442–477, 2007.

58. Bouret S, Richmond BJ. Ventromedial and orbital prefrontal neurons differentially encode internally and externally driven motivational values in monkeys. *J Neurosci* 30: 8591–8601, 2010.

59. Bouton ME. Context and behavioural processes in extinction. *Learning Memory* 11: 485–494, 2004.

60. Bowman EM, Aigner TG, Richmond BJ. Neural signals in the monkey ventral striatum related to motivation for juice and cocaine rewards. *J Neurophysiol* 75: 1061–1073, 1996.

61. Bredfeldt CE, Ringach DL. Dynamics of spatial frequency tuning in macaque V1. *J Neurosci* 22: 1976–1984, 2002.

62. Breton YA, James C, Marcus JC, Shizgal P. *Rattus psychologicus*: construction of preferences by self-stimulating rats. *Behav Brain Res* 202: 77–91, 2009.

63. Brzezniak Z, Capinski M, Flandoli F. Pathwise global attractors for stationary random dynamical systems. *Probab Theory Relat Fields* 95: 87–102, 1993.

64. Brischoux F, Chakraborty S, Brierley DI, Ungless MA. Phasic excitation of dopamine neurons in ventral VTA by noxious stimuli. *Proc Natl Acad Sci USA* 106: 4894–4899, 2009.

65. Brody CD, Hernandez A, Zainos A, Romo R. Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cereb Cortex* 13: 1196–1207, 2003.

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

67. Bromberg-Martin ES, Hikosaka O. Lateral habenula neurons signal errors in the prediction of reward information. *Nat Neurosci* 14: 1209–1216, 2011.

68. Bromberg-Martin ES, Matsumoto M, Hon S, Hikosaka O. A pallidus-habenula-dopamine pathway signals inferred stimulus values. *J Neurophysiol* 104: 1068–1076, 2010.

69. Brosnan SF, de Waal FBM. Monkeys reject unequal pay. *Nature* 425: 297–299, 2003.

70. Brown JR, Arbuthnott GW. The electrophysiology of dopamine (D_2) receptors: a study of the actions of dopamine on corticostriatal transmission. *Neuroscience* 10: 349–355, 1983.

71. Brown JW, Bullock D, Grossberg S. How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *J Neurosci* 19: 10502–10511, 1999.

72. Brown JW, Bullock D, Grossberg S. How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks* 17: 471–510, 2004.

73. Brown MTC, Henny P, Bolam JP, Magill PJ. Activity of neurochemically heterogeneous dopaminergic neurons in the substantia nigra during spontaneous and driven changes in brain state. *J Neurosci* 29: 2915–2925, 2009.

74. Bruce CJ, Goldberg ME. Primate frontal eye fields. I. Single neurons discharging before saccades. *J Neurophysiol* 53: 603–635, 1985.

75. Budygin EA, Park J, Bass CE, Grinevich VP, Bonin KD, Wightman RM. Aversive stimulus differentially triggers subsecond dopamine release in reward regions. *Neuroscience* 201: 331–337, 2012.

76. Bunney BS, Grace AA. Acute and chonic haloperidol treatment: comparison of effects on nigral dopaminergic cell activity. *Life Sci* 23: 1715–1728, 1978.

77. Burkart JM, Fehr E, Efferson C, van Schaik CP. Other-regarding preferences in a non-human primate: common marmosets provision food altruistically. *Proc Natl Acad Sci USA* 104: 19762–19766, 2007.

78. Caggiano V, Fogassi L, Rizzolatti G, Casile A, Giese MA, Thier P. Mirror neurons encode the subjective value of an observed action. *Proc Natl Acad Sci USA* 109: 11848–11853, 2012.

79. Cai X, Kim S, Lee D. Heterogeneous coding of temporally discounted values in the dorsal and ventral striatum during intertemporal choice. *Neuron* 69: 170–182, 2011.

80. Cai X, Padoa-Schioppa C. Neuronal encoding of subjective value in dorsal and ventral anterior cingulate cortex. *J Neurosci* 32: 3791–3808, 2012.

81. Cai X, Padoa-Schioppa C. Contributions of orbitofrontal and lateral prefrontal cortices to economic choice and the good-to-action transformation. *Neuron* 81: 1140–1151, 2014.

82. Calabresi P, Gubellini P, Centonze D, Picconi B, Bernardi G, Chergui K, Svenningsson P, Fienberg AA, Greengard P. Dopamine and cAMP-regulated phosphoprotein 32 kDa controls both striatal long-term depression and long-term potentiation, opposing forms of synaptic plasticity. *J Neurosci* 20: 8443–8451, 2000.

83. Camerer CF. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton Univ. Press, 2003.

84. Camille N, Tsuchida A, Fellows LK. Double dissociation of stimulus-value and action-value learning in humans with orbitofrontal or anterior cingulate cortex damage. *J Neurosci* 31: 15048–15052, 2001.

85. Caplin A, Dean M, Martin D. Search and satisficing. *Am Econ Rev* 101: 2899–2922, 2011.

86. Caraco T, Blankenhorn WU, Gregory GM, Newman JA, Recer GM, Zwicker SM. Risk-sensitivity: ambient temperature affects foraging choice. *Anim Behav* 39: 338–345, 1990.

87. Caraco T, Martindale S, Whitham TS. An empirical demonstration of risk-sensitive foraging preferences. *Anim Behav* 28: 820–830, 1980.

88. Cardinal RN, Howes NJ. Effects of lesions of the nucleus accumbens core on choice between small certain and large uncertain rewards in rats. *BMC Neurosci* 6: 37, 2005.

89. Chang SWC, Gariépy JF, Platt ML. Neuronal reference frames for social decisions in primate frontal cortex. *Nat Neurosci* 16: 243–250, 2013.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

90. Chang SWC, Winecoff AA, Platt ML. Vicarious reinforcement in rhesus macaques (*Macaca mulatta*). *Front Neurosci* 5: 27, 2011.

91. Chen MK, Lakshminarayanan V, Santos LR. How basic are behavioral biases? Evidence from capuchin trading behavior. *J Pol Econ* 114: 517–537, 2006.

92. Chesselet MF. Presynaptic regulation of neurotransmitter release in the brain: facts and hypothesis. *Neuroscience* 12: 347–375, 1984.

93. Chevalier G, Vacher S, Deniau JM, Desban M. Disinhibition as a basic process in the expression of striatal functions. I. The striato-nigral influence on tecto-spinal/tecto-diencephalic neurons. *Brain Res* 334: 215–226, 1985.

94. Chow CC, White JA. Spontaneous action potentials due to channel fluctuations. *Biophys J* 71: 3013–3023, 2000.

95. Christoph GR, Leonzio RJ, Wilcox KS. Stimulation of the lateral habenula inhibits dopamine-containing neurons in the substantia nigra and ventral tegmental area of the rat. *J Neurosci* 6: 613–619, 1986.

96. Churchland AK, Kiani R, Chaudhuri R, Wang XJ, Pouget A, Shadlen MN. Variance as a signature of neural computations during decision making. *Neuron* 69: 818–831, 2011.

97. Churchland AK, Kiani R, Shadlen MN. Decision-making with multiple alternatives. *Nat Neurosci* 11: 693–702, 2008.

98. Cisek P, Kalaska JF. Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron* 45: 801–814, 2005.

99. Cisek P, Kalaska JF. Neural mechanisms for interacting with a world full of action choices. *Annu Rev Neurosci* 33: 269–298, 2010.

100. Clarke HF, Dalley JW, Crofts HS, Robbins TW, Roberts AC. Cognitive inflexibility after prefrontal serotonin depletion. *Science* 304: 878–880, 2004.

101. Coe B, Tomihara K, Matsuzawa M, Hikosaka O. Visual and anticipatory bias in three cortical eye fields of the monkey during an adaptive decision-making task. *J Neurosci* 22: 5081–5090, 2002.

102. Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* 482: 85–88, 2012.

103. Corbett D, Wise RA. Intracranial self-stimulation in relation to the ascending dopaminergic systems of the midbrain: a moveable microelectrode study. *Brain Res* 185: 1–15, 1980.

104. Crespi LP. Quantitative variation in incentive and performance in the white rat. *Am J Psychol* 40: 467–517, 1942.

105. Critchley HG, Rolls ET. Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. *J Neurophysiol* 75: 1673–1686, 1996.

106. Cromwell HC, Hassani OK, Schultz W. Relative reward processing in primate striatum. *Exp Brain Res* 162: 520–525, 2005.

107. Cromwell HC, Schultz W. Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. *J Neurophysiol* 89: 2823–2838, 2003.

108. Croxson PL, Walton ME, O'Reilly JX, Behrens TEJ, Rushworth MFS. Effort-based cost-benefit valuation and the human brain. *J Neurosci* 29: 4531–4541, 2009.

109. Cui H, Andersen RA. Posterior parietal cortex encodes autonomously selected motor plans. *Neuron* 56: 552–559, 2007.

110. d'Acremont M, Fornari E, Bossaerts P. Activity in inferior parietal and medial prefrontal cortex signals the accumulation of evidence in a probability learning task. *PLoS Comput Biol* 9: e1002895, 2013.

111. d'Ardenne K, McClure SM, Nystrom LE, Cohen JD. BOLD Responses reflecting dopaminergic signals in the human ventral tegmental area. *Science* 319: 1264–1267, 2008.

112. Darwin C. *On the Origin of Species by Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray, 1859.

113. Daw ND, Courville AC, Touretsky DS. Representation and timing in theories of the dopamine system. *Neur Comput* 18: 1637–1677, 2006.

114. Dawkins R. *The Selfish Gene*. London: Oxford Univ. Press, 1976.

115. Day JJ, Jones JL, Carelli RM. Nucleus accumbens neurons encode predicted and ongoing reward costs in rats. *Eur J Neurosci* 33: 308–321, 2011.

116. Day JJ, Jones JL, Wightman RM, Carelli RM. Phasic nucleus accumbens dopamine release encodes effort- and delay-related costs. *Biol Psychiat* 68: 306–309, 2010.

117. Day JJ, Roitman MF, Wightman RM, Carelli RM. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat Neurosci* 10: 1020–1028, 2007.

118. De Araujo IE, Oliveira-Maia AJ, Sotnikova TD, Gainetdinov RR, Caron MC, Nicolelis MAL, Simon SA. Food reward in the absence of taste receptor signaling. *Neuron* 57: 930–941, 2008.

119. De Araujo IE, Rolls ET, Velazco MI, Margot C, Cayeux I. Cognitive modulation of olfactory processing. *Neuron* 46: 671–679, 2005.

120. De Lafuente O, Romo R. Neuronal correlates of subjective sensory experience. *Nat Neurosci* 8: 1698–1703, 2005.

121. De Lafuente O, Romo R. Neural correlate of subjective sensory experience gradually builds up across cortical areas. *Proc Natl Acad Sci USA* 103: 14266–14271, 2006.

122. De Lafuente O, Romo R. Dopamine neurons code subjective sensory experience and uncertainty of perceptual decisions. *Proc Natl Acad Sci USA* 49: 19767–19771, 2011.

123. De Waal FB, Davis JM. Capuchin cognitive ecology: cooperation based on projected returns. *Neuropsychologia* 41: 221–228, 2003.

124. Deaner RO, Khera AV, Platt ML. Monkeys pay per view: adaptive valuation of social images by rhesus monkeys. *Curr Biol* 15: 543–548, 2005.

125. Deco G, Rolls ET, Romo R. Stochastic dynamics as a principle of brain function. *Prog Neurobiol* 88: 1–16, 2009.

126. Delamater AR. Outcome selective effects of intertrial reinforcement in a Pavlovian appetitive conditioning paradigm with rats. *Anim Learn Behav* 23: 31–39, 1995.

127. Deniau JM, Chevalier G. Disinhibition as a basic process in the expression of striatal function. II. The striato-nigral influence on thalamocortical cells of the ventromedial thalamic nucleus. *Brain Res* 334: 227–233, 1985.

128. Dennett DC. *Elbow Room. The Varieties of Free Will Worth Wanting*. Boston: MIT Press, 1984.

129. Destexhe A, Contreras D. Neuronal computations with stochastic network states. *Science* 314, 85–88, 2006.

130. Di Ciano P, Cardinal RN, Cowell RA, Little SJ, Everitt B. Differential involvement of NMDA, AMPA/kainate, and dopamine receptors in the nucleus accumbens core in the acquisition and performance of Pavlovian approach behavior. *J Neurosci* 21: 9471–77, 2001.

131. Di Loreto S, Florio T, Scarnati E. Evidence that a non-NMDA receptor is involved in the excitatory pathway from the pedunculopontine region to nigrostriatal dopaminergic neurons. *Exp Brain Res* 89: 79–86, 1992.

132. Dickinson A. *Contemporary Animal Learning Theory*. Cambridge, UK: Cambridge Univ. Press, 1980, p. 43.

133. Dickinson A, Balleine B. Motivational control of goal-directed action. *Anim Learn Behav* 22: 1–18, 1994.

134. Ding L, Gold JI. Caudate encodes multiple computations for perceptual decisions. *J Neurosci* 30: 15747–15759, 2010.

135. Ding L, Hikosaka O. Comparison of reward modulation in the frontal eye field and caudate of the macaque. *J Neurosci* 26: 6695–6703, 2006.

136. Dormont JF, Conde H, Farin D. The role of the pedunculopontine tegmental nucleus in relation to conditioned motor performance in the cat. I. Context-dependent and reinforcement-related single unit activity. *Exp Brain Res* 121: 401–10, 1998.

137. Dorris MC, Glimcher PW. Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron* 44: 365–378, 2004.

138. Doucet G, Descarries L, Garcia S. Quantification of the dopamine innervation in adult rat neostriatum. *Neuroscience* 19: 427–445, 1986.

139. Doya K, Samejima K, Katagiri K, Kawato M. Multiple model-based reinforcement learning. *Neural Comput* 14: 1347–1369, 2002.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

140. Drevets WC, Gautier C, Price JC, Kupfer DJ, Kinahan PE, Grace AA, Price JL, Mathis CA. Amphetamine-induced dopamine release in human ventral striatum correlates with euphoria. *Biol Psychiatry* 49: 81–96, 2001.

141. Eblen F, Graybiel AM. Highly restricted origin of prefrontal cortical inputs to striosomes in the macaque monkey. *J Neurosci* 15: 5999–6013, 1995.

142. Edgeworth F. *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*. New York: Augustus M. Kelly, 1881.

143. Ekman P. An argument for basic emotions. *Cogn Emot* 6: 169–200, 1992.

144. Enomoto K, Matsumoto N, Nakai S, Satoh T, Sato TK, Ueda Y, Inokawa H, Haruno M, Kimura M:. Dopamine neurons learn to encode the long-term value of multiple future rewards. *Proc Natl Acad Sci USA* 108: 15462–15467, 2011.

145. Estes WK. Discriminative conditioning. I: a discriminative property of conditioned anticipation. *J Exp Psychol* 32: 150–155, 1943.

146. Everitt BJ, Stacey P. Studies of instrumental behavior with sexual reinforcement in male rats (*Rattus norvegicus*). II. Effects of preoptic area lesions, castration, and testosterone. *J Com Psychol* 101: 407–419, 1987.

147. Everling S, Munoz DP. Neuronal correlates for preparatory set associated with prosaccades and anti-saccades in the primate frontal eye field. *J Neurosci* 20: 387–400, 2000.

148. Fairhall AL, Lewen GD, Bialek W, de Ruyter van Steveninck RR. Efficiency and ambiguity in an adaptive neural code. *Nature* 412: 787, 2001.

149. Fehr E, Camerer CF. Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn Neurosci* 11: 419–427, 2007.

150. Fehr E, Schmidt KM. A theory of fairness, competition, and cooperation. *Q J Econ* 114: 817–688, 1999.

151. Fehr E, Schmidt KM. On inequity aversion: a reply to Binmore and Shaked. *J Econ Behav Organiz* 73: 101–108, 2010.

152. Fehr-Duda H, Bruhin A, Epper T, Schubert R. Rationality on the rise: Why relative risk aversion increases with stake size. *J Risk Uncertain* 40: 147–180, 2010.

153. Feierstein CE, Quirk MC, Uchida N, Sosulski DL, Mainen ZF. Representation of spatial goals in rat orbitofrontal cortex. *Neuron* 51: 495–507, 2006.

154. Fiala J, Grossberg S, Bullock D. Metabotropic glutamate receptor activation in cerebellar purkinje cells as substrate for adaptive timing of the classically conditioned eye-blink response. *J Neurosci* 16: 3760–3774, 1996.

155. Fibiger HC, LePiane FG, Jakubovic A, Phillips AG. The role of dopamine in intracranial self-stimulation of the ventral tegmental area. *J Neurosci* 7: 3888–3896, 1987.

156. Fiorillo CD. Transient activation of midbrain dopamine neurons by reward risk. *Neuroscience* 197: 162–171, 2011.

157. Fiorillo CD. Two dimensions of value: dopamine neurons represent reward but not aversiveness. *Science* 341: 546–549, 2013.

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

159. Fiorillo CD, Song MR, Yun SR. Diversity and homogeneity in responses of midbrain dopamine neurons. *J Neurosci* 33: 4693–4709, 2013.

160. Fiorillo CD, Song MR, Yun SR. Multiphasic temporal dynamics in responses of midbrain dopamine neurons to appetitive and aversive stimuli. *J Neurosci* 33: 4710–4725, 2013.

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

162. Fiorillo CD, Tobler PN, Schultz W. Evidence that the delay-period activity of dopamine neurons corresponds to reward uncertainty rather than backpropagating TD errors. *Behav Brain Funct* 1: 7, 2005.

163. Flagel SB, Clark JJ, Robinson TE, Mayo L, Czuj A, Willuhn I, Akers CA, Clinton SM, Phillips PE, Akil H. A selective role for dopamine in stimulus-reward learning. *Nature* 469: 53–57, 2011.

164. Flaherty AW, Graybiel A. Output architecture of the primate putamen. *J Neurosci* 13: 3222–3237, 1993.

165. Fliessbach K, Weber B, Trautner P, Dohmen T, Sunde U, Elger CE, Falk A. Social comparison affects reward-related brain activity in the human ventral striatum. *Science* 318: 1305–1308, 2007.

166. Floresco SB, West AR, Ash B, Moore H, Grace AA. Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission. *Nat Neurosci* 6: 968–973, 2003.

167. Fourcaud-Trocme N, Brunel N. Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neur Comp* 14: 2057–2110, 2002.

168. Fox RF, Lu Y. Emergent collective behavior in large numbers of globally coupled independently stochastic ion channels. *Phys Rev E* 49: 3421–3425, 1994.

169. Frank MJ, Seeberger LC, O'Reilly RC. By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science* 306: 1940–1943, 2004.

170. Frémaux N, Sprekeler H, Gerstner W. Functional requirements for reward-modulated spike-timing-dependent plasticity. *J Neurosci* 30: 13326–13337, 2010.

171. Freund TF, Powell JF, Smith AD. Tyrosine hydroxylase-immunoreactive boutons in synaptic contact with identified striatonigral neurons, with particular reference to dendritic spines. *Neuroscience* 13: 1189–1215, 1984.

172. Fried I, Mukamel R, Kreiman G. Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron* 69: 548–562, 2011.

173. Friedman M, Savage LJ. The utility analysis of choices involving risk. *J Polit Econ* 56: 279–304, 1948.

174. Gaffan D, Murray EA, Fabre-Thorpe M. Interaction of the amygdala with the frontal lobe in reward memory. *Eur J Neurosci* 5: 968–975, 1993.

175. Gallistel CR, Gibbon J. Time, rate and conditioning. *Psych Rev* 107: 289–344, 2000.

176. Gao WY, Lee TH, King GR, Ellinwood EH. Alterations in baseline activity and quinpirole sensitivity in putative dopamine neurons in the substantia nigra and ventral tegmental area after withdrawal from cocaine pretreatment. *Neuropsychopharmacology* 18: 222–232, 1998.

177. Gauthier J, Parent M, Lévesque M, Parent A. The axonal arborization of single nigrostriatal neurons in rats. *Brain Res* 834: 228–232, 1999.

178. Gdowski MJ, Miller LE, Parrish T, Nenonene EK, Houk JC. Context dependency in the globus pallidus internal segment during targeted arm movements. *J Neurophysiol* 85: 998–1004, 2001.

179. Genovesio A, Tsujimoto S, Navarra G, Falcone R, Wise SP. Autonomous encoding of irrelevant goals and outcomes by prefrontal cortex neurons. *J Neurosci* 34: 1970–1978, 2014.

180. Gerstein G, Mandelbrot B. Random walk models for the spike activity of a single neuron. *Biophys J* 4: 41–68, 1964.

181. Gietzen DW, Hao S, Anthony TG. Mechanisms of food intake repression in indispensable amino acid deficiency. *Annu Rev Nutr* 27: 63–78, 2007.

182. Gilbert PFC, Thach WT. Purkinje cell activity during motor learning. *Brain Res* 128: 309–328, 1977.

183. Glimcher PW. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proc Natl Acad Sci USA* 108: 15647–15654, 2011.

184. Gnadt JW, Andersen RA. Memory related motor planning activity in posterior parietal cortex of macaque. *Exp Brain Res* 70: 216–220, 1988.

185. Gold JI, Shadlen MN. The neural basis of decision making. *Annu Rev Neurosci* 30: 535–574, 2007.

186. Goldman-Rakic PS, Leranth C, Williams MS, Mons N, Geffard M. Dopamine synaptic complex with pyramidal neurons in primate cerebral cortex. *Proc Natl Acad Sci USA* 86: 9015–9019, 1989.

187. Gonzalez R, Wu G. On the shape of the probability weighting function. *Cogn Psychol* 38: 129–166, 1999.

188. Grabenhorst F, Hernadi I, Schultz W. Prediction of economic choice by primate amygdala neurons. *Proc Natl Acad Sci USA* 109: 18950–18955, 2012.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

189. Grabenhorst F, Rolls ET, Bilderbeck A. How cognition modulates affective responses to taste and flavor: top-down influences on the orbitofrontal and pregenual cingulate cortices. *Cereb Cortex* 18: 1549–1559, 2008.

190. Grabenhorst F, Rolls ET, Margot C, da Silva MAAP, Velazco MI. How pleasant and unpleasant stimuli combine in different brain regions: odor mixtures. *J Neurosci* 27: 13532–13540, 2007.

191. Grether WF. Pseudo-conditioning without paired stimulation encountered in attempted backward conditioning. *Comp Psychol* 25: 91–96, 1938.

192. Groves PM, Garcia-Munoz M, Linder JC, Manley MS, Martone ME, Young SJ. Elements of the intrinsic organization and information processing in the neostriatum. In: *Models of Information Processing in the Basal Ganglia*, edited by Houk JC, Davis JL, Beiser DG. Cambridge, MA: MIT Press, 1995, p. 51–96.

193. Guarraci FA, Kapp BS. An electrophysiological characterization of ventral tegmental area dopaminergic neurons during differential pavlovian fear conditioning in the awake rabbit. *Behav Brain Res* 99: 169–179, 1999.

194. Gurden H, Takita M, Jay TM. Essential role of D1 but not D2 receptors in the NMDA receptor-dependent long-term potentiation at hippocampal-prefrontal cortex synapses in vivo. *J Neurosci* 106: 1–5, 2000.

195. Gutkin BS, Ermentrout GB. Dynamics of membrane excitability determine interspike interval variability: a link between spike generation mechanisms, and cortical spike train statistics. *Neur Comp* 10: 1047–1065, 1998.

196. Haber SN, Fudge JL, McFarland NR. Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *J Neurosci* 20: 2369–2382, 2000.

197. Haber SN, Kim KS, Mailly P, Calzavara R. Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, providing a substrate for incentive-based learning. *J Neurosci* 26: 8368–8376, 2006.

198. Hanes DP, Schall JP. Neural control of voluntary movement initiation. *Science* 274: 427–430, 1996.

199. Hanks T, Kiani R, Shadlen MN. A neural mechanism of speed-accuracy tradeoff in macaque area LIP. *eLife* 3: e02260, 2014.

200. Hansel D, Sompolinsky H. Synchronization and computation in a chaotic neural network. *Phys Rev Lett* 68: 718–721, 1992.

201. Hariri AR, Mattay VS, Tessitore A, Kolachana B, Fera F, Goldman D, 2 Egan MF, Weinberger DR. Serotonin transporter genetic variation and the response of the human amygdala. *Science* 297: 400–403, 2002.

202. Harlow HF. The formation of learning sets. *Psychol Rev* 56: 51–65, 1949.

203. Harnett MT, Bernier BE, Ahn KC, Morikawa H. Burst-timing-dependent plasticity of NMDA receptor-mediated transmission in midbrain dopamine neurons. *Neuron* 62: 826–838, 2009.

204. Hart AS, Rutledge RB, Glimcher PW, Phillips PEM. Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *J Neurosci* 34: 698–704, 2014.

205. Hassani OK, Cromwell HC, Schultz W. Influence of expectation of different rewards on behavior-related neuronal activity in the striatum. *J Neurophysiol* 85: 2477–2489, 2001.

206. Hayden BY, Heilbronner SR, Pearson JM, Platt ML. Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. *J Neurosci* 31: 4178–4187, 2011.

207. Hayden BY, Heilbronner SR, Platt ML. Ambiguity aversion in rhesus macaques. *Front Neurosci* 4: 166: 1–7, 2010.

208. Hayden BY, Pearson JM, Platt ML. Fictive reward signals in the anterior cingulate cortex. *Science* 324: 948–950, 2009.

209. Hayden BY, Pearson JM, Platt ML. Neuronal basis of sequential foraging decisions in a patchy environment. *Nat Neurosci* 14: 933–939, 2011.

210. Hayden BY, Platt ML. Neurons in anterior cingulate cortex multiplex information about reward and action. *J Neurosci* 30: 3339–3346, 2010.

211. Henry DJ, Greene MA, White FJ. Electrophysiological effects of cocaine in the mesoaccumbens dopamine system: Repeated administration. *J Pharm Exp Ther* 251: 833–839, 1989.

212. Hernández A, Zainos A, Romo R. Temporal evolution of a decision-making process in medial premotor cortex. *Neuron* 33: 959–972, 2002.

213. Hernández-López S, Bargas J, Surmeier DJ, Reyes A, Galarraga E. D1 receptor activation enhances evoked discharge in neostriatal medium spiny neurons by modulating an L-type Ca^{2+} conductance. *J Neurosci* 17: 3334–3342, 1997.

214. Hernández-López S, Tkatch T, Perez-Garci E, Galarraga E, Bargas J, Hamm H, Surmeier DJ. D2 dopamine receptors in striatal medium spiny neurons reduce L-type Ca^{2+} currents and excitability via a novel PLCβ1-IP_3-calcineurin-signaling cascade. *J Neurosci* 20, 8987–8995, 2000.

215. Hikosaka O, Sakamoto M, Usui S. Functional properties of monkey caudate neurons. III. Activities related to expectation of target and reward. *J Neurophysiol* 61: 814–832, 1989.

216. Hikosaka K, Watanabe M. Long-range and short-range reward expectancy in the primate orbitofrontal cortex. *Eur J Neurosci* 19: 1046–1054, 2004.

217. Histed MH, Pasupathy A, Miller EK. Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. *Neuron* 63: 244–253, 2009.

218. Ho MY, Mobini S, Chinang TJ, Bradshaw CM, Szabadi E. Theory and method in the quantitative analysis of "impulsive choice" behaviour: implications for psychopharmacology. *Psychopharmacology* 146: 362–372, 1999.

219. Hofstadter DR. *Gödel, Escher, Bach: The Eternal Golden Braid*. New York: Basics Books, 1979, p 711.

220. Holland PC. CS-US interval as a determinant of the form of Pavlovian appetitive conditioned responses. *J Exp Psychol Anim Behav Process* 6: 155–174, 1980.

221. Hollerman JR, Schultz W. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat Neurosci* 1: 304–309, 1998.

222. Hollerman JR, Tremblay L, Schultz W. Influence of reward expectation on behavior-related neuronal activity in primate striatum. *J Neurophysiol* 80: 947–963, 1998.

223. Hollis KL, Pharr VL, Dumas MJ, Britton GB, Field J. Classical conditioning provides paternity advantage for territorial male blue gouramis *(Trichogaster trichopterus)*. *J Comp Psychol* 111: 219–225, 1997.

224. Hollon NG, Arnold MM, Gan JO, Walton ME, Phillips PEM. Dopamine-associated cached values are not sufficient as the basis for action selection. *Proc Natl Acad Sci USA* 111: 18357–18362, 2014.

225. Holt CA, Laury SK. Risk aversion and incentive effects. *Am Econ Rev* 92: 1644–1655, 2002.

226. Hong S, Jhou TC, Smith M, Saleem KS, Hikosaka O. Negative reward signals from the lateral habenula to dopamine neurons are mediated by rostromedial tegmental nucleus in primates. *J Neurosci* 31: 11457–11471, 2011.

227. Hong S, Hikosaka O. The globus pallidus sends reward-related signals to the lateral habenula. *Neuron* 60, 720–729, 2008.

228. Hong S, Hikosaka O. Pedunculopontine tegmental nucleus neurons provide reward, sensorimotor, and alerting signals to midbrain dopamine neurons. *Neuroscience* 282: 139–155, 2014.

229. Horvitz JC, Stewart Jacobs BL. Burst activity of ventral tegmental dopamine neurons is elicited by sensory stimuli in the awake cat. *Brain Res* 759: 251–258, 1997.

230. Horwitz GD, Batista AP, Newsome WT. Representation of an abstract perceptual decision in macaque superior colliculus. *J Neurophysiol* 91: 2281–2296, 2004.

231. Hosokawa T, Kato K, Inoue M, Mikami A. Neurons in the orbitofrontal cortex code both visual shapes and reward types. *NeuroReport* 15: 1493–1496, 2004.

232. Hosokawa T, Kato K, Inoue M, Mikami A. Neurons in the macaque orbitofrontal cortex code relative preference of both rewarding and aversive outcomes. *Neurosci Res* 57: 434–445, 2007.

233. Hosokawa T, Kennerley SW, Sloan J, Wallis JD. Single-neuron mechanisms underlying cost-benefit analysis in frontal cortex. *J Neurosci* 33: 17385–17397, 2013.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

234. Hosokawa T, Watanabe M. Prefrontal neurons represent winning and losing during competitive video shooting games between monkeys. *J Neurosci* 32: 7662–7671, 2012.

235. Hosoya T, Baccus SA, Meister M. Dynamic predictive coding by the retina. *Nature* 436: 71–77, 2005.

236. Houk JC, Adams JL, Barto AG. A model of how the basal ganglia generate and use neural signals that predict reinforcement. In: *Models of Information Processing in the Basal Ganglia*, edited by Houk JC, Davis JL, Beiser DG. Cambridge, MA: MIT Press, 1995, p. 249–270.

237. Howe MW, Tierney PL, Sandberg SG, Phillips PEM, Graybiel AM. Prolonged dopamine signaling in striatum signals proximity and value of distant rewards. *Nature* 500: 575–579, 2013.

238. Huang CF, Litzenberger RH. *Foundations for Financial Economics*. Upper Saddle River, NJ: Prentice-Hall, 1988.

239. Hrupka BJ, Lin YM, Gietzen DW, Rogers QR. Small changes in essential amino acid concentrations alter diet selection in amino acid-deficient rats. *J Nutr* 127: 777–784, 1997.

240. Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF. Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310: 1680–1683, 2005.

241. Hsu M, Krajbich I, Zhao C, Camerer CF. Neural response to reward anticipation under risk is nonlinear in probabilities. *J Neurosci* 29: 2237–2231, 2009.

242. Hull CL. *Principles of Behavior*. New York: Appleton-Century-Crofts, 1943.

243. Ikeda T, Hikosaka O. Reward-dependent gain and bias of visual responses in primate superior colliculus. *Neuron* 39: 693–700, 2003.

244. Ilango A, Kesner AJ, Keller KL, Stuber GD, Bonci A, Ikemoto S. Similar roles of substantia nigra and ventral tegmental dopamine neurons in reward and aversion. *J Neurosci* 34: 817–822, 2014.

245. Ito M, Doya K. Validation of decision-making models and analysis of decision variables in the rat basal ganglia. *J Neurosci* 29: 9861–9874, 2009.

246. Ito S, Stuphorn V, Brown JW, Schall JD. Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science* 302: 120–122, 2003.

247. Izhikevich EM. Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cereb Cortex* 17: 2443–2452, 2007.

248. Janssen P, Shadlen MN. A representation of the hazard rate of elapsed time in macaque area LIP. *Nat Neurosci* 8: 234–241, 2005.

249. Jhou TC, Fields HL, Baxter MB, Saper CB, Holland PC. The rostromedial tegmental nucleus (RMTg), a GABAergic afferent to midbrain dopamine neurons, encodes aversive stimuli and inhibits motor responses. *Neuron* 61: 786–800, 2009.

250. Ji H, Shepard PD. Lateral habenula stimulation inhibits rat midbrain dopamine neurons through a GABAA receptor-mediated mechanism. *J Neurosci* 27: 6923–6930, 2007.

251. Jimenez-Castellanos J, Graybiel AM. Subdivisions of the dopamine-containing A8-A9-A10 complex identified by their differential mesostriatal innervation of striosomes and extrastriosomal matrix. *Neuroscience* 23: 223–242, 1987.

252. Jimenez-Castellanos J, Graybiel AM. Evidence that histochemically distinct zones of the primate substantia nigra pars compacta are related to patterned distributions of nigrostriatal projection neurons and striatonigral fibers. *Exp Brain Res* 74: 227–238, 1989.

253. Jog MS, Kubota Y, Connolly CI, Hillegaart V, Graybiel AM. Building neural representations of habits. *Science* 286: 1745–1749, 1999.

254. Johnson JG, Busemeyer JR. A dynamic, stochastic, computational model of preference reversal phenomena. *Psychol Rev* 112: 841–861, 2005.

255. Joshua M, Adler A, Mitelman R, Vaadia E, Bergman H. Midbrain dopaminergic neurons and striatal cholinergic interneurons encode the difference between reward and aversive events at different epochs of probabilistic classical conditioning trials. *J Neurosci* 28: 11673–11684, 2008.

256. Joshua M, Adler A, Prut Y, Vaadia E, Wickens JR, Hagai Bergman H. Synchronization of midbrain dopaminergic neurons is enhanced by rewarding events. *Neuron* 62: 695–704, 2009.

257. Kagel JH, Battalio RC, Green L. *Economic Choice Theory: An Experimental Analysis of Animal Behavior*. Cambridge, UK: Cambridge Univ. Press, 1995.

258. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. *Econometrica* 47: 263–291, 1979.

259. Kahneman D, Wakker PP, Sarin R. Back to Bentham? Explorations of experienced utility. *Q J Econ* 112: 375–405, 1997.

260. Kawagoe R, Takikawa Y, Hikosaka O. Expectation of reward modulates cognitive signals in the basal ganglia. *Nat Neurosci* 1: 411–416, 1998.

261. Kennerley SW, Behrens TEJ, Wallis JD. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat Neurosci* 14: 1581–1589, 2011.

262. Kennerley SW, Wallis JD. Reward-dependent modulation of working memory in lateral prefrontal cortex. *J Neurosci* 29: 3259–3270, 2009.

263. Kennerley SW, Wallis JD. Evaluating choices by single neurons in the frontal lobe: outcome value encoded across multiple decision variables. *Eur J Neurosci* 29: 2061–2073, 2009.

264. Kepecs A, Uchida N, Zariwala H, Mainen ZF. Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455: 227–231, 2008.

265. Kerr JN, Wickens JR. Dopamine D-1/D-5 receptor activation is required for long-term potentiation in the rat neostriatum in vitro. *J Neurophysiol* 85: 117–124, 2001.

266. Kettner RE, Mahamud S, Leung HC, Sitkoff N, Houk JC, Peterson BW, Barto AG. Prediction of complex two-dimensional trajectories by a cerebellar model of smooth pursuit eye movements. *J Neurophysiol* 77: 2115–2130, 1997.

267. Khamassi M, Quilodran R, Pierre Enel P, Dominey PF, Procyk E. Behavioral regulation and the modulation of information coding in the lateral prefrontal and cingulate cortex. *Cereb Cortex*. In press.

268. Kheramin S, Body S, Ho MY, Velazquez-Martinez DN, Bradshaw CM, Szabadi E, Deakin JFW, Anderson IM. Effects of orbital prefrontal cortex dopamine depletion on inter-temporal choice: a quantitative analysis. *Psychopharmacology* 175: 206–214, 2004.

269. Kiani R, Cueva CJ, Reppas JB, Newsome WT. Dynamics of neural population responses in prefrontal cortex indicate changes of mind on single trials. *Curr Biol* 24: 1542–1547, 2014.

270. Kiani R, Hanks TD, Shadlen MN. Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *J Neurosci* 28: 3017–3029, 2008.

271. Kim KM, Baratta MV, Yang A, Lee D, Boyden ES, Fiorillo CD. Optogenetic mimicry of the transient activation of dopamine neurons by natural reward is sufficient for operant reinforcement. *PLoS One* 7: e33612, 2012.

272. Kim S, Hwang J, Lee D. Prefrontal coding of temporally discounted values during intertemporal choice. *Neuron* 59: 161–172, 2008.

273. Kim JJ, Krupa DJ, Thompson RF. Inhibitory cerebello-olivary projections and blocking effect in classical conditioning. *Science* 279: 570–573, 1998.

274. Kim JN, Shadlen MN. Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nat Neurosci* 2: 176–185, 1999.

275. Kim H, Sul JH, Huh N, Lee D, Jung MW. Role of striatum in updating values of chosen actions. *J Neurosci* 29: 14701–14712, 2009.

276. Kimura M, Rajkowski J, Evarts E. Tonically discharging putamen neurons exhibit set-dependent responses. *Proc Natl Acad Sci USA* 81: 4998–5001, 1984.

277. Kirby KN, Marakovic NN. Delay-discounting probabilistic rewards: rates decrease as amounts increase. *Psychonom Bull Rev* 3: 100–104, 1996.

278. Kitazawa S, Kimura T, Yin PB. Cerebellar complex spikes encode both destinations and errors in arm movement. *Nature* 392: 494–497, 1998.

279. Klein JT, Deaner RO, Platt ML. Neural correlates of social target value in macaque parietal cortex. *Curr Biol* 18: 419–424, 2008.

280. Knutson B, Adams CM, Fong GW, Hommer D. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J Neurosci* 21: 1–5, 2001.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

281. Kobayashi Y, Inoue Y, Yamamoto M, Isa T, Aizawa H. Contribution of pedunculopontine tegmental nucleus neurons to performance of visually guided saccade tasks in monkeys. *J Neurophysiol* 88: 715–731, 2002.

282. Kobayashi S, Lauwereyns J, Koizumi M, Sakagami M, Hikosaka O. Influence of reward expectation on visuospatial processing in macaque lateral prefrontal cortex. *J Neurophysiol* 87: 1488–1498, 2002.

283. Kobayashi S, Nomoto K, Watanabe M, Hikosaka O, Schultz W, Sakagami M. Influences of rewarding and aversive outcomes on activity in macaque lateral prefrontal cortex. *Neuron* 51: 861–870, 2006.

284. Kobayashi S, Pinto de Carvalho O, Schultz W. Adaptation of reward sensitivity in orbitofrontal neurons. *J Neurosci* 30: 534–544, 2010.

285. Kobayashi S, Schultz W. Influence of reward delays on responses of dopamine neurons. *J Neurosci* 28: 7837–7846, 2008.

286. Kobayashi S, Schultz W. Reward contexts extend dopamine signals to unrewarded stimuli. *Curr Biol* 24: 56–62, 2014.

287. Koch C. *Biophysics of Computation*. New York: Oxford Univ. Press, 1999.

288. Komura Y, Nikkuni A, Hirashima N, Uetake T, Miyamoto A. Responses of pulvinar neurons reflect a subject's confidence in visual categorization. *Nat Neurosci* 16: 749–755, 2013.

289. Konorski J. *Integrative Activity of the bBrain*. Chicago: Univ. of Chicago Press, 1967.

290. Köszegi B, Rabin M. A model of reference-dependent preferences. *Q J Econ* 121: 1133–1165, 2006.

291. Krajbich I, Armel C, Rangel R. Visual fixations and the computation and comparison of value in simple choice. *Nat Neurosci* 13: 1292–1298, 2010.

292. Krauzlis RJ, Basso MA, Wurtz RH. Shared motor error for multiple eye movements. *Science* 276: 1693–1695, 1997.

293. Kravitz AV, Tye LD, Kreitzer AC. Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nat Neurosci* 15: 816–818, 2012.

294. Kreitzer AC, Malenka RC. Endocannabinoid-mediated rescue of striatal LTD and motor deficits in Parkinson's disease models. *Nature* 445: 643–647, 2007.

295. Kreps DM. *A Course in Microeconomic Theory*. Harlow: Pearson Education, 1990.

296. Kreps DM, Porteus E. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica* 46: 185–200, 1978.

297. Kuhn A, Aertsen A, Rotter S. Neuronal integration of synaptic input in the fluctuation-driven regime. *J Neurosci* 24, 2345–2356, 2004.

298. Kurata K, Wise SP. Premotor cortex of rhesus monkeys: set-related activity during two conditional motor tasks. *Exp Brain Res* 69: 327–343, 1988.

299. Kurata K, Wise SP. Premotor and supplementary motor cortex in rhesus monkeys: neuronal activity during externally- and internally-instructed motor tasks. *Exp Brain Res* 72: 237–248, 1988.

300. Lak A, Arabzadeh E, Harris JA, Diamond ME. Correlated physiological and perceptual effects of noise in a tactile stimulus. *Proc Natl Acad Sci USA* 107: 7981–7986, 2010.

301. Lak A, Stauffer WR, Schultz W. Dopamine prediction error responses integrate subjective value from different reward dimensions. *Proc Natl Acad Sci USA* 111: 2343–2348, 2014.

302. Laming DRJ. *Information Theory of Choice Reaction Time*. New York: Wiley, 1968.

303. Lammel S, Lim BK, Ran C, Huang KW, Betley MJ, Tye KM, Deisseroth K, Malenka RC. Input-specific control of reward and aversion in the ventral tegmental area. *Nature* 491: 212–217, 2012.

304. Lammel S, Steinberg E, Földy C, Wall NR, Beier K, Luo L, Malenka RC. Diversity of transgenic mouse models for selective Ttrgeting of midbrain dopamine neurons. *Neuron* 85: 429–438, 2015.

305. Lattimore PK, Baker JR, Witte AD. The influence of probability on risky choice: a parametric examination. *J Econ Behav Organ* 17: 377–400, 1992.

306. Lau B, Glimcher PW. Value representations in the primate striatum during matching behavior. *Neuron* 58: 451–463, 2008.

307. Laughlin S. A simple coding procedure enhances a neuron's information capacity. *Z Naturforsch* 36: 910–912, 1981.

308. Lauwereyns J, Takikawa Y, Kawagoe R, Kobayashi S, Koizumi M, Coe B, Sakagami M, Hikosaka O. Feature-based anticipation of cues that predict reward in monkey caudate nucleus. *Neuron* 33: 463–473, 2002.

309. Leathers ML, Olson CR. In monkeys making value-based decisions, LIP neurons encode cue salience and not action value. *Science* 338: 132–135, 2012.

310. Lecourtier L, DeFrancesco A, Moghaddam B. Differential tonic influence of lateral habenula on prefrontal cortex and nucleus accumbens dopamine release. *Eur J Neurosci* 27: 1755–1762, 2008.

311. Lee IH, Assad JA. Putaminal activity for simple reactions or self-timed movements. *J Neurophysiol* 89: 2528–2537, 2003.

312. Lemus L, Hernández A, Luna R, Zainos A, Nácher V, Romo R. Neural correlates of a postponed decision report. *Proc Natl Acad Sci USA* 104: 17174–17179, 2007.

313. Leon MI, Shadlen MN. Effect of expected reward magnitude on the responses of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron* 24: 415–425, 1999.

314. Leung PMB, Rogers QR. Importance of prepyriform cortex in food-intake response of rats to amino acids. *Am J Physiol* 221: 929–935, 1971.

315. Levy H, Markowitz HM. Approximating expected utility by a function of mean and variance. *Am Econ Rev* 69: 308–317, 1979.

316. Libet B, Gleason CA, Wright EW, Pearl DK. Time of conscious intention to act in relation to onset of cerebral activities (readiness-potential): the unconscious initiation of a freely voluntary act. *Brain* 106: 623–642, 1983.

317. Lishman WA. *Organic Psychiatry*. Oxford: Blackwell, 1998.

318. Liu QS, Pu L, Poo MM. Repeated cocaine exposure in vivo facilitates LTP induction in midbrain dopamine neurons. *Nature* 437: 1027–1031, 2005.

319. Liu Z, Richmond BJ. Response differences in monkey TE and perirhinal cortex: stimulus association related to reward schedules. *J Neurophysiol* 83: 1677–1692, 2000.

320. Livio M. *The Golden Ratio: The Story of PHI, the World's Most Astonishing Number*. New York: Random House, 2002.

321. Ljungberg T, Apicella P, Schultz W. Responses of monkey midbrain dopamine neurons during delayed alternation performance. *Brain Res* 586: 337–341, 1991.

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

323. Lo CC, Wang XJ. Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nat Neurosci* 9: 956–963, 2006.

324. Loewenstein G, Thompson L, Bazerman MH. Social utility and decision making in interpersonal contexts. *J Personality Soc Psych* 57: 426–441, 1989.

325. Loewenstein G, Prelec D. Anomalies in intertemporal choice: evidence and an interpretation. *Q J Econ* 107: 573–597, 1992.

326. Leszczuk MH, Flaherty CF. Lesions of nucleus accumbens reduce instrumental but not consummatory negative contrast in rats. *Behav Brain Res* 116: 61–79, 2000.

327. Logan GD, Cowan WB. On the ability to inhibit thought and action: a theory of an act of control. *Psychol Rev* 91: 295–327, 1984.

328. Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412: 150–157, 2001.

329. Lohrenz T, McCabe K, Camerer CF, Montague PR. Neural signature of fictive learning signals in a sequential investment task. *Proc Natl Acad Sci USA* 104: 9493–9498, 2007.

330. Longtin A. Stochastic resonance in neuron models. *J Stat Phys* 70: 309, 1993.

331. Louie K, Glimcher PW. Separating value from choice: delay discounting activity in the lateral intraparietal area. *J Neurosci* 30: 5498–5507, 2010.

332. Louie K, Grattan LE, Glimcher PW. Reward value-based gain control: divisive normalization in parietal cortex. *J Neurosci* 31: 10627–10639, 2011.

333. Luce RD. *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley, 1959.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

334. Machens CK, Romo R, Brody CD. Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science* 307: 1121–1124, 2005.

335. Mackintosh NJ. *The Psychology of Animal Learning*. London: Academic Press, 1974.

336. Mackintosh NJ. A theory of attention: variations in the associability of stimulus with reinforcement. *Psychol Rev* 82: 276–298, 1975.

337. Maimon G, Assad JA. Parietal area 5 and the initiation of self-timed movements versus simple reactions. *J Neurosci* 26: 2487–2498, 2006.

338. Markowitsch HJ, Pritzel M. Reward-related neurons in cat association cortex. *Brain Res* 111: 185–188, 1976.

339. Markowitz H. The utility of wealth. *J Polit Econ* 6: 151–158, 1952.

340. Martínez-García M, Rolls ET, Deco G, Romo R. Neural and computational mechanisms of postponed decisions. *Proc Natl Acad Sci* 108: 11626–11631, 2011.

341. Mas-Colell A, Whinston M, Green J. *Microeconomic Theory*. New York: Oxford Univ. Press, 1995.

342. Matsuda W, Furuta T, Nakamura KC, Hioki H, Fujiyama F, Arai R, Kaneko T. Single nigrostriatal dopaminergic neurons form widely spread and highly dense axonal arborizations in the neostriatum. *J Neurosci* 29: 444–453, 2009.

343. Matsuda Y, Marzo A, Otani S. The presence of background dopamine signal converts long-term synaptic depression to potentiation in rat prefrontal cortex. *J Neurosci* 26: 4803–4810, 2006.

344. Matsumoto M, Hikosaka O. Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature* 447: 1111–1115, 2007.

345. Matsumoto M, Hikosaka O. Two types of dopamine neuron distinctively convey positive and negative motivational signals. *Nature* 459: 837–841, 2009.

346. Matsumoto M, Hikosaka O. Representation of negative motivational value in the primate lateral habenula. *Nat Neurosci* 12: 77–84, 2009.

347. Matsumoto M, Matsumoto K, Abe H, Tanaka K. Medial prefrontal cell activity signaling prediction errors of action values. *Nat Neurosci* 10: 647–656, 2007.

348. Matsumoto K, Suzuki W, Tanaka K. Neuronal correlates of goal-based motor selection in the prefrontal cortex. *Science* 301: 229–232, 2003.

349. Matsumoto M, Takada M. Distinct representations of cognitive and motivational signals in midbrain dopamine neurons. *Neuron* 79: 1011–1024, 2013.

350. Mauritz KH, Wise SP. Premotor cortex of the rhesus monkey: neuronal activity in anticipation of predictable environmental events. *Exp Brain Res* 61: 229–244, 1986.

351. McClure SM, Berns GS, Montague PR. Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 38: 339–346, 2003.

352. McClure SM, Laibson DI, Loewenstein G, Cohen JD. Separate neural systems value immediate and delayed monetary rewards. *Science* 306: 503–507, 2004.

353. McClure SM, Li J, Tomlin D, Cypert KS, Montague LM, Montague PR. Neural correlates of behavioral preference for culturally familiar drinks. *Neuron* 44: 379–387, 2004.

354. McCoy AN, Crowley JC, Haghighian G, Dean HL, Platt ML. Saccade reward signals in posterior cingulate cortex. *Neuron* 40: 1031–1040, 2003.

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

356. McEchron MD, Bouwmeester H, Tseng W, Weiss C, Disterhoft JF. Hippocampectomy disrupts auditory trace fear conditioning and contextual fear conditioning in the rat. *Hippocampus* 8: 638–646, 1998.

357. Medina JF, Nores WL, Mauk MD. Inhibition of climbing fibers is a signal for the extinction of conditioned eyelid responses. *Nature* 416: 330–333, 2002.

358. Meijer JH, Robbers Y. Wheel running in the wild. *Proc R Soc B* 281: 2014.330–0210, 2014.

359. Melis AP, Hare B, Tomasello M. Engineering cooperation in chimpanzees: tolerance constraints on cooperation. *Anim Behav* 72: 275–286, 2006.

360. Merten K, Nieder A. Active encoding of decisions about stimulus absence in primate prefrontal cortex neurons. *Proc Natl Acad Sci USA* 109: 6289–6294, 2012.

361. Merten K, Nieder A. Comparison of abstract decision encoding in the monkey prefrontal cortex, the presupplementary, and cingulate motor areas. *J Neurophysiol* 110: 19–32, 2013.

362. Mileykovskiy B, Morales M. Duration of inhibition of ventral tegmental area dopamine neurons encodes a level of conditioned fear. *J Neurosci* 31: 7471–7476, 2011.

363. Mill JS. *Utilitarianism.* London: Parker, Son and Bourn, 1863.

364. Miller LA. Cognitive risk-taking after frontal or temporal lobectomy. I: Synthesis of fragmented visual information. *Neuropsychologia* 23: 359–369, 1985.

365. Mirenowicz J, Schultz W. Importance of unpredictability for reward responses in primate dopamine neurons. *J Neurophysiol* 72: 1024–1027, 1994.

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

367. Mitchell DS, Gormezano I. Effects of water deprivation on classical appetitive conditioning of the rabbit's jaw movement response. *Learn Motivat* 1: 199–206, 1970.

368. Mink JW. The basal ganglia: focused selection and inhibition of competing motor programs. *Prog Neurobiol* 50: 381–425, 1996.

369. Mizuhiki T, Richmond BJ, Shidara M. Encoding of reward expectation by monkey anterior insular neurons. *J Neurophysiol* 107: 2996–3007, 2012.

370. Mobini S, Body S, Ho MY, Bradshaw CM, Szabadi E, Deakin JFW, Anderson IM. Effects of lesions of the orbitofrontal cortex on sensitivity to delayed and probabilistic reinforcement. *Psychopharmacology* 160: 290–298, 2002.

371. Mogami T, Tanaka K. Reward association affects neuronal responses to visual stimuli in macaque TE and perirhinal cortices. *J Neurosci* 26: 6761–6770, 2006.

372. Monosov IE, Hikosaka O. Regionally distinct processing of rewards and punishments by the primate ventromedial prefrontal cortex. *J Neurosci* 32: 10318–10330, 2012.

373. Monosov IE, Hikosaka O. Selective and graded coding of reward uncertainty by neurons in the primate anterodorsal septal region. *Nat Neurosci* 16: 756–762, 2013.

374. Montague PR, Dayan P, Sejnowski TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* 16: 1936–1947, 1996.

375. Montague PR, Sejnowski TJ. The predictive brain: temporal coincidence and temporal order in synaptic learning mechanisms. *Learn Mem* 1: 1–33, 1994.

376. Morris G, Arkadir D, Nevet A, Vaadia E, Bergman H. Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron* 43: 133–143, 2004.

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

378. Morrison SE, Saez A, Lau B, Salzman CD. Different time courses for learning-related changes in amygdala and orbitofrontal cortex. *Neuron* 71, 1127–1140, 2011.

379. Munoz DP, Wurtz RH. Saccade-related activity in monkey superior colliculus. I. Characteristics of burst and buildup cells. *J Neurophysiol* 73: 2313–2333, 1995.

380. Murakami M, Vicente MI, Costa GM, Mainen ZF. Neuronal antecedents of self-initiated actions in secondary motor cortex. *Nat Neurosci* 17: 1574–1582, 2014.

381. Musallam S, Corneil BD, Greger B, Scherberger H, Andersen RA. Cognitive control signals for neural prosthetics. *Science* 305: 258–262, 2004.

382. Nakahara H, Itoh H, Kawagoe R, Takikawa Y, Hikosaka O. Dopamine neurons can represent context-dependent prediction error. *Neuron* 41: 269–280, 2004.

383. Nakamura K, Mikami A, Kubota K. Activity of single neurons in the monkey amygdala during performance of a visual discrimination task. *J Neurophysiol* 67: 1447–1463, 1992.

384. Nassar MR, Wilson RC, Heasly B, Gold JI. An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J Neurosci* 30: 12366–12378, 2010.

385. Newsome WT, Britten KH, Movshon JA. Neuronal correlates of a perceptual decision. *Nature* 341: 52–54, 1989.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

386. Niki H, Watanabe M. Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain Res* 171: 213–224, 1979.

387. Nishijo H, Ono T, Nishino H. Single neuron responses in amygdala of alert monkey during complex sensory stimulation with affective significance. *J Neurosci* 8: 3570–3583, 1988.

388. Niv Y, Duff MO, Dayan P. Dopamine, uncertainty and TD learning. *Behav Brain Func* 1:6, 2005.

389. Nomoto K, Schultz W, Watanabe T, Sakagami M. Temporally extended dopamine responses to perceptually demanding reward-predictive stimuli. *J Neurosci* 30: 10692–10702, 2010.

390. O'Doherty J, Dayan P, Friston K, Critchley H, Dolan RJ. Temporal difference models and reward-related lNiv earning in the human brain. *Neuron* 28: 329–337, 2003.

391. O'Neill M, Schultz W. Coding of reward risk distinct from reward value by orbitofrontal neurons. *Neuron* 68: 789–800, 2010.

392. O'Neill M, Schultz W. Risk prediction error coding in orbitofrontal neurons. *J Neurosci* 33: 15810–15814, 2013.

393. Ogawa M, van der Meer MAA, Esber GR, Cerri DH, Stalnaker TA, Schoenbaum G. Risk-responsive orbitofrontal neurons track acquired salience. *Neuron* 77: 251–258, 2013.

394. Ohyama K, Sugase-Miyamoto Y, Matsumoto N, Shidara M, Sato C. Stimulus-related activity during conditional associations in monkey perirhinal cortex neurons depends on upcoming reward outcome. *J Neurosci* 32: 17407–17419, 2012.

395. Ojakangas CL, Ebner TJ. Purkinje cell complex and simple spike changes during a voluntary arm movement learning task in the monkey. *J Neurophysiol* 68: 2222–2236, 1992.

396. Okada KI, Toyama K, Inoue Y, Isa T, Kobayashi Y. Different pedunculopontine tegmental neurons signal predicted and actual task rewards. *J Neurosci* 29: 4858–4870, 2009.

397. Okano K, Tanji J. Neuronal activities in the primate motor fields of the agranular frontal cortex preceding visually triggered and self-paced movement. *Exp Brain Res* 66: 155–166, 1987.

398. Olds J, Milner P. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J Comp Physiol Psychol* 47: 419–427, 1954.

399. Ostlund SB, Balleine BW. Differential involvement of the basolateral amygdala and mediodorsal thalamus in instrumental action selection. *J Neurosci* 28: 4398–4405, 2008.

400. Otani S, Blond O, Desce JM, Crepel F. Dopamine facilitates long-term depression of glutamatergic transmission in rat prefrontal cortex. *Neuroscience* 85: 669–676, 1998.

401. Otmakhova NA, Lisman JE. D1/D5 dopamine receptor activation increases the magnitude of early long-term potentiation at CA1 hippocampal synapses. *J Neurosci* 16: 7478–7486, 1996.

402. Oyama K, Hernádi I, Iijima T, Tsutsui KI. Reward prediction error coding in dorsal striatal neurons. *J Neurosci* 30: 11447–11457, 2010.

403. Padoa-Schioppa C. Range-adapting representation of economic value in the orbitofrontal cortex. *J Neurosci* 29: 14004–14014, 2009.

404. Padoa-Schioppa C. Neuronal origins of choice variability in economic decisions. *Neuron* 80: 1322–1336, 2013.

405. Padoa-Schioppa C, Assad JA. Neurons in the orbitofrontal cortex encode economic value. *Nature* 441: 223–226, 2006.

406. Padoa-Schioppa C, Assad JA. The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nat Neurosci* 11: 95–102, 2008.

407. Paladini CA, Celada P, Tepper JM. Striatal, pallidal, and pars reticulata evoked inhibition of nigrostriatal dopaminergic neurons is mediated by GABAa receptors in vivo. *Neuroscience* 89: 799–812, 1998.

408. Pan WX, Brown J, Dudman JT. Neural signals of extinction in the inhibitory microcircuit of the ventral midbrain. *Nat Neurosci* 16: 71–78, 2013.

409. Pan X, Fan H, Sawa K, Tsuda I, Tsukada M, Sakagami M. Reward inference by primate prefrontal and striatal neurons. *J Neurosci* 34: 1380–1396, 2014.

410. Pan WX, Hyland BI. Pedunculopontine tegmental nucleus controls conditioned responses of midbrain dopamine neurons in behaving rats. *J Neurosci* 25: 4725–4732, 2005.

411. Pan X, Sawa K, Tsuda I, Tsukada M, Sakagami M. Reward prediction based on stimulus categorization in primate lateral prefrontal cortex. *Nat Neurosci* 11: 703–712, 2008.

412. Pan WX, 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* 25: 6235–6242, 2005.

413. Parker JG, Beutler LR, Palmiter RD. The contribution of NMDA receptor signaling in the corticobasal ganglia reward network to appetitive Pavlovian learning. *J Neurosci* 31: 11362–11369, 2011.

414. Parker JG, Wanat MJ, Soden ME, Ahmad K, Zweifel LS, Bamford NS, Palmiter RD. Attenuating $GABA_A$ receptor signaling in dopamine neurons selectively enhances reward learning and alters risk preference in mice. *J Neurosci* 31: 17103–17112, 2011.

415. Parker JG, Zweifel LS, Clark JJ, Evans SB, Phillips PEM, Palmiter RD. Absence of NMDA receptors in dopamine neurons attenuates dopamine release but not conditioned approach during Pavlovian conditioning. *Proc Nat Acad Sci USA* 107: 13491–13496, 2010.

416. Parthasarathy HB, Schall JD, Graybiel AM. Distributed but convergent ordering of corticostriatal projections: analysis of the frontal eye field and the supplementary eye field in the macaque monkey. *J Neurosci* 12: 4468–4488, 1992.

417. Pascal B. *Pensées 1658–1662*, translated by Ariew R. Indianapolis: Hackett, 2004.

418. Pasquereau B, Nadjar A, Arkadir D, Bezard E, Goillandeau M, Bioulac B, Gross CE, Boraud T. Shaping of motor responses by incentive values through the basal ganglia. *J Neurosci* 27: 1176–1183, 2007.

419. Pasquereau B, Turner RS. Limited encoding of effort by dopamine neurons in a cost-benefit trade-off task. *J Neurosci* 33: 8288–8300, 2013.

420. Pastor-Bernier A, Cisek P. Neural correlates of biased competition in premotor cortex. *J Neurosci* 31: 7083–7088, 2011.

421. Pasupathy A, Miller EK. Different time courses of learning-related activity in the prefrontal cortex and striaum. *Nature* 433: 873–876, 2005.

422. Paton JJ, Belova MA, Morrison SE, Salzman CD. The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* 439: 865–870, 2006.

423. Pavlov PI. *Conditioned Reflexes*. London: Oxford Univ. Press, 1927.

424. Pawlak V, Kerr JND. Dopamine receptor activation is required for corticostriatal spike-timing-dependent plasticity. *J Neurosci* 28: 2435–2446, 2008.

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

426. Pearson JM, PLatt ML. Dopamine: burning the candle at both ends. *Neuron* 79: 831–833, 2013.

427. Peck CJ, Jangraw DC, Suzuki M, Efem R, Gottlieb J. Reward modulates attention independently of action value in posterior parietal cortex. *J Neurosci* 29: 11182–11191, 2009.

428. Peck CJ, Lau B, Salzman CD. The primate amygdala combines information about space and value. *Nat Neurosci* 16: 340–348, 2013.

429. Penrose R. *The Emperor's New Mind*. Oxford, UK: Oxford Univ. Press, 1989.

430. Percheron G, Yelnik J, Francois C. A Golgi analysis of the primate globus pallidus. III. Spatial organization of the striopallidal complex. *J Comp Neurol* 227: 214–227, 1984.

431. Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* 442: 1042–1045, 2006.

432. Plassmann H, O'Doherty J, Shiv B, Rangel A. Marketing actions can modulate neural representations of experienced pleasantness. *Proc Natl Acad Sci USA* 105: 1050–1054, 2009.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

433. Platt ML, Glimcher PW. Neural correlates of decision variables in parietal cortex. *Nature* 400: 233–238, 1999.

434. Pooresmaeili A, Poort J, Roelfsema PR. Simultaneous selection by object-based attention in visual and frontal cortex. *Proc Natl Acad Sci USA* 111: 6467–6472, 2014.

435. Prelec D. The probability weighting function. *Econometrica* 66: 497–527, 1998.

436. Prelec D, Loewenstein G. Decision making over time and under uncertainty: a common approach. *Management Sci* 37: 770–786, 1991.

437. Preuschoff K, Bossaerts P, Quartz SR. Neural differentiation of expected reward and risk in human subcortical structures. *Neuron* 51: 381–390, 2006.

438. Preuschoff Bossaerts P. Adding prediction risk to the theory of reward learning. *Ann NY Acad Sci* 1104: 135–146, 2007.

439. Prévost C, Pessiglione M, Météreau E, Cléry-Melin ML, Dreher JC. Separate valuation subsystems for delay and effort decision costs. *J Neurosci* 30: 14080–14090, 2010.

440. Puig MV, Miller EK. The role of prefrontal dopamine D1 receptors in the neural mechanisms of associative learning. *Neuron* 74: 874–886, 2012.

441. Purcell BA, Heitz RP, Cohen JY, Schall JD, Logan GD, Palmeri TJ. Neurally constrained modeling of perceptual decision making. *Psychol Rev* 117: 1113–1143, 2010.

442. Rabin M. Incorporating fairness into game theory and economics. *Am Econ Rev* 83: 1281–1302, 1993.

443. Raby CR, Alexis DM, Dickinson A, Clayton NS. Planning for the future by western scrub-jays. *Nature* 445: 919–921, 2007.

444. Raghuraman AP, Padoa-Schioppa C. Integration of multiple determinants in the neuronal computation of economic values. *J Neurosci* 34: 11583–11603, 2014.

445. Rahman S, Sahakian BJ, Hodges JR, Rogers RD, Robbins TW. Specific cognitive deficits in mild frontal variant frontotemporal dementia. *Brain* 122: 1469–1493, 1999.

446. Rao RPN, Sejnowski TJ. Self-organizing neural systems based on predictive learning. *Phil Trans R Soc A* 361: 1149–1175, 2003.

447. Ratcliff R. A theory of memory retrieval. *Psychol Rev* 83: 59–108, 1978.

448. Ratcliff R, Cherian A, Segraves M. A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions. *J Neurophysiol* 90: 1392–1407, 2003.

449. Ratcliff R, Frank MJ. Reinforcement-based decision making in corticostriatal circuits: mutual constraints by neurocomputational and diffusion models. *Neural Comput* 24: 1186–1229, 2012.

450. Ratcliff R, Hasegawa YT, Hasegawa RP, Smith PL, Segraves MA. Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *J Neurophysiol* 97: 1756–1774, 2007.

451. Ravel S, Legallet E, Apicella P. Tonically active neurons in the monkey striatum do not preferentially respond to appetitive stimuli. *Exp Brain Res* 128: 531–534, 1999.

452. Ravel S, Legallet E, Apicella P. Responses of tonically active neurons in the monkey striatum discriminate between motivationally opposing stimuli. *J Neurosci* 23: 8489–8497, 2003.

453. Ravel S, Richmond BJ. Dopamine neuronal responses in monkeys performing visually cued reward schedules. *Eur J Neurosci* 24: 277–290, 2006.

454. Reddi BA, Asrress KN, Carpenter RH. Accuracy, information, and response time in a saccadic decision task. *J Neurophysiol* 90: 3538–3546, 2003.

455. Redgrave P, Prescott TJ, Gurney K. Is the short-latency dopamine response too short to signal reward? *Trends Neurosci* 22: 146–151, 1999.

456. Redgrave P, Gurney K. The short-latency dopamine signal: a role in discovering novel actions? *Nat Rev Neurosci* 7: 967–975, 2006.

457. Redish AD. Addiction as a computational process gone awry. *Science* 306: 1944–1947, 2004.

458. Reed P, Mitchell C, Nokes T. Intrinsic reinforcing properties of putatively neutral stimuli in an instrumental two-lever discrimination task. *Anim Learn Behav* 24: 38–45, 1996.

459. Rescorla RA. Pavlovian conditioning and its proper control procedures. *Psychol Rev* 74: 71–80, 1967.

460. Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: *Classical Conditioning II: Current Research and Theory*, edited by Black AH, Prokasy WF. New York: Appleton Century Crofts, 1972, p. 64–99.

461. Reynolds JNJ, Hyland BI, Wickens JR. A cellular mechanism of reward-related learning. *Nature* 413: 67–70, 2001.

462. Richards JB, Mitchell SH, de Wit H, Seiden LS. Determination of discount functions in rats with an adjusting-amount procedure. *J Exp Anal Behav* 67: 353–366, 1997.

463. Richards CD, Shiroyama T, Kitai ST. Electrophysiological and immunocytochemical characterization of GABA and dopamine neurons in the substantia nigra of the rat. *Neuroscience* 80: 545–557, 1997.

464. Richardson KA, Aston-Jones G. Lateral hypothalamic orexin/hypocretin neurons that project to ventral tegmental area are differentially activated with morphine preference. *J Neurosci* 32: 3809–3817, 2012.

465. Richardson MJE, Gerstner W. Statistics of subthreshold neuronal voltage fluctuations due to conductance based synaptic shot noise. *Chaos* 16: 026106, 2006.

466. Richardson RT, DeLong MR. Context-dependent responses of primate nucleus basalis neurons in a go/no-go task. *J Neurosci* 10: 2528–2540, 1990.

467. Riehle A, Requin J. Monkey primary motor and premotor cortex: single-cell activity related to prior information about direction and extent of an intended movement. *J Neurophysiol* 61: 534–549, 1989.

468. Ringach DL, Hawken MJ, Shapley R. Dynamics of orientation tuning in macaque primary visual cortex. *Nature* 387: 281–284, 1997.

469. Roberts S. Isolation of an internal clock. *J Exp Psychol Anim Behav Proc* 7: 242–268, 1981.

470. Robbins TW, Arnsten AFT. The neuropsychopharmacology of fronto-executive function: monoaminergic modulation. *Annu Rev Neurosci* 32: 267–287, 2009.

471. Robinson TE, Berridge KC. The neural basis for drug craving: an incentive-sensitization theory of addiction. *Brain Res Rev* 18: 247–291, 1993.

472. Robinson MJF, Berridge KC. Instant transformation of learned repulsion into motivational "wanting." *Curr Biol* 23: 282–289, 2013.

473. Rodriguez ML, Logue AW. Adjusting delay to reinforcement: comparing choice in pigeons and humans. *J Exp Psychol Anim Behav Proc* 14: 105–117, 1988.

474. Roelfsema PR, Tolboom M, Khayat PS. Different processing phases for features, figures, and selective attention in the primary visual cortex. *Neuron* 56: 785–792, 2007.

475. Roesch MR, Calu DJ, Esber GR, Schoenbaum G. Neural correlates of variations in event processing during learning in basolateral amygdala. *J Neurosci* 30: 2464–2471, 2010.

476. Roesch MR, Olson CR. Impact of expected reward on neuronal activity in prefrontal cortex, frontal and supplementary eye fields and premotor cortex. *J Neurophysiol* 90: 1766–1789, 2003.

477. Roesch MR, Olson CR. Neuronal activity dependent on anticipated and elapsed delay in macaque prefrontal cortex, frontal and supplementary eye fields, and premotor cortex. *J Neurophysiol* 94: 1469–1497, 2005.

478. Roesch MR, Olson CR. Neuronal activity in orbitofrontal cortex reflects the value of time. *J Neurophysiol* 94: 2457–2471, 2005.

479. Roesch MR, Singh T, Brown PL, Mullins SE, Schoenbaum G. Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards. *J Neurosci* 29: 13365–13376, 2009.

480. Rogers QR, Harper AE. Selection of a solution containing histidine by rats fed a histidine-imbalanced diet. *J Comp Physiol Psychol* 72: 66–71, 1970.

481. Roiser JP, de Martino B, Tan GCY, Kumaran D, Seymour B, Wood NW, Dolan RJ. A genetically mediated bias in decision making driven by failure of amygdala control. *J Neurosci* 29: 5985–5991, 2009.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

482. Roitman JD, Roitman MF. Risk-preference differentiates orbitofrontal cortex responses to freely chosen reward outcomes. *Eur J Neurosci* 31: 1492–1500, 2010.

483. Roitman JD, Shadlen MN. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci* 22: 9475–9489, 2002.

484. Roitman MF, Wheeler RA, Carelli RM. Nucleus accumbens neurons are innately tuned for rewarding and aversive taste stimuli, encode their predictors, and are linked to motor output. *Neuron* 45: 587–597, 2005.

485. Rolls ET, Critchley HD, Mason R, Wakeman EA. Orbitofrontal cortex neurons: role in olfactory and visual association learning. *J Neurophysiol* 75: 1970–1981, 1996.

486. Rolls ET, Grabenhorst F, Parris BA. Warm pleasant feelings in the brain. *NeuroImage* 41: 1504–1513, 2008.

487. Rolls ET, Yaxley S, Sienkiewicz ZJ. Gustatory responses of single neurons in the caudolateral orbitofrontal cortex of the macaque monkey. *J Neurophysiol* 64: 1055–1066, 1990.

488. Romo R, Brody CD, Hernández A, Lemus L. Neuronal correlates of parametricworking memory in the prefrontal cortex. *Nature* 399: 470–473, 1999.

489. Romo R, Hernández A, Zainos A. Neuronal correlates of a perceptual decision in ventral premotor cortex. *Neuron* 41: 165–173, 2004.

490. Romo R, Hernández A, Zainos A, Lemus L, Brody CD. Neuronal correlates of decision-making in secondary somatosensory cortex. *Nat Neurosci* 5: 1217–1225, 2002.

491. Romo R, Schultz W. Neuronal activity preceding self-initiated or externally timed arm movements in area 6 of monkey cortex. *Exp Brain Res* 67: 656–662, 1987.

492. Romo R, Schultz W. Dopamine neurons of the monkey midbrain: contingencies of responses to active touch during self-initiated arm movements. *J Neurophysiol* 63: 592–606, 1990.

493. Romo R, Schultz W. Role of primate basal ganglia and frontal cortex in the internal generation of movements. III. Neuronal activity in the supplementary motor area. *Exp Brain Res* 91: 396–407, 1992.

494. Rorie AE, Gao J, McClelland JL, Newsome WT. Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP) of the macaque monkey. *PLoS One* 5: e9308, 2010.

495. Rosenkranz JA, Grace AA. Dopamine-mediated modulation of odour-evoked amygdala potentials during pavlovian conditioning. *Nature* 417: 282–287, 2002.

496. Rossi MA, Fan D, Barter JW, Yin HH. Bidirectional modulation of substantia nigra activity by motivational state. *PLoS One* 8: e71598, 2013.

497. Rothschild M, Stiglitz IA. Increasing risk: definition. *J Econ Theory* 2: 225–243, 1970.

498. Rutledge RB, Skandalia N, Dayan P, Dolan RJ. A computational and neural model of momentary subjective well-being. *Proc Natl Acad Sci USA* 111: 12252–12257, 2014.

499. Salamone JD. The involvement of nucleus accumbens dopamine in appetitive and aversive motivation. *Behav Brain Res* 61: 117–133, 1994.

500. Samejima K, Ueda Y, Doya K, Kimura M. Representation of action-specific reward values in the striatum. *Science* 310: 1337–1340, 2005.

501. Samuelson P. A note on measurement of utility. *Rev Econ Stud* 4: 155–161, 1937.

502. Sato M, Hikosaka O. Role of primate substantia nigra pars reticulata in reward-oriented saccadic eye movement. *J Neurosci* 22: 2363–2373, 2002.

503. Satoh T, Nakai S, Sato T, Kimura M. Correlated coding of motivation and outcome of decision by dopamine neurons. *J Neurosci* 23: 9913–9923, 2003.

504. Savage LJ. *The Foundations of Statistics*. New York: Wiley, 1954.

505. Sawaguchi T, Goldman-Rakic PS. D1 dopamine receptors in prefrontal cortex: involvement in working memory. *Science* 251: 947–950, 1991.

506. Schall JD. Neuronal activity related to visually guided saccadic eye movements in the supplementary motor area of rhesus monkeys. *J Neurophysiol* 66: 530–558, 1991.

507. Schall JD, Stuphorn V, Brown JW. Monitoring and control of action by the frontal lobes. *Neuron* 36: 309–322, 2002.

508. Schall JD, Thompson KG. Neural selection and control of visually guided eye movements. *Annu Rev Neurosci* 22: 241–259, 1999.

509. Scherberger H, Andersen RA. Target selection signals for arm reaching in the posterior parietal cortex. *J Neurosci* 27: 2001–2012, 2007.

510. Schmalfuss B. The random attractor of the stochastic Lorenz system. *Z Angew Math Physik* 48: 951–975, 1997.

511. Schmidt R, Leventhal DK, Mallet N, Chen F, Berke JD. Canceling actions involves a race between basal ganglia pathways. *Nat Neurosci* 16: 1118–1124, 2013.

512. Schoenbaum G, Chiba AA, Gallagher M. Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning. *J Neurosci* 19: 1876–1884, 1999.

513. Schreiber CA, Kahneman D. Determinants of the remembered utility of aversive sounds. *J Exp Psych* 129: 27–42, 2000.

514. Schroeder T. *Three Faces of Desire*. Boston: MIT Press, 2004.

515. Schurger A, Jacobo Sitt JD D, Dehaene S. An accumulator model for spontaneous neural activity prior to self-initiated movement. *J Proc Natl Acad Sci USA* 109: E2904–E2913, 2012.

516. Schultz W. Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. *J Neurophysiol* 56: 1439–1462, 1986.

517. Schultz W. Predictive reward signal of dopamine neurons. *J Neurophysiol* 80: 1–27, 1998.

518. Schultz W. Multiple dopamine functions at different time courses. *Annu Rev Neurosci* 30: 259–288, 2007.

519. Schultz W. Midbrain dopamine neurons: a retina of the reward system? In: *Neuroeconomics: Decision Making and the Brain*, edited by Glimcher PW, Camerer CF, Fehr E, Poldrack RA. New York: Academic, 2009, p. 323–329.

520. Schultz W. Potential vulnerabilities of neuronal reward, risk, and decision mechanisms to addictive drugs. *Neuron* 69: 603–617, 2011.

521. 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* 13: 900–913, 1993.

522. Schultz W, Apicella P, Romo R, Scarnati E. Context-dependent activity in primate striatum reflecting past and future behavioral events. In: *Models of Information Processing in the Basal Ganglia*, edited by Houk JC, Davis JL, Beiser DG. Cambridge, MA: MIT Press, 1995, p. 11–28.

523. Schultz W, Apicella P, Scarnati E, Ljungberg T. Neuronal activity in monkey ventral striatum related to the expectation of reward. *J Neurosci* 12: 4595–4610, 1992.

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

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

526. Schultz W, Romo R. Neuronal activity in the monkey striatum during the initiation of movements. *Exp Brain Res* 71: 431–436, 1988.

527. Schultz W, Romo R. Dopamine neurons of the monkey midbrain: contingencies of responses to stimuli eliciting immediate behavioral reactions. *J Neurophysiol* 63: 607–624, 1990.

528. Schultz W, Romo R. Role of primate basal ganglia and frontal cortex in the internal generation of movements. I. Preparatory activity in the anterior striatum. *Exp Brain Res* 91: 363–384, 1992.

529. Searle JR. *Intentionality.* Cambridge, UK: Cambridge Univ. Press, 1983.

530. Seo H, Barraclough DJ, Lee D. Dynamic signals related to choices and outcomes in the dorsolateral prefrontal cortex. *Cereb Cortex* 17: -i110–i117, 2007.

531. Seo H, Lee D. Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J Neurosci* 27: 8366–8377, 2007.

532. Seo H, Lee D. Behavioral and neural changes after gains and losses of conditioned reinforcers. *J Neurosci* 29: 3627–3641, 2009.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

533. Seo M, Lee E, Averbeck BB. Action selection and action value in frontal-striatal circuits. *Neuron* 74: 947–960, 2012.

534. Seymour B, Daw ND, Roiser JP, Dayan P, Dolan R. Serotonin selectively modulates reward value in human decision-making. *J Neurosci* 32: 5833–5842, 2012.

535. Shadlen MN, Newsome WT. Neural basis of a perceptual decision in the parietal cortex (Area LIP) of the rhesus monkey. *J Neurophysiol* 86: 1916–1936, 2001.

536. Sheafor PJ. Pseudoconditioned jaw movements of the rabbit reflect associations conditioned to contextual background cues. *J Exp Psychol Anim Behav Proc* 104: 245–260, 1975.

537. Sheafor PJ, Gormezano I. Conditioning the rabbit's (*Oryctolagus cuniculus*) jaw-movement response: US magnitude effects on URs, CRs, and pseudo-CRs. *J Comp Physiol Psychol* 81: 449–456, 1972.

538. Shefrin HM, Thaler RH. The behavioral life-cycle hypothesis. *Econ Inq* 26: 609–643, 1988.

539. Sheffield FD, Roby TB. Reward value of a non-nutritive sweet taste. *J Comp Physiol Psychol* 43: 471–481, 1950.

540. Shen W, Flajolet M, Greengard P, Surmeier DJ. Dichotomous dopaminergic control of striatal synaptic plasticity. *Science* 321: 848–851, 2008.

541. Shidara M, Aigner TG, Richmond BJ. Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. *J Neurosci* 18: 2613–2625, 1998.

542. Shidara M, Richmond BJ. Anterior cingulate: single neuron signals related to degree of reward expectancy. *Science* 296: 1709–1711, 2002.

543. Shima K, Tanji J. Role for cingulate motor area cells in voluntary movement selection based on reward. *Science* 282: 1335–1338, 1998.

544. Simmons JM, Richmond BJ. Dynamic changes in representations of preceding and upcoming reward in monkey orbitofrontal cortex. *Cereb Cortex* 18: 93–103, 2008.

545. Simon HA. Rational choice and the structure of the environment. *Psychol Rev* 63: 129–138, 1956.

546. Singh S, Lewis RL, Barto AG. Where do rewards come from? In: *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, edited by Taatgen NAvan Rijn H. Austin, TX: Cogn. Sci. Soc., 2009, p. 2601–2606.

547. Singh S, Lewis RL, Barto AG, Sorg J. Intrinsically motivated reinforcement learning: an evolutionary perspective. *IEEE Trans Autonom Mental Dev* 2: 70–82, 2010.

548. Smith AD, Bolam JP. The neural network of the basal ganglia as revealed by the study of synaptic connections of identified neurones. *Trends Neurosci* 13: 259–265, 1990.

549. Smith PL, Ratcliff R. Psychology and neurobiology of simple decisions. *Trends Neurosci* 27: 161–168, 2004.

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

551. So NY, Stuphorn V. Supplementary eye field encodes reward prediction error. *J Neurosci* 32: 2950–2963, 2012.

552. Solomon RL, Corbit JD. An opponent-process theory of motivation. *Psychol Rev* 81: 119–145, 1974.

553. Solomon PR, Vander Schaaf ER, Thompson RF, Weisz DJ. Hippocampus and trace conditioning of the rabbit's classically conditioned nictitating membrane response. *Behav Neurosci* 100: 729–744, 1986.

554. Soltani A, Lee D, Wang XJ. Neural mechanism for stochastic behaviour during a competitive game. *Neur Networks* 19: 1075–1090, 2006.

555. Soon CS, Brass M, Heinze HJ, Haynes JD. Unconscious determinants of free decisions in the human brain. *Nat Neurosci* 11: 543–545, 2008.

556. Stalnaker TA, Calhoon GG, Ogawa M, Roesch MR, Schoenbaum G. Reward prediction error signaling in posterior dorsomedial striatum is action specific. *J Neurosci* 32: 10296–10305, 2012.

557. Stanisor L, van der Togt C, Cyriel MA, Pennartz CMA, Roelfsema PR. A unified selection signal for attention and reward in primary visual cortex. *Proc Natl Acad Sci USA* 110: 9136–9141, 2013.

558. Stanton SJ, Mullette-Gillman O'DA, McLaurin RE, Kuhn CM, LaBar KS, Platt ML, Huettel SA. Low- and high-testosterone individuals exhibit decreased aversion to economic risk. *Psychol Sci* 22: 447–453, 2011.

559. Stauffer WR, Lak A, Bossaerts P, Schultz W. Economic choices reveal probability distortion in monkeys. *J Neurosci* 35: 3146–3154, 2015.

560. Stauffer WR, Lak A, Schultz W. Dopamine reward prediction error responses reflect marginal utility. *Curr Biol* 24: 2491–2500, 2014.

561. Stein RB. Some models of neuronal variability. *Biophys J* 7: 37–68, 1967.

562. Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH. A causal link between prediction errors, dopamine neurons and learning. *Nat Neurosci* 16: 966–973, 2013.

563. Steinfels GF, Heym J, Strecker RE, Jacobs BL. Behavioral correlates of dopaminergic unit activity in freely moving cats. *Brain Res* 258: 217–228, 1983.

564. Stevens CF. Inferences about membrane properties from electrical noise measurements. *Biophys J* 12: 1028–1047, 1972.

565. Stone M. Models for choice reaction time. *Psychometrika* 25: 251–260, 1960.

566. Stopper CM, Tse MTL, David Montes DR R, Wiedman CR, Floresco SB. Overriding phasic dopamine signals redirects action selection during risk/reward decision making. *Neuron* 84: 177–189, 2014.

567. Strassberg AF, DeFelice LJ. Limitations of the Hodgkin-Huxley formalism: effects of single channel kinetics on transmembrane voltage dynamics. *Neur Comp* 5: 843–855, 1993.

568. Stuber GD, Klanker M, de Ridder B, Bowers MS, Joosten RN, Feenstra MG, Bonci A. Reward-predictive cues enhance excitatory synaptic strength onto midbrain dopamine neurons. *Science* 321: 1690–1692, 2008.

569. Stuphorn V, Taylor TL, Schall JD. Performance monitoring by the supplementary eye field. *Nature* 408: 857–860, 2000.

570. Sugam JA, Day JJ, Wightman RM, Carelli RM. Phasic nucleus accumbens dopamine encodes risk-based decision-making behavior. *Biol Psychiat* 71: 199–215, 2012.

571. Sugrue LP, Corrado GS, Newsome WT. Matching behavior and the representation of value in the parietal cortex. *Science* 304: 1782–1787, 2004.

572. Sul JH, Kim H, Huh N, Lee D, Jung MW. Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. *Neuron* 66: 449–460, 2010.

573. Suri R, Schultz W. A neural network with dopamine-like reinforcement signal that learns a spatial delayed response task. *Neuroscience* 91: 871–890, 1999.

574. Sutton RS, Barto AG. Toward a modern theory of adaptive networks: expectation and prediction. *Psychol Rev* 88: 135–170, 1981.

575. Sutton RS, Barto AG. *Reinforcement Learning*. Cambridge, MA: MIT Press, 1998.

576. Tai LH, Lee AM, Benavidez N, Bonci A, Wilbrecht L. Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. *Nat Neurosci* 15: 1281–1289, 2012.

577. Takahashi T. A mathematical framework for probabilistic choice based on information theory and psychophysics. *Med Hypoth* 67: 183–186, 2006.

578. Takahashi YK, Roesch MR, Stalnaker TA, Haney RZ, Calu DJ, Taylor AR, Burke KA, Schoenbaum G. The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes. *Neuron* 62: 269–280, 2009.

579. Takikawa Y, Kawagoe R, Hikosaka O. A possible role of midbrain dopamine neurons in short- and long-term adaptation of saccades to position-reward maping. *J Neurophysiol* 92: 2520–2529, 2004.

580. Talwar SK, Xu S, Hawley ES, Weiss SA, Moxon KA, Chapin JK. Rat navigation guided by remote control. *Nature* 417: 37–38, 2002.

581. Tan CO, Bullock D. A local circuit model of learned striatal and dopamine cell responses under probabilistic schedules of reward. *J Neurosci* 28: 10062–10074, 2008.

582. Tan KR, Yvon C, Turiault M, Mirzabekov JJ, Doehner J, Labouèbe G, Deisseroth K, Tye KM, Lüscher C. GABA neurons of the VTA drive conditioned place aversion. *Neuron* 73: 1173–1183, 2012.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

583. Tang KC, Low MJ, Grandy DK, Lovinger DM. Dopamine-dependent synaptic plasticity in striatum during in vivo development. *Proc Natl Acad Sci USA* 98: 1255–1260, 2001.

584. Tanimoto H, Heisenberg M, Gerber B. Event timing turns punishment to reward. *Nature* 430: 983, 2004.

585. Teodorescu TR, Usher M. Disentangling decision models: from independence to competition. *Psychol Rev* 120: 1–38, 2013.

586. Tesauro G. TD-Gammon, a self-teaching backgammon program, achieves master-level play. *Neural Comp* 6: 215–219, 1994.

587. Thompson KG, Bichot NP, Schall JD. Dissociation of visual discrimination from saccade programming in macaque frontal eye field. *J Neurophysiol* 77:1046–1050, 1997.

588. Thompson KG, Hanes DP, Bichot NP, Schall JD. Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. *J Neurophysiol* 76: 4040–4055, 1996.

589. Thorndike EL. *Animal Intelligence: Experimental Studies*. New York: MacMillan, 1911.

590. Thut G, Schultz W, Roelcke U, Nienhusmeier M, Maguire RP, Leenders KL. Activation of the human brain by monetary reward. *NeuroReport* 8: 1225–1228, 1997.

591. Tindell AJ, Smith KS, Peciña S, Berridge KC, Aldridge JW. Ventral pallidum firing codes hedonic reward: when a bad taste turns good. *J Neurophysiol* 96: 2399–2409, 2006.

592. Thorpe SJ, Rolls ET, Maddison S. The orbitofrontal cortex: neuronal activity in the behaving monkey. *Exp Brain Res* 49: 93–115, 1983.

593. Tinklepaugh OL. An experimental study of representation factors in monkeys. *J Comp Psychol* 8: 197–236, 1928.

594. Toan DL, Schultz W. Responses of rat pallidum cells to cortex stimulation and effects of altered dopaminergic activity. *Neuroscience* 15: 683–694, 1985.

595. Tobler PN, Christopoulos GI, O'Doherty JP, Dolan RJ, Schultz W. Neuronal distortions of reward probability without choice. *J Neurosci* 28:11703–11711, 2008.

596. Tobler PN, Christopoulos GI, O'Doherty JP, Dolan RJ, Schultz W. Risk-dependent reward value signal in human prefrontal cortex. *Proc Natl Acad Sci USA* 106: 7185–7190, 2009.

597. Tobler PN, Dickinson A, Schultz W. Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. *J Neurosci* 23: 10402–10410, 2003.

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

599. Tobler PN, O'Doherty JP, Dolan R, Schultz W. Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *J Neurophysiol* 97: 1621–1632, 2007.

600. Tremblay L, Hollerman JR, Schultz W. Modifications of reward expectation-related neuronal activity during learning in primate striatum. *J Neurophysiol* 80: 964–977, 1998.

601. Tremblay L, Schultz W. Relative reward preference in primate orbitofrontal cortex. *Nature* 398: 704–708, 1999.

602. Tremblay L, Schultz W. Reward-related neuronal activity during go-nogo task performance in primate orbitofrontal cortex. *J Neurophysiol* 83: 1864–1876, 2000.

603. Tremblay L, Schultz W. Modifications of reward expectation-related neuronal activity during learning in primate orbitofrontal cortex. *J Neurophysiol* 83: 1877–1885, 2000.

604. Tsai CT, Nakamura S, Iwama K. Inhibition of neuronal activity of the substantia nigra by noxious stimuli and its modification by the caudate nucleus. *Brain Res* 195: 299–311, 1980.

605. Tsai HC, Zhang F, Adamantidis A, Stuber GD, Bonci A, de Lecea L, Deisseroth K. Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science* 324: 1080–1084, 2009.

606. Tsujimoto S, Genovesio A, Wise SP. Monkey orbitofrontal cortex encodes response choices near feedback time. *J Neurosci* 29: 2569–2574, 2009.

607. Umeno MM, Goldberg ME. Spatial processing in the monkey frontal eye field. I. Predictive visual responses. *J Neurophysiol* 78: 1373–1383, 1997.

608. Ungless MA, Magill PJ, Bolam JP. Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. *Science* 303: 2040–2042, 2004.

609. Usher M, McClelland JL. The time course of perceptual choice: the leaky, competing accumulator model. *Psych Rev* 108: 550–592, 2011.

610. Valenza E, Simion F, Macchi Cassia V, Umiltà C. Face Preference at Birth. *J Exp Psych Hum Percept Perform* 22: 892–903, 1996.

611. Van Vreeswijk C, Sompolinsky H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274, 1724–1726, 1996.

612. Van Wolkenten M, Brosnan SF, de Waal F. Inequity responses of monkeys modified by effort. *Proc Nat Acad Sci USA* 104: 18854–18859, 2007.

613. Van Zessen R, Phillips JL, Budygin EA, Stuber GD. Activation of VTA GABA neurons disrupts reward consumption. *Neuron* 73: 1184–1194, 2012.

614. Vickers D. Evidence for an accumulator model of psychophysical discrimination. *Ergonomics* 13: 37–58, 1970.

615. Vickery TJ, Chun MM, Lee D. Ubiquity and specificity of reinforcement signals throughout the human brain. *Neuron* 72: 166–177, 2011.

616. Vijayraghavan S, Wang M, Birnbaum SG, Williams GV, Arnsten AFT. Inverted-U dopamine D1 receptor actions on prefrontal neurons engaged in working memory. *Nat Neurosci* 10: 376–384, 2007.

617. Von Neumann J, Morgenstern O. *The Theory of Games and Economic Behavior*. Princeton, NJ: Princeton Univ. Press, 1944.

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

619. Wald A, Wolfowitz J. Optimum character of the sequential probability ratio test. *Ann Math Statist* 19: 326–339, 1947.

620. Wallis JD, Miller EK. Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Eur J Neurosci* 18: 2069–2081, 2003.

621. Wanat MJ, Kuhnen CM, Phillips PEM. Delays conferred by escalating costs modulate dopamine release to rewards but not their predictors. *J Neurosci* 30: 12020–12027, 2010.

622. Wang XJ. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* 36: 955–968, 2002.

623. Wang XJ. Decision making in recurrent neuronal circuits. *Neuron* 60: 215–234, 2008.

624. Wang LP, Li F, Wang D, Xie K, Wang D, Shen X, Tsien JZ. NMDA receptors in dopaminergic neurons are crucial for habit learning. *Neuron* 72: 1055–1066, 2011.

625. Watabe-Uchida M, Zhu L, Ogawa SK, Vamanrao A, Uchida N. Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* 74, 858–873, 2012.

626. Watanabe M. Prefrontal unit activity during associative learning in the monkey. *Exp Brain Res* 80: 296–309, 1990.

627. Watanabe M. Reward expectancy in primate prefrontal neurons. *Nature* 382: 629–632, 1996.

628. Watanabe M, Hikosaka K, Sakagami M, Shirakawa SI. Coding and monitoring of behavioral context in the primate prefrontal cortex. *J Neurosci* 22: 2391–2400, 2002.

629. Watson KK, Platt ML. Social signals in primate orbitofrontal cortex. *Curr Biol* 22: 2268–2273, 2012.

630. Weber BJ, Chapman GB. Playing for peanuts: why is risk seeking more common for low-stakes gambles? *Organiz Behav Human Dec Proc* 97: 31–46, 2005.

631. Weber EU, Johnson EJ. Decisions under uncertainty: psychological, economic, and neuroeconomic explanations of risk preference. In: *Neuroeconomics* , edited by Glimcher PW, Camerer CF, Fehr E, Poldrack RA. London: Academic, 2009.

632. Weber EU, Milliman RA. Perceived risk attitudes: relating risk perception to risky choice. *Management Sci* 43: 123–144, 1997.

633. Weber EU, Shafir S, Blais AR. Predicting risk sensitivity in humans and lower animals: risk as variance or coefficient of variation. *Psychol Rev* 111: 430–445, 2004.

634. Wegner DM. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press, 2002.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93

635. West EA, Forcelli PA, McCuea DL, Malkova L. Differential effects of serotonin-specific and excitotoxic lesions of OFC on conditioned reinforcer devaluation and extinction in rats. *Behav Brain Res* 246: 10–14, 2013.

636. Weinrich M, Wise SP. The premotor cortex of the monkey. *J Neurosci* 2: 1329–1345, 1982.

637. Wickens J, Kötter R. Cellular models of reinforcement. In: *Models of Information Processing in the Basal Ganglia*, edited by Houk JC, Davis JL, Beiser DG. Cambridge, MA: MIT Press, 1995, p. 187–214.

638. Wightman RM, Robinson DL. Transient changes in mesolimbic dopamine and their association with "reward." *J Neurochem* 82: 721–735, 2002.

639. Williams ZM, Eskandar EN. Selective enhancement of associative learning by microstimulation of the anterior caudate. *Nat Neurosci* 4: 562–568, 2006.

640. Wise RA. Brain reward circuitry: insights from unsensed incentives. *Neuron* 36: 229–240, 2002.

641. Witten IB, Steinberg EE, Lee SY, Davidson TJ, Zalocusky KA, Brodsky M, Yizhar O, Cho SL, Gong S, Ramakrishnan C, Stuber GD, Tye KM, Janak PH, Deisseroth K. Recombinase-driver rat lines: tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron* 72: 721–733, 2011.

642. Yagishita S, Hayashi-Takagi A, Ellis-Davies GCR, Urakubo H, Ishii S, Kasai H. A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science* 345: 1616–1620, 2014.

643. Yang T, Shadlen MN. Probabilistic reasoning by neurons. *Nature* 447: 1075–1080, 2007.

644. Yamada H, Tymula A, Louie K, Glimcher PW. Thirst-dependent risk preferences in monkeys identify a primitive form of wealth. *Proc Natl Acad Sci USA* 110: 15788–15793, 2013.

645. Yasuda M, Yamamoto S, Hikosaka O. Robust representation of stable object values in the oculomotor basal ganglia. *J Neurosci* 32: 16917–16932, 2012.

646. Yin HH, Ostlund SB, Knowlton BJ, Balleine BB. The role of the dorsomedial striatum in instrumental conditioning. *Eur J Neurosci* 23: 513–523, 2005.

647. Yoshida K, Saito N, Iriki A, Isoda M. Social error monitoring in macaque frontal cortex. *Nat Neurosci* 15, 1307–1312, 2012.

648. Zhang JC, Lau PM, Bi GQ. Gain in sensitivity and loss in temporal contrast of STDP by dopaminergic modulation at hippocampal synapses. *Proc Natl Acad Sci USA* 106: 1328–1333, 2009.

649. Zweifel LS, Argilli E, Bonci A, Palmiter R. Role of NMDA receptors in dopamine neurons for plasticity and addictive behaviors. *Neuron* 59: 486–496, 2008.

650. Zweifel LS, Parker JG, Lobb CJ, Rainwater A, Wall VZ, Fadok JP, Darvas M, Kim MJ, Mizumori SJ, Paladini CA, Philipps PEM, Palmiter R. Disruption of NMDAR-dependent burst firing by dopamine neurons provides selective assessment of phasic dopamine-dependent behavior. *Proc Natl Acad Sci USA* 106, 7281–7288, 2009.

Gustavo Antonio Augusto Pereira de Almeida - gustavo.3.5.p@gmail.com - CPF: 461.327.018-93