https://doi.org/10.20378/irb-50371

An evolutionary model of reinforcer value

Dr. Matthias Borgstede

University of Bamberg

Markusplatz 3

D-96047 Bamberg

matthias.borgstede@uni-bamberg.de

An evolutionary model of reinforcer value

Highlights

- Reinforcer value is formally defined in the context of fitness maximization
- The presented model explicitly links reinforcer value to evolutionary fitness
- The model is applied to matching behavior, yielding new empirical hypotheses
- The results are valid for any mechanism that leads to maximal reinforcement

Abstract

Within the field of evolutionary biology, natural selection is often thought to favor traits that lead to individuals behaving as if they were maximizing their evolutionary fitness. The concept of the individual as a maximizer is also popular in behavioral psychology, especially when it comes to theories of operant learning. Here, the individual is taken to adapt its behavior to the local environment, such that the expected amount of reinforcer value is maximized.

Whereas there is a considerable consensus concerning the formal properties of an evolutionary maximand (‘fitness’), there is no generally accepted conceptualization of a corresponding behavioral maximand (‘reinforcer value’). However, such theoretical clarification is crucial to the development and empirical testing of learning theories, since it is impossible to decide whether the concept of reinforcer maximization is adequate, as long as the maximand is not well defined.

This paper presents a formal model of reinforcer value that is consistent with existing work on the nature of reinforcement and provides an explicit link between behavioral psychology and evolutionary biology. The main result is that *the reinforcer value of an additional time unit spent at a behavior equals its expected marginal effects on evolutionary fitness*. Applying the model to matching behavior, it is further demonstrated how the established link between reinforcer value and evolutionary fitness can be used to derive new hypotheses.

Keywords: *selection by consequences, evolutionary fitness, reinforcer value, phylogenetically important event, generalized matching law*

1 Introduction

Several authors have proposed an analogy between reinforcement learning and natural selection (e.g. Broadbent, 1961; Campbell, 1956; Gilbert, 1970; Herrnstein, 1964; Pringle, 1951). Staddon and Simmelhag (1971), for example, state that the 'Law of Effect […] can best be understood by analogy with evolution by means of natural selection' (p. 40). The analogy between natural selection and reinforcement learning was further developed by Skinner (1981) who even claims that natural selection and reinforcement learning are two instances of the same underlying causal principle: selection by consequences. The selectionist account of learning has also been adopted in neuronal accounts of reinforcement (Donahoe, Burgos, & Palmer, 1993), and Bayesian learning models (Richerson, 2019), and remains popular among behavioral psychologists (Baum, 2017, 2018a; A. M. Becker, 2019; Hull, Langman, & Glenn, 2001; Simon & Hessen, 2019; Staddon, 2016)[1].

However, the analogy between natural selection and reinforcement learning has also been subject to substantial criticism (e.g. the open peer commentaries to Skinner, 1984; also Burgos, 2019; Pennypacker, 1992; Tonneau & Sokolowski, 2000 for more recent accounts). Furthermore, Burgos (2019) argues that re-conceptualizing reinforcement as a special kind of selection by consequences is not necessary to link theories of learning to the broader scope of evolutionary theory. Instead, learning mechanisms can be regarded as being subject to natural selection themselves, yielding the notion of a common underlying causal principle redundant. This is in line with biological approaches to the evolution of learning from the perspective of niche-construction (Timberlake, 2001) that address the evolutionary adaptiveness of specific learning algorithms (see e.g. Aoki & Feldman, 2014 for an overview). The idea of natural selection favoring learning algorithms is plausible because the

[1] In fact, several variants of this analogy can be found throughout the literature. The position taken in this paper is that 'natural selection' and 'reinforcement learning' are analogues, in the way that they are both mechanisms of adaptation that can be conceptualized in terms of an implicit optimization principle.

world is too complex to produce rules that behave optimally in every possible circumstance (Frank, 1997; McNamara & Houston, 2009). On a more general level, attempts have been made to link learning (among other ontogenetic factors) to the process of natural selection to form an 'extended evolutionary synthesis' (Laland et al., 2015). The core idea of this approach is that natural selection acts on *phenotypic* variation which is only partly caused by genetic variation. Hence, behavioral changes due to reinforcement learning are an integral part of evolutionary change.

A similar argument has been proposed by Frankenhuis, Panchanathan, and Barto (2019), who argue that learning both results from and contributes to biological evolution. They further propose an integrative approach to behavioral adaptations by introducing formal models of reinforcement learning (namely Markov Decision Processes, MDP) into evolutionary theory. Treating both, learning and natural selection, as optimization processes, they replace the conceptual analogy between natural selection and learning by a functional relation: if both, evolutionary fitness and reinforcer value (i.e. 'reward' in the language of MDP) are maximized, they need to be intrinsically linked such that behavior that is optimal with regard to obtained reinforcer value yields the greatest average evolutionary fitness. This idea is further developed and formally modelled by Singh, Lewis, and Barto (2010) and Sorg (2011) who search for optimal reward functions with respect to a specified fitness function in a varying spatial environment. While presenting some interesting insights, their simulations rely entirely on numerical methods and thus do not provide general results. Moreover, Singh et al. (2010) model natural selection based on the amount of food intake. Even though food intake is commonly used as a proxy for evolutionary fitness – for example in optimal foraging theory (see Stephens & Krebs, 1986 for a review) - its effect on evolutionary fitness may vary between different environmental or individual characteristics, depending on the life-cycle of the species. This is true for any proxy of evolutionary fitness (McGraw & Caswell, 1996).

In this paper, I present definitions of evolutionary fitness and reinforcer value in a very general framework that is consistent with existing work from evolutionary biology and behavioral psychology. I further apply these definitions to choice behavior and derive a modified version of the generalized matching law. The aim is to explore the functional relation between natural selection and reinforcement learning by establishing a formal link between the maximands of the underlying processes. In order to achieve the desired level of abstraction, this paper does not deal with specific mechanisms of reinforcement learning or their neural basis, nor does it incorporate genetic mechanisms like transmission or mutation. Instead, natural selection and reinforcement learning are both conceived as problems of optimization, leaving open the proximate causes that lead (under certain circumstances) to 'optimal' behavior. Treating behavioral adaptations as problems of optimization comes with the assumption that there is some quantitative measure that is maximized (the maximand of the optimization program). On the level of the individual, this quantity is taken to be the value of the received reinforcers within a comparably small time scale and is thus called 'reinforcer value' (even though the model does not explicitly deal with mechanisms of reinforcement) [2]. On the level of the population, the quantity to be maximized is taken to be evolutionary fitness. The relation between these two maximization processes is by no means trivial since, first, they act on different time scales (relatively short intervals vs. lifetime) and, second, they act on different levels of aggregation (individual vs. population). In this paper, I integrate the principles of reinforcement learning and natural selection within a single model that is consistent with current behavioral theory.

2 What is a reinforcer?

One of the most influential accounts of reinforcement goes back to the 'law of effect', formulated by Thorndike (2010/1911), who links reinforcement to the level of satisfaction experienced by an animal following a certain behavior. The notion of an internal state like satisfaction has been criticized by behavioral psychologists (e.g. Watson, 1930) since there is no procedure to measure satisfaction,

[2] Throughout this paper, the term 'reinforcer value' is used in a purely technical way, referring to the effectiveness of reinforcement, or – more precisely – the objective of maximization by behavioral adaptations on an individual level. I do not presumed that this 'value' has a direct cognitive or neural counterpart in the organism. Therefore, whether or not individuals actually assign values to different behavioral options (or objects) and make decisions based on these values lies without the scope of this article.

thus an empirical test of the law of effect is impossible. One way to avoid the problem of measurement is to define reinforcement by its behavioral consequences, rather than by relating it to an internal state (Skinner, 1969). The reinforcer value of an event is thus equated with the amount of behavioral change it produces in an experimental setup. While unproblematic in a laboratory context, Skinner's account of reinforcement is not easily scaled to 'real life' behavior since reinforcers often have to be inferred ex post facto. Applied in this way, the law of effect becomes a tautology (Meehl, 1950). This circularity also arises when dealing with choice behavior as described in the matching equation (Herrnstein, 1970, 1974)[3]. As long as one deals with response rate in a very specific experimental environment, the matching law provides an adequate description of behavior. However, when applied to more general settings (or when applied to behavior outside the laboratory), the matching law often fails to describe behavioral allocation over response options. This is especially true when the reinforcers are of different 'quality' leading to a biased distribution towards the preferred option. Some deviations from strict matching can be accounted for by two additional parameters for 'bias' and 'sensitivity' (Baum, 1974). However, these parameters are defined solely within the matching equation itself. Consequently, the only meaning of the bias parameter is that it quantifies the bias towards one choice option, leaving open the question of *why* there is bias in the first place. Accordingly, the sensitivity parameter quantifies the amount of overmatching or undermatching, but does not provide an explanation for these phenomena. Thus, even though the generalized matching law performs well in describing deviations from strict matching, it says nothing about when to expect such deviations or how they come about. There have been attempts to modify the matching equation by adding theoretically meaningful parameters, as well, including the concept of reinforcer value (cf. Rachlin, 1971). Some of these parameters (e.g.

[3] Note that a circular definition of reinforcement is not a problem per se – as long as the definition is part of an overarching theory that can be subject to empirical test (compare, e.g. the definition of 'mass' in Newtonian mechanics). From a pragmatic point of view, 'reinforcement' as defined by the law of effect or the matching equation may be used to derive useful empirical protocols or applications. However, this is only reasonable if one assumes these 'laws' to be part of an overarching theory. I argue that this is indeed the case in most applications of behavioral psychology and that this overarching theory is (at least implicitly) Darwin's theory of evolution by natural selection.

delay of reinforcement) are well defined and can easily be measured or manipulated in an experiment. However, this is not the case for reinforcer value. Hence, it would be desirable to have a reliable and valid procedure to measure reinforcer value.

Measuring reinforcer value has proven notoriously difficult. Various procedures have been developed to infer reinforcer values from behavioral data. Usually, individual choice behavior is used to construct an individual's preference structure, which is then scaled in order to yield relative reinforcer value (Johnson & Bickel, 2006; Madden, Smethells, Ewan, & Hursh, 2007; Premack, 1963; Schwartz, Silberberg, Casey, Paukner, & Suomi, 2016; Timberlake & Allison, 1974; Tustin, 2000). Whilst providing valuable empirical protocols, these procedures do not explain why certain events are more reinforcing than others. Therefore, reinforcer values are arbitrary in the sense that there are no constraints on whether individuals of the same species should prefer similar reinforcers, or whether reinforcer values are constant over time and/or situations.

This paper aims to address this issue. Linking reinforcer values to evolutionary fitness will provide an explanation for the observed preference structures by giving an ultimate cause for the amount of reinforcer value obtained from an activity. An explicit model of this relation further yields general results on the quantitative nature of reinforcement.

Reinforcer values have previously been informally related to evolutionary fitness. Perhaps the most elaborated account is the concept of the *Phylogenetically Important Event* (PIE), which is defined as '…an event that directly affects survival and reproduction. […] Those individuals for whom these events were unimportant produced fewer surviving offspring than their competitors, for whom they were important, and are no longer represented in the population.' (Baum, 2012 p.106). By linking PIEs to survival and reproduction, Baum anticipates a quantitative relation with evolutionary fitness. However, in its present form, the account is strictly qualitative, leaving open the question of how exactly PIEs are related to survival and reproduction and how this affects evolutionary fitness.

2.1 Reinforcement learning and maximization

Reinforcement learning has been characterized as a mechanism (or a collection of mechanisms) to maximize obtained reinforcer value (Rachlin, Battalio, Kagel, & Green, 1981; Rachlin & Burkhard, 1978; Rachlin, Green, Kagel, & Battalio, 1976). Formally, this understanding of reinforcer value is equivalent to the concept of subjective additive utility from behavioral economics (Herrnstein, Loewenstein, Prelec, & Vaughan, 1993; Loewenstein, Prelec, & Seung, 2009). The maximization principle has been applied to various behavioral settings. It has been argued, for example, that maximizing reinforcer rate coincides with matching in concurrent variable interval schedules (Baum, 1981; Staddon & Motheral, 1978; Vaughan, 1981). However, if the schedule of reinforcement is influenced by past behavior, this is not the case (Sakai & Fukai, 2008). From an empirical point of view, the concept of reinforcer maximization has also been questioned (Herrnstein, 1970, 1990).

Interestingly, the question of whether reinforcer value is maximized or not continues to be discussed, even though reinforcer value has not yet been defined in a satisfactory manner. Hence, different authors may refer to slightly different concepts when they speak of 'reinforcer maximization' and (at least implicitly) pose their own restrictions to the concept of reinforcer value. It is therefore essential to give an explicit account of what can possibly be the maximand of reinforcement learning. Hence, in this paper, I do not argue for or against reinforcement maximization as an empirical principle, but try to provide an explicit account of what exactly could be meant by 'reinforcer value' if it was the quantity to be maximized in reinforcement learning. Note that it is not necessary to assume reinforcer maximization as an actual mechanism of behavioral adaptation. The concept of maximization still makes sense, even if an animal's behavior is governed by a much simpler mechanism such as matching or melioration, as long as these mechanisms lead – on average – to the maximization of obtained reinforcer value in the environment to which the animal is adapted by natural selection. While being neutral about the ontological status of reinforcer maximization, in this paper I build on the assumption that organisms (as long as they face situations that occur in the environment to which they are adapted by evolutionary processes) behave *as if maximizing* the amount of reinforcer value they receive.

3 What is evolutionary fitness?

The concept of fitness plays a crucial role in evolutionary biology. It serves as a unifying concept for theories of adaptation and evolutionary change. However, there is no generally accepted way to measure fitness. Although there is a broad consensus that fitness somehow refers to the reproductive schedules of individuals, only few researchers have attempted to construct a direct measure of fitness (McGraw & Caswell, 1996). Instead, many studies rely on proxies for fitness, like the number of descendants produced by an individual, reproductive success, or the product of fertility and survival. However useful these fitness surrogates may be when conducting empirical studies, for the purpose of this paper it is necessary to give an explicit formal account of fitness that is consistent with evolutionary theory.

In order to achieve this, the fitness concept is approached from two different directions. The first one operates on the level of quantitative genetics, linking the dynamics of changing allele frequencies in a population to mean phenotypic traits. The second one deals with the conditions under which mutant phenotypes can spread in a population. I focus on these two approaches because they are both well established in evolutionary biology and can be integrated to form a unified fitness concept as the maximand of natural selection.

From the perspective of population genetics, evolution is primarily concerned with the dynamics of changing genotypes, or more specifically, with the change of allele frequencies within a population from one generation to the next generation. This change of allele frequencies can be formally linked to the concept of evolutionary fitness using the covariance arithmetic of Price (1970, 1972). The Price equation has been stated in various forms, incorporating different genetic architectures, class structured populations, stochasticity and inclusive fitness (Grafen, 2000). In its simplest form, the Price equation states that (given perfect transmission) the change in an arbitrary allele frequency equals the covariance between the allele and evolutionary fitness. Applying the equation to arbitrary weighted sums of alleles ('additive genetic values'), changes in allele frequencies can be linked to quantitative phenotypic traits. The change in mean trait value then equals the covariance between the trait and evolutionary fitness:

$$\Delta\bar{z} = cov\left(z_j, \frac{w_j}{\bar{w}}\right)$$

Here, $\Delta\bar{z}$ is the change in the mean value of a quantitative trait z in a given population, z_j is the trait value for individual j, and $\frac{w_j}{\bar{w}}$ is the fitness of an individual j divided by the mean fitness in the population. If population size remains constant over generations, mean fitness equals one, and w_j refers to the number of descendants of an individual j[4]. Fitness, defined this way, denotes the relative genetic contribution of an individual to the future population.

Although natural selection is generally understood to act on the level of genotypes rather than phenotypes, evolutionary adaptations are often analyzed on the level of phenotypic traits (Grafen, 1982). From this perspective, models of natural selection are primarily concerned with the potential outcomes of natural selection rather than with the process of selection itself. Following the logic of evolutionary game theory (Maynard Smith, 1982), this line of research is primarily concerned with the question, which phenotypes can spread in a given population. The focus of analysis therefore lies on demographic properties of the population and its dependence on certain phenotypic traits like height, weight, mating preferences, parental investment etc., without considering the underlying genetics. For example, an organism may face the problem of when to start reproduction: although reproduction is essential for evolutionary success, reproduction in early life stages often reduces offspring survival; hence, a delayed reproduction may contribute more to an individual's fitness than early reproduction, because if offspring survive at a higher rate, on average, there will be more grandchildren. Another example is parental care: if parents do not invest in their offspring, they can

[4] Note that this holds only for haploid organisms with perfect clonal reproduction. Incorporating sexual reproduction and recombination for diploid species is possible, but requires a more complicated formal treatment.

continue to reproduce and will have a higher lifetime reproductive success. However, if parents provide care to their offspring, they raise the probability that offspring survive into adulthood. Hence, parental care may ultimately enhance evolutionary fitness. Demographic models can describe these trade-offs by specifying the effects of the evolving trait (e.g. start of reproduction or parental care) on survival and reproduction in different stages of the life-cycle of the species.

The simplest demographic analysis consists in modelling the dynamics of an age structured population using age-specific data on survival and reproduction (so called 'vital rates'). These data are usually presented in the form of a life table (Keyfitz & Caswell, 2005)[5]. Given these vital rates remain constant over time, the population eventually will approach a stable age distribution and grow (or decrease) at a constant rate λ (with $\lambda = 1$ meaning the population size remains constant). Of course, in a real environment, the population will not grow indefinitely due to limited resources. However, for most evolutionary analyses, the exponential model provides a reasonable simplification, because the aim is not to predict actual population growth, but to explore, which phenotype would spread at the highest rate (i.e. which phenotype produces the highest λ). Following Fisher (1930), the vital rates can be used to define the *reproductive value* v_x of an individual of age x as the sum of all expected future offspring, discounted by the amount of population growth until they are produced. In discrete time notation, this can be written as:

$$v_x = \frac{F_x}{\lambda} + \frac{P_x F_{x+1}}{\lambda^2} + \frac{P_x P_{x+1} F_{x+2}}{\lambda^3} + \cdots \frac{P_x P_{x+1} \ldots\ P_{k-1} F_k}{\lambda^{k-x+1}}$$

Here, k denotes the number of age classes, F_x denotes the birth rate at age x and P_x denotes the probability to survive from age x to age $x+1$. Hence, the numerators in this weighted sum correspond to the expected number of offspring produced at each age, starting at x until maximum age k is reached. The denominators denote the corresponding discount factors. Given the population grows exponentially by the growth rate λ, future offspring are discounted according to the number

[5] Usually, only female individuals are tracked in the model. However, it is possible to include different sexes (Caswell and Weeks (1986).

of individuals that will be alive in the next time step. The reproductive value of an individual thus refers to the present value of future offspring (Fisher, 1930).

In biologically motivated demographic models, the population is sometimes better characterized by different developmental stages (like ‘juvenile’, ’adult’), which leads to a stage-structured (or class-structured) population model. The concept of ‘class’ can also be applied to describe different sizes, quantitative variation in one or more phenotypic traits or even the spatial distribution of individuals (Caswell, 2001). In this more general framework, ‘age’ is just a special case of how to structure the population based on individual characteristics that are relevant for survival and reproduction. Note that even though the classes in a demographic model are defined as individual state variables, it is possible to describe the dynamics of the population without keeping track of each individual separately as long as, within classes, the vital rates are identical for all individuals and the effects of the individuals on the class transition rates can be captured by the sum of the contribution of all individuals (compare Metz & Diekmann, 1986). This is always the case, if a) the class structure captures the relevant factors influencing survival and reproduction probabilities for every individual and b) interactions between individuals do not affect these probabilities.

Within this framework, the reproductive value of an individual in each class can be calculated from a demographic population model using matrix algebra (see Appendix 1 for details). Following Grafen (2015), the fitness (W_j) of an individual j can now be formally defined as:

$$W_j = \sum_y c_{jy} v_y$$

Here, c_{jy} denotes the number of class y offspring produced by individual j. v_y is the corresponding reproductive value of the produced offspring. Following this definition, fitness refers to the reproductive value weighted sum of an individual’s offspring. If the class structure is given by age, all offspring are of age zero, resulting in the same weights for each offspring. However, if offspring vary in some characteristic that is relevant to their survival or reproductive rate, an individual’s fitness may be differentially affected depending on offspring characteristics such as size, weight, agility etc.

It has been shown that, under a wide range of conditions, the demographic definition of fitness coincides with the population genetic account given by the Price equation (Batty, Crewe, Grafen, & Gratwick, 2014; Grafen, 2015; Taylor, 1990). This means that for a broad class of cases, allele frequency change can be explicitly linked to the demography of the corresponding population. Hence, it is often possible to calculate empirical estimates of evolutionary fitness from demographic data.

3.1 Natural selection and maximization

It has been a long held belief that natural selection acts as if individuals were maximizing their fitness. While many empirical studies rely on the implicit assumption of fitness maximization, the theoretical status of a maximization principle in natural selection is not quite as straightforward. Even though Fisher's so called 'fundamental theorem of natural selection' has sometimes been interpreted in the way that organisms have a general tendency to maximize their fitness, from a mathematical point of view, this is not what the theorem implies (Ewens, 2004). Nevertheless, the concept of fitness maximization has gained some substantial theoretical support in recent years (Grafen, 2014).

In a series of papers, Grafen argues that natural selection can indeed be characterized as a form of fitness maximization, as long as the maximand is defined in such a way that it conforms to the laws of allele frequency change that drive natural selection on a genetic level. The papers deal with a variety of scenarios, generalizing the population genetic definition of fitness to arbitrary class structures (Grafen, 2006b), stochastic environments (Batty et al., 2014), and inclusive fitness (Grafen, 2006a), presenting a corresponding Price equation for each of these special cases. These genetic models are formally linked to maximization programs on the individual level. The maximand of natural selection is then derived and proven to exist under a wide variety of conditions. The main result of this work is that fitness as a maximand of natural selection corresponds to the reproductive value weighted sum of offspring as defined above.

4 Integrating reinforcement and natural selection

Having identified the maximand of natural selection, the question of reinforcer value can now be addressed from a quantitative perspective. The core argument rests on the assumption that behavior is chosen such that both evolutionary fitness and reinforcer value are maximized simultaneously. Since fitness relies on survival and birth rates, it can be calculated from the vital rates in a demographic model. Therefore, behavior affecting survival or birth rates changes an individual's fitness. Following the concept of a PIE (Baum, 2005), behaviors that co-vary with events that affect survival or reproduction are selected by reinforcement (Baum, 2018b), leading to a behavioral adaptation on an individual level[6]. If the process of behavioral selection can be properly described by a maximization principle in the above sense, reinforcer value has to be directly related to the effects of the PIEs on future survival or reproduction. Figure 1 illustrates the resulting relation between reinforcer value and evolutionary fitness for an age-structured population. Within each (age-)class the individual is assumed to move through a state space that may vary with the environment. This individual state space describes the possible states of an individual at each class in each environment and is not to be confused with the class structure of the population. The trajectory of an individual through this lower level state space describes the behavior within population classes, i.e. they capture the dynamics of local behavioral adaptations. Modelling these dynamics would require the inclusion of specific mechanisms of adaptation e.g. reinforcement learning or associative learning. These specific mechanisms may vary between species with regard to their general structure (i.e. the evolved learning model), as well as their fine-tuning (i.e. the evolved parameter values in the learning model) (cf. McNamara & Houston, 2009). Instead of focusing on these 'molecular' mechanisms (Baum, 2002), I focus on the resulting 'molar' patterns of behavior that manifest on the level of the population classes.

[6] Note that if a PIE has a negative effect on survival or reproduction, we would usually call it a 'punishing event'. Moreover, the covariance between behavior and PIE may be either positive or negative, resulting in what is traditionally referred to as 'positive' or 'negative' reinforcement or punishment, respectively.

The allocation of behavior within each class is assumed to be optimal with regard to the total amount of reinforcer value obtained by moving through the individual state space within this class. The gain in reinforcer value by moving through this state space is symbolized by the plus and minus signs in Figure 1. Note that the values of these reinforcer values are not arbitrary, if we assume survival and birth rates to be functions of behavior. They are intrinsically linked to evolutionary fitness by their effects on future reproduction and survival. In the next section I shall make this link explicit by providing a definition of reinforcer value that is consistent with the assumption of simultaneous maximization of fitness and reinforcer value.

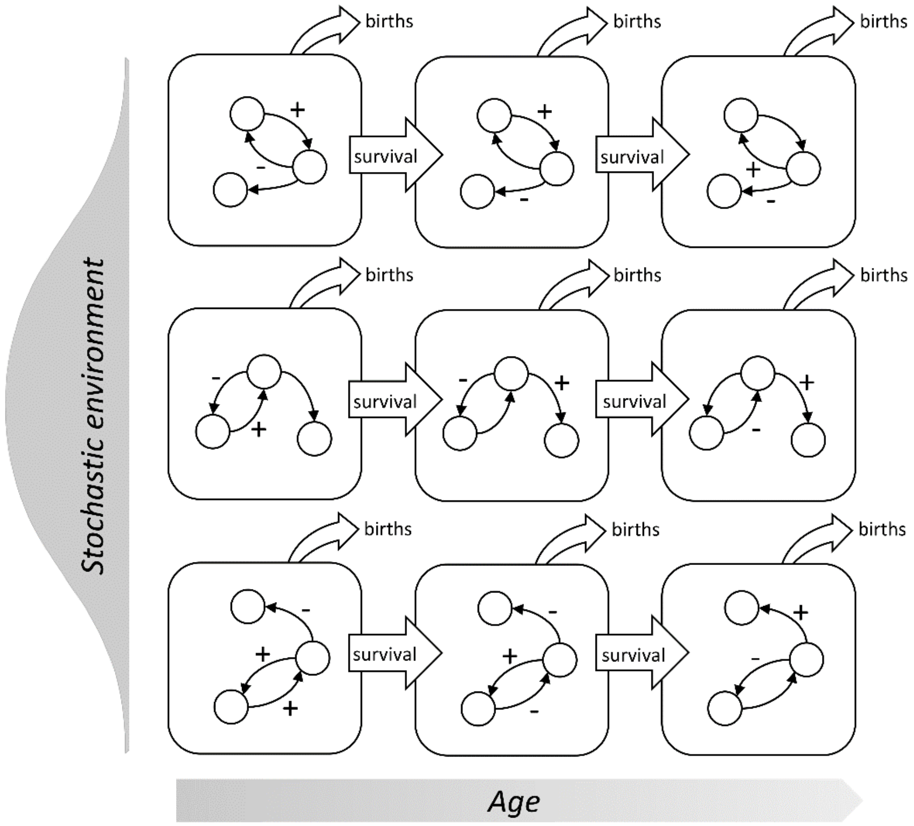


Figure 1: Conceptual model linking maximization of evolutionary fitness and reinforcer value. Each row corresponds to an individual in a different environment. The columns represent different age classes in ascending order from left to right. Within each age class, behavior is conceptualized as a trajectory through an individual's state space (states are symbolized by circles, possible transitions by solid curved arrows connecting circles). Depending on the environment, moving through the individual state space is associated with certain class specific reinforcer values (this is symbolized by the plus and minus signs next to the curved arrows, corresponding to reinforcement and punishment, respectively). Learning is assumed to maximize the expected sum of these reinforcer values at each class. The resulting allocation of behavior affects birth rates and survival rates. Because evolutionary fitness relies on birth rates and survival rates, reinforcer value needs to be defined within the constraints posed by the principle of fitness maximization.

5 Model

The following model tries to capture the conditions under which learning can lead to simultaneous maximization of reinforcer value and evolutionary fitness. Actual learning dynamics or processes of evolutionary adaptation are not included in the model. Instead, the focus lies on the relation between the maximands of reinforcement and natural selection given maximization is accomplished on both levels. The model is presented in a way that the core concepts can be defined explicitly, i.e. some formal notation is introduced, including functions and their derivatives. In order to enhance accessibility to mathematically less trained readers, I leave aside most of the technical details on how to perform the corresponding calculations (but see Appendix 1 for a brief introduction).

Evolutionary fitness W is conceptualized as the reproductive value weighted sum of offspring. In order to bridge the different time scales of evolution and learning, surviving individuals are formally treated as a special kind of offspring. This means that, for example, an individual of age x that survives to the next time step is counted as a 'child' of age $x + 1$. Hence, survival can be treated as a special kind of reproduction[7]. As a result, fitness consists of two components: one survival component, consisting of the individual itself weighted by its reproductive value in the next time step; and one fertility component, consisting of the number of offspring weighted by the corresponding offspring reproductive values. Because survival and transitioning from one class to another is treated as a special kind of reproduction, the vital rates of the corresponding population model can all be treated as if they were birth rates. Hence, in accordance with the above definition of fitness, the vital rates are written as c_{xy}, designating the rate at which a class x individual produces class y offspring.

Reproductive value v is defined within a population with an arbitrary class structure and refers to the long term contribution of an individual to the future population. The model equally applies to age-

[7] Even though this treatment of survival may be a little counter-intuitive, it is not uncommon in biological population models, since it provides a mathematically convenient way to model overlapping generations (cf. Taylor, 1990; Batty et al., 2014).

structured populations, as well as populations differing in one or more phenotypic traits (like size or weight), developmental stages (like juvenile and adult) or to populations in varying environments (like different food patches). The population is assumed to have reached an equilibrium state within its environment, such that the relative distribution of individuals over classes is stable. Under these conditions, the class specific reproductive values can be treated as constants (Caswell, 2001).

Within classes, individuals are assumed equal (i.e. they behave identically conditioned on their age, condition or local environment[8]). Hence, individual behavior can be analyzed on a population level, as long as it is conditioned on class. The behavior of individuals within each class x is written as a vector b_x, capturing the allocation of the possible behaviors at class x over time. Formally speaking, each behavior b_{xi} is represented by a number between 0 and 1 that designates the relative amount of time spent doing b_i in class x. Consequently, the sum of all b_{xi} is always 1 within each class.

Each class specific allocation of behavior b_{xi} is linked to the total reinforcer value in class x via a real-valued function $R(b_{xi})$. The sum of the total reinforcer values of the individual behaviors is referred to as class reinforcer value R_x.

$$R_x = \sum_i R(b_{xi})$$

The reinforcer value of an additional time unit spent with a behavior within class x is designated as $r(b_{xi})$. In accordance with the definition of total reinforcer value, $r(b_{xi})$ is defined as the marginal amount of reinforcer value given the current allocation of behavior – or, more formally, the partial derivative[9] of R_x with respect to b_{xi}:

$$r(b_{xi}) = \frac{\partial R_x}{\partial b_{xi}}$$

[8] Note that this does not rule out variation between individuals, as long as all demographically relevant factors are captured by the class structure of the population model.

[9] A partial derivative describes the change in one variable per unit change in another variable, keeping everything else constant.

This means that the reinforcer value of an additional time unit spent at a behavior is equal to its marginal change in total reinforcer value within the current class.

The behavior in each class is assumed to influence survival and fertility in this class (note that influences on future survival or reproduction are captured by the reproductive value of an individual in the next time step). Therefore, the vital rates c_{xy} are taken to be functions of the allocation of behavior within the corresponding class (i.e. each c_{xy} functionally depends on the allocation of behavior b_x in class x). Environmental variation is assumed to produce random fluctuations in these vital rates, such that that the expected values of the class transitions (and hence the expected value of future survival and reproduction) are determined by the allocation of behavior over time within each class. Since long term population dynamics are determined by the vital rates c_{xy}, this implies that the fitness of a class x individual (as measured by W_x) is a function of the allocation of behavior over time within class x.

Within this formal framework, it is now possible to link evolutionary fitness W with reinforcer value r such that fitness will be maximized, if every individual maximizes reinforcer value within each class. To accomplish this, it is necessary to give an explicit account of what makes behavior reinforcing. Following Baum (2012), behavior is reinforced because it affects expected future survival and reproduction via a Phylogenetically Important Event (PIE). This also includes negative reinforcement and punishment, since a PIE may have either positive or negative effects on survival and reproduction (hence, the term 'reinforcement' is used in a very broad sense here). Within the present model, this means that a PIE has to be defined via its effects on the vital rates c_{xy}. This means that the vital rates c_{xy} are treated as functions of the rates at which PIEs occur in each class x. The class specific PIE rates are referred to as Π_x. The corresponding functions are the survival functions and fertility functions of the class specific PIE rates. Since for reinforcement to occur the PIE rate needs to co-vary with the rate of a certain behavior (Baum, 2018b), I treat the (average) PIE rate Π_x as a function of behavioral allocation b_x in the corresponding class. The functional relation between behavior and PIE rate is usually specified by the feedback function of the corresponding schedule of reinforcement. I further assume that the effects of the behavioral allocation on the vital rates (and hence, evolutionary fitness) are fully mediated by PIEs (see Fig. 2). Consequently, any event that affects future survival and fertility and co-varies with the behavioral allocation in at least one class x is, by definition, a PIE. This is a simple yet effective way to formalize the reinforcing aspects of PIEs as specified by Baum (2018b). I do not incorporate multiple simultaneous schedules of reinforcement acting on the same behavior here, because this would require several feedback functions for each behavior, possibly linking them to more than one PIE. Although the model naturally includes these cases, keeping track of the indices for PIEs, schedules of reinforcement, behaviors and offspring classes would inflate the notation in a way that might distract the reader from the general structure of the model. Therefore, all following equations are restricted to the special case of one schedule of reinforcement and one PIE for each behavior. Generalizations to more complex scenarios are straightforward and do not change the general structure of the model. Furthermore, as a simplifying model assumption, I focus on the reinforcing nature of PIEs only, neglecting effects of behavioral induction.

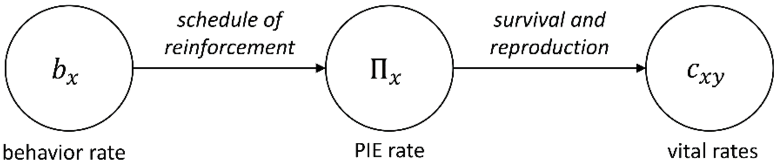


Figure 2: Functional relation between behavioral allocation in class x, PIE rate Π_x, and vital rates c_{xy} for a single behavior b_x. Note that it is possible to include several schedules of reinforcement acting on multiple behavior rates via multiple PIEs and that PIE rates may have multiple effects on different vital rates.

The functional relation between PIE rate and future survival and reproduction as specified above allows for a quantification of the reinforcing power of a PIE[10]. I fist define the effect of PIE rate on future survival and reproduction (π_{xy}) as the marginal change in the vital rates (c_{xy}) per unit change in PIE rate (Π_x):

[10] I use the term 'reinforcing power' here to avoid confusion with the concept of reinforcer value. By definition, reinforcer value is a function of behavior, whereas here I refer to the effectiveness of a PIE as a reinforcer.

$$\pi_{xy} = \frac{\partial c_{xy}}{\partial \Pi_x}$$

This is the partial derivative of the vital rates c_{xy} (i.e. the rate at which a class x individual produces class y offspring) with respect to PIE rate Π_x. Statistically, the PIE effects π_{xy} can be interpreted as partial regression coefficients of a regression of the vital rates on PIE rate[11]. It is reasonable to assume that the reinforcing power of a PIE is proportional to its effects on the vital rates (i.e. it is proportional to the π_{xy}). However, survival and fertility in different classes usually vary in their effects on evolutionary fitness depending on the life cycle of the species. For example, an individual in good physical condition will produce offspring of higher reproductive value when compared to an individual in bad condition. Therefore, a PIE affecting fertility should have a higher reinforcing power if the individual is in good condition, whereas a PIE affecting survival should have a higher reinforcing power if the individual is in bad condition. This differential weighting of the PIE effects π_{xy} can be formally captured by the class reproductive values v_y (with y referring to offspring class). Hence, the reinforcing power of a PIE is defined as the reproductive value weighted sum of all effects on survival and reproduction, i.e. $\sum_y \pi_{xy} v_y$.

To retrieve the reinforcer value of a behavior b_{xi} from the reinforcing power of a PIE it is necessary to consider the effect of the behavior on the PIE, i.e. the schedule of reinforcement or, more formally, the feedback function relating PIE rate to behavioral allocation. The change in PIE rate per unit change in behavioral allocation (written as p_x) corresponds to the slope of the feedback function and is captured formally by the partial derivative of Π_x with respect to b_{xi}:

$$p_x = \frac{\partial \Pi_x}{\partial b_{xi}}$$

Like above, there is a statistical interpretation of p_x, linking it to the partial regression coefficients of a regression of PIE rate on the set of behaviors. However, since the schedule of reinforcement is

[11] Note that these regression coefficients are only exact in the region of the evaluated PIE rate. If PIE effects are non-linear, the π_{xy} will change depending on PIE rate.

usually under the control of the experimenter the feedback function will be known in most cases. Therefore, in general p_x can be calculated analytically by taking the derivative of the feedback function.

The reinforcer value $r(b_{xi})$ of an additional time unit spent at behavior b_{xi} can now be defined as the resulting change in PIE rate (i.e. p_x) times the reinforcing power of a unit change in PIE rate (i.e. $\sum_y \pi_{xy} v_y$) evaluated at the point of the current behavioral allocation (i.e. p_x and π_{xy} are treated as functions of b_{xi}):

$$r(b_{xi}) = p_x(b_{xi}) \sum_y v_y \pi_{xy}\,(b_{xi})$$

This means that in order to get the reinforcer value of a change in behavioral allocation the derivative of the feedback function needs to be multiplied with the reproductive value weighted sum of the effects of a PIE on future survival and reproduction.

From this definition it follows that the reinforcer value of a unit change in behavioral allocation equals the marginal fitness change due to this behavior:

$$r(b_{xi}) = \frac{\partial W_x}{\partial b_{xi}}$$

This implies that evolutionary fitness is affected by a change in behavioral allocation if and only if the corresponding reinforcer value is different from zero. Therefore, reinforcer value (as defined above) is maximized if and only if evolutionary fitness is maximized (see Appendix 2 for a formal proof).

6 Application of the model

I will now demonstrate how the above model can be applied to matching behavior using a numerical example. The example is kept as simple as possible, yet rich enough to illustrate the implications of the model. I first derive a modified version of the matching law for choice behavior in an age structured population. I then use a numerical simulation to demonstrate how observed deviations from the matching law can be explained by an evolutionary account of reinforcer value.

6.1 Matching behavior

In the classical matching paradigm, individuals face a choice situation between one or more behavioral options, each associated with a certain average reinforcer rate. In this context, individuals often match their response rates to the expected reinforcer rate. To account for deviations from the matching law, Baum (1974) introduced two parameters that correspond to 'bias' (i.e. preference for one type of reinforcer) and 'sensitivity' (i.e. undermatching or overmatching). However, there are instances where individuals deviate from the matching law, even if these parameters are added (e.g. Simon & Baum, 2017)

In order to apply the above model to matching behavior, I identify the class structure of the population with a set of possible choice situations that an organism can encounter during its lifetime. Each choice situation x is characterized by a set of possible behaviors b_{xi} with the above restriction that $\sum_i b_{xi} = 1$ (i.e. there is a fixed time budget in each class). Each behavioral option comes with a total reinforcer value $R(b_{xi})$ with class reinforcer value being $R_x = \sum_i R(b_{xi})$. As an additional assumption the $R(b_{xi})$ are taken to be monotone increasing functions of behavior with a concave curvature (i.e. the slope is high for low values of b_{xi} and becomes smaller for higher values of b_{xi}).

Given these assumptions, if every individual optimizes the allocation of behavior over time in every class with respect to reinforcer value, all marginal reinforcer values are equal at the point of the behavioral optimum (cf. G. S. Becker, 1976; Friedman, 1953). This means that if individuals maximize R_x, it follows for all x:

$$\frac{\partial R_x}{\partial b_{x1}} = \frac{\partial R_x}{\partial b_{x2}} = \frac{\partial R_x}{\partial b_{x3}} \ldots = \frac{\partial R_x}{\partial b_{xn}}$$

For reasons of simplicity, the following analysis focuses on a single class containing only two behavioral options. In order to simplify notation, I drop the class index and only refer to b_1 and b_2, respectively. The result easily generalizes to more than one class and several behaviors. Assuming optimal allocation of behavior within class, it holds that:

$$r(b_1) = r(b_2)$$

Since each behavioral option is subject to its own schedule of reinforcement, I use p_1 and p_2 to designate the corresponding feedback function derivatives. The effects of the PIEs on the vital rates are labeled π_{1y} and π_{2y}, respectively (note that the first index does not refer to class here but to the two different schedules of reinforcement). Inserting this in the above definition of reinforcer value yields:

$$p_1(b_1)\sum_y v_y\pi_{1y}(b_1) = p_2(b_2)\sum_y v_y\pi_{2y}(b_2)$$

Since p_1 and p_2 are the change in PIE rate per unit change of behavioral allocation, they can be derived from the feedback functions of the corresponding schedules of reinforcement. The reproductive values v_y can be calculated analytically from a demographic model. The effects of PIE rates on survival and reproduction (π_{1y} and π_{2y}) depend on the functional relations between PIE rates and the vital rates. In most cases these will not be known a priori and hence need to be estimated by statistical regression coefficients of survival and fertility on PIE rate.

Let us now assume that we are dealing with an age structured population, such that all transition rates are either birth rates or survival rates from the present age to the next age. This can be illustrated by a so called life cycle graph, with the nodes designating the population classes and arrows denoting transitions between classes. For an age structured population with four age classes, the corresponding graph consists only of survival rates P_x and birth rates F_x (see Fig. 3).

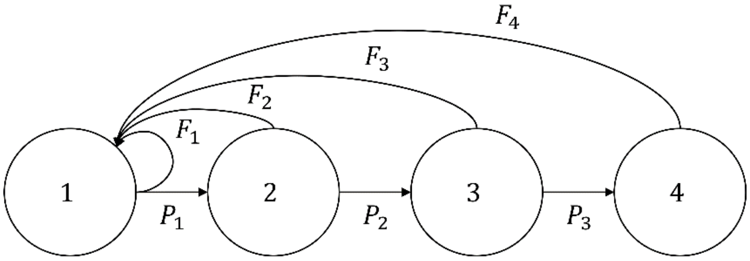


Figure 3: Life cycle graph of an age structured population with four age classes. Each node represents a population class. The arrows designate the vital rates (c_{xy}) with P_1, P_2, P_3 being the survival rates from one age to the next and F_1, F_2, F_3, F_4 being

the birth rates in each age class. All newborn offspring are of age 1, therefore the arrows designating birth rates all point to class 1.

If we focus on one age class, we need only consider the transition rates originating from this one class (thereby assuming that all other transition rates remain constant). I further restrict the analysis to PIE effects on survival only (which is a reasonable assumption if the PIEs consist in the availability of food). In this case the reinforcing power of a PIE reduces to its marginal effect on survival, weighted by the reproductive value of the following age class. Since there is only one class to be considered, it is possible to drop the class subscripts x and y entirely. Moreover, since the only reproductive value to be considered in this case belongs to the following age class (resulting in the same weighting factor for both PIEs), the condition for maximized reinforcer value simplifies to:

$$p_1(b_1)\pi_1(b_1) = p_2(b_2)\pi_2(b_2)$$

The behavioral options b_1 and b_2 can be interpreted as different food patches. The patches are modelled such that they contain two different types of food differing in nutritional quality and abundance. Formally, variation in food availability within patches is modelled as a variable interval (VI) schedule of reinforcement. The corresponding feedback function for food rate is:

$$F(b_i) = \frac{a_i b_i}{a_i + b_i}$$

With $\frac{1}{a_i} = T_i$ being the minimum average time between reinforcements. To account for fluctuations in the environment, the parameters a_i are taken to vary at random. Different nutritional values of the food types should change the marginal effects of food consumption on survival. Given the survival functions of the two food options are steep for low values (i.e. low PIE rates) and gradually approach an asymptote at 1, the survival functions may be described by the following equation:

$$P(f_i) = \frac{f_i}{s_i + f_i}$$

with f_i being the obtained food rate and s_i being a parameter describing the steepness of the survival function. The derivative of the feedback function $F(b_i)$ provides the marginal change in PIE rate per unit change in behavioral allocation $p_i(b_i)$:

$$p_i(b_i) = F'(b_i) \; = \frac{a_i^2}{(a_i + b_i)^2}$$

The derivative of the survival function $P(f_i)$ provides the marginal change in survival per unit change in behavioral allocation $\pi_i(b_i)$:

$$\pi_i(b_i) = P'(f_i) \; = \frac{s_i}{(s_i + f_i)^2}$$

with $f_i = F(b_i)$. Multiplying both terms and simplifying yields the marginal reinforcer value of b_i:

$$r(b_i) = p_i(b_i)\pi_i(b_i) = \frac{s_i a_i^2}{(b_i(s_i + a_i) + a_i s_i)^2}$$

Substituting the maximization condition with the corresponding values for $r(b_1)$ and $r(b_2)$ yields:

$$\frac{s_1 a_1^2}{(b_1(s_1 + a_1) + a_1 s_1)^2} = \frac{s_2 a_2^2}{(b_2(s_2 + a_2) + a_2 s_2)^2}$$

Due to the budget constraint ($b_1 + b_2 = 1$), b_2 can be replaced by $1 - b_1$. Solving for b_1 gives the optimal behavioral allocation b_1^* with respect to reinforcer value:

$$b_1^* = \frac{a_1(a_2 + s_2 + a_2 s_2)\sqrt{s_1} - a_1 a_2 s_1 \sqrt{s_2}}{a_1(a_2 + s_2)\sqrt{s_1} + a_2(a_1 + s_1)\sqrt{s_2}}$$

Unfortunately, it is hard to give an intuitive interpretation to the terms in this equation. However, it is possible to transform the result to get a more meaningful separation of terms. We start by taking the ratio of the optimal behavioral allocations:

$$\frac{b_1^*}{b_2^*} = \frac{a_1(a_2 + s_2 + a_2 s_2)\sqrt{s_1} - a_1 a_2 s_1 \sqrt{s_2}}{a_2(a_1 + s_1 + a_1 s_1)\sqrt{s_2} - a_1 a_2 s_2 \sqrt{s_1}}$$

Next, we retrieve the total reinforcer values $R(b_i)$ by taking the integral over the marginal reinforcer values $r(b_i)$:

$$R(b_i) = \int_0^{b_i} r(b_i)\mathrm{d}b_i = \frac{a_i b_i}{b_i(a_i + s_i) + a_i s_i}$$

Taking the ratios of these total reinforcer values gives:

$$\frac{R(b_1)}{R(b_2)} = \frac{a_1 b_1(b_2(s_2 + a_2) + a_2 s_2)}{a_2 b_2(b_1(s_1 + a_1) + a_1 s_1)}$$

We now evaluate this at the point of optimal behavioral allocation, i.e. we replace b_1 and b_2 by the corresponding expressions for b_1^* and b_2^* and simplify to get:

$$\frac{R(b_1^*)}{R(b_2^*)} = \frac{a_1(a_2 + s_2 + a_2 s_2) - a_1 a_2 \sqrt{s1}\sqrt{s_2}}{a_2(a_1 + s_1 + a_1 s_1) - a_1 a_2 \sqrt{s1}\sqrt{s_2}}$$

This ratio of total reinforcer values looks very similar to the ratio of behavioral allocations. In fact, if we multiply $\frac{R(b_1^*)}{R(b_2^*)}$ by $\frac{\sqrt{s1}}{\sqrt{s2}}$, both expressions become equivalent:

$$\frac{b_1^*}{b_2^*} = \frac{R(b_1^*)\sqrt{s_1}}{R(b_2^*)\sqrt{s_2}}$$

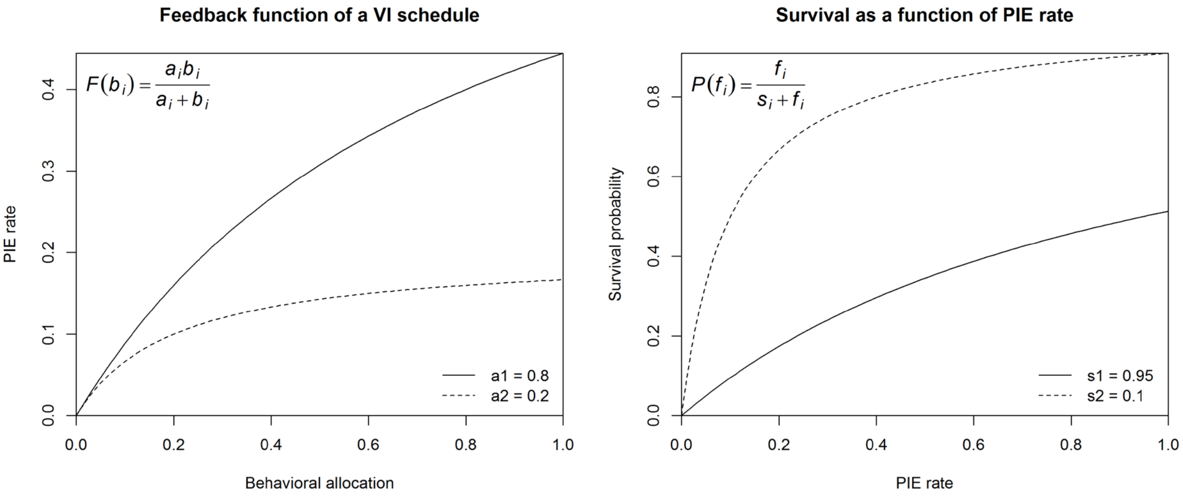
Writing B_i instead of b_i^* and R_i instead of $R(b_i^*)$ we finally arrive at the generalized matching law, with a bias parameter of $\beta = \frac{\sqrt{s_1}}{\sqrt{s_2}}$ and a sensitivity parameter of $\alpha = 1$:

$$\frac{B_1}{B_2} = \beta\left(\frac{R_1}{R_2}\right)^{\alpha}$$

Note that α and β are no longer free parameters (as introduced by Baum, 1974) but are derived from first principles here. Moreover, it is not reinforcer *rate* (or PIE rate) that is matched, but reinforcer *value*. Apparently, transforming PIE rate into reinforcer value fully accounts for undermatching and overmatching (i.e. the sensitivity parameter is one) but introduces a bias depending on the steepness parameters of the PIE's survival functions (i.e. s_1 and s_2).

6.2 Numerical Example

I will now illustrate the above derivation with a numerical example. Like in the previous section, I focus on the optimal allocation of behavior within one single class in an age structured population. I assume two mutually exclusive behaviors b_1 and b_2 with the usual time budget constraint $b_1 + b_2 = 1$. Individuals are subject to a variable interval schedule for each of the two behaviors with parameters $a_1 = 0.8$ and $a_2 = 0.2$ (compare left panel of Fig. 4). Each schedule of reinforcement links the behavioral allocation of an individual to food rate. The two schedules provide different foods, each linked to the probability to survive till the next age via a survival function with corresponding steepness parameters $s_1 = 0.95$ and $s_2 = 0.1$ (compare right panel of Fig. 4). Hence, the first schedule provides a rich environment with a high average rate of reinforcement (i.e. high PIE rate) but a low quality food (i.e. low survival gain), whereas the second schedule provides a low average reinforcement (i.e. low PIE rate) but a high quality food (i.e. high survival gain).


Feedback function of a VI schedule

Survival as a function of PIE rate

Figure 4: Graphs of feedback function and survival function for two different parameterizations. The parameter a_i specifies the minimum average time between reinforcements in a VI schedule. The parameter s_i refers to the steepness of the survival curve.

Substituting the maximization condition with this parameterization yields:

$$\frac{0.95\cdot 0.8^2}{(b_1(0.95+0.8)+0.8\cdot 0.95)^2}=\frac{0.1\cdot 0.2^2}{(b_2(0.1+0.2)+0.2\cdot 0.1)^2}$$

Replacing b_2 by $1-b_1$ and solving for b_1 results in:

$$b_1=\frac{0.8(0.2+0.1+0.2\cdot 0.1)\sqrt{0.95}-0.8\cdot 0.2\cdot 0.95\sqrt{0.1}}{0.8(0.2+0.1)\sqrt{0.95}+0.2(0.8+0.95)\sqrt{0.1}}$$

Within the constraint $b_1 \in [0,1]$ there is exactly one solution to this equation, yielding an optimal value of:

$$b_1^* = 0.585$$

Equating the marginal reinforcer values is equivalent to finding the point of intersection between the corresponding graphs. This graphical approach is illustrated in Figure 5. The figure also depicts the prediction derived from the (strict) matching law at $b_1 = 0.8$. Obviously, for the given parameterization the evolutionary model predicts a substantial deviation from matching.

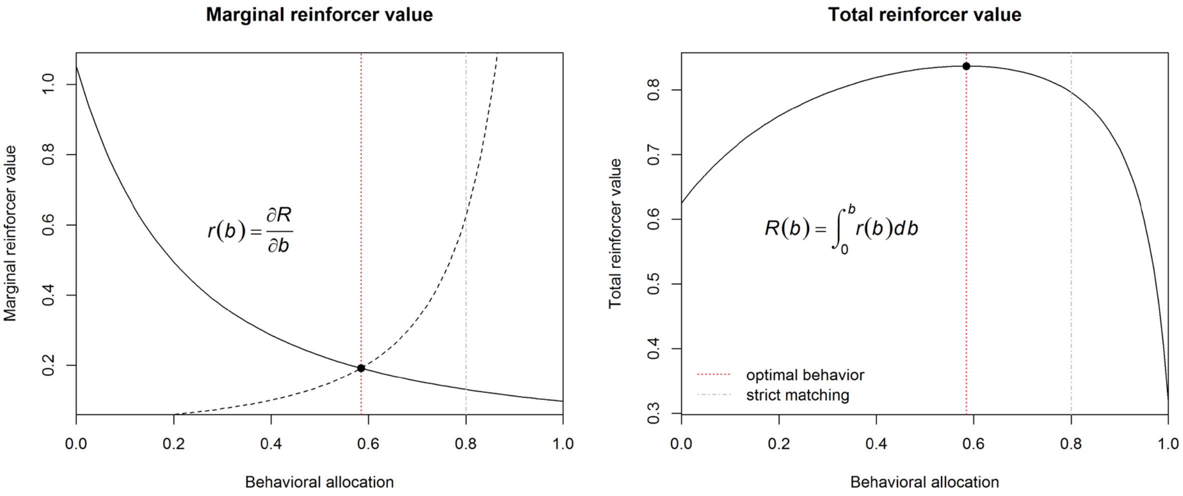


Figure 5: Optimal allocation of behavior with respect to total reinforcer value R. The left panel depicts the marginal reinforcer values of the two choice options. The point of intersection marks the behavioral optimum. The right panel shows the total reinforcer value as a function of behavioral allocation (i.e. the sum of the integrals over the marginal reinforcer values of the two behaviors). The grey line marks the predicted behavioral allocation derived from the matching law. For the given parameterization, there is a clear deviation between maximal reinforcer value and matching.

In order to investigate whether this is a general pattern, a random environment was simulated by letting the average time between reinforcements vary independently for the behavioral options (i.e. the a_i were treated as normally distributed random variables with expectation 0.5 and a variance of 0.1). The survival functions for the two PIEs were treated as constant ($s_1 = 0.95$, $s_2 = 0.1$). Figure 6 depicts the result of a simulation with 100 randomly generated behavioral samples[12]. The resulting behavioral allocations were then treated as a dependent variable in two linear regression models: the

[12] The simulation was repeated several times to make sure the results are reliable.

first model predicts the observed behavior from the ratio of reinforcer rate; the second model predicts the observed behavior from the ratio of reinforcer value. Using logarithmic axes the resulting linear regressions correspond to the fit of the generalized matching law (Baum, 1974).

The left panel of Figure 6 shows the fit of the generalized matching law using reinforcer rate (this is the common way to analyze matching experiments). Even though the regression approximates the simulated data to a certain degree (89% of variance explained), there is considerable deviation from matching when using reinforcer rate, even after accounting for bias and sensitivity. Note that even though the variation around the regression line looks like random noise, the data actually result from a deterministic model. This means, the deviation cannot be explained by sampling error or unreliable measurement. The reason why the generalized matching law fails in this case is that it does not account for the effects the PIEs have on survival.

The right panel of Figure 6 shows the fit of the generalized matching law applied to reinforcer *values* using the same data. As expected, the regression has a perfect fit. Moreover, in accordance with the above derivation, the slope of the regression line is exactly 1. This means that there is neither undermatching nor overmatching when the equation is fitted using reinforcer values instead of reinforcer rates. In order to check whether the evolutionary model correctly predicts the bias parameter, we calculate:

$$\log\left(\frac{\sqrt{0.95}}{\sqrt{0.1}}\right) = 1.13$$

which is exactly the intercept of the fitted regression line. This demonstrates that even though maximizing reinforcer value produces perfect matching (with regard to reinforcer value) empirical deviations from the matching law are to be expected if reinforcer rate is not properly scaled to account for the PIEs' effects on evolutionary fitness.

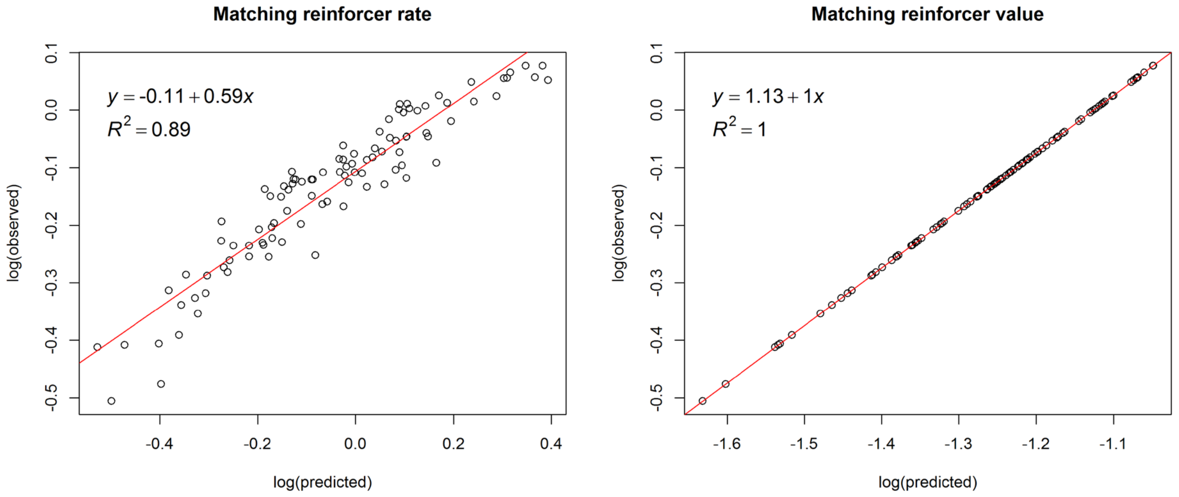


Figure 6: Fit of the generalized matching equation to the data of 100 randomly generated matching experiments assuming individual maximization of reinforcer value. The left panel shows the fit of the generalized matching law using the ratio of the untransformed reinforcer rates as a predictor. The right panel shows the fit of the generalized matching law using the ratio of the corresponding reinforcer values a predictor. Whereas there is a clear deviation from matching with regard to reinforcer rate, there is perfect matching with regard to reinforcer value.

7 Discussion

This paper dealt with the question of how reinforcement learning may be related to natural selection on a quantitative level. Assuming natural selection and reinforcement learning can both be described by an 'individual-as-maximizer' analogy, it was argued that the maximands of reinforcement learning and natural selection must be related by a quantitative law. Building on the theory of Phylogenetically Important Events (PIE) by Baum (2005, 2012, 2018b), a formal model was introduced that specifies the relation between reinforcing events and vital rates. The model further incorporates the concept of evolutionary fitness, which is derived from a demographic population model, and an explicit definition of reinforcer value. The model was then used to establish a formal link between reinforcer values and evolutionary fitness, which consists in a very general quantitative law: *the reinforcer value of an additional time unit spent at a behavior equals its expected marginal effects on evolutionary fitness*. The model was then applied to explore the relation between maximization of reinforcer value and maximization of fitness with regard to matching behavior. It was found that empirical deviations from matching may in some cases be explained by the lack of a proper scaling of reinforcer rate before fitting the matching equation. The scaling factor to be applied is the reinforcing power of a PIE, defined as the reproductive value weighted sum of all effects of the PIE on survival and reproduction.

One distinctive feature of reinforcer value, as compared to existing accounts, is that within the current model, reinforcer value is defined independently of an empirical preference structure. This means that reinforcer value is not inferred ex post facto from observed choice behavior like in most studies on choice behavior. Instead, reinforcer value is formally linked to survival and future reproduction via the concept of a PIE. The evolutionary model of reinforcer value can be used to extend existing theories of choice (as demonstrated with the generalized matching law), yielding new empirical hypotheses. Moreover, it provides a reasonable explanation why the maximization principle seems to fail in some experimental paradigms: the variables that are usually used as proxies for reinforcer value (like food intake) are only valid as far as they actually represent the marginal fitness effects of a PIE. Because the reinforcing power of a PIE is the sum of its reproductive value weighted effects on survival and reproduction, reinforcer value depends not only on the PIE itself, but also on the demography of the species. Therefore, when we observe behavior to be 'non-optimal' with regard to food intake or reinforcer rate, we might actually be wrong about what is being optimized (cf. Houston, McNamara, & Steer, 2007). Apart from this, and perhaps most importantly, the model presented in this study provides an important step to integrating behavioral psychology with a general theory of natural selection and gene frequency change. Even though the theoretical link between quantitative genetics and behavioral optimization has not yet been fully developed, there is good reason to believe that the results produced so far may generalize to more complex scenarios (Grafen, 2006a, 2009; Taylor, 1990).

Of course, such a strong result comes at a price. Therefore, in the following, the scope of the model shall be explored before its underlying assumptions. First, the model makes rather restrictive assumptions regarding the life-cycle of the species under study. As mentioned, the above equations only hold for a population with discrete classes and reinforcement leading to a behavioral optimum in each class (at least when averaging over the environment). This implies that local adaptations due to learning occur rather fast in comparison to the time scale of the class structure imposed on the demographic model (i.e. behavior is assumed stable – post learning). Effectively, deviations from this assumption will result in behavior being slightly suboptimal in each class. As long as this bias is more or less constant for all behaviors (i.e. learning occurs at a constant rate), however, this should not change the validity of the model. Moreover, learning mechanisms are assumed to be constructed such that they lead to optimization of expected reinforcement. It remains an open question whether this a plausible assumption since there is reason to doubt a general tendency to maximize reinforcement (Herrnstein & Prelec, 1991). However, in the context of this study, experimental results can only yield evidence against reinforcer maximization to the extent that they present valid models of the environment to which the organism has been adapted by natural selection. In many behavioral experiments, this is most likely not the case (c.f. examples in Herrnstein, 1990).

Another simplifying model assumption is that reinforcer values of different behaviors are additive (i.e. class reinforcer value is the sum of the total reinforcer values of all behaviors). This may turn out to be problematic when the associated PIEs are not completely substitutable. Such non-additive effects have been demonstrated in the laboratory (Hursh, 1978; Willis, van Hartesveldt, Loken, & Hall, 1974). However, as long as these effects are small in naturalistic environments, they should not pose a major problem for the modelling approach chosen here.

On the demographic level, the restriction to a one-sex model leads to omitting the effects of different mating strategies. Even though it has been demonstrated that different mating strategies may indeed change population dynamics considerably if the life-cycles of the sexes differ (Caswell & Weeks, 1986), the use of one-sex models is common practice in demographic studies since, empirically, one-sex models are usually a good approximation to the more complex two-sex models (Keyfitz & Caswell, 2005). Nevertheless, in order to generalize the results of this study to learning in the context of mating strategies, it may turn out to be necessary to incorporate a two-sex model.

Perhaps the most important restriction of this study is the assumption of a non-social species in the sense of Hamilton (1964). This means that fitness is only affected by an individual's own behavior, excluding effects of kin selection or inclusive fitness. As long as inclusive fitness effects are independent of reinforcement learning, this simplification may be justified. On a formal level, the model is still valid if interactions between individuals do not affect the vital rates of the population (Metz & Diekmann, 1986). However, if there is social learning, like imitation or shaping of behavior by another individual's behavior (i.e. verbal behavior in the sense of Skinner, 1957), the predictions of the model presented in this study may be unreliable. There are at least two strategies to address this shortcoming. First, one could try to provide an inclusive fitness formulation of the above definition of reinforcer value (similar to Rogers, 1994). This would require that a PIE may be reinforcing because it affects the fitness of an individual different from the actor. In addition, PIEs may be reinforcing because of the effect they have on other individuals' behavior, thereby affecting the actor's fitness indirectly. Disentangling these social effects requires a careful handling of weighting factors and a more complicated mathematical treatment. An alternative way to incorporate the effects of interactions between individuals might consist in a group selection approach as proposed by Rachlin (2019). The concept of group selection has recently gained attention in the context of cultural evolution (Richerson et al., 2016; Smaldino, 2014). However, on a formal level, models of group selection have repeatedly been shown to be equivalent to inclusive fitness models (Lehmann, Keller, West, & Roze, 2007; Marshall, 2011; Queller, 1992). Therefore, a group selection approach to formalizing reinforcer value is unlikely to resolve the mathematical challenges posed by inclusive fitness theory. Nevertheless, incorporating sociality in a general theory of reinforcement and linking this to the concept of evolutionary fitness remains an important task for future research.

Of the several attempts to integrate theories of operant learning with the principles of natural selection, this study is the first one to provide an explicit formal link between reinforcer value and evolutionary fitness. On a theoretical level, this is an important step towards an integrative account of behavioral adaptations on different levels of selection. Whether behavior *actually* maximizes either reinforcement, evolutionary fitness or both remains an empirical question and cannot be decided on theoretical grounds. However, substantial theory is essential to guide empirical research in this field. The model presented here may provide such a theoretical framework.

8 Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

9 Competing interests

The author declares no competing interests.

10 Appendix

10.1 Calculating reproductive values using matrix algebra

It is common practice to describe the dynamics of a class structured population using matrix notation: The transition rates from each class to the remaining classes are written into the columns of a projection matrix A. In the case of an age structured population, the corresponding projection matrix contains the birth rates in the first line and the survival rates in the sub-diagonal, representing the probability of surviving from one age class to the next one. The resulting matrix is called a Leslie-matrix (Leslie, 1945). For example, for an age structured population consisting of four classes the projection matrix is:

$$A = \begin{bmatrix} F_1 & F_2 & F_3 & F_4 \\ P_1 & 0 & 0 & 0 \\ 0 & P_2 & 0 & 0 \\ 0 & 0 & P_3 & 0 \end{bmatrix}$$

In order to describe the dynamics of the population, the projection matrix is multiplied with a vector n, representing the distribution of individuals over classes. This results in a system of linear equations describing the change of the population distribution from time t to the next time step:

$$n(t+1) = An(t)$$

Writing out the matix entries and the elements of the vector, the equation becomes:

$$\begin{bmatrix} n_1(t+1) \\ n_2(t+1) \\ n_3(t+1) \\ n_4(t+1) \end{bmatrix} = \begin{bmatrix} F_1 & F_2 & F_3 & F_4 \\ P_1 & 0 & 0 & 0 \\ 0 & P_2 & 0 & 0 \\ 0 & 0 & P_3 & 0 \end{bmatrix} * \begin{bmatrix} n_1(t) \\ n_2(t) \\ n_3(t) \\ n_4(t) \end{bmatrix}$$

This expression can be expanded to a system of linear equations, describing how the number of individuals changes from one point in time to the next one in each age class:

$$n_1(t+1) = F_1 n_1(t) + F_2 n_2(t) + F_3 n_3(t) + F_4 n_4(t)$$

$$n_2(t+1) = P_1 n_1(t)$$

$$n_3(t+1) = P_2 n_2(t)$$

$$n_4(t+1) = P_3 n_3(t)$$

Iterating the model equation for a sufficiently large amount of time steps projects the distribution into the distant future. In general, the relative distribution of individuals over classes (i.e. the vector n) eventually converges to a stable class distribution u. Conveniently, the long term dynamics of the system can be calculated analytically from the projection matrix. The long term growth rate λ is equal to the largest eigenvalue of the matrix A. The stable class distribution u equals the corresponding right eigenvector (Caswell, 2001)[13]:

$$Au = \lambda u$$

Following the same rationale, the transpose of the projection matrix can be used to project the population back into the distant past, resulting in the relative long-term contributions of the different

[13] An *eigenvector* is a vector that only changes by a constant factor when multiplied with a matrix. The corresponding constant factor is called an *eigenvalue*.

classes to the population distribution, i.e. the class reproductive values v. Using matrix calculus, this corresponds to the left eigenvector of the largest eigenvalue (Caswell, 2001):

$$v^T A = \lambda v^T$$

10.2 Proof that maximizing reinforcer value coincides with maximizing evolutionary fitness

Marginal reinforcer value is defined as:

$$r(b_{xi}) = p_x(b_{xi}) \sum_y v_y \pi_{xy}(b_{xi})$$

with $p_x(b_{xi}) = \frac{\partial \Pi_x}{\partial b_{xi}}$ and $\pi_{xy}(b_{xi}) = \frac{\partial c_{xy}}{\partial \Pi_x}$. Furthermore, from the definition of fitness it follows that:

$$v_y = \frac{\partial W_x}{\partial c_{xy}}$$

Substituting this into the definition of marginal fitness yields:

$$r(b_{xi}) = \frac{\partial \Pi_x}{\partial b_{xi}} \sum_y \frac{\partial W_x}{\partial c_{xy}} \frac{\partial c_{xy}}{\partial \Pi_x}$$

Assuming all Π_x are independent, it further holds that:

$$\frac{\partial W_x}{\partial \Pi_x} = \sum_y \frac{\partial W_x}{\partial c_{xy}} \frac{\partial c_{xy}}{\partial \Pi_x}$$

By substitution one obtains:

$$r(b_{xi}) = \frac{\partial W_x}{\partial \Pi_x} \frac{\partial \Pi_x}{\partial b_{xi}}$$

Applying the chain rule of differential calculus results in:

$$\frac{\partial W_x}{\partial b_{xi}} = \frac{\partial W_x}{\partial \Pi_x} \frac{\partial \Pi_x}{\partial b_{xi}}$$

Therefore:

$$r(b_{xi}) = \frac{\partial W_x}{\partial b_{xi}}$$

Since $r(b_{xi})$ is defined as the partial derivative of total reinforcer value R_x with respect to b_{xi}this is equivalent to:

$$\frac{\partial R_x}{\partial b_{xi}} = \frac{\partial W_x}{\partial b_{xi}}$$

In order to find the behavioral allocation that maximizes evolutionary fitness and total reinforcer value, respectively, one takes the total derivatives of W_x and R_x and sets them equal to zero. Since the total derivatives are completely determined by the partial derivatives and the functional relations between the b_{xi}, maximization of W_x necessarily coincides with maximization of R_x.

References

Aoki, K., & Feldman, M. W. (2014). Evolution of learning strategies in temporally and spatially variable environments: A review of theory. *Theoretical Population Biology*, *91*, 3–19. https://doi.org/10.1016/j.tpb.2013.10.004

Batty, C. J. K., Crewe, P., Grafen, A., & Gratwick, R. (2014). Foundations of a mathematical theory of darwinism. *Journal of Mathematical Biology*, *69*(2), 295–334. https://doi.org/10.1007/s00285-013-0706-2

Baum, W. M. (1974). On two types of deviation from the matching law: bias and undermatching. *Journal of the Experimental Analysis of Behavior*, *22*(1), 231–242.

Baum, W. M. (1981). Optimization and the matching law as accounts of instrumental behavior. *Journal of the Experimental Analysis of Behavior*, *36*, 387–403.

Baum, W. M. (2002). From molecular to molar: A paradigm shift in behavior analysis. *Journal of the Experimental Analysis of Behavior*, *78*(1), 95–116. https://doi.org/10.1901/jeab.2002.78-95

Baum, W. M. (2005). *Understanding behaviorism: Behavior, culture, and evolution* (2nd edition). Malden, MA: Blackwell Publishing.

Baum, W. M. (2012). Rethinking reinforcement: Allocation, induction, and contingency. *Journal of the Experimental Analysis of Behavior*, *97*(1), 101–124. https://doi.org/10.1901/jeab.2012.97-101

Baum, W. M. (2017). Selection by consequences, behavioral evolution, and the price equation. *Journal of the Experimental Analysis of Behavior*, *107*(3), 321–342. https://doi.org/10.1002/jeab.256

Baum, W. M. (2018a). Multiscale behavior analysis and molar behaviorism: An overview. *Journal of the Experimental Analysis of Behavior*, *110*(3), 302–322. https://doi.org/10.1002/jeab.476

Baum, W. M. (2018b). Three laws of behavior: Allocation, induction, and covariance. *Behavior Analysis: Research and Practice*, *18*(3), 239–251. https://doi.org/10.1037/bar0000104

Becker, A. M. (2019). The flight of the locus of selection: Some intricate relationships between evolutionary elements. *Behavioural Processes*, *161*, 31–44. https://doi.org/10.1016/j.beproc.2018.01.002

Becker, G. S. (1976). *The economic approach to human behavior*. *Phoenix books: Vol. 803*. Chicago: Univ. of Chicago Press.

Broadbent, D. E. (1961). *Behaviour*. London: Methuen.

Burgos, J. E. (2019). Selection by reinforcement: A critical reappraisal. *Behavioural Processes*, *161*, 149–160. https://doi.org/10.1016/j.beproc.2018.01.019

Campbell, D. T. (1956). Adaptive behavior from random response. *Behavioral Science*, *1*(2), 105–110. https://doi.org/10.1002/bs.3830010204

Caswell, H. (2001). *Matrix Population Models. Construction, analysis, and interpretation. 2nd ed.* Sunderland: Sinauer Associates.

Caswell, H., & Weeks, D. E. (1986). Two-Sex Models: Chaos, Extinction, and Other Dynamic Consequences of Sex. *The American Naturalist*, *128*(5), 707–735. https://doi.org/10.1086/284598

Donahoe, J. W., Burgos, J. E., & Palmer, D. C. (1993). A selectionist approach to reinforcement. *Journal of the Experimental Analysis of Behavior*, *60*(1), 17–40. https://doi.org/10.1901/jeab.1993.60-17

Ewens, W. J. (2004). *Mathematical Population Genetics: I. Theoretical Introduction* (Second Edition). *Interdisciplinary Applied Mathematics: Vol. 27*. New York, NY: Springer. Retrieved from http://dx.doi.org/10.1007/978-0-387-21822-9 https://doi.org/10.1007/978-0-387-21822-9

Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford: Clarendon Press.

Frank, S. A. (1997). The design of adaptive systems: optimal parameters for variation and selection in learning and development. *Journa of Theoretical Biology*, *184*, 31–39.

Frankenhuis, W. E., Panchanathan, K., & Barto, A. G. (2019). Enriching behavioral ecology with reinforcement learning methods. *Behavioural Processes*. (161), 94–100. https://doi.org/10.1016/j.beproc.2018.01.008

Friedman, M. (1953). *Essays in positive economics* (7. impr). Chicago, Ill.: Univ. of Chicago Press.

Gilbert, R. M. (1970). Psychology and biology. *Canadian Psychologist/Psychologie Canadienne*, *11*(3), 221–238. https://doi.org/10.1037/h0082574

Grafen, A. (1982). How not to measure inclusive fitness. *Nature*, *298*(5873), 425. https://doi.org/10.1038/298425a0

Grafen, A. (2000). Developments of the Price equation and natural selection under uncertainty. *Proceedings. Biological Sciences*, *267*, 1223–1227.

Grafen, A. (2006a). Optimization of inclusive fitness. *Journal of Theoretical Biology*, *238*(3), 541–563. https://doi.org/10.1016/j.jtbi.2005.06.009

Grafen, A. (2006b). A theory of Fisher's reproductive value. *Journal of Mathematical Biology*, *53*(1), 15–60. https://doi.org/10.1007/s00285-006-0376-4

Grafen, A. (2009). Formalizing Darwinism and inclusive fitness theory. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1533), 3135–3141. https://doi.org/10.1098/rstb.2009.0056

Grafen, A. (2014). The formal darwinism project in outline. *Biology & Philosophy*, *29*(2), 155–174. https://doi.org/10.1007/s10539-013-9414-y

Grafen, A. (2015). Biological fitness and the Price Equation in class-structured populations. *Journal of Theoretical Biology*, *373*, 62–72. https://doi.org/10.1016/j.jtbi.2015.02.014

Hamilton, W. D. (1964). The genetical evolution of social behaviour, 1. *Journal of Theoretical Biology*, *7*, 1–16.

Herrnstein, R. J. (1964). Will. *Proceedings of the American Philosophical Society,*, *108*(6), 455–458.

Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior*, *13*, 243–266.

Herrnstein, R. J. (1974). Formal properties of the matching law. *Journal of the Experimental Analysis of Behavior*, *21*, 159–164.

Herrnstein, R. J. (1990). Behavior, Reinforcement and Utility. *Psychological Science*, *1*(4), 217–224. https://doi.org/10.1111/j.1467-9280.1990.tb00203.x

Herrnstein, R. J., Loewenstein, G. F., Prelec, D., & Vaughan, W. (1993). Utility maximization and melioration: Internalities in individual choice. *Journal of Behavioral Decision Making*, *6*(3), 149–185. https://doi.org/10.1002/bdm.3960060302

Herrnstein, R. J., & Prelec, D. (1991). Melioration: A Theory of Distributed Choice. *Journal of Economic Perspectives*, *5*(3), 137–156. https://doi.org/10.1257/jep.5.3.137

Houston, A. I., McNamara, J. M., & Steer, M. D. (2007). Do we expect natural selection to produce rational behaviour? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1485), 1531–1543. https://doi.org/10.1098/rstb.2007.2051

Hull, D. L., Langman, R. E., & Glenn, S. S. (2001). A general account of selection: Biology, immunology, and behavior. *The Behavioral and Brain Sciences*, *24*(3), 511-28; discussion 528-73.

Hursh, S. R. (1978). The economics of daily consumption controlling food- and water-reinforced responding1. *Journal of the Experimental Analysis of Behavior*, *29*(3), 475–491. https://doi.org/10.1901/jeab.1978.29-475

Johnson, M. W., & Bickel, W. K. (2006). Replacing relative reinforcing efficacy with behavioral economic demand curves. *Journal of the Experimental Analysis of Behavior*, *85*(1), 73–93.

Keyfitz, N., & Caswell, H. (2005). *Applied mathematical demography* (3rd ed.). *Statistics for biology and health*. New York NY: Springer.

Laland, K. N., Uller, T., Feldman, M. W., Sterelny, K., Müller, G. B., Moczek, A., . . . Odling-Smee, J. (2015). The extended evolutionary synthesis: Its structure, assumptions and predictions. *Proceedings. Biological Sciences*, *282*(1813), 20151019. https://doi.org/10.1098/rspb.2015.1019

Lehmann, L., Keller, L., West, S., & Roze, D. (2007). Group selection and kin selection: Two concepts but one process. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(16), 6736–6739. https://doi.org/10.1073/pnas.0700662104

Leslie, P. H. (1945). On the Use of Matrices in Certain Population Mathematics. *Biometrika*, *33*(3), 183. https://doi.org/10.2307/2332297

Loewenstein, Y., Prelec, D., & Seung, H. S. (2009). Operant matching as a Nash equilibrium of an intertemporal game. *Neural Computation*, *21*(10), 2755–2773. https://doi.org/10.1162/neco.2009.09-08-854

Madden, G. J., Smethells, J. R., Ewan, E. E., & Hursh, S. R. (2007). Tests of behavioral-economic assessments of relative reinforcer efficacy: Economic substitutes. *Journal of the Experimental Analysis of Behavior*, *87*(2), 219–240.

Marshall, J. A. R. (2011). Group selection and kin selection: Formally equivalent approaches. *Trends in Ecology & Evolution*, *26*(7), 325–332. https://doi.org/10.1016/j.tree.2011.04.008

Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge: Cambridge University Press.

McGraw, J., & Caswell, H. (1996). Estimation of individual fitness from life-history data. *American Naturalist*, *147*(1), 47–64.

McNamara, J. M., & Houston, A. I. (2009). Integrating function and mechanism. *Trends in Ecology & Evolution*, *24*(12), 670–675. https://doi.org/10.1016/j.tree.2009.05.011

Meehl, P. E. (1950). On the circularity of the law of effect. *Psychological Bulletin*, *47*(1), 52–75. https://doi.org/10.1037/h0058557

Metz, J. A. J., & Diekmann, O. (1986). *The Dynamics of Physiologically Structured Populations*. *Lecture Notes in Biomathematics.* https://doi.org/10.1007/978-3-662-13159-6

Pennypacker, H. S. (1992). Is behavior analysis undergoing selection by consequences? *The American Psychologist*, *47*(11), 1491–1498.

Premack, D. (1963). Rate differential reinforcement in monkey manipulation. *Journal of the Experimental Analysis of Behavior*, *6*, 81–89. https://doi.org/10.1901/jeab.1963.6-81

Price, G. R. (1970). Selection and Covariance. *Nature*, *227*(5257), 520–521. https://doi.org/10.1038/227520a0

Price, G. R. (1972). Extension of covariance selection mathematics. *Annals of Human Genetics*, *35*(4), 485–490. https://doi.org/10.1111/j.1469-1809.1957.tb01874.x

Pringle, J.W.S. (1951). On the Parallel Between Learning and Evolution. *Behaviour*, *3*(1), 174–214. https://doi.org/10.1163/156853951X00269

Queller, D. C. (1992). Quantitative Genetics, Inclusive Fitness, and Group Selection. *The American Naturalist*, *139*(3), 540–558. https://doi.org/10.1086/285343

Rachlin, H. (1971). On the tautology of the matching law. *Journal of the Experimental Analysis of Behavior*, *15*(2), 249–251.

Rachlin, H. (2019). Group selection in behavioral evolution. *Behavioural Processes*. (161), 65–72. https://doi.org/10.1016/j.beproc.2017.09.005

Rachlin, H., Battalio, R. C., Kagel, J. H., & Green, L. (1981). Maximization theory in behavioral psychology. *The Behavioral and Brain Sciences*, *4*, 371–417.

Rachlin, H., & Burkhard, B. (1978). The temporal triangle: Response substitution in instrumental conditioning. *Psychological Review*. (85), 22–47.

Rachlin, H., Green, L., Kagel, J. H., & Battalio, R. C. (1976). Economic demand theory and psychological studies of choice. *Psychology of Learning and Motivation*, *10*, 129–154.

Richerson, P. J. (2019). An integrated bayesian theory of phenotypic flexibility. *Behavioural Processes*, *161*, 54–64. https://doi.org/10.1016/j.beproc.2018.02.002

Richerson, P. J., Baldini, R., Bell, A. V., Demps, K., Frost, K., Hillis, V., . . . Zefferman, M. (2016). Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. *The Behavioral and Brain Sciences*, *39*, e30. https://doi.org/10.1017/S0140525X1400106X

Rogers, A. R. (1994). Evolution of Time Preference by Natural Selection. *American Economic Review*, *84*(3), 460–481.

Sakai, Y., & Fukai, T. (2008). When does reward maximization lead to matching law? *PloS One*, *3*(11), e3795. https://doi.org/10.1371/journal.pone.0003795

Schwartz, L. P., Silberberg, A., Casey, A. H., Paukner, A., & Suomi, S. J. (2016). Scaling reward value with demand curves versus preference tests. *Animal Cognition*, *19*(3), 631–641. https://doi.org/10.1007/s10071-016-0967-4

Simon, C., & Baum, W. M. (2017). Allocation of speech in conversation. *Journal of the Experimental Analysis of Behavior*, *107*(2), 258–278. https://doi.org/10.1002/jeab.249

Simon, C., & Hessen, D. O. (2019). Selection as a domain-general evolutionary process. *Behavioural Processes*, *161*, 3–16. https://doi.org/10.1016/j.beproc.2017.12.020

Singh, S., Lewis, L., & Barto, A. G. (2010). Intrinsically Motivated Reinforcement Learning: An Evolutionary Perspective. *IEEE Transactions on Autonomous Mental Development*, *2*(2), 70–82.

Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1969). *Contingencies of reinforcement: A theoretical analysis*. *The century psychology series*. Englewood Cliffs, NJ: Prentice-Hall.

Skinner, B. F. (1981). Selection by consequences. *Science (New York, N.Y.)*, *213*(4507), 501–504.

Skinner, B. F. (1984). Selection by consequences. *The Behavioral and Brain Sciences*, *7*(04), 477. https://doi.org/10.1017/S0140525X0002673X

Smaldino, P. E. (2014). The cultural evolution of emergent group-level traits. *The Behavioral and Brain Sciences*, *37*(3), 243–254. https://doi.org/10.1017/S0140525X13001544

Sorg, J. D. (2011). *The Optimal Reward Problem: Designing Effective Reward for Bounded Agents* (Doctoral Dissertation). University of Michigan, Michigan.

Staddon, J. E. R. (2016). *Adaptive behavior and learning* (Second edition). Cambridge: Cambridge University Press. Retrieved from http://dx.doi.org/10.1017/CBO9781139998369 https://doi.org/10.1017/CBO9781139998369

Staddon, J. E. R., & Motheral, S. (1978). On matching and maximizing in operant choice experiments. *Psychological Review*, *85*(5), 436–444. https://doi.org/10.1037/0033-295X.85.5.436

Staddon, J. E. R., & Simmelhag, V. L. (1971). The "supersitition" experiment: A reexamination of its implications for the principles of adaptive behavior. *Psychological Review*, *78*(1), 3–43. https://doi.org/10.1037/h0030305

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory*. *Monographs in behavior and ecology*. Princeton, NJ: Princeton Univ. Pr.

Taylor, P. D. (1990). Allele-Frequency Change in a Class-Structured Population. *The American Naturalist*, *135*(1), 95–106.

Thorndike, E. L. (2010/1911). *Animal intelligence; experimental studies*. New Brunswick, NJ: Transaction Publishers. https://doi.org/10.5962/bhl.title.55072

Timberlake, W. (2001). Integrating niche-related and general process approaches in the study of learning. *Behavioural Processes*, *54*(1-3), 79–94. https://doi.org/10.1016/S0376-6357(01)00151-6

Timberlake, W., & Allison, J. (1974). Response deprivation: An empirical approach to instrumental performance. *Psychological Review*, *81*, 146–164.

Tonneau, F., & Sokolowski, M. B. C. (2000). Pitfalls of Behavioral Selectionism. In F. Tonneau & N. S. Thompson (Eds.), *Perspectives in Ethology: Vol. 13. Perspectives in Ethology: Evolution, Culture, and Behavior* (Vol. 13, pp. 155–180). Boston, MA: Springer. https://doi.org/10.1007/978-1-4615-1221-9_6

Tustin, D. (2000). Revealed preference between reinforcers used to examine hypotheses about behavioral consistencies. *Behavior Modification*, *24*(3), 411–424. https://doi.org/10.1177/0145445500243007

Vaughan, W. (1981). Melioration, matching, and maximization. *Journal of the Experimental Analysis of Behavior*, *36*(2), 141–149. https://doi.org/10.1901/jeab.1981.36-141

Watson, J. B. (1930). *Behaviorism*. New York: W. W. Norton.

Willis, R. D., van Hartesveldt, C., Loken, K. K., & Hall, D. C. (1974). Motivation in concurrent variable-interval schedules with food and water reinforcers. *Journal of the Experimental Analysis of Behavior*, *22*(2), 323–331.