Mathematical Models of Behavior of Individual Animals

Vladimir L. Tsibulsky* and Andrew B. Norman

Division of Neuroscience, Department of Psychiatry, College of Medicine, University of Cincinnati, Cincinnati, OH 45267-0559, USA

Abstract: This review is focused on mathematical modeling of behaviors of a whole organism with special emphasis on models with a clearly scientific approach to the problem that helps to understand the mechanisms underlying behavior. The aim is to provide an overview of old and contemporary mathematical models without complex mathematical details. Only deterministic and stochastic, but not statistical models are reviewed. All mathematical models of behavior can be divided into two main classes. First, models that are based on the principle of teleological determinism assume that subjects choose the behavior that will lead them to a better payoff in the future. Examples are game theories and operant behavior models both of which are based on the matching law. The second class of models are based on the principle of causal determinism, which assume that subjects do not choose from a set of possibilities but rather are compelled to perform a predetermined behavior in response to specific stimuli. Examples are perception and discrimination models, drug effects models and individual-based population models. A brief overview of the utility of each mathematical model is provided for each section.

INTRODUCTION

There are two major goals of studying behavior of living organisms: to determine the mechanisms underlying a specific behavioral act and to determine mechanisms of learning (i.e., changing of behavioral act). These two mechanisms represent two sides of a coin – adaptation to changing environment. Traditionally, ethologists focused on the relative roles of nature (genotype, innate or inherited responses, consummatory behavior, unconditioned reflex) and behaviorists focused on the relative role of nurture (individual differences, learned or acquired responses, appetitive behavior, conditioned reflex) in animal behavior.

Behavior is the manner in which something functions or operates. Behavior of living organisms includes foraging (the act of looking or searching for food, water or provisions), defense (protecting the physical or functional integrity of body), reproduction (mating, parenting), navigation, and social cooperation and communication. The study of behavior aims to help us to understand when and why animals, including humans, act.

Simply defined, behavior is the organized movement of an organism. Even plants behave in a simple sense, and the study of animal behavior includes the behavior of protists and even bacteria which are not animals by definition. To achieve movements, there must be effectors - contractile elements, cilia, flagellae, or muscles. To initiate and stop movements, there must be an internal controlling and coordinating system - the nervous system or something analogous to it. To organize behavior according to inner or outer environment, there must be receptors – sensitive elements. Information flows from sensors, through the nervous system to effectors. Sherrington's Principles of the Common Path and of Reciprocal Innervation state that in order for behavior to be adaptive and effective the number of effectors should be smaller than the number of receptors and that antagonistic effectors (for example, flexors and extensors of the same joint) should not be excited simultaneously.

The study of behavior, though, extends beyond looking at just the outward physical manifestation (movement). Behavior of organisms can be better understood if we know how parts of an organism behave (sub-organismic level) and how community (population, colony, flock etc.) behaves as a whole (supra-organismic level). Behavior at the lower level is typically called function, for example, one can study the function of the hypophysis or the function of a worker honeybee.

A major pitfall in the study of behavior is uncritical adaptationist thinking [1]. There's a temptation, when a behavior is observed, to conjecture about the adaptive value of the behavior. The appropriate approach is to formulate hypotheses for the function of the behavior, and to test the hypotheses in a careful, controlled manner. On the other hand, there is a reasonable expectation that with appropriate testing, much of animal behavior is adaptive, and completely abandoning a search for adaptive explanations for behavior may cause one to ignore important underpinnings of behavior. Some behavior, though, is clearly maladaptive. The assumption that all of an animal's behavior is in its own self-interest is clearly false. We have to accept that not all behavior has an adaptive explanation and that some behavior can be pathological or maladaptive. Examples of behavior that is maladaptive include a parasite exerting behavioral control over its host, aggressive or sexual behavior targeted at an inappropriate object, drug addiction etc.

*Address correspondence to this author at the Dept. of Psychiatry, University of Cincinnati, College of Medicine, PO Box 670559, Cincinnati, OH. 45267-0559, USA; Tel: (513) 558 1740; Fax: (513) 558 0042; E-mail: vladimir.tsibulsky@uc.edu

1381-6128/07 $50.00+.00

Adaptationism is a form of teleology. The German philosopher Immanuel Kant said, that we cannot describe nature without referring to a teleological cause. Kant however thought that teleology is not a science in the strict sense. Teleology and causality are complimentary concepts of explanation. The idea of a science in the strict sense is to explain a phenomenon with a natural-physical cause. Teleology, adaptationism, for example, helps us to describe the function in a meaningful way but we should continue looking for a scientific causal explanation of it. As ethologists are tempted to think that all behavior is adaptive, similarly behaviorists have a temptation to think that all behavior is motivated to seek pleasure or to avoid pain. Hedonism is another form of teleology. Why do I eat? Proximally, because I feel hungry, I feel the urge to eat, and it feels better to have eaten; distally, it is because such tendencies led to better survival and reproduction than indifference to food in my ancestors.

Modeling is an investigative technique that uses a mathematical or physical representation of a system that accounts for all or some of its known properties. The simpler the model the weaker is the temptation to use teleological explanations. Therefore modeling is a powerful tool of scientific analysis. If a simple model can account for essential properties of a system, i.e., it passes some form of the Turing test (the mathematical model gives the same answers as the modeled system), then there is little room left for teleology. It is important to distinguish rational mathematical models as opposed to empirical models. Empirical equations are simply equations which happen to fit certain data. They are not derived from scientific hypotheses, and consequently the use of empirical equations does not enhance our knowledge of any scientific theory. A rational equation is derived from certain basic theories and represents adequately the relationships that will be found in the experimental data, if the theory is true. If the equation does not fit the data, it proves that the tested theory cannot be applied to the particular data.

Models can be also deterministic and stochastic. Let's assume that we are analyzing a typical behavioral experiment that is the result of a sequence of trials or observations. Each observation can be described in terms of a finite set of possible outcomes. This set is the same for each experiment. Then a possible history for the entire experiment may be represented by a sequence of different outcomes. If our goal is to assign probabilities to each event then our approach is stochastic. If all probabilities are close to unity then the process and the model are deterministic. In this case our goal is to determine the magnitude of the effect.

The aim of this review is to draw a landscape of old and contemporary models without complex mathematical details. All existing models of effects of drugs on behavior are presented to the best of our knowledge. This review is focused on mathematical modeling of behaviors of a whole organism with special emphasis on models with a clearly scientific approach to the problem that helps to understand the mechanisms underlying behavior. Only deterministic and stochastic, but not statistical models are reviewed. A brief overview of the utility of each mathematical model is provided for each section.

1. PERCEPTION, DETECTION AND DISCRIMINATION

Psychophysics is the oldest and perhaps the most controversial area of psychology. As early as 1825, Johann Herbart attempted to employ mathematical equations to describe the interaction between ideas in consciousness. Although his mathematical model is obsolete, his concept of an absolute threshold happened to be very useful. In the 1850s, Gustav von Fechner proposed that corresponding to the physical world there is a psychological world. In the physical world, material objects can be measured in physical units, and in the psychological world, sensations can be measured in psychological units. Based on the ideas of D. Bernoulli and Ernst H. Weber, Fechner concluded that psychological sensations (R) are a logarithmic function of physical values (S). A century later, Stevens [2] argued that the relationship between physical and psychological variables is exponential.

Because the two "laws" do not agree, equations (1.a.4) *vs.* (1.a.5), Stevens said, "honor Fechner, but repeal his law." Magnitude estimations are not a linear function of category ratings, nor are they linearly related to scales based on the subtractive signal detection theory. Instead, magnitude estimations tend to be approximately an exponential function of category ratings. This conflict of scales caused much disagreement in psychophysical scaling.

The relationship between the two scales depends on the stimulus range, spacing, and frequency distribution and on other contextual features of the study. Some psychophysicists concluded [3] that the conflicts of scales might be due to contextual effects, because the exact results for either category ratings or magnitude estimations could be altered by changing the procedures and stimuli used in the experiment.

a. Weber-Fechner's Law

After measuring how accurately the intensity of a stimulus is estimated by a subject, experimenters usually report results in percentages. Humans can discriminate the energy of stimuli with approximately the same accuracy (about 5%) regardless of the quality of the stimulus. Over a wide range of magnitudes, this constancy is true and it is called Weber's Law of *perception threshold*:

$$\Delta S = k\, S_0 \qquad (1.a.1)$$

where ΔS is the minimal difference (threshold) between two signals S_1 and S_0, which is just noticeable, and k is the Weber constant. Weber's Law (1834) states that the ratio of the increment threshold to the background intensity is relatively constant. So when you are in a noisy environment you must shout to be heard while a whisper works in a quiet room. And when you measure increment thresholds on various intensity backgrounds, the thresholds increase in proportion

to the background. This equation (1.a.1) does not usually hold either for very weak or very strong stimuli. Weber constant should be called Weber fraction because it is not really constant.

Assume that all along the scale, variability of the signal which is barely distinguishable from the standard signal S is proportional to the variability of the standard signal. It means that the difference must be some constant proportion C_1 of σ_S, i.e., $\Delta S = C_1\sigma_S$. Suppose now that a stimulus has n units. If each unit is thought of as a random component with mean μ and variance σ^2, and if the various units vary independently, then the mean and variance of the stimulus will be given as $\mu_n = n\mu$ and $\sigma_n^2 = n\sigma^2$. The standard deviation of the stimulus will be $\sigma\sqrt{n}$. Since the number of unit stimuli is proportional to the physical intensity S of the stimulus $\sigma_S = C_2\sqrt{S}$. Combining the two equations ($\Delta S = C_1\ \sigma_S$, $\sigma_S = C_2\sqrt{S}$) and assuming $C = C_1C_2$ gives:

$$\Delta S = C\sqrt{S} \qquad (1.a.2)$$

This is obviously a very different equation and sometimes it's called the Rose-DeVries law [4, 5].

Little significance was attached to Weber's Law, which was not even stated in mathematical terms originally, until Fechner theorized in 1860 [6] that if equation (1.a.1) holds for any standard stimulus (S_0) and for any definition of "just noticeable difference" (ΔS) which are physiologically equal then Weber's Law actually implies that:

$$\mathrm{d}R = k\ \mathrm{d}S / S_0 \qquad (1.a.3)$$

that is physiological response R is proportional to the difference between signals S_1 and S_0 ($\mathrm{d}S = S_1 - S_0$). Integrating both sides of equation (1.a.3) gives Fechner's Law which states that perceptual response (R) or sensation is proportional to the logarithm of the stimulus:

$$R = K\log(S) + b \text{ or } \qquad \Delta R = K\log(S / S_o) + b \qquad (1.a.4)$$

where K is coefficient of proportionality different from k in Weber's Law. The response to the standard stimulus ($S = S_0$) is equal to b. If the standard stimulus is equal to the absolute threshold ($S = \Delta S$) then the response is infinitely small ($b \rightarrow 0$).

According to the logarithmic function, when stimuli bear the same physical ratio, the psychological differences will be equal (i.e., if $S_1 / S_2 = S_3 / S_4$, then $R_1 - R_2 = R_3 - R_4$). In 1957, S.S. Stevens showed [2] that for many different sensations Fechner's law is not really adequate. He assumed that it was because Weber and Fechner defined the ability to discriminate among stimuli as the basic unit of sensation. He devised a way of measuring the psychological perception of a response on a more absolute scale, and found that so-called *prothetic* stimuli (questions concerning how bright a surface appears on its own, or how bright a flash of light appears) are better described by a power law:

$$R = K\,(S / S_o)^{a} \qquad (1.a.5)$$

This relation is called Stevens' power law [2]. According the power law, equal physical ratios produce equal sensation ratios, not equal sensation differences (i.e., if $S_1 / S_2 = S_3 / S_4$, then $R_1 / R_2 = R_3 / R_4$).

One of the most versatile of the functions proposed as "the psychophysical law" is the power law, not only because of the variety of curves it yields, but also because of the variety of interpretations of terms. Guilford [7] was one of the first to formulate the power law ($\Delta R = KR^n$) and to see its wide applicability to psychophysical data, and Stevens believes it to be the only valid psychophysical law. If n = 1, it becomes the Weber Law (1.a.1), and if n = ½, it becomes the Rose-DeVries law (1.a.2). Stevens introduced a new task, called cross-modality matching, in which the subject was instructed to report the sensation produced on one continuum by adjusting the value of a stimulus on another dimension whose sensation matched in intensity. For example to express the heaviness of a lifted weight, the subject could adjust the loudness of a tone so that the loudness matched the heaviness. If subjective values are power functions of physical values on all continua, then cross-modality matches should be power functions of one another, with exponents predictable from the ratio of exponents determined by the method of magnitude estimation.

Both logarithmic and exponential laws of psychophysics represent deterministic, static models with elements of stochastic approach (the Rose-DeVries law) to sensations or responses to physical stimuli. Although the independent variable has a physical dimension, the dependent variable is non-physical by nature and therefore does not have a physical dimension. Therefore, the dependent variable is subjective and can be measured by a questionnaire in humans and cannot be measured in animals.

b. Theory of Perceptual Relativity

The simplest "physical' approach to perception of some object corresponds to the pattern of energy transmitted from that object to receptors. This may lead to some erroneous predictions which can be corrected only by considering special psychological factors like "illusion", "bias" etc. From the physical point of view, perception is either correct or incorrect. In contrast, the theory of perceptual relativity [3] states that perception depends on the overall pattern or flux of energy received by the observer:

$$J = S / A \qquad (1.b.1)$$

where J is the judged magnitude of a stimulus of physical magnitude S, and A is the adaptation level. According to this idea, all factors in the perceptual field, including the past perceptual experience, contribute to the adaptation level A. The founder of the theory, Harry Helson, assumed that contributions are weighted as geometric mean of the physical magnitudes of all the effective stimuli:

$$A = [\ S_1^{\omega_1} S_2^{\omega_2} \ldots S_N^{\omega_N}\ ]^{\,1/(\omega_1+\omega_2+\ldots\omega_N)} \qquad (1.b.2)$$

Taking logarithms yields:

$$\log A = \left[\sum_{i=1}^{N} \omega_i \log S_i\right] \Big/ \sum_{i=1}^{N} \omega_i \qquad (1.b.3)$$

In experiments when two stimuli should be compared, properties of the adaptation level can be studied. For two physical stimuli of the same modality but different magnitude, at the point of subjective equality $J_1 = J_2$ the following equality must hold:

$$S_1 / A_1 = S_2 / A_2 \text{ or } S_2 = S_1 (A_2 / A_1) \quad (1.b.4)$$

By changing the background two stimuli may be judged as equal. However, people can also make judgments of magnitude when there is no obvious standard of comparison present. In the theory of category scales given by Michels and Helson [8] it is assumed that the category of any stimulus depends on the ratio of that stimulus to the adaptation level. A stimulus with magnitude equal to the adaptation level is judged in the middle category C_0, and the remaining categories are located so that there are equal ratios between the stimuli that are classified in adjacent categories.

Category	C_{-j}	…	C_{-2}	C_{-1}	C_0	C_{+1}	C_{+2}	C_{+3}	…	C_{+n}
Typical Stimulus	$S_0 r^{-j}$	…	$S_0 r^{-2}$	$S_0 r^{-1}$	S_0	$S_0 r^{1}$	$S_0 r^{2}$	$S_0 r^{3}$	…	$S_0 r^{n}$

The relationship between stimulus magnitude and categories can be given in two ways. First, the midpoint of category C_i, I steps from A, depends on the adaptation level A and the ratio r according to:

$$M_i = Ar^j \quad \text{or} \quad \log(M_i) = \log(A) + i\log(r) \quad (1.b.5)$$

According to the Michels-Helson theory, the subject considers the stimulus M_0 at adaptation level A, and realizes that his scale has $N = n + j$ categories above and below that level. Then he fixes the category just below the adaptation level so that:

$$M_0 - M_{-1} = A / N \qquad \text{or} \qquad M_{-1} = A - A / N$$

Bearing in mind that $i = -1$ and substituting in equation (1.b.5) yields

$$r = N / (N - 1) \quad (1.b.6)$$

This simple equation has interesting implications. Suppose that a stimulus having physical measurement $S = 20$ is presented, and the subject gives a judgment $i = -3$, i.e., puts it in the category –3. Then, according to equation (1.b.5):

$$\log A = \log(20) - (-3)\log(1.25) \text{ or } A = 39.3$$

Thus we can estimate the adaptation level. However, the adaptation level depends not only on environment but also on the subject itself. It has been demonstrated that the adaptation level shifts with each presentation of a stimulus, decreasing with low level of stimulus and increasing when an intense stimulus is presented. A corresponding self-adaptation constant was introduced. With some further refinement, it may be possible to establish a formal theory of adaptation level on a firm mathematical foundation [9].

The perceptual relativity model represents a deterministic, dynamic model which is far more sophisticated and advanced compared to any of those described in Section 1.a. Nevertheless, it has to deal with the parameters which are intrinsically subjective with all unavoidable limitations.

2. CHOICE AND PREFERENCE

The concept of choice in philosophy, a corollary of the proposition of free will, i.e., the ability voluntarily to decide to perform one of several possible acts or to avoid action entirely, is one of the most controversial concepts applied in behavioral sciences. One standard approach to the problem of choice is to say that motivational, experiential, and perceptual factors combine to yield strength of a given response, and the process of responding involves some sort of comparison of these strengths. It implies that the strength of behavior can be best measured by the probability to choose this behavior over many alternatives. The classical strength theory explains variability in choice by assuming that response strength oscillates. A theory of the behavior ratio assumes that response strengths are constant, but there is variability in the process of choosing. Interestingly, the two theories lead to approximately the same predictions, despite the mathematical difference.

a. Behavior Competition

The strength of habit R was denoted by Hull [10] as ${}_SE_{Ri}$ and it was assumed constant for each habit. However, there is a special oscillating factor ${}_SO_R$ subtracted from each response strength. The value of ${}_SO_{Ri}$ varies according to a normal distribution with mean μ_i and standard deviation σ_i. If we assume that two responses R_1 and R_2 oscillate with equal standard deviation from their corresponding means and the threshold for response is very low, then the probability of response R_1 is the same as the probability that ${}_SE_{R1}$ is greater than ${}_SE_{R2}$. The difference between strengths of these two responses is normally distributed with mean $\mu_1 - \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$. The response R_1 will be chosen when the difference between mean $\mu_1 - \mu_2$ is positive and *vice versa*.

Because the difference between any two independent normally distributed variables is itself normally distributed, then the moment-generating function of a normal distribution of the difference will be a product of the moment-generating functions of the two separate distributions:

$$M_{a+b}(t) = M_a(t) M_b(t) = \left(e^{t\mu_a + \frac{1}{2}t^2\sigma_a^2} \right) \left(e^{t(-\mu_b) + \frac{1}{2}t^2\sigma_b^2} \right) = e^{t(\mu_a - \mu_b) + \frac{1}{2}t^2(\sigma_a^2 + \sigma_b^2)} \quad (2.a.1)$$

Equation (2.a.1) holds for the situation of the choice between two alternatives, but this theory may be extended to the analysis of more than two responses if they are presented two at a time. The case of three choices at a time can be found in Restle and Greeno [9].

An alternative to the classical theory where the response strengths oscillate is the behavior ratio theory where the process of choosing oscillates. The probability of choice is proportional to the strength of response [11]. The theory is based on the assumption that the relative probability of any two alternatives would remain unchanged as other alternatives are introduced or eliminated. If $v(x)$ is a value of an alternative x then probability of choosing a over b is:

$$P(a,b) = v(a) / (v(a) + v(b)) \quad (2.a.2)$$

dividing this equation by $P(b,a) = v(b) / (v(a) + v(b))$ yields:

$$v(a) / v(b) = P(a,b) / P(b,a) \quad (2.a.3)$$

In situation with three alternatives the same ratio should hold and therefore we can derive:

$$P(a,c) / P(c,a) = P(a,b)P(b,c) / P(b,a)P(c,b) \quad (2.a.4)$$

The behavior ratio theory, however, cannot explain why people will choose the ten-dollar bill over the five-dollar bill with a probability close to unity instead of a probability close to 0.66 as this theory predicts. Another paradox appears if we compare the probability of choice between alternatives a and b before and after their values are changed by the same amount $v(d)$. According to Luce's theory [11] the new probability should become

$$P(a,b) = (v(a) + v(d)) / (v(a) + v(b) + 2v(d))$$

which is obviously not equal to the initial probability. These paradoxes can be resolved [12] if we assume that common elements of two complex alternatives equally contribute to the choice of both alternatives and therefore do not influence it. Therefore, equation (2.a.2) should be rewritten as:

$$P(a,b) = v(a - b) / (v(a - b) + v(b - a)) \quad (2.a.5)$$

Models presented in this section are derived from the idea that motivational behavior is deterministic behavior in Hull's tradition. Competition between behaviors is a competition between stimuli inducing these behaviors. In the two following sections, models will be based on the teleological determinism of choice.

b. Game Theory

Game theory allows biologists to model how evolution affects strategies and tactics. When thinking about an evolutionary game, the biologist precisely defines the participants, their goals, and the actions they can take to meet the goals. Generally, there are many different possible strategies but some are obviously faulty. An animal will never win a competition for food, for example, if its strategy is to not eat. Choosing the right strategy (or gaining the winning strategy via natural selection) is the center point of success in either games or life. By analogy, then, the theory of human games applies to animal behavior. Game theory does not answer all of our questions about animal behavior, but it gives a solid framework in which we can understand the behavioral choices made by animals.

A new branch of mathematics was founded by John von Neumann and Oskar Morgenstern in their book "Theory of Games and Economic Behavior" [13]. Although von Neumann had laid the mathematical foundation of the theory of games in his earlier paper [14], it was largely due to his collaboration with Morgenstern that economists learned of this new tool for analyzing economic problems. John Nash proposed [15] a kind of optimal collective strategy in a game involving two or more players, where no player has anything to gain by changing only his or her own strategy. If each player has chosen a strategy and no player can benefit by changing his or her strategy while the other players keep theirs unchanged, then the current set of strategy choices and the corresponding payoffs constitute *Nash equilibrium* (NE).

Let (S, f) be a game, where S is the set of strategy profiles and f is the set of payoff profiles. When each player i in n-player game ($i \in [1, n]$) chooses strategy x_i ($x_i \in S$) resulting in strategy profile $x = (x_1, ..., x_n)$ then player i obtains payoff $f_i(x)$. A strategy profile $x^* \in S$ is a Nash equilibrium if no deviation in strategy by any single player is profitable, that is, if for all i

$$f_i(x_i, x^*_{-i}) \leq f_i(x^*) \quad (2.b.1)$$

A game can have a pure strategy NE or a NE in its mixed extension (that of choosing a pure strategy stochastically with a fixed frequency). Nash proved that, if we allow *mixed strategies* (players choose strategies randomly according to pre-assigned probabilities), then every n-player game in which every player can choose from a finite number of strategies admits at least one Nash equilibrium.

Let σ_{-i} be a mixed strategy profile of all players except for player i. We can define a best response correspondence for player i, b_i. b_i is a relation from the set of all probability distributions over opponent player profiles to a set of player i's strategies, such that each element of $b_i(\sigma_{-i})$ is a best response to σ_{-i}. Define $b(\sigma) = b_1(\sigma_{-1}) \cdot b_2(\sigma_{-2}) \cdot \ldots \cdot b_n(\sigma_{-n})$. One can use the Kakutani fixed point theorem to prove that b has a fixed point. That is, there is a σ^* such that $\sigma^* \in b(\sigma^*)$. Since $b(\sigma^*)$ represents the best response for all players to σ^*, the existence of the fixed point proves that there is some strategy set which is a best response to itself. No player could do any better by deviating, and it is therefore a Nash equilibrium.

Driving on a road, and having to choose either to drive on the left or to drive on the right of the road, is an example of a simple coordination game with payoffs 1 meaning no crash and payoffs 0 meaning a crash. This coordination game can be defined with the following payoff matrix:

	Drive on the left	Drive on the right
Drive on the left	1,1 (*A,A*)	0,0 (*B,C*)
Drive on the right	0,0 (*C,B*)	1,1 (*D,D*)

where the payoffs are according to $A = D$ and $B = C$. In this case there are two pure strategy Nash equilibria, when both choose to either drive on the left or on the right. If we admit

mixed strategies (where a pure strategy is chosen at random, subject to some fixed probability), then there are three Nash equilibria for the same case: two we have seen from the pure-strategy form, where the probabilities are (0%, 100%) for player A, (0%, 100%) for player B; and (100%, 0%) for player A, (100%, 0%) for player B, respectively. We add another where the probabilities for each player are equal (50%, 50%).

An animal's choice of strategy may be based on genetic background (essentially, information gained by the action of natural selection on previous generations), experience, and/or an assessment of conditions. Natural selection provides a testing ground for strategies; computer simulations allow biologists to simulate the interaction of strategies. A strategy that cannot be dislodged by other strategies is called an "Evolutionarily Stable Strategy", or an ESS. The concept of the ESS, is an important part of game theory. An ESS is a strategy which, over evolutionary time, is able to withstand the invention of new strategies. Although Maynard-Smith and Price [16], who introduced the concept, visualized strategies as being genetically encoded, this same logic applies to strategies which are learned during the course of an animal's life. In most models of the *prisoner's dilemma* the "tit for tat" strategy is evolutionarily stable; over time it can beat any other strategy that you might invent for this game.

The classical prisoner's dilemma is as follows: Two suspects A and B are arrested by the police. The police have insufficient evidence for a conviction, and having separated both prisoners, visit each of them and offer the same deal: if one testifies for the prosecution against the other and the other remains silent, the silent accomplice receives the full 10-year sentence and the betrayer goes free. If both stay silent, the police can only give both prisoners 6 months for a minor charge. If both betray each other, they receive a 2-year sentence each. It can be summarized thus:

	Prisoner A Stays Silent	Prisoner A Betrays
Prisoner B Stays Silent	Both serve six months (*A,A*)	Prisoner B serves ten years; Prisoner A goes free (*B,C*)
Prisoner B Betrays	Prisoner A serves ten years; Prisoner B goes free (*C,B*)	Both serve two years (*D,D*)

The prisoner's dilemma has the same payoff matrix as depicted for the driving coordination game, but now $C > A > D > B$. Because $C > A$ and $D > B$, each player improves his situation by switching from strategy #1 (stay silent) to strategy #2 (betray), no matter what the other player decides. The prisoner's dilemma thus has a single Nash equilibrium: both players choosing strategy #2 ("betraying"). What has long made this an interesting case to study is the fact that $2D < 2A$ ("both betray" is globally inferior to "both remain loyal"). The globally optimal strategy is unstable; it is not an equilibrium.

Games become more complicated if strategies contain conditional responses. In the prisoner's dilemma, for example, player "A" (human or animal) might respond in one way if "B" has just defected, and another if "B" has just cooperated. The most complicated conditional strategies involve learning, so that a player can change its behavior depending on the personality of its opponent.

An ESS, is a strategy which cannot be dislodged by the presence of another strategy in the population of animals. In some games there is only one ESS, and all successful animals play that ESS.

In his book Robert Axelrod [17] explored an extension to the classical prisoner's dilemma scenario, which he called the iterated prisoner's dilemma (IPD). In IPD, participants have to choose their mutual strategy again and again, and have memory of their previous encounters. Axelrod invited academic colleagues all over the world to devise computer strategies to compete in an IPD tournament. The programs that were entered varied widely in algorithmic complexity; initial hostility; capacity for forgiveness; and so forth. Axelrod discovered that when these encounters were repeated over a long period of time with many players, each with different strategies, "greedy" strategies tended to do very poorly in the long run while more "altruistic" strategies did better, as judged purely by self-interest. He used this to show a possible mechanism for the evolution by natural selection of altruistic behavior from mechanisms that are initially purely selfish.

The best deterministic strategy for IPD was found to be "Tit for Tat", which Anatol Rapoport developed and entered into the tournament. Based on the English saying meaning "equivalent retaliation" ("tip for tap"), it was the simplest of any program entered, containing only four lines of BASIC, and won the contest. The strategy is simply to cooperate on the first iteration of the game; after that, do what your opponent did on the previous move. If the opponent previously was cooperative, the subject is cooperative. If not, the subject is not. A slightly better strategy is "Tit for Tat with forgiveness". When the opponent defects, on the next move the subject sometimes cooperates anyway with small probability (around 1%-5%). This allows for occasional recovery from getting trapped in a cycle of defections. The exact probability depends on the line-up of opponents. "Tit for Tat with forgiveness" is best when miscommunication is introduced to the game. That means that sometimes your move is incorrectly reported to your opponent: you cooperate but your opponent hears that you defected. Despite the fact that only "always-defect" strategy is only strict Nash equilibrium (not "always-cooperate" and "tit-for-tat" strategies), evolutionary oscillations can be observed among all three strategies [18].

Analyses of fighting behavior are particularly well suited for game theory analysis [19-21]. In fights, each animal has clear-cut tactics which it can employ, and these can be analyzed as a series of interchanges within the fight. The choices of which tactic to play, when to escalate, and when to submit or flee are all moves in a game. Assessment of the opponent's strength, size, and commitment to winning the fight are important in many animal conflicts, and assessment can be included in the model.

In general, size and possession of a resource give a fighter considerable advantages. Most studies of animal fighting show that the winner can be easily predicted by these factors. The effect of resource possession is somewhat surprising, as the resource (a territory, a mate, or a prey item) should have equal value to the two contestants. The holder of the resource may be not so much defending the resource itself as defending the previous effort put into gaining the resource and defending it. The ability of an animal to maintain possession of a resource is termed its resource holding power (RHP); measurement of RHP can give a way of comparing animals' prospects for winning contests.

Applications of the Game Theory to analysis of animal behavior are probably the most intriguing approaches to mathematical modeling of behavior. Teleological determinism of these approaches is well justified and fruitful. Mathematics of these models is too complex to be presented in this review.

c. Matching Law and Melioration Theory

Herrnstein's melioration theory is a theory of choice [22]. However, it views behavior, in contrast to game theory, not as a result of analysis of information and selection between alternative behaviors. The theory has been developed on the basis of the matching law which was found empirically. According to the matching law, a choice is neither an internal decision nor an isolated output of an internal decision process. Herrnstein assumed that a choice is a rate of overt events strung out over time. Therefore, the model predicts not whether the animal will mate, drink, or sleep at this particular moment but the proportion of the time it will spend mating, drinking, or sleeping. So, the relative frequency of responding is a dependent variable. A pigeon, for example, may face two alternative keys that are identical in every way except that the one delivers more pellets per peck than the other. Experiments might determine how the pigeon's distribution of pecks on the keys depends on how often the keys deliver pellets. The matching law says that if an interval of time may be divided into several alternative activities (in example with a pigeon it will be pecks on left and right keys, P_L and P_R), animals will allocate their behavior to the activities in exact proportion to the relative rate of reinforcement (R_L and R_R) actually obtained during these activities. This equality is called matching, and the statement of the equality is called the matching law:

$$P_L / (P_L + P_R) = R_L / (R_L + R_R) \quad (2.c.1)$$

As long as the alternative responses are equivalent in topography (symmetrical choices) and the reinforcements vary only in frequency or amount, the distribution of choices is well predicted. A symmetric choice is the choice among "nearly-equivalent" alternatives depending on the differences in the alternatives. The "near-equivalence" of alternatives defines the "symmetry" of the problem; the differences, the way the symmetry is broken or restricted. The equation should be reinterpreted for the responses or reinforcements varying significantly in quality (such as apples versus oranges or sex versus eating).

Generally, confronted with choices differing only in the frequency of reinforcement, subjects match the distribution of response (P) alternatives to the distribution of reinforcements (R):

$$P_1 / (P_1 + P_2 + P_3 + \ldots + P_n) = R_1 / (R_1 + R_2 + R_3 + \ldots + R_n) \quad (2.c.2)$$

For the behavior in the situation of symmetric choice, there is a fixed quantity that is to be distributed symmetrically among N similar alternative states S_1, S_2, S_3, …, S_n. The fractional share, $f(S_i)$, appointed to the ith state, S_i, may depend on all the numbers in the experiment, as seen from a reference state S_i. The crucial requirement, however, is that $f(S_i)$ depend only on the pattern of those numbers as seen from S_i. One can expect that there is a fixed response to a given pattern that is the same no matter on which key that pattern appears. Therefore, there is a unique formula, $f(S_i)$, for all the states, without the restriction of symmetry, each of the n states might have a different one:

$$f(S_1) + f(S_2) + \ldots + f(S_n) = 1 \quad (2.c.3)$$

The initial analysis of symmetrical choice in terms of group symmetry was developed by Natapoff [23].

The matching law was a pure empirical finding. The question became what kind of relationship should exist between the rate of response and the rate of reinforcement. Herrnstein demonstrated [24] that the absolute rate of responding at each alternative can obey one or the other of two simple functions and still produce matching when concatenated. The simpler of the two principles is a direct proportionality between responding and reinforcement:

$$P = kR \quad (2.c.4)$$

If each of the two alternatives were to obey this proportionality and if each had the same value of k (which is to assume comparable, "symmetrical" responses), then matching would be the result:

$$P_L / (P_L + P_R) = kR_L / (kR_L + kR_R) = R_L / (R_L + R_R) \quad (2.c.5)$$

The second principle is a direct proportionality between responding and relative reinforcement, as follows:

$$P_L = kR_L / (R_L + R_R) \quad (2.c.6)$$

It, too, can readily be shown to imply matching across a pair of comparable alternative responses:

$$P_L / (P_L + P_R) = [kR_L / (kR_L + kR_R)] / [kR_L / (kR_L + kR_R) + kR_R / (kR_L + kR_R)] = R_L / (R_L + R_R) \quad (2.c.7)$$

It has been shown in many experiments that when there is more than one alternative, any change in reinforcement affects all responses in the directions described by equation (2.c.6) whereas equation (2.c.4) stands only in situations of choice between response and no-response (no alternative reinforcements). In the case of n choices the following hyperbolic equation can be written:

$$P_1 = kR_1 / \sum_{i=0}^{n} R_i \quad (2.c.8)$$

where k is the sum of all alternative behavior distributions ($P_0 + P_1 + P_2 + \ldots + P_n$). This equation (2.c.8) seems to collapse in the simplest case of only one defined alternative because numerator and denominator cancel. However, there are likely to be other sources of reinforcement in real situations anyway, for example, for exploratory behavior or sleep. As a practical matter, the denominator will exceed the explicitly programmed by the experimenter reinforcement by some amount R_0, but by how much can be only inferred. Later this parameter was called an extraneous reinforcement and was notated as r_e. The distribution (probability) of response P was then interpreted as the rate of response R [25]. The conventional forms of the matching law for the single-alternative behavior are:

$$R / (R + R_e) = r / (r + r_e) \text{ or } R = kr / (r + r_e) \qquad (2.c.9)$$

Over a hundred experiments showed that equation (2.c.8) can account for the absolute frequency of responding to the individual alternatives in a variety of choice procedures and in a variety of species [26,27]. Nevertheless, questions continue to be raised about matching as an exact principle. Staddon [28] showed that equation (2.c.9) can be derived as the equilibrium solution. Assume that an animal can perform two different behaviors with the probability p and q, where $p + q = 1$, and rates of responses P_0 and P_1 are proportional to these probabilities, i.e., $P_0 = Kq$ and $P_1 = Kp$. Then we can write the scheme of the reversible process:

$$p \underset{R_1}{\overset{R_0}{\rightleftharpoons}} q;\ \mathrm{d}p/\mathrm{d}t = -R_0 p \text{ and } \mathrm{d}q/\mathrm{d}t = -R_1 q$$

For equilibrium we can write the equation like the binding Hill equation, which is identical to equation (2.c.9):

$$P = KR_1 / (R_1 + R_0)$$

Some researchers believe that the matching law is a tautology, because the reinforcement is defined as a factor controlling the rate of response (Skinner's definition of reinforcement, [29]) and, respectively, the rate of response is proportional to the rate of reinforcement (Herrnstein's matching law, [24]). As a tautology, the matching law is not subject to empirical test [30] and is an implicit definition of utility, i.e., reinforcing value [31]. In fact, Herrnstein postulated that "reinforcement serves only to govern the distribution of behavior over the alternatives and nothing more, as regards its effects on frequencies of behavior" [32]. This definition of reinforcement does not contradict the definition given by Skinner but puts it in context of multiple reinforcements/ behaviors which compete for the time. The total amount of behavior equals the number of time units in a period of observation; however, they were apportioned across alternatives. This means that when an increase in reinforcement produces an increase in the frequency of one alternative (Skinner' definition), other alternatives must decrease in frequency so as to hold constant the total amount of behavior (Herrnstein's definition). This invariance requirement is given by the equation:

$$\sum R = k \qquad (2.c.10)$$

The invariance assumption states that it is in principle possible to cut through differences in topography to a level of behavior measurement at which the effects of reinforcement leave total behavior unchanged – the condition for the matching law to hold.

When total behavior is measured invariantly according to (2.c.10), the ratio between all behavior, k, and all reinforcement, $\sum r$, defines a quantity, $k/\sum r$, of behavior per unit reinforcement. Since total reinforcement is in principle unbounded and total behavior is assumed to be invariant, there must be more behavior per reinforcement the less the total reinforcement is, and *vice versa*. This ratio looks like a Weber's Law for reinforcement: a change in effect (reinforcement) on behavior is inversely proportional to total sensor stimulation (reinforcement). Herrnstein [32] expressed relative reinforcement as $1 - r/\sum r$. This quantity varies between 0, when $r = \sum r$, and 1, when $r = 0$. Then, he assumed that the change in behavior ($\mathrm{d}R_1$) associated with reinforcement (r_1) should be proportional to reinforcement change ($\mathrm{d}r_1$) and also to invariant and relativity factors. The resulting differential equation is:

$$\mathrm{d}R_1 = (k/\sum r)\,(1 - r_1/\sum r)\,\mathrm{d}r_1 \qquad (2.c.11)$$

Bearing in mind that $\sum r = r_1 + r_e$ where r_1 is the variable and r_e is constant, the solution of equation (2.c.11) is:

$$R_1 = \int_0^{r1} (k/\sum r)\,(1 - r_1/\sum r)\,\mathrm{d}r_1 = \int_0^{r1} [k/(r_1 + r_e)]\,[r_e/(r_1 + r_e)]\,\mathrm{d}r_1 =$$

$$= [-kr_e/(r_1 + r_e) + c] - (-k + c) = kr_1/(r_1 + r_e) \qquad (2.c.12)$$

This equation is invariant to the equation (2.c.9), i.e., is the matching law thus derived from a relativistic conception of reinforcement, given invariance in the measurement of behavior. De Villiers and Herrnstein [25] reviewed the literature on absolute response strength confirming that in many cases the matching law equation holds and, therefore, r_e is constant (relativity assumption) and k is constant (invariance assumption). A substantial body of literature supports the matching principle at least as an approximation for steady-state responding under the control of reinforcement. There is little doubt that the hyperbola describes the relation between reinforcement and responding, at least on some schedules (see, for example, [33]). There is considerable doubt, however, about the interpretations of the parameters.

Dallery and Soto [34] reviewed published research that applied Herrnstein's equation to distinguish the motoric (k) from motivational (r_e) effects of drugs, and to identify additional independent variables responsible for drug effects, such as extraneous reinforcement. The validity of inferences about drug effects depends on the constancy of k and r_e response to environmental manipulations: k should change only with response or motoric variables, and r_e should change with reinforcer or motivational variables and with the rate of extraneous reinforcement. Empirical tests of these predictions, however, have produced inconsistent results. The review [34] suggests that Herrnstein's theory has not

fulfilled its promise of identifying the behavioral mechanisms of drug action. Modifications to the equation, known as bias and sensitivity, may explain some of these inconsistent results, and the modified equation may have utility in behavioral pharmacology.

In some experiments, concurrent schedules, for example, the classic Herrnstein's equation $R_1/R_2 = r_1/r_2$ does not fit the data [35-37]. The goodness of fit can be significantly improved by incorporating two additional parameters into this classic equation producing a new set of equations [35,38,39], which were called the generalized matching law or the modern theory of matching:

$$R_1 / R_2 = b\,(r_1 / r_2)^a \text{ or}$$

$$R = kr^a / (r^a + r_e{}^a / b) \text{ etc.} \qquad (2.c.13)$$

where a represents sensitivity and b represents bias. If there are no variations from strict matching in the situation both parameters equal 1 and equation (2.c.13) reduces to equation (2.c.9). The classic theory can be considered a special case of the modern theory, with limited applicability [39]. It is, of course, trivial that the more parameters in the equation the better is the goodness of fit. More work is necessary to thoroughly test the utility of the modern theory of matching.

The melioration model represents a stochastic (not deterministic) model and as such has limitations. The model has a solid empirical basis, but an uncertain theoretical foundation. While the equations fit the data, the meaning of its parameters is not clear.

3. ACQUISITION

Failure to distinguish between the process of learning and the criteria for learning may result in misunderstanding. By learning it is commonly understood that an individual is able to do something that he/she was not able to do earlier. An acquisition or learning curve gives a measure of performance P as a function of training time or number of trials. The independent variable may be continuous (t) or discrete (n). The dependent variable may be a measure of production (words typed per minute), it may be a physical measurement (strength or distance), or it may be rate or latency of response. However, the fundamental measure is the probability of success (P), in other words a probability of a correct response in a multiple-choice problem.

L.L. Thurstone [40] was the first to try about forty different equations on published learning curves and selected a form of the hyperbola as giving the best fit to the data:

$$P = aX / (X + K_L) \text{ or} \quad P = a / (1 + K_L / X) \qquad (3.1)$$

where P is attainment or performance in terms of the number of successful acts per unit time, X is the practice in terms of total number of trials, K_L is rate of learning which indicates how many trials are required to reach a half of the upper limit of performance (a). One of other equations with high goodness of fit was:

$$P = a\,(1 - e^{-K_L X}) \qquad (3.2)$$

This equation in Thurston's study gave a fair approximation to the acquisition curves but it did not fit nearly as well as hyperbolic form (3.1). Why? Thurston applied these equations to learning curves obtained in experiments on typewriter learning. However, there is another class of learning curves. The difference between models can be understood better if we ask in each case whether the subject is accumulating habits (accumulation model, equation 3.1), or replacing wrong tendencies by correct ones (replacement model, equation 3.2). Gullisken reviewed propositions contained in Thorndike's law of effect [41] which have been stated in mathematical form [42]. It was shown that these same functions first empirically found by Thurston [40] may be obtained be generalizing the learning theory. Detailed description of stochastic and statistical models of learning can be found in [43,44].

a. Replacement Theory

In a replacement learning model, some responses or expectancies are replaced by others leaving the total number of entities constant by substituting one for another. Restle and Greeno [9] explain this model using a simple process in which some number k of total M marbles are taken from the pool S on each trial, discarded, and replaced with an equal number of marbles from the infinitely large training or experimental pool E with a certain proportion a of red marbles. Both samples from S and E are chosen at random. Let R_n and W_n be the numbers of red and white marbles (correct and incorrect responses, respectively) in S on Trial n. Then, on the average, the marbles taken from S will include kR_n/M red marbles and kW_n/M white marbles. The marbles put into S will include ka red marbles and $k(1 - a)$ white ones. On Trial $n + 1$, then, the number of red marbles will be on the average:

$$R_{n+1} = R_n - kR_n / M + ka$$

The probability of a correct response on Trial n is equal to the proportion of red marbles in the pool S:

$$P_n = R_n / M$$

If we let θ stand for the proportion sampled from S, k / M, then rearrangement gives:

$$P_{n+1} = P_n + \theta\,(a - P_n) \qquad (3.a.1)$$

Equation (3.a.1) provides the basis for a linear model of the learning process [45], since P_{n+1} is a linear function of P_n. Assuming that $P_1 = 1$, performance as a function of number of trials will be:

$$P_n = a\,[1 - (1 - \theta)^{n-1}] \qquad (3.a.2)$$

According to the linear model, performance exponentially grows to the asymptote a, and coefficient θ characterizes the rate of learning. Suppose that $P_1 = b$, then:

$$P_2 = b + (a - b)\,\theta$$

now applying mathematical induction we have:

$$P_n = a - (a - b)(1 - \theta)^{n-1} \qquad (3.a.3)$$

This is the general solution to a single linear operation applied $n - 1$ times.

Matching law is usually tested against the steady state of reinforced behavior. But if we assume that behavior during acquisition is also guided by reinforcement, and then differential equation (2.c.11) should be applicable to learning curves. Mazur and Hastie [46] surveyed the literature on simple, negatively accelerated acquisition curves for human subjects and concluded that most simple learning curves fit the linear-operator form of equation (2.c.11):

$$\Delta R_1 = (k / \sum r) (1 - r_1 / \sum r) \Delta r_1$$

as $R_1 / k = r_1 / \sum r$ then after simple rearrangements:

$$\Delta R_1 = \theta (k - R_1) \quad (3.a.4)$$

where $\theta = \Delta r_1 / \sum r$. It is easy to derive by substitution from equation (2.c.9) that $\theta = \Delta R_1 / \sum R$. This coefficient of proportionality represents a proportion of behavior changed after the amount of reinforcement was changed.

Linear models are usually expressed in terms of the probability of occurrence of a particular response on Trial $n + 1$, that is, P_{n+1}, after n reinforced trials:

$$P_{n+1} = P_n + \theta (1 - P_n) \quad (3.a.5)$$

where P_n is the probability of response on Trial n and θ is the learning-rate parameter. This coefficient of proportionality represents a proportion of stimulus elements conditioned to the correct response on any trial. It is easy to show that acquisition curve must follow an exponential form if individual trials obey equation (3.a.5). Bearing in mind that $\Delta P = P_{n+1} - P_n$, we can write an equation identical to (3.a.4) bearing in mind that k represents the total behavior and corresponds to the probability of behavior equal to unity:

$$\Delta P = \theta (1 - P_n) \quad (3.a.6)$$

b. Accumulation Theory

In accumulative learning models [9], habits or knowledge are added to those that were there before. Suppose that a constant number of marbles k are transferred on each trial from the experimenter pool E to the subject pool S. If a is a proportion of the red (correct) marbles in an inexhaustible supply of marbles in the pool E the number of red marbles in the pool S will increase so that on the average,

$$R_{n+1} = R_n + ka \text{ or } R_n = R_1 + (n - 1)ka$$

Meanwhile, the number of white marbles also increases:

$$W_{n+1} = W_n + \mathrm{k}(1 - a) \text{ or } W_n = W_1 + (n - 1)k(1 - a)$$

The proportion of red marbles in S is again assumed to be equal to P_n:

$$P_n = R_n / (R_n + W_n)$$

Substituting yields:

$$P_n = [R_1 + (n - 1)ka] / [R_1 + W_1 + (n - 1)k] \quad (3.b.1)$$

Introducing constants $b = P_1 = R_1 / (R_1 + W_1)$ and $\theta = k / (R_1 + W_1)$ and rearranging yields

$$P_n = [b + \theta a(n - 1)] / [1 + \theta(n - 1)] \quad (3.b.2)$$

which can be compared with equation (3.a.3) for the linear model. The two theories thus give learning curves starting at b and approaching an asymptote of performance equal to a, and θ is a characteristics of the rate of learning. The two curves differ, however, in the form between b and a.

c. All-or-None Theory

In the simple all-or-none model [9], the subject moves from state "unlearned" (U) to state "learned" (L) at once, in one trial. The probability of a correct response in state U is b, and the probability of a correct response in state L is a. In general, initial performance should be more than zero, $b > 0$, and final performance should approach unity, $a = 1$. The model assumes that the learning rate θ is the same for each trial, i.e., the probability of learning is constant on all trials. What is the probability that learning occurs on Trial n? In order to occur on Trial n, it must fail to occur $n - 1$ times. Bearing in mind that probability to fail is $1 - \theta$ we can calculate the probability of a correct response on Trial n, in case of initial probability of correct response $b = 0$, as:

$$P_n = \theta (1 - \theta)^{n-1} \quad (3.c.1)$$

As for $P_1 = b$ then

$$P_n = 1 - (1 - b) (1 - \theta)^{n-1} \quad (3.c.2)$$

This equation is exactly the same as (3.a.3) for the linear model of replacement theory where $a = 1$ and $b = 1$. This model can be also applied to learning in a group of subjects when every subject can be considered as trained or untrained according to a certain criterion (threshold). The learning curve described by equation (3.c.2) will represent the number of trained subjects as a function of the number of trials.

4. CLASSICAL CONDITIONING

Classical (Pavlovian) conditioning is the process whereby an organism learns the associations between stimuli. Two stimuli are involved. The first stimulus, called the unconditioned stimulus (US), has the ability to elicit an unconditioned response (UR) when presented. The UR is a reflexive response and is unconditionally elicited by the US. For example, food placed in the mouth is an US which will cause an UR of salivation. The second stimulus, termed the conditioned stimulus (CS), differs in that prior to conditioning it is essentially neutral. Pavlovian conditioning occurs when the CS predicts the US. For example, if the CS is presented repeatedly and is immediately followed by the US, the organism will learn that the CS predicts the occurrence of the US. As a result, responses to presentations of the CS itself, called conditioned responses (CR), develop. Parameters in mathematical models of classical conditioning are dimensionless and typically have values between zero and unity.

a. Compound Signal Models

Earlier models of acquisition treated each cue of a multicue compound independently, erroneously assuming no interaction between cues. In contrast, Rescorla and Wagner (1972) came up with an idea how to provide a trial-by-trial description of changes in the associative status of a conditioned stimulus when the stimulus is paired with an unconditioned stimulus in the presence of other conditioned stimuli [47].

The Rescorla-Wagner (RW) model assumes that the change in the associative strength (V) of conditioned stimulus (X), as a result of a pairing with unconditioned stimulus on Trial $n + 1$, is proportional to the difference between the maximum associative strength that this unconditioned stimulus can support (λ) and the sum of associative strengths (V^n_{total}) of all conditioned stimuli including X that are present on Trial $n + 1$:

$$\Delta V_X^{n+1} = \alpha_X \beta(\lambda - V^n_{total}) \quad (4.a.1)$$

where α_x is the associability of conditioned stimulus X ($0 < \alpha_x < 1$), β is the associability of unconditioned stimulus ($0 < \beta < 1$). Both coefficients are closely related to the intensity of the corresponding stimuli. The associative strengths of conditioned stimuli before Trial 1 are assumed to be zero, $V^0_{total} = 0$.

The absolute value of the parenthetical term, $\lambda - V_{total}$, is interpreted in RW model as a surprise defined as the difference between the unconditioned stimulus that is actually presented on the trial and the unconditioned stimulus that is expected on the basis of the assumed predictive value of all of the cues that are present on the trial. Equation (4.a.1) describes the degree of associative learning that occurs on each trial.

According to theories of selective attention, learning about a stimulus depends on attending to that stimulus [48]. The first assumption is that α is not a fixed consequence of such physical characteristics of a stimulus as its intensity or modality, but rather that α may vary with the subject's experience before and even in the course of an experiment. The probability of attending to relevant stimuli typically increases, while the probability of attending to irrelevant stimuli typically decreases. The second assumption is that the total amount of attention is limited. The extent to which a reinforcer is predicted by the cue A is represented by the absolute value of the term $|\lambda - V_A|$. It is assumed that α_A increases/decreases whenever the outcome of a trial is better/worse predicted by A than by all other events X on that trial:

$$\Delta\alpha_A > 0 \text{ if } |\lambda - V_A| < |\lambda - V_X| \text{ and}$$
$$\Delta\alpha_A < 0 \text{ if } |\lambda - V_A| > |\lambda - V_X| \quad (4.a.2)$$

Stimuli can be established, of course, as signals for non-reinforcement and the value of λ can be set to zero.

RW model predicts changes in associative strength only for those stimuli that are presented on a particular trial. A modification of the original RW model was proposed by Van Hamme and Wassermann [49] in order to explain the associative strength of element A in a compound stimulus AX was changed after another compound stimulus BX was presented in non-reinforced trials (backward blocking effect). Van Hamme and Wassermann proposed that stimuli indirectly activated through within-compound-associations have a negative learning parameter. Thus, phenomena of retrospective reevaluation can be explained. Let's consider the following simple example, indicative of retrospective reevaluation, where AB is the compound stimulus $A+B$:

Phase 1: AB-US

Phase 2: A-US

Test trials: Group 1, which received both Phase 1- and 2-trials, elicits a weaker Conditioned Response (CR) compared to the Control group, which only received Phase 1-trials.

The original RW model cannot account for this effect. But the revised model can: In phase 2, stimulus B is indirectly activated through within-compound-association with A. But instead of a positive learning parameter α when physically present, during Phase 2, B has a negative learning parameter. Thus during the second phase, B's associative strength declines whereas A's value increases because of its positive learning parameter.

$$\Delta V_A^{n+1} = \alpha_1\beta_1(\lambda - V^n_{total})$$
$$\Delta V_A^{n+1} = \alpha_1\beta_2(0 - V^n_{total}) \quad (4.a.3)$$
$$\Delta V_A^{n+1} = \alpha_2\beta_1(\lambda - V^n_{total})$$
$$\Delta V_A^{n+1} = \alpha_2\beta_2(0 - V^n_{total})$$

where

α_1 = learning rate parameter for cue A present

α_2 = learning rate parameter for cue A absent

β_1 = learning rate parameter for outcome present

β_2 = learning rate parameter for outcome absent

This modification of the RW model allows CS and US nonoccurrence to be treated in a parallel manner assuming that missing input cues, as well as missing output cues, are actively encoded as absent.

Review of the RW model's predictive successes and failures can be found in [50]. The RW model provides a causal explanation of associative process and has become one of the most influential and stimulating new research models of learning. The real-time mathematical equivalent of RW model will be given in the end of the following section [see equation (4.b.10)].

b. Theory of Automatic Processing

This theory was primary designed to address the learning and retention processes in infrahuman subjects in circumstances of simple conditioning and habituation. It assumes that information processing in animals proceeds mechanically according to relatively invariant operating characteristics and stable structural features. This model is also called SOP - the *standard operating procedures* in memory [51]. There is an additional justification for the acronym "SOP". Because the more comprehensive treatment of the mechanism of priming (isolated presentation of cues) of short-term memory that is included leads to theoretical effects that are only *sometimes opponent processes*. SOP model is based

on three groups of assumptions: stimulus representation, network characteristics, and rules for learning and performance.

Memory is conceptualized in SOP as a connectionist system with unitized, representative nodes connected with directional links. It is assumed that each node, corresponding in some manner to an isolable experimental stimulus, consists of a large but finite set of informational elements. These elements are proposed to be in one of three states of activation: a primary ($A1$) state that is elicited upon presentation of the corresponding stimulus event according to the momentary probability $p1$; a secondary ($A2$) state into which elements passively decay from $A1$ according to the momentary probability $pd1$, and an inactive state (I), into which elements passively decay from $A2$ with momentary probability $pd2$, and in which they remain for as long as the corresponding stimulus remains absent. It is also assumed that $pd1 > pd2$. The rules for momentary activation and decay are:

$$\Delta p_{A1} = p1(p_I) - pd1(p_{A1}) \quad (4.b.1)$$

$$\Delta p_{A2} = pd1(p_{A1}) - pd2(p_{A2}) \quad (4.b.2)$$

$$\Delta p_{A1} = pd2(p_{A2}) \quad (4.b.3)$$

where p_I, p_{A1}, and p_{A2} are the proportion of elements that are inactive, active in $A1$, and active in $A2$, respectively.

Although there are no practical limits on the number of nodes, it is assumed that there is a limitation on the number that can be active at any particular moment, especially, on the number that can be in focal, $A1$ memory, with less limitation on the number of nodes that can be in peripheral, $A2$ memory. The SOP model makes no assumptions about which stimuli can function effectively as CSs and which can function as USs. With these assumptions about the dynamic nature of the stimulus representation across time, it is possible to derive numerous behavioral outcomes [52].

SOP theory uses associative network principles: when one node is activated by peripheral stimulation, a spread of activation occurs, *via* associative links. If a node is activated associatively, it is only to the $A2$ state, not to the $A1$ state. What SOP specifies is that the memory of an event is a weaker version of the original experience, because it is a decayed version of the experience, but contains whatever effective properties that latter may have. We can define $p2$ for associative activation of US node as:

$$p_{2,\mathrm{US}/\Sigma\mathrm{CS}} = \sum_i [V_{CS/US}(r_1 p_{A1,CSi} + r_2 p_{A2,CSi})] \quad (4.b.4)$$

where $0 \le p_2 \le 1$ and $r_1 > r_2$. That is, the momentary $p2$ value for a US node is the product of the associative strength from each CS to the US and the weighted sum of the proportion of CS elements momentarily in the $A1$ and $A2$ states, summed over the concurrent CSs. The summation across CSs reflects the componential nature of any activation stimulus.

SOP has an explicit performance rule stating how response output is determined by which nodes: the temporal course of responding (R_j) is a weighted function of the proportions of nodal elements in $A1$ and $A2$, each with a separate multiplicative weighting, according to:

$$R_j = f_j(w_{1,jp_{A1,US}} + w_{2,jp_{A2,US}}) \quad (4.b.5)$$

where $w_1 \ge 0$, w_1 may have any value, and the mapping function f_j is a subject to empirical determination.

The nature of the assumed stimulus representation in SOP results in the strong prediction that behaviors assumed to be indicative of nodal activity are biphasic and sometimes opponent process, as USs often are (for example, the burst of activity followed by freezing after foot shock or sedation followed by hyperactivity after a dose of morphine). The rule for increment of CS-US excitatory association and inhibitory association are:

$$\Delta V_i^+ = L^+ \sum_i (p_{A1,CSi} \cdot p_{A2,USj}) \quad (4.b.6)$$

$$\Delta V_i^- = L^- \sum_i (p_{A1,CSi} \cdot p_{A2,USj}) \quad (4.b.7)$$

where L is the learning rate parameter and $L^+ > L^-$. The net CS-US strength is the addition of excitatory and inhibitory strengths:

$$\Delta V_i = \Delta V_i^+ + \Delta V_i^- \quad (4.b.8)$$

Although the rules shown above are stated for the most common instance of Pavlovian conditioning, they have more general applicability. Moreover, because there is no assumption about different rules for excitatory and inhibitory processes, so the SOP's learning rule can be stated in terms of a single equation:

$$\Delta V_i = \sum_i [p_{A1,CSi}(L^+_{A1,USj} - L^-_{A2,USj})] \quad (4.b.9)$$

Equation (4.b.9) is similar to a real-time rendition of the predecessor, the RW model rule which can be written as follows:

$$\Delta V_i = \sum_i [\alpha_i^t \beta_i^+ (\lambda_j^t - \sum V_i) - \beta_i^- \sum V_i] \quad (4.b.10)$$

Both models of classical conditioning presented here are strictly deterministic models explaining the behavior as caused by stimuli according to the rules determined by the network. These models belong to the group of replacement theory models and the linear-operator rule [see equation (3.a.1)] is an algorithm used to compute variations in the associative strength of the CS as a function of the discrepancy between the value of the reinforcement provided by the US (λ) and the current associative level of the CS element.

5. OPERANT BEHAVIOR

Several mathematical models were developed within the framework of Skinnerian theory of operant behavior. An independent variable in these models is the amount of reinforcement (food, water, locomotor activity etc.). A dependent variable, according to the definition of reinforcement given by Skinner, is supposed to be a probability of a behavior to occur, i.e., a dimensionless parameter.

However, both theoretically and practically it is difficult to measure a probability. That is why all these models use the rate of behavior, incorrectly called rate of response, instead of rate of operant (s^{-1} dimension) or the reciprocal variable – inter-response interval (s dimension).

a. Mathematical Principles of Reinforcement

Peter Killeen put three main constructs in the basis of his work on Mathematical Principles of Reinforcement (MPR): motivation, association and constraint [53]. "Reinforcement empowers and directs behavior. It empowers it by generating a heightened state of arousal, one that can become conditioned to the context in which reinforcement occurs. It directs behavior by selecting from among the responses it instigates those that immediately precede and predict it. That arousal and direction acts under the constraint of finite time and energy available to execute a response" [54]. The following are the three assumptions/principles of the theory: (1) An incentive activates a seconds of responding; (2) A response requires δ seconds for its completion; (3) Reinforcement occurs when an incentive enters a memory that also contains a response.

Activation principle assumes that arousal induced by the incentive is not permanent and decays according to the first order kinetic process:

$$b(t) = b_1 e^{-t/\alpha} \quad (5.a.1)$$

where b_1 is the y-intercept corresponding to the highest rate of response b, and α is the time constant. The total number of responses after such incitement is the integral of the exponential function, which is αb_1. This total number of responses is a measure of total arousal induced by a single presentation of a reinforcer $a = \alpha b_1$. In MPR a is called *specific activation*. Arousal level is proportional to both specific activation and rate of presentation of the reinforcer $A = ar$. Both arousal level (A) and specific activation (a) are hypothetical constructs referring to motivational states (A) and the power of a single reinforcer, of a particular type, applied to a particular organism, at some level of deprivation (a), to induce responding. The above considerations are epitomized in *The first principle of reinforcement*: arousal (A) is proportional to the rate of reinforcement: $A = ar$.

The simplest relation between responding level and arousal level was assumed – proportionality with the coefficient k which depends on a feature of the measuring apparatus and incorporated into a. Experiments show that this assumption is working at least at low response rates. However, competition from other responses may cause response rate to fall short of its theoretical asymptote A. Another reason it may fall short is because it takes time to make a response, and therefore, responses cannot be emitted as fast as they are elicited. The inter-response time (IRT) consists of two parts:

$$IRT = \delta + \tau \quad \text{or} \quad 1/IRT = b = 1/(\delta + \tau) \quad (5.a.2)$$

where δ is the time required to emit the response and τ is the literal inter-response interval. There is an obvious maximal rate of responding (b_{max}) when $\tau = 0$:

$$b_{max} = 1 / \delta \quad (5.a.3)$$

This equation represents *The second principle of reinforcement*: response rate (b) is constrained by the time required to emit a response (δ).

Coupling is a parameter representing an association between responses of the same operant class and reinforcement. The magnitude of coupling depends on the schedule of reinforcement and rate of responding. *The third principle of reinforcement*: the coupling between a response and reinforcer decreases with the distance between them. The distance may be measured in space, in time, or in the number of intervening events. In MPR theory this decrease in coupling is assumed to be exponential because the probability of whatever responses are presented in memory at the time of reinforcement decays according to equation (5.a.1). The coupling coefficient $c.$ should be an integral of distances between this particular reinforcer and all previous responses:

$$c. = \int_0^\infty b(t) e^{-\lambda t} \, dt \quad (5.a.4)$$

Where $b(t) = 1$ while a target response is in progress and $b(t) = 0$ otherwise, λ is the rate of decay of response traces in memory. The dot in the symbol for coupling coefficient is a place holder for the designation of the particular contingencies under study.

Killeen *et al.* [55] developed a map between response rate and probability. They showed that if the probability P of observing a response in a small interval of time is proportional to arousal level, then the instantaneous response rate $1/\tau$ will also be proportional to A. By definition $A = ar$, so:

$$1/\tau = ar \text{ or } \tau = 1/ar \quad (5.a.5)$$

This proportionality between the expected time between responses, τ, and the expected time between reinforcers, $1/r$, is a premise of linear systems theory (McDowell *et al.*, 1993). Substituting equation (5.a.5) into equation (5.a.2) yields:

$$b = 1/(\delta + 1/ar) \text{ or } b = r / (\delta r + 1/a) \quad (5.a.6)$$

Coupling tells us how arousal is associated with the target response. But reinforcement can only act on behavior which is already manifested and so it must modulate response rates, not arousal level. The simplest implementation is to multiply equation (5.a.6) by the coupling coefficient c.

$$b = c.r / (\delta r + 1/a) \quad (5.a.7)$$

Equation (5.a.7) is the fundamental model of MPR.

A similar mathematical model was developed by John Nevin [56], called the Behavioral Momentum Theory (BMT), who proposed to measure *response strength* in terms of resistance to change. In Nevin's theory a construct similar to arousal level A is called *momentum* (detailed comparison between MPR and BMT can be found in [57]). In Skinner's theory a specific arousal a was called the *reflex reserve* [29], however, a mathematical model was not developed by Skinner.

The principle of temporal constraint produced in MPR the equation (5.a.7) which becomes similar to Herrnstein's equation (2.c.9) if we denote $1/\delta$ as k and $1/\delta a$ as r_e. It shows that melioration principle is also based on the same temporal constraint: behaviors compete for time if they are not compatible. More important is the difference between MPR and melioration theory – the coupling coefficient. Equation (2.c.9) is a monotonic function and does not fit well data when rate of response decreases with continuous increase in rate of reinforcement. The coupling coefficient in equation (5.a.7) may become less than 1 and better describes a decrease in rate of response.

On fixed ratio (FR) schedules by definition $r = b/n$, where n is the ratio requirement. This relation is called the schedule feedback function (FF) [58]. Substituting in equation (5.a.7) yields:

$$b = c. / \delta - n / \delta a \qquad (5.a.8)$$

For FR schedules the coupling coefficient c_{FRn} is according to equation (5.a.4):

$$c_{FRn} = 1 - (1 - \beta)^n = 1 - e^{-\lambda\delta n} \qquad (5.a.9)$$

where β is the proportion of the maximum association conferred on the response that immediately precedes reinforcement $\beta = 1 - e^{-\lambda\delta}$. Equation (5.a.9) implies that consumption of a reinforcer completely erases the traces of prior responses. This often seems to be the case for pigeons because the amount of food given is usually big and takes a long time to consume. In rats it is not typically true and there remains some residual trace of the preceding reinforcer. To describe this, the coupling coefficient should be modified to allow for the possibility of a reinforcer strengthening by prior presentations. For FR schedules it is simplest to add a free parameter n_0 signifying the nominal number of additional responses beyond the prior reinforcer that contribute to response strength. The coupling coefficient for FR schedules becomes:

$$c_{FRn} = 1 - (1 - \beta)^{n + n_0} = 1 - \varepsilon(1 - \beta)^n = 1 - \varepsilon e^{-\lambda\delta n} \qquad (5.a.10)$$

where $\varepsilon = (1 - \beta)^{n_0}$ can range from 0 to 1 and is 1 if there is no savings of memory trace ($n_0 = 0$). This is an example of analysis of schedules of reinforcement according to MPR. Equations (5.a.8) and (5.a.10) combined provide very high goodness of fit to the experimental data obtained with the FR schedules.

MPR represents a deterministic model of operant behavior in the Skinnerian school of behaviorism. It is one of the most sophisticated mathematical models attempting to account for a wide variety of behaviors, schedules, types of reinforcers and other conditions.

b. General Models of Animal Choice

After Herrnstein [24] the idea that all behavior involves choice was widely accepted. A variety of different procedures have been developed in operant conditioning experiments on choice, but most of them can be divided into three main categories. (1) In *discrete-trial* procedures, the subject chooses between different alternatives by making a single response after which the chosen alternative is given. The next trial begins in a fixed or variable interval. Thus the choice period is restricted to a short segment of each trial. (2) In *concurrent* procedures, each experimental session can be viewed as one continuous choice period, because at each moment the subject can respond on one alternative, switch to the other alternative, or make no response at all. Each alternative delivers reinforcers according to its own schedule of reinforcement. (3) In *concurrent-chain* procedures, there are extended periods of concurrent schedules and periods during which the subject receives the consequences of its previous choices but cannot switch to the other alternative.

James Mazur argued [59] that a worthwhile goal for research on choice is to develop a general model that can account for the results obtained with these different procedures. Taking on this challenge, three models were developed and compared. Grace [60,61] demonstrated that his *contextual-choice model* (CCM) can account for 90% and more of variability. Mazur compared this model with Fantino's [62] *delay-reduction theory* (DRT) and his new *hyperbolic valued-added* (HVA) model [59] and demonstrated that they are almost as good as CCM in describing empirical results obtained using these three procedures.

DRT was developed [62] to account for choice with variable interval (VI) schedules:

$$B_L / (B_L + B_R) = r_L(T - t_L) / [r_L(T - t_L) + r_R(T - t_R)] \qquad (5.b.1)$$

where B_L and B_R are the choice responses on the left and right keys, respectively, r_L and r_L are the overall rates of primary reinforcement on the left and right keys, respectively, T is the average overall time to primary reinforcement measured from the onset of the choice phase, and t_L and t_R are the average times (or delays) during the terminal links (or outcome phase) on the left and right keys, respectively. The term $(T - t_x)$ represents the degree to which a terminal-link stimulus is correlated with reduction in time to primary reinforcement. Reinforcement strength of stimulus A in general theory of delay reduction is:

$$A = f[(T - t_A)/T] \qquad (5.b.2)$$

The inclusion of the terms r_L and r_L in equation (5.b.1) acknowledges the fact that reinforcement experienced more frequently has more impact on the behavior on which it is contingent. When the durations of the terminal links are zero ($t_L = t_R = 0$) then equation (5.b.1) reduces to the matching law of equation (2.c.1). The viability of DRT was demonstrated by the ability of the theory to explain some anomalous findings (e.g., preference for less reliable reinforcement) and to explain choice in chained and tandem schedules (see review [63]).

According to CCM [60], relative responding in the initial links is controlled by relative rate of conditioned reinforcement, as described be the generalized matching law (see also quation (2.c.13)), and terminal-link value can be concatenated multiplicatively:

$$B_L / B_R = b \prod_{i=1}^{n} (x_{iL}/x_{iR})^{ai} \qquad (5.b.3)$$

where b is a bias term, B_L and B_R are the responses emitted on the left and right keys, a_i is a parameter to reflect sensitivity to the initial-link schedules, and n is the number of factors x_i that can affect preference. This means that relative initial-link responding in concurrent chains equals the relative value of the terminal-link stimuli as conditioned reinforcers. Terminal-link sensitivity is a function of the ratio of the average terminal-link to initial-link durations:

$$B_1 / B_2 = b\,(\mu_{1R} / \mu_{1L})^{a_1}\,[(\mu_{2R} / \mu_{2L})^{a_2}\,(x_{iR} / x_{iL})^{a_i}]^{(T_t / T_i)^k} \quad (5.b.4)$$

where μ_{1R} and μ_{1L} are the relative intervals between reinforcers on the left and right schedules, k is a scaling parameter, (x_{iR} / x_{iL}) represents a possible additional variable, such as reinforcement magnitude, T_t is the average terminal-link duration, including any postreinforcer blackouts, T_i is the average initial-link duration, per reinforcement. This equation (5.b.4) is the most general form of the contextual choice model (CCM). There is an important distinction in the way in which conditioned reinforcement value is defined in CCM and in DRT. The latter states that values of stimuli as conditioned reinforcers are determined by context. CCM states that these values are determined independently of context, but that differential effectiveness of the stimuli is context dependent.

HVA begins with the assumption [59] that the hyperbolic-decay model describes how reinforcer value decreases with increasing delay (see equation (5.c.3) in the next section). When applied to concurrent-chain schedules, HVA assumes (1) the value of each terminal-link depends on the time from the onset of that link to primary reinforcement, (2) the value of the initial-links depends on the variable time from the onset of the initial-links to primary reinforcement, (3) choice proportions are based on the amount of value added when a terminal link is entered (i.e., on the amount of increase in value when terminal-link is entered):

$$B_L / B_R = b\,(r_{iR} / r_{iL})^{a_i}\,[(V_{iL} - a_t V_i) / (V_{iR} - a_t V_i)] \quad (5.b.5)$$

where the left portions of the equation are identical to CCM, including the ratio of initial-link reinforcement rates (r_{iR} and r_{iL} were replaced with relative intervals between reinforcers μ_{1R} and μ_{1L} in equation (5.b.4)). All values (V_{iL}, V_{iR} and V_i) are calculated using the hyperbolic-decay equation (5.c.3). Equation of HVA is only applicable when V_{iL} and V_{iR} are both greater than $a_t V_i$. HVA assumes that if the value of either terminal-link is less than $a_t V_i$, the subject will show exclusive preference for the other alternative.

In behavioral momentum theory (BMT), the effects of reinforcement on resistance to change follow from an analogy between the change in velocity of a moving body in relation to its mass when an external force is applied, and the change in response rate in relation to reinforcement when a disruptor is applied [56,64]. The behavioral equivalent of Newton's second law is:

$$\log(B_X / B_0) = -\,x / m \quad (5.b.6)$$

where B_0 is baseline response rate, B_X is response rate during disruption, x (with a minus sign) represents the decremental force-like effects of the disruptor, and m represents behavioral mass. Thus, B_X / B_0 is the proportion of baseline during disruption. Equation (5.b.6) states that for a given value of x, the relative change in response rate is inversely related to m, or equivalently, resistance to change is directly related to m. In terms of analogy to Newtonian mechanics, behavioral mass is a positive function of reinforcer rate.

All models of animal choice are based on the matching law and therefore are stochastic, not deterministic models. They all fit experimental data which is not a surprise because each of these models has at least three floating parameters.

c. Behavioral Economics

One goal of behavioral economics is to suggest mathematical models with some psychological foundations. Because economics is the science of how resources are allocated by individuals and by collective institutions like firms and markets, the psychology of individual behavior should provide the basis for economics. However, economists routinely use models that are grossly inconsistent with findings from psychology. A relatively recent approach, "behavioral economics", is trying to bridge studies of *homo economicus* on one side and *homo sapiens* and *bestia vulgaris* on the other. Mathematical models should be based on four principles of economic behavior [65].

Utility principle. *Expected utility theory* assumes that rationally behaving people value (u) an outcome (x_i) and integrate the outcomes into their overall wealth proportionally to probabilities (P_i) of those outcomes to occur $\sum_i P_i u(X_i)$. Psychology adds that real people adapt to what they have experienced and weigh probabilities nonlinearly. For example, people overweigh low probability outcomes or they spend more from a tax refund than from an increase in the value of their stocks or their homes, which standard rational theory does not recognize. *Prospect theory* corrects this by introducing a nonlinear function $\pi(P_i)$ and coefficient r in the equation for utility $\sum_i \pi(P_i)u(X_i - r)$. Studies in animals [66] suggest that unit price may serve as a measure of the utility of a reinforcer for animals. Unit price is a function of the schedule of reinforcement divided by reinforcer magnitude.

Delay discounting principle. All organisms prefer immediate over delayed rewards. Rational evaluation of future consequences assumes "exponential discounting", that is, future utilities $u(X_i)$ are discounted by a weight δ^t, where $0 < \delta < 1$, which is an exponentially declining function of time. Therefore the expected utility function can be written as [67]:

$$U = \sum_i P_i u(X_i) e^{-\pi t} \quad (5.c.1)$$

The rationally behaving individual will choose a consumption sequence (c_1, c_2, …, c_n) over a sequence (d_1, d_2, …, d_n) if and only if:

$$\sum U(c)\delta^t > \sum U(d)\delta^t \quad (5.c.2)$$

Exaggerated preference for immediate reward is particularly evident in animals and young children. There are also theories to explain adult behavior like addiction and procrastination on the basis of equation (5.c.2).

Four types of discrepancies were found between experimental data obtained in humans and discounted utility model [68]. The common difference effect implies that discount rates should decrease as a function of time delay over which they are estimated. The absolute magnitude effect implies that discount rates should decrease as a function of reward amount. The gain-loss asymmetry implies that losses are discounted at a lower rate than gains. The delay-speedup asymmetry demonstrates that the amount required to compensate for delaying a reward by a given interval, from t to $t + s$, is from two to four times greater than the amount subjects were willing to sacrifice to speed consumption up by the same interval, that is, from $t + s$ to t. In animals, however, rate of discounting did not vary systematically as a function of the amount of the delayed reward [69].

Mazur [70] found a different equation to describe pigeon's choices among delayed rewards of various amounts:

$$v = V / (1 + kd) \quad (5.c.3)$$

where v is discounted value of the reward V, d is the delay between choice and acquisition of reward and k is a constant determining the degree of discounting. This hyperbolic discounting model can be derived from matching law [71] if we assume that subjects' relative valuation of reward (V) may be assessed by frequency of choice. The reward will be proportional to the average rate of reinforcement (R), in a set amount of reward (A), and inversely proportional to an average delay after the successful behavior $Z + \Gamma(T - t)$, where T is the time each reward is due, t is the time of the behavior that obtains it, Z is an empirical constant that limits the maximum value at zero delay, and Γ determines the delay gradient. Matching law can be written as:

$$V / V' = R / R' \cdot A / A' \cdot [Z + \Gamma(T' - t)] / [Z + \Gamma(T - t)] \quad (5.c.4)$$

Assuming $R = R'$ and rearranging yields:

$$V = A / [Z + \Gamma(T - t)] \quad (5.c.5)$$

Mazur (1987) empirically found that Z and Γ can be set at 1.0 because data about their proper values at long intervals are not available. Consider two alternative rewards, one of which, A, will be available at time T, and a greater one A', will be available at time $T + \Delta$. Equation (5.c.4) suggests that subjects will prefer them equally when:

$$V / V' = A / A' \cdot [1 + (T' + \Delta) - t] / [1 + (T - t)] = 1$$

where t is the time the choice is made. Solving for t yields:

$$t_{indifference} = [A(1 + T + \Delta) - A'(1 + T)] / (A - A') \quad (5.c.6)$$

Equation (5.c.6) suggests that there is time of indifference or ambivalence when the subject cannot make a rational choice because discounted values of rewards are equal.

Social utilities principle. Most applications of economic theory assume individuals care only about their own wealth and won't sacrifice to help or hurt others. Laboratory experiments demonstrate these social utilities both in humans and animals. Mathematical theories of social utility explain altruistic and vengeance behavior by assuming that people dislike allocations in which they earn a different amount than others or that people like to reciprocate. These theories are parsimonious, can explain a surprisingly wide range of different experiments (such as frequent cooperation in one-iteration prisoners' dilemma games), and predict some new patterns as well.

Equilibrium principle. Mathematical models of behavioral economics typically study systems in equilibrium: prices are stable, supply meets demand, all subjects have chosen optimal strategies etc. However, new theories of learning from others [72] and theories of population evolution [73], for example, suggest how an equilibrium comes about. There are two classes of learning rules: (1) reinforcement, mostly studied in psychology, and (2) belief learning, studied in game theory. These classes, thought to be fundamentally different, are now assumed to be closely related.

A major concern among economists is that the ideas of behavioral economics are too informal and fragmented to serve as a basis for economic theory. In fact, recent research has already produced theories which can replace standard theory and still maintain formal structure and reasonable parsimony in seven crucial areas: (1) Utility maximization could be replaced by theories of reference-dependent preference (in which preferences exist, but are sensitive to current consumption or another reference point) and by theories of preference "construction"; (2) Expected utility theory can be replaced by prospect theory; (3) Subjective expected utility theory (in which "personal" probabilities are expressed by judgments, rather than derived from objective evidence) can be replaced by theories with non-additive probability; (4) Discounted utility can be replaced by "hyperbolic discounting", in which very short-term discount rates are much higher than future discount rates, reflecting a temporary impatience or impulsiveness; (5) Bayesian updating could be replaced by "support theory" or by formalizations of cognitive heuristics like availability (easily retrievable information is overweighted) and representativeness (hypotheses which are well-represented by evidence are thought to be likely); (6) theories of self-interest can be replaced by theories of "social preference", e.g., [74]; (7) and theories of equilibrium behavior can be replaced by (or perhaps justified by) theories of adaptive learning (e.g., [75]).

6. EFFECTS OF PSYCHOACTIVE DRUGS

A basic principle of pharmacology states that drug effects are proportional to the concentration of the drug at a site of action. At the receptor level, drug effects are believed to be proportional to the relative occupancy of receptors as described by the law of mass action. However, many effects of agonists do not follow this law and the relationship between drug dose and behavioral effects are even more obscure at the level of the whole animal. There are a few mathematical models dealing with behavioral effects of drugs and they will be presented in this and in the next section.

a. Rate-Dependent Effect of Drugs

The sensitivity of operant behavior to drugs acting on the central nervous system was demonstrated by Skinner and Heron [76]. In the 1950's, Dews did systematic analysis of the behavioral effects of psychoactive drugs using the procedures developed by Skinner and showed that the behavioral effects of drugs depended critically on the reinforcement contingencies maintaining the behavior being measured. It was the beginning of Behavioral Pharmacology as a science. In particular, Dews speculated that the drug effect depends on the rate of responding which is under the control of a schedule of reinforcement [77]. What has come to be termed the "rate-dependency hypothesis" is now a well established generalization, and the phenomenon has been widely replicated using an array of drugs [78]. Ironically, mathematical analysis of this hypothesis which appeared to occupy in 1977 "the status of a low-level law" [79] in behavioral pharmacology was published in the same year [80] and undermined the basic premise of the rate-dependence hypothesis.

Drug induced changes in response rate, a ratio between response rate in the presence of the drug (R_d) to the baseline response rate in the absence of the drug (R_c), plotted against the baseline rate (R_c) on logarithmic scales often falls along a straight line [81]. Gonzalez and Byrd, [80] put this statement in the mathematical form:

$$\log(R_d / R_c) = \log k + j \log R_c \qquad (6.a.1)$$

Rearranging yields:

$$R_d = k R_c^{\,j+1} \quad \text{or} \quad R_d = k R_c \, R_c^{\,j} \qquad (6.a.2)$$

where k is a linear proportionality coefficient and j is an exponential coefficient determining the shape of the curve. Values for the rates of response have limits. The rate of response cannot be below zero and higher than the maximal for each subject and procedure rate R_m, i.e., $0 \leq (R_c, R_d) \leq R_m$. These limits constrain the relationship between real values for R_d and R_c. It is, therefore, useful to incorporate R_m in equation (6.a.2). Gonzalez and Byrd derived this new equation in a different and more complex way than presented herein: at the maximal control rate, $R_c = R_m$, the rate in the presence of the drug will be maximal or minimal, $R_d = R_{d,m}$, then according to equation (4,a,2) we will have $R_{d,m} = k R_m \, R_m^{\,j}$, therefore $k = (R_{d,m} / R_m) \, R_m^{\,-j}$. Let's denote the ratio between the two maximal effects as a new constant $A = R_{d,m} / R_m$ and the new equation becomes:

$$R_d = A R_m^{-j} \, R_c^{\,j+1} \qquad (6.a.3)$$

A cursory examination of publications reporting rate-dependent effects of drugs revealed that the slope j in equation (6.a.1) is seldom more negative than –1 or slightly positive. If $j \approx 0$, then ordinate-intercept log$k$ is typically close to 1 meaning that the effect of the drug is very weak. If $j \approx -1$, then the effect of the drug does not depend on the control rate of response at all. According to the analysis conducted by Gonzalez and Byrd, the experimental results suggest that as drug dose increases, the rate of responding becomes stronger and stronger (log$k \rightarrow$ maximal) and more independent of baseline ($j \rightarrow -1$) and when $j = -1$ the drug effect reaches the maximal level according to equation (6.a.3). Therefore, much of the data interpreted as demonstrating rate-dependent drug effects on log-log scales of equation (6.a.1) should instead be interpreted as indicating that drugs effect is increasingly independent of the baseline approaching the constant level at the maximal effect. Initial misinterpretation is a result of the traditional use of the ratio R_d / R_c as the dependent variable which is obviously dependent on the control response rate.

From the point of view of modeling, Gonzalez and Byrd demonstrated that the drug dose represents another independent variable determining the relationship between the response rates in the presence and the absence of the drug. Ksir analyzed this dose-response relationship and showed that the range of rates displayed by animals under no-drug condition decreases and sometimes converges into one single rate which is obviously independent from the control rates. He called the entire phenomena the rate-converging effects of drugs [82]. To the best of our knowledge, there is no mathematical model of the rate-convergent effect. Attempts to develop a mathematical model of so-called "rate-dependent-effect" or "rate-convergent-effect" are not to revise experimental data showing again and again that the rate of the response in the presence of the drug is different compared with the rate of response in the absence of the drug but to provide accurate and parsimonious account of drug effects. For more details refer to the previous discussions [83,84,85].

b. Onset of Psychoactive Drug Effects

It is well known that certain classes of psychoactive drugs require considerable time to produce a maximal effect. The full-scale effect of antidepressants and antipsychotic drugs takes up to two weeks. Tardive dyskinesia is a neurological syndrome caused by the long-term use of neuroleptic drugs. Syndrome of withdrawal of addicting drugs also develops in the course of many days. Kuhar and Joyce recently [86] hypothesized that some newly synthesized protein(s) could account for the action of these slow-onset effects of psychoactive drugs.

The basic pharmacological principle states that the effect of the drug (E) is proportional to the drug concentration (D_e) at the site of action:

$$E = k_e \, D_e \qquad \text{and} \qquad D_e = D_s - D_d \qquad (6.b.1)$$

The concentration of the endogenous drug is a net balance between synthesis (D_s) and degradation (D_d). The widely accepted assumption is that the rate of synthesis for most drugs is a zero-order process:

$$\mathrm{d}D_s / dt = r_s \qquad \text{or} \qquad D_s = r_s \, t \qquad (6.b.2)$$

whereas the degradation rate for most drugs is a first-order process:

$$\mathrm{d}D / \mathrm{d}t = k \, D \qquad \text{or} \qquad D = \mathrm{e}^{-kt} \, D_0 \qquad (6.b.3)$$

where k is the elimination constant and D_0 is the drug concentration at the beginning of the process. Kuhar and Joyce assumed that the initial concentration of the drug responsible

for the delayed psychoactive effects is zero ($D_0 = 0$) then we can write the following differential equation and its solution for the effective drug concentration (D_e) at any time (t) after the process of synthesis began:

$$dD / dt = r - kD \quad \text{or} \quad D_e = r_s / k\,(1 - e^{-kt}) \qquad (6.b.4)$$

According to this model the effect of the psychoactive drug, if mediated by newly synthesized endogenous drug, should increase monotonously and asymptotically reaching the maximal level (E_{max}) at the maximal concentration of the drug $D_{max} = r_s / k$

$$E = k_e\, r_s / k\,(1 - e^{-kt}) \quad \text{or} \quad E = E_{max}\,(1 - e^{-kt}) \qquad (6.b.5)$$

Drug elimination constant k defines the half-life of the process $t_{1/2} = \ln(2) / k$. In time equal to five half-lives ($t = 5\ t_{1/2}$) the effect will reach almost 97% of maximal effect E_{max}.

Nonlinear regression analysis of data derived from Cole and Davis [87] showed that the rate of clinical improvement following treatment with chlorpromazine after subtracting the effect of the placebo was well described by equation (6.b.5) with half-life of approximately 18 days [86]. The time-course of clinical improvement for early responders treated with clozapine (data from [88]) showed a half-life in the range of 30 days [86]. However, it should be noted that some factors may affect the shape of the curve. For example, the effect occurs only after some minimal concentration of the endogenous drug has to be accumulated – the "threshold" of the effect or the maximal effect occurs before the maximal concentration of the drug has been reached – the "ceiling" of the effect. These effects may be taken care of by simple modifications of the equation (6.a.5); however, Kuhar and Joyce did not advance their mathematical model in this direction.

In case of several (n) endogenous drugs contributing to the same effect additively the equation becomes:

$$E = k_{e1}\, r_{s1} / k_1\,(1 - e^{-k1t}) + k_{e2}\, r_{s2} / k_2\,(1 - e^{-k2t}) + \ldots + k_{en}\, r_{sn} / k_n\,(1 - e^{-knt}) \qquad (6.b.6)$$

The total number of independent parameters in this equation compared to equation (6.b.5) is multiplied by a number of contributing drugs, therefore, the utility of equation (6.b.6) decreases considerably in case of three or more drugs because of increasing interdependency of all parameters.

In case when a pharmaceutical agent activates a cascade of the synthesis of two or more endogenous drugs sequentially, i.e., the concentration of the next one is dependent on the concentration of the previous one in the cascade, but the effect depends only on the concentration of the last drug, the following equation can be written:

$$D_n = D_{max,n}\,(1 - e^{-k1t})\,(1 - e^{-k2t})\,\ldots\,(1 - e^{-knt}) \qquad (6.b.7)$$

where D_n is the concentration of the endogenous drug inducing the clinically relevant effect at the time t after injection of the pharmaceutical agent, $D_{max,n}$ is the maximal concentration of this drug, n is a number of drugs in the cascade, and k_1, k_2, …, k_n are elimination constants of the those drugs [86]. It is clear from equation (6.b.7) that the limiting step of the entire cascade will be dominant in the time-course of the effect. An example of such a cascade would be the transcription factor, mRNA, i.e., the gene product. Kuhar and Joyce [86,89] suggest that proteins are good candidates for putative endogenous drugs inducing clinically relevant effects of antidepressants, neuroleptics, drugs of abuse etc, because many proteins found in the brain have half-lives of elimination in the range of days [90,91]. Recently it was demonstrated that the slow increase in concentration of protein p11 may be responsible for the increase in serotonin 5-HT_{1B} receptors in brain after chronic administration of antidepressants [92].

7. DRUG SELF-ADMINISTRATION

Drug self-administration represents perhaps the best experimental model to study drug addiction. Drug self-administration differs from other areas of behavioral pharmacology in that the amount of the drug administered is partially under the control of the organism. Therefore, the experimenter controls the unit dose of administered drug but not the level of accumulated drug in the body. Self-administration of many drugs of abuse in animals has a remarkable property – the intervals between intravenous injections are regular and depend on the unit dose of the drug. There are several attempts to describe the relationship between the drug unit dose and the rate of self-administration.

a. Dose-Effect Function for Self-Administered Drugs

Based on a conventional theory that maintained drug self-administration represents the behavior which is reinforced by drugs of abuse, Sizemore and Martin [93] attempted to develop a mathematical model of maintained self-administration. A theory was constructed in which the ability of a dose to maintain responding was described in terms of receptor theory and a function relating rate of responding to amount of drug self-administered. They assumed that the effect of the drugs of abuse on behavior, including the rate of self-administration, should be related to the occupancy of receptors at which this drugs bind. The simplest form of this function is called the Hill equation:

$$P = 1 / (1 + K_d / D) \qquad (7.a.1)$$

where P is fractional occupancy, D is the drug concentration, and K_d is the dissociation constant. This function has a sigmoid shape if plotted against the logarithm of the drug concentration. The authors proposed, for the assessment of self-administration behavior, to use so-called pharmacological reinforcement function (PRF) which allows us to understand the phenomenon in "a dynamic system sense" because the behavior of an organism determines the amount of drug in the body. Specifically, for self-administration, they plotted response rate against the amount of drug/time and assumed that it will be also a sigmoid function. The second function they used was the so-called feedback function (FF) [58] which gives rate of reinforcement as a function of rate of response for a given schedule of reinforcement. In the case of the self-administration paradigm, the rate of reinforcement is the amount of drug per unit time. Thus the characteristics of the FF will depend not

only on the type of schedule and schedule parameters, but also on the unit dose of the drug. In contrast to the sigmoidal PRF, the FF is a linear function under ratio schedules:

$$r = pd / F \quad (7.a.2)$$

where r is reinforcement rate, p is a response (lever pressing, for example) rate, d is a drug unit dose, and F is a number of responses required for one injection (fixed ratio schedule). It should be noted, that response rate is a "run rate", i.e., the rate within the burst of activity following the long pause induced by the drug injection. Under the condition of stable self-administration; both of the relevant functions must be simultaneously satisfied, the PRF by assumption, and FF by definition of the particular schedule.

There are a maximum of three points on the graphs where the straight line crosses the sigmoid and, therefore, both functions can be simultaneously satisfied: the origin (at the dose zero), somewhere around the flexion point, and somewhere on the asymptotic part of the sigmoid. The following is a short description of some qualitative aspects of the system, given by Sizemore and Martin. At the lower part of the curves, between the origin and the second intersection point, the initial rate of response will not be stable (because functions deviate significantly from each other) and will have a trend to decrease. The system must eventually be attracted to the steady state that lies at the origin. On the other hand, if rate of response is initially high (higher than the second intersection point) the system will be attracted to the steady state that lies at the sigmoid asymptote. It was concluded, although it is not clear how they arrived to this conclusion, that under ratio schedules there should typically be no steady state response rates observed above saline rate and below the peak rate at the minimal dose maintaining stable self-administration response. These intermediate rates, which constitute the ascending limb of the dose response curve, can be obtained only after averaging individual rates across animals.

Behavior maintained under interval schedules would be different in certain aspects from that under ratio schedules.

$$r = d / (I + 0.5 / p) \quad (7.a.3)$$

where r is reinforcement rate, p is a response rate, d is a drug unit dose, I is the interval schedule parameter. This feedback function, in contrast to the fixed ratio function (7.a.2), is monotonically accelerating and the authors conclude somehow that the points of intersection are always stable states. Therefore, the theory predicts that under interval schedules the response rate should increase as a function of dose (ascending limb), but then level off at the asymptote of the PRF.

In the model proposed by Sizemore and Martin, the time spent pausing does not enter into the feedback portion of the theory. Theoretically, the feedback operates only during the time that responding is occurring post-pause (run rate, not mean rate per session). This position is consistent with the view that organisms "titrate" levels of the self-administered drug [94]. The duration of inter-injection pauses on the drug unit dose (pause function, PF) is not expected to have an asymptote because higher dose is supposed to induce longer interval up to the point of lethal doses (infinite pause). The shape of the PF is close to the following exponential function:

$$PD = a\mathrm{e}^{bd} \quad (7.a.4)$$

where PD is a pause duration, d is a drug unit dose, a and b are constants. Although a substantial amount of variance (88% – 96% for individual rats) was accounted for, it is possible that the use of a three-parameter exponential function is justifiable:

$$PD = PD_0 + a\mathrm{e}^{bd} \quad (7.a.4)$$

The PD_0 parameter serves to correct the meaningless durations at low doses that do not maintain self-administration when responding ceases. Because the feedback is assumed to operate exclusive of the pause duration, the feedback function must be modified. That is, the functional interval value is the interval parameter minus the pause generated by the dose. Interval schedules with parameter values less than the duration of the pause are equivalent to fixed ratio 1 (FR1) schedules. With larger schedule parameter values or lower unit doses, the interval properties of the schedule begin to predominate. As mentioned, the run rate dependence on the dose should be truly biphasic (having ascending and descending limbs) in contrast to the fixed ratio schedules when run rate is relatively independent on the unit dose.

b. Satiety Threshold Model of Drug Self-Administration

Tsibulsky and Norman [95] based their mathematical model of cocaine self-administration on two typical characteristics of this behavior: (i) self-injections occur with regularity, (ii) the rate of self-injections is inversely related to the unit dose of the drug. If inter-injection intervals are stable then one can assume that the level of the drug in the body is stable. According to first-order kinetics this level in a one compartment model will be:

$$D = D_0\, \mathrm{e}^{-kT} \quad (7.b.1)$$

where D_0 is the peak amount of the drug in the body produced by the injection, D is the total amount of drug in the body at the beginning of the next injection, T is the inter-injection interval, and k is the elimination rate constant. Then, it was hypothesized that the same minimum level is maintained across different drug unit doses:

$$D_{ST} = (D_{ST} + D_U)\, \mathrm{e}^{-kT} \quad (7.b.2)$$

where D_{ST} is the minimum level of the drug, called the satiety threshold, maintained at different unit doses (D_U). Rearrangements give the inter-injection interval as a function of the unit dose, the satiety threshold and the elimination half-life ($t_{1/2} = \ln(2) / k$):

$$T = \ln(1 + D_U / D_{ST})\, t_{1/2} / \ln(2) \quad (7.b.3)$$

According to the satiety threshold model inter-injection intervals are proportional to the drug unit dose (nonlinearly), to the half-life of the drug (linearly), and inversely proportional to the satiety threshold (nonlinearly). The half-life

constitutes the pharmacokinetic parameter of the process; the satiety threshold constitutes the pharmacodynamic parameter of the process, and the unit dose represents the environmental characteristics of the process. Given the inter-injection intervals at different unit doses, both pharmacodynamic/pharmacokinetic parameters can be estimated. Nonlinear regression analysis applied to the individual animal dose-interval data [95] showed that the satiety threshold was in the range of 1.6 – 1.9 mg/kg and the elimination half-life was in the range of 7 – 12 min.

The variance in the inter-injection intervals implies that the parameters determining the interval are subject to fluctuation. If a rat initiates self-administration at the time when the cocaine level reaches the mean D_{ST}, then following the injection the cocaine level will be $D_{ST} + D_U$. If the next self-administration occurs at an interval T_1, the cocaine level will be $(D_{ST} + D_U)\ e^{-kT_1} = D_{ST} + \Delta D_{ST}$, where ΔD_{ST} is the difference between the mean D_{ST} and the D_{ST} at the interval T_1. In addition, any fluctuation of the elimination half-life ($\Delta t_{1/2}$) during the session would also produce a change of the observed inter-injection interval T_1. Therefore, the difference (ΔT) between the mean inter-injection interval T and the observed interval T_1 can be described by:

$$\Delta T = [\ln(1+D_U / D_{ST})\ t_{1/2} / \ln(2)] - [\ln((D_{ST} + D_U) / (D_{ST} + \Delta D_{ST}))\ \Delta\ (t_{1/2} + \Delta t_{1/2}) / \ln(2)] \quad (7.b.4)$$

The standard deviation of the mean inter-injection interval encompasses 96% of the observed ΔT. Using equation (7.b.4), the best fit to the mean S.D. across the range of cocaine unit doses generated values for the fluctuation of the satiety threshold ΔD_{ST} of + 0.14 mg/kg and – 0.13 mg/kg from the mean of 1.7 mg/kg. The fluctuation of the elimination half-life $\Delta t_{1/2}$ was calculated to be ± 20 s from the mean of 8.2 min (492 s). If there were no fluctuations of $t_{1/2}$ then the time for cocaine level to drop from D_{ST} + 0.14 mg/kg to D_{ST} – 0.13 mg/kg will be equal to $((D_{ST} + 0.14) – (D_{ST} – 0.13)) \cong t_{1/2} / \ln(2) / D_{ST} = 112$ sec, where $D_{ST} = 1.74$ mg/kg and $t_{1/2} = 8.2$ min. It will take the same 56 s for cocaine levels to drop from the upper boundary to the mean satiety threshold and from the mean to the lower boundary. Because the rate of elimination at the higher levels of cocaine is higher, cocaine levels will drop further in the same time resulting in an asymmetric variance from the mean level (0.14 *vs* 0.13 mg/kg). Thus, the total value for the S.D. of the mean inter-injection interval consists of the deviation (± 56 s) caused by the fluctuations of the satiety threshold, which does not depend on the unit dose of cocaine, plus the deviation caused by the fluctuations of half-life, which is $T / t_{1/2} \cong \Delta t_{1/2}$, i.e., proportional to the cocaine unit dose.

Norman and coworkers [96] also demonstrated that the cocaine-induced priming effect (the reinstatement of self-administration behavior following its extinction) takes place if cocaine level was above a certain minimal level, they called the priming threshold. Therefore, at least three zones of cocaine levels in the body can be defined in their relationship with self-administration behavior: (1) the basal level below priming threshold [96] where cocaine does not have any behavioral effects, (2) the compulsion zone above priming but below satiety threshold where cocaine induces self-administration behavior (lever pressing, for example), and (3) the satiety zone above the satiety threshold where cocaine selectively suppresses self-administration behavior and induces satiety [97].

The concept of the satiety zone predicts that any cocaine level fractionally above the satiety threshold the self-administration behavior will cease. Infusion of cocaine at a rate V_{in} that is equal or greater than the rate of elimination ($V_{out,ST}$) of cocaine at the satiety threshold should suppress self-administration completely ($T = \infty$):

$$V_{in} >= k\ D_{ST} \quad \text{or} \quad V_{in} >= \ln(2) / t_{1/2}\ D_{ST} \quad (7.b.5)$$

Supplemental continuous infusion of cocaine in animals self-administering cocaine will increase the inter-injection intervals proportionally to the rate of infusion:

$$T = \ln[1 + k\ D_U / (V_{out,ST} - V_{in})]\ t_{1/2} / \ln(2) \quad (7.b.6)$$

Equation (7.b.6) is a hyperbolic logarithmic function of V_{in}, i.e., the relationship between the rate of infusion and inter-injection intervals is nonlinear with significant effect of infusion only at rates approaching the rate of elimination at the satiety threshold, $V_{out,ST}$.

Equation (7.b.3) assumes that the injection is instant. In fact, an injection takes time (T_{inj}) which represents the minimal Time-Out (TO, T_{out}) and may reach many seconds. For the case when the rate of injections equals $V_{out,ST}$ (there is no pauses between injections) the following equations may be written:

$$V_{in} = D_{U,min,ST} / T_{out} = V_{out,ST} \quad (7.b.7)$$

Equation (7.b.7) predicts that there is a minimal drug unit dose ($D_{U,min,ST}$) required to maintain cocaine level at the satiety threshold and injections should follow each other without any pauses:

$$D_{U,min,ST} = V_{out,ST}\ T_{out} \quad (7.b.8)$$

The duration of injection represents the absolutely minimal TO, i.e., the period when the following injection cannot be initiated. In many studies of drug self-administration, experimenters impose TO period even longer than the duration of injection itself. The longer T_{out} period makes the minimal dose $D_{U,min,ST}$ proportionally higher. Equation (7.b.8) simply states that the unit dose lower than $D_{U,min,ST}$ will be self-administered without pauses between TO periods but the steady state level of cocaine will be maintained below the satiety threshold.

Furthermore, the following equation predicts that there is a unit dose which cannot maintain self-administration behavior at all because the steady state level of cocaine will be below the priming threshold:

$$D_{U,min,PT} = V_{out,PT}\ T_{out} \quad (7.b.9)$$

This dose $D_{U,min,PT}$ represents the self-administration threshold dose. Noteworthy, this threshold is not absolute and will change up or down proportionally to the TO period and inversely proportionally to the priming threshold. Equations (7.b.7) – (7.b.9) are applicable not only if TO is a

limiting factor of self-administration. There is another limit imposed by how fast animals can perform the required response (lever press, for example). These equations hold also if T_{out} represents minimal time between injections defined by the fastest rate of performance.

The satiety threshold model of drug self-administration represents a deterministic model based on the rational assumption that behavior is under the control of physicochemical processes in the body related to pharmacokinetics/pharmacodynamics of the drug. It is the first model rejecting the teleological approach that assumes that self-administration behavior is under the control of cognitive processes guided by the motivation to optimize the balance between positive and negative subjective effects of the drug.

c. Allostatic Decrease of Reward Threshold Model of Addiction

Escalation of drug use in response to increased drug availability is well documented [98]. The important aspect of this escalation is that it goes along with developing tolerance to most drug effects suggesting that some kind of homeostatic process is involved. Recently, Ahmed and Koob reviewed the results of studies in rats regarding the dependency of cocaine intake on the duration of self-administration session and proposed a mathematical model [98]. According to the hedonic allostatic hypothesis of drug addiction [99] the precipitation of compulsive drug use by increased drug availability would result from a chronic decrease in reward system responsivity that is produced by an overactivation of brain antireward processes.

The model assumes that behavior is under the control of a reward system. There is a reward set point (T_S) which represents an ideal level of the activity of the reward system. Real activity (T_0) is measured by the sensor and compared to the set point. The error $T_0 - T_S$ is used by the controller to turn on and off the drug-seeking behavior. Injected drug of abuse, for example, cocaine decreases activity of the reward system by

$$T = T_0 - T_{max}C / (T_{50} + C) \qquad (7.c.1)$$

where T is the threshold-decreasing effect of cocaine below baseline threshold T_0. This lowering effect of cocaine is a hyperbolic function of T_{max} which is the maximal effect of cocaine and C which is the cumulative concentration of cocaine in the brain. T_{50} represents the concentration of the drug that produces half of the maximal effect, i.e., the index of drug potency.

The threshold decreasing effect of cocaine changes as a function of cocaine concentration (C) which in turn changes with time (t) according to first order kinetic process for a two-compartment model:

$$C = DK(e^{-\beta t} - e^{-\alpha t}) \qquad (7.c.2)$$

where D is the drug unit dose and K, α and β are complex pharmacokinetic parameters:

$$K = k_{12} / V_b(\alpha - \beta)$$

$$\alpha = (k_{12}+k_{21}+k_{el} + \sqrt{(k_{12} + k_{21} + k_{el})^2 - 4k_{21}k_{el}}\ )/2$$

$$\beta = (k_{12}+k_{21}+k_{el} - \sqrt{(k_{12} + k_{21} + k_{el})^2 - 4k_{21}k_{el}}\ )/2$$

where k_{el} is the rate constant of elimination, k_{12} is the rate constant of distribution from the blood to the brain, k_{21} is the rate constant of redistribution from the brain to the blood, V_b is the apparent volume of distribution in the brain. All parameters in this model can be divided into three categories: pharmacokinetic parameters (k_{el}, k_{12}, k_{21} and V_b), pharmacodynamic (T_{max} and T_{50}), and motivational (T_0 and T_S).

The temporal distribution of self-injections may be explained by this model as follows: the first injection is triggered by the initial error ($T_0 - T_S > 0$). These injections are repeated until cocaine concentration is loaded to the level which decreases the activity of the reward system to the set point T_S. After loading, the abrupt increase of self-injection intervals occurs because cocaine threshold-decreasing effect brings the reward system activity to the level below the set point. Injections then will resume each time when cocaine level decreases to the concentration C_S which brings the reward system activity exactly to the set point T_S. Taking values for the pharmacokinetic parameters from classical pharmacokinetic studies, we can assess the inter-injection intervals.

In order to verify this model some simplification are needed. It was assumed that infinitely high concentrations of cocaine bring the reward threshold to zero:

if $C \to \infty$ then $T_{max}C / (T_{50} + C) \to T_0$ therefore

$$T = T_0 [1 - C / (T_{50} + C)]$$

Cocaine concentration which brings the reward threshold to the set point can be written as

$$T_S = T_0 [1 - C_S / (T_{50} + C_S)]$$

Rearranging yields

$$C_S = T_{50} (T_0 - T_S) / T_S \qquad (7.c.3)$$

The second simplification was made that the distribution of cocaine from the blood to the brain is very likely to be finished before the entire injected dose is eliminated and therefore this distribution component does not influence the duration of drug effects above the set point. Indeed, experimental pharmacokinetic studies of cocaine demonstrate that $\alpha >> \beta$. If we designate the inter-injection interval as I then the cocaine level at the time of the next injection is

$$C_S = (DK + C_S) (e^{-\beta I} - e^{-\alpha I})$$

Simplification and rearranging yields

$$C_S = DK / (e^{\beta I} - 1) \qquad (7.c.4)$$

Now we can combine two equations for C_S and write the following equality

$$DK / (e^{\beta I} - 1) = T_{50} (T_0 - T_S) / T_S$$

Rearranging yields

$$I = \frac{1}{\beta} \ln \frac{1 + DKT_S}{T_{50}(T_0 - T_S)} \quad (7.c.5)$$

Equation (7.c.5) relates inter-injection intervals with the drug unit dose D which is under experimenter control and also with six more parameters some of them are pharmacokinetic constants (k_{el}, k_{12}, k_{21} and V_b) values for which can be taken from pharmacokinetic studies. Three other parameters are pharmacodynamic (T_{50}) and motivational (T_0 and T_S) which cannot be determined by the regression analysis of dose-response relationship because of their interdependency.

In general, the allosteric decrease of reward threshold model is similar to the satiety threshold model of self-administration. The most significant difference is that the former contains three immeasurable pharmacodynamic/-motivational parameters which have the dimension of drug concentration. The second difference is that the allosteric model assumes a two-compartment model of cocaine elimination and distribution instead of a simplified one-compartment model. However, the allosteric model still has to ignore the influence of the distribution phase on the inter-injection interval to avoid insurmountable difficulties of derivation of equation (7.c.4) for C_S.

8. POPULATION BEHAVIOR

Mathematical models in Section 2.b. described behavior of an individual animal in the complex environment where some objects of the environment can change their behavior (can learn). In this section, mathematical models of an individual as a member of population will be presented.

a. Individual-Based Modeling of Population Dynamics

All ways of approaches to the modeling of ecological systems can be categorized as either "top-down" or "bottom-up". In the first group of models, pioneered by Nicholson's [100] model of competition for resources, variables in differential equations captured the global properties of a system. In contrast, bottom-up models start from a description of local interactions. For example, how and when do parasites of a honeybee colony reproduce? What happens when two or more parasites compete for the same resource? The experimental data can then be used in an individual-based model [101-103].

The population dynamics of tree-dwelling aphids (*Myzocallis boerneri*) sharply increases in number during the first 15 – 40 days in spring, then sharply declines to a plateau of low numbers in summer and recovers in autumn [101]. They assumed that behavior of individual aphids can explain this dynamics. The shape of the reproduction curve (number of off-spring, $F(t)$, born per mother per day) depends of the relative size of the soma and gonads:

$$F(t) = 2Rg(D) / (s_0 + g_0)(2 - t/D) \quad (8.a.1)$$

where t is the age of the individual, s_0 is an initial size of the soma, g_0 is an initial size of the gonads, R is the constant of exponential gonadal growth rate ($dg/dt = Rg$), $g(D)$ is the size of the gonads at maturity, D is developmental time (age of maturity). The second factor controlling the number of individuals is migration. The proportion of aphids that fly away from a colony, m, is assumed to be a linear function of the number of aphids (larvae + adults) present in the colony, x, with positive slope S and positive intercept i, times a linear function depending on parameter a:

$$m = (S_x + i)(A_{max} - a) / (A_{max} - A_{min}) \quad (8.a.2)$$

The second term in equation (8.a.2) simulates the reluctance of aphids to migrate, when living on a good host plant, and their willingness when on a poor quality plant. The success of an aphid in finding another host plant is assumed to be in the range of 0.01, meaning that the immigration from other colonies equals $0.01m$. This is simply an illustration of an individual-based population model.

Classical theory of population dynamics ignored the individual approach for historical and technical reasons [102]. Historically, classical ecological theory developed in the first half of 20th century, when the implications of the Darwinian theory were much more vaguely perceived than today. It was commonly believed that biological characteristics are for the good of species. Most biologists thought of the evolutionary theory of natural selection only as a metaphysical research program. A population, a community and even an ecosystem were considered to be like a single organism or cell. Therefore, this view implied that a population can be studied, understood and manipulated without worrying about the interrelations between individuals or about how adaptations good for populations can evolve. As a result, developed models were only epiphenomenological descriptions of some population processes. The logistic equation is a classical example of such a description. The technical difficulties are related to methods of how to obtain data about individual animal living in population in the field and not in isolated experimental settings.

Advanced techniques and development of population genetics helped to overcome these difficulties. Population genetics does not contain epiphenomenological descriptions that are not derived from mechanisms of inheritance. The theoretical structure of population genetics reveals that any model of a complex system must be derived from the properties of its elements, i.e., individuals.

b. Language

Theories of the evolution of language explain the evolution of the simplest possible communication systems, role of natural selection in the language evolution, and how human language changes [104-106].

In the model proposed by Trapa and Nowak [107], a language L can be described by two matrices. Denote by P the n x m matrix whose entries p_{ij} represent the probabilities that an individual will send signal j when wanting to transmit the information *"object i"*. Let Q be the m x n matrix whose entries q_{ji} are the probabilities that an individual will conceive of *"object i"* when receiving signal j. Therefore, there are n objects and m signals. We have the following constrains if an individual may not always signal seeing an

object, and not always conceive of an object when receiving a signal:

$$\sum_{j=1}^{m} p_{ij} \le 1 \qquad \text{and} \qquad \sum_{i=1}^{n} q_{ji} \le 1 \tag{8.b.1}$$

Consider two individuals that use languages $L(P, Q)$ and $L'(P', Q')$. Payoffs can be defined as:

$$F(L, L') = (1/2) \sum_{i=1}^{n} \sum_{j=1}^{m} \left(p_{ij} q'_{ji} + p'_{ij} q_{ji} \right) \tag{8.b.2}$$

The probability of transmitting object I from L to L' is given by $\sum_j p_{ij} q'_{ji}$. The payoff function sums these probabilities for all objects and then takes the average over the two situations, L signals to L' and L' signals to L. The specific assumptions of the payoff function are that sending and receiving yield equal payoffs and that all objects contribute the same amount of the payoff, $F(L, L') = F(L', L)$. Following the notions of classical game theory, a language L can be defined as a strict Nash equilibrium if $F(L, L) > F(L', L)$ for all languages $L' \neq L$. Furthermore, a language L in a Nash equilibrium if $F(L, L) \geq F(L', L)$ holds for all languages L'. A language L is an evolutionarily stable strategy (ESS) if $F(L, L) > F(L', L)$ holds for all $L' \neq L$, or if $F(L, L) = F(L', L)$ then we must have $F(L, L') > F(L', L')$.

Nash equilibria or ESS are fixed points of evolutionary dynamics: if a whole population uses a language that is either Nash or ESS then evolution will normally not change this situation, i.e., mutant strategies – as long as they are rare – cannot invade the population by natural selection. Random drift is also very improbable. This is not the case for general Nash equilibria: if $F(L, L) = F(L', L) = F(L', L')$ then L and L' are neutral variants. There is no selection, but random drift can replace L by L' [106].

The human language dynamical equation is a game dynamical equation with learning. Mitchener and Nowak [108] showed that complex limit cycles and chaos can arise even for very simple choices of the payoff and learning matrices. They considered five languages (strategies), each of which is a strict Nash equilibrium. Pure game dynamics would have five stable equilibria corresponding to linguistically homogeneous populations. However, a carefully structured learning matrix is sufficient to induce chaos. Thus very conservative natural choices of payoff and learning matrices lead to deterministic chaos. Simple learning errors can lead to complex, unpredictable and seemingly stochastic changes in languages over time.

SUMMARY

Mathematical models of behavior of a whole animal can be divided into two main classes. First, models are based on the principle of teleological determinism assuming that subjects choose the behavior which will lead them to a better payoff in the future. Two groups of theories belong to this class: game theories relying upon the Nash equilibria principle and reinforcement theories relying upon the Skinnerian reinforcement principle or matching law.

The second class consists of models which are based on the principle of causal determinism assuming that subjects do not choose from a set of possibilities but rather are compelled to do what is predetermined and they do so in response to specific stimuli. In behavioral experiments, an investigator usually deals with a chain of cause-effect events A → B → C → A → B → C →…, which can be viewed as circular $\ulcorner$A → B → C →$\urcorner$. Therefore, these events are said to be in a *temporal causality loop*.

Different mathematical models explain the same chain of events starting from different points and moving from event to event differently. For example, the sequence of events in the self-administration paradigm according to causal determinism is as follows: stimuli A (cues or self-administered drugs) induce behavior B (lever press) which induces event C (drug injection) which maintains effects of stimuli A and so on. In contrast, a teleologically determined sequence of events is as follows: injected drugs or memories of previous injections (event C) induce euphoria or similar kinds of subjective effects (event A) which induce lever-pressing (event B). The principle of causation says that behavior occurs because necessary and sufficient stimuli were presented. The principle of teleology says that behavior occurs in order to achieve a certain goal or for a defined purpose. Mathematically these two classes of models may be very similar and based on the same fundamental equations (linear, exponential, hyperbolic etc.). However, the parameters in equations describing causal relationships typically are defined in terms of SI (System Internationale or Metric System) derived units, while parameters in teleological models are typically dimensionless or have a dimension of s^{-1}. Moreover, these rate terms (number of events per unit time) merely substitute for more theoretically appropriate probability terms (dimensionless) when the latter cannot be practically determined.

The circularity of teleological models is inevitable because these models use tautological definitions of the basic explanatory principle. Thus, Skinnerian reinforcement is defined as an event resulting in an increase of the probability of behavior which occurred before the reinforcement took place. This definition is obviously tautological (compare to "all crows are either black, or they are not black"): reinforcement either increases the probability of behavior (positive reinforcement) or decreases it (negative reinforcement).

The understanding of mathematical modeling in biology and particularly in behavioral and social sciences fluctuates between a Platonian view that an equation is the essence, pure ideal design of a complex biological process and a more contemporary view that a simple equation is a crude approximation, a formal presentation of a hypothesis which is only valid if its assumptions agree with empirical data obtained under certain conditions. Equations are inherent in Nature, but in the real world the ideal design is corrupted and obscured be many unimportant and superficial processes which have to be ignored. The mathematical discovery that strictly deterministic systems may exhibit chaotic behavior has changed our understanding of how relatively simple mathematical equations can describe apparently complex

behavior. Some progress has been made in modeling of electrical activity of the brain, for example. We can expect that the theory of deterministic chaos will be applied in future to model behavior of an individual animal.

REFERENCES

References 109-111 are related articles recently published in Current Pharmaceutical Design.

[1] Gould SJ, Lewontin RC. The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. Proc R Soc Lond B Biol Sci 1979; 205(1161): 581-98.

[2] Stevens SS. On the psychophysical law. Psychol Rev 1957; 64(3): 153-81.

[3] Helson H. Adaptation-level theory. An experimental and systematic approach to behavior. New York, Harper and Raw, 1964.

[4] DeVries HL. The quantum character of light and its bearing upon threshold of vision, the differential sensitivity and visual acuity of the eye. Physica 1943; 10: 553-64.

[5] Rose A. The sensitivity performance of the human eye on an absolute scale. J Optical Soc Am 1948; 38: 196-208.

[6] von Fechner GT. Elemente der Psychophysik. Bristol, Thoemmes, 1998.

[7] Guilford JP. A generalized psychophysical law. Psychol Rev 1932; 39(1): 73-85.

[8] Michels WC, Helson H. A reformulation of the Fechner law in terms of adaptation-level applied to rating-scale data. Am J Psychol 1949; 62: 355-68.

[9] Restle F, Greeno JG. Introduction to mathematical psychology. reading, MA, Addison-Wesley, 1970.

[10] Hull CL. Principles of behavior: An introduction to behavior theory. New York, Appleton-Century Co, 1943.

[11] Luce RD. Individual choice behavior. New York, NY, Wiley, 1959.

[12] Restle F. Psychology of judgment and choice. New York, NY, Wiley, 1961.

[13] Von Neumann J, Morgenstern O. Theory of games and economic behavior. Princeton, Princeton University Press, 1944.

[14] von Neumann J. Zur theorie der gesellschaftsspiele. Mathematische Annalen 1928; 100: 295-320.

[15] Nash J. Equilibrium points in n-person games. Proc Natl Acad Sci USA 1950; 36(1): 48-9.

[16] Maynard-Smith J, Price GR. The logic of animal conflict. Nature 1973; 246(5427): 15-8.

[17] Axelrod RM. The evolution of cooperation. New York, NY, Basic Books 1984.

[18] Imhof LA, Fudenberg D, Nowak MA. Evolutionary cycles of cooperation and defection. Proc Natl Acad Sci USA 2005; 102(31): 10797-800.

[19] Breed MD, Rasmussen CD. Behavioral strategies during intermale agonistic encounters in a cockroach. Animal Behav 1980; 28: 1063-9.

[20] Riechert SE. Games spiders play: III. Cues underlying context associated changes in agonistic behavior. Anim Behav 1984; 32(1): 1-15.

[21] Riechert SE. Spider fights as a test of evolutionary game theory. Am Sci 1986; 74(6): 604-10.

[22] Herrnstein RJ, Prelec D. Melioration - A theory of distributed choice. J Econ Perspect 1991; 5(3): 137-56.

[23] Natapoff A. How symmetry restricts symmetric choice. J Math Psychol 1970; 7(3): 444-65.

[24] Herrnstein RJ. On the law of effect. J Exp Anal Behav 1970; 13(2): 243-66.

[25] de Villiers PA, Herrnstein RJ. Toward a law of response strength. Psychol Bull 1976; 83(6): 1131-53.

[26] Herrnstein RJ. Quantitative hedonism. J Psychiatr Res 1971; 8(3-4): 399-412.

[27] McDowell JJ. Matching theory in natural human environments. The Behavior Analyst 1988; 11(2): 95-109.

[28] Staddon JE. On Herrnstein's equation and related forms. J Exp Anal Behav 1977; 28(2): 163-70.

[29] Skinner BF. The behavior of organisms: An experimental analysis. New York, Appleton-Century Company, 1938.

[30] Rachlin H. On the tautology of the matching law. J Exp Anal Behav 1971; 15(2): 249-51.

[31] Killeen P. The matching law. J Exp Anal Behav 1972; 17(3): 489-95.

[32] Herrnstein RJ. Derivatives of matching. Psychol Rev 1979; 86(5): 486-95.

[33] McDowell JJ. A computational model of selection by consequences. J Exp Anal Behav 2004; 81(3): 297-317.

[34] Dallery J, Soto PL. Herrnstein's hyperbolic matching equation and behavioral pharmacology: Review and critique. Behav Pharmacol 2004; 15(7): 443-59.

[35] Baum WM. On two types of deviation from the matching law: Bias and undermatching. J Exp Anal Behav 1974; 22(1): 231-42.

[36] Baum WM. Matching, undermatching, and overmatching in studies of choice. J Exp Anal Behav 1979; 32(2): 269-81.

[37] Staddon JE. In: Gilbert RM, Millenson JR Eds, Reinforcement: Behavioral analyses. New York, Academic Press 1972; 212-62.

[38] McDowell JJ. On the falsifiability of matching theory. J Exp Anal Behav 1986; 45(1): 63-74.

[39] McDowell JJ. On the classic and modern theories of matching. J Exp Anal Behav 2005; 84(1): 111-27.

[40] Thurstone LL. The learning curve equation. Psychological Monographs 1919; Vol. 26.

[41] Thorndike EL. The Fundamentals of Learning. New York, NY, Columbia Univ. Teach Coll Bur Publ 1932.

[42] Gulliksen H. A rational equation of the learning curve based on Thorndike's law of effect. J Gen Psychol 1934; 11: 395-434.

[43] Bush RR, Mosteller F. Stochastic models for learning. New York, NY, John Willey & Sons, 1955.

[44] Estes WK. In: Koch S Ed, Study I. Conceptual and systematic. volume 2. General systematic formulations, learning, and special processes. New York, NY, McGraw-Hill 1959; 380-491.

[45] McDowell JJ, Bass R, Kessel R. A new understanding of the foundation of linear system analysis and an extension to nonlinear cases. Psychol Rev 1993; 100(3): 407-19.

[46] Mazur JE, Hastie R. Learning as accumulation: A reexamination of the learning curve. Psychol Bull 1978; 85(6): 1256-74.

[47] Rescorla RA, Wagner AR. A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black AH, Prokasy WF Eds, Classical conditioning II: current research and theory. New York: Apple-Century-Crofts 1972; 64-99.

[48] Mackintosh NJ. A theory of attention: Variations in the associability of stimuli with reinforcement. Psychol Rev 1975; 82(4): 276-98.

[49] Van Hamme LJ, Wasserman EA. Cue competition in causality judgements: The role of nonrepresentation of compound stimulus elements. Learn Motiv 1994; 25: 127-51.

[50] Miller RR, Barnet RC, Grachame NJ. Assessment of rescorla-wagner model. Psychol Bull 1995; 117(3): 363-86.

[51] Wagner AR. SOP: A model of automatic memory processing in animal behavior. In: Spear NE, Miller RR Eds, Information processing in animals: Memory mechanisms. Hillsdale, NJ: Lawrence Erlbaum Assoc., Publishers 1981; 5-47.

[52] Brandon SE, Vogel EH, Wagner AR. Stimulus representation in SOP: I. Theoretical rationalization and some implications. Behav Process 2003; 62(1): 5-25.

[53] Killeen PR. Mathematical principles of reinforcement. Behav Brain Sci 1994; 17: 105-72.

[54] Killeen PR, Sitomer MT. MPR. Behav Processes 2003; 62(1-3): 49-64.

[55] Killeen PR, Hall SS, Reilly MP, Kettle LC. Molecular analyses of the principal components of response strength. J Exp Anal Behav 2002; 78(2): 127-60.

[56] Nevin JA. An integrative model for the study of behavioral momentum. J Exp Anal Behav 1992; 57(3): 301-16.

[57] Nevin JA. Mathematical principles of reinforcement and resistance to change. Behav Processes 2003; 62(1-3): 65-73.

[58] Baum WM. The correlation-based law of effect. J Exp Anal Behav 1973; 20(1): 137-53.

[59] Mazur JE. Hyperbolic value addition and general models of animal choice. Psychol Rev 2001; 108(1): 96-112.

[60] Grace RC. A contextual model of concurrent-chains choice. J Exp Anal Behav 1994; 61(1): 113-29.

[61] Grace RC. Choice between fixed and variable delays to reinforcement in the adjusting-delay procedure and concurrent chains. J Exp Anal Behav 1996; 22(3): 362-83.
[62] Fantino E. Choice and rate of reinforcement. J Exp Anal Behav 1969; 12: 723-30.
[63] Fantino E, Preston RA, Dunn R. Delay reduction: Current status. J Exp Anal Behav 1993; 60(1): 159-69.
[64] Nevin JA, Mandell C, Atak JR. The analysis of behavioral momentum. J Exp Anal Behav 1983; 39(1): 49-59.
[65] Camerer C. Behavioral economics: Reunifying psychology and economics. Proc Natl Acad Sci USA 1999; 96(19): 10575-7.
[66] DeGrandpre RJ, Bickel WK, Hughes JR, Layng MP, Badger G. Unit price as a useful metric in analyzing effects of reinforcer magnitude. J Exp Anal Behav 1993; 60(3): 641-66.
[67] Samuelson PA. A note on measurement of utility. Rev Econom Stud 1937; 4(2): 155-61.
[68] Loewenstein G, Prelec D. In: Loewenstein G, Elster J Eds, Choice over time. New York, NY, Russell Sage Foundation 1992; 119-45.
[69] Green L, Myerson J, Holt DD, Slevin JR, Estle SJ. Discounting of delayed food rewards in pigeons and rats: Is there a magnitude effect? J Exp Anal Behav 2004; 81(1): 39-50.
[70] Mazur JE. In: Commons ML, Mazur JE, Nevin JA, Rachlin H Eds, The effect of delay and intervening events on reinforcement value. Inc, Hillsdale, NJ, USA, Lawrence Erlbaum Associates 1987; 55-73.
[71] Ainslie G, Haslam N. In: Loewenstein G, Elster J Eds, Choice over time. New York, NY, Russell Sage Foundation 1992; 57-92.
[72] Bikhchandani S, Hirshleifer D, Welch I. Learning from the behavior of others: Conformity, fads, and informational cascades. J Econom Perspect 1998; 12(3): 151-70.
[73] Weibull J. Evolutionary game theory. Cambridge, MA, MIT Press, 1995.
[74] Rabin M. Incorporating fairness into game theory and economics. Am Econom Rev 1993; 1993(5): 1281-302.
[75] Camerer CF, Ho TH. Experienced-weighted attraction learning in normal form games. Econometrica 1999; 67(4): 827-74.
[76] Skinner BF, Heron WT. Effects of caffeine and benzedrine upon conditioning and extinction. Psychol Rec 1937; 1: 340-6.
[77] Dews PB. Effects of chlorpromazine and promazine on performance on a mixed schedule of reinforcement. J Exp Anal Behav 1958; 1(1): 73-82.
[78] Kelleher RT, Morse WH. Determinants of the specificity of behavioral effects of drugs. Ergebnisse der Physiologie 1968; 60: 1-56.
[79] Thompson T. Current status of behavioral pharmacology. In: Thompson T, Dews PB Eds, Advances in behavioral pharmacology. New York, Academic Press 1977; vol 1: 259-64.
[80] Gonzalez FA, Byrd LD. Mathematics underlying the rate-dependency hypothesis. Science 1977; 195(4278): 546-50.
[81] Dews PB. A behavioral effect of amobarbital. Naunyn Schmiedebergs Arch Exp Pathol Pharmakol 1964; 248: 296-307.
[82] Ksir C. Rate-convergent effects of drugs. In: Thompson T, Dews PB, McKim WA Eds, New York, Academic Press. Advances in behavioral pharmacology 1981; vol 3: 39-59.
[83] Dews PB. Rate-dependency hypothesis. Science 1977; 198(4322): 1182-3.
[84] Gonzalez FA, Byrd LD. Rate-dependency hypothesis - Reply. Science 1977; 198(4322): 1183.
[85] Thompson T, Dews PB, McKim WA. In: Thompson T, Dews PB, McKim WA Eds, Advances in behavioral pharmacology. New York, Academic Press 1981; Vol 3.
[86] Kuhar MJ, Joyce AR. Slow onset of CNS drugs: can changes in protein concentration account for the delay? Trends Pharmacol Sci 2001; 22(9): 450-6.
[87] Cole JO, Davis JM. In: Bellak L, Loeb L Eds, The Schizophrenic Syndrome. New York, Grune & Stratton 1969; 478-568.
[88] Lieberman JA, Safferman AZ, Pollack S, Szymanski S, Johns C, Howard A, *et al*. Clinical effects of clozapine in chronic schizophrenia: Response to treatment and predictors of outcome. Am J Psychiatry 1994; 151(12): 1744-52.
[89] Kuhar MJ, Joyce AR. Is the onset of psychoactive drug effects compatible with a protein-synthesis mechanism? Neuropsychopharmacol 2003; 28(Suppl 1): S94-7.
[90] Norman AB, Battaglia G, Creese I. Differential recovery rates of rat D2 dopamine receptors as a function of aging and chronic reserpine treatment following irreversible modification: a key to receptor regulatory mechanisms. J Neurosci 1987; 7(5): 1484-91.
[91] Kuhar MJ, Joyce A, Dominguez G. Genes in drug abuse. Drug Alcohol Depend 2001; 62(3): 157-62.
[92] Svenningsson P, Chergui K, Rachleff I, Flajolet M, Zhang X, Yacoubi ME, *et al*. Alterations in 5-HT1B receptor function by p11 in depression-like states. Science 2006; 311(5757): 77-80.
[93] Sizemore GM, Martin TJ. Toward a mathematical description of dose-effect functions for self-administered drugs in laboratory animal models. Psychopharmacology 2000; 153(1): 57-66.
[94] Yokel RA, Pickens R. Drug level of d- and l-amphetamine during intravenous self-administration. Psychopharmacologia 1974; 34(3): 255-64.
[95] Tsibulsky VL, Norman AB. Satiety threshold: A quantitative model of maintained cocaine self-administration. Brain Res 1999; 839(1): 85-93.
[96] Norman AB, Norman MK, Hall JF, Tsibulsky VL. Priming threshold: A novel quantitative measure of the reinstatement of cocaine self-administration. Brain Res 1999; 831(1-2): 165-74.
[97] Norman AB, Tsibulsky VL. The compulsion zone: A pharmacological theory of acquired cocaine self-administration. Brain Res 2006; 1116:143-52.
[98] Ahmed SH, Koob GF. Transition to drug addiction: A negative reinforcement model based on an allostatic decrease in reward function. Psychopharmacology 2005; 180(3): 473-90.
[99] Koob GF, Le Moal M. Drug abuse: Hedonic homeostatic dysregulation. Science 1997; 278(5335): 52-8.
[100] Nicholson AJ. An outline of the dynamics of animal populations. Austr J Zool 1954; 2(1): 9-65.
[101] Kindlmann P, Dixon AFG. Population dynamics of a tree-dwelling aphid: Individual to populations. Ecol Model 1996; 89(1-3): 23-30.
[102] Lomnicki A. Individual-based models and the individual-based approach to population ecology. Ecol Model 1999; 115(2-3): 191-8.
[103] Sumpter DJ, Broomhead DS. Relating individual behaviour to population dynamics. Proc Biol Sci 2001; 268(1470): 925-32.
[104] Pinker S, Bloom P. Natural language and natural selection. Behav Brain Sci 1990; 13: 707-84.
[105] Bickerton D. Language and species. Chicago, IL, University of Chicago Press 1990.
[106] Deacon T. The symbolic species. London, UK, Allen Lane, The Pinguin Press 1997.
[107] Trapa PE, Nowak MA. Nash equilibria for an evolutionary language game. J Math Biol 2000; 41(2): 172-88.
[108] Mitchener WG, Nowak MA. Chaos and language. Proc Biol Sci 2004; 271(1540): 701-4.
[109] Pattij T, Olivier B, Waldinger MD. Animal models of ejaculatory behavior. Curr Pharm Des 2005; 11(31): 4069-77.
[110] Salvioli S, Capri M, Valensin S, Tieri P, Monti D, Ottaviani E, *et al*. Inflamm-aging, cytokines and aging: state of the art, new hypotheses on the role of mitochondria and new perspectives from systems biology. Curr Pharm Des 2006; 12(24): 3161-71.
[111] Valverde O. Participation of the cannabinoid system in the regulation of emotional-like behaviour. Curr Pharm Des 2005; 11(26): 3421-9.