This manuscript has been accepted for publication in Behavioural Processes. The final publication can be retrieved from:
https://doi.org/10.1016/j.beproc.2021.104370

The formal foundation of an evolutionary theory of reinforcement

Borgstede, M., Eggert, F.***

*Corresponding author:

University of Bamberg

Markusplatz 3

D-96047 Bamberg

matthias.borgstede@uni-bamberg.de

**Technische Universität Braunschweig

Spielmannstrasse 19

D-38106 Braunschweig

f.eggert@tu-braunschweig.de

This manuscript has been accepted for publication in Behavioural Processes. The final publication can be retrieved from:
https://doi.org/10.1016/j.beproc.2021.104370

The formal foundation of an evolutionary theory of reinforcement

Abstract

Reinforcement learning is often described by analogy to natural selection. However, there is no coherent theory relating reinforcement learning to evolution within a single formal model of selection. This paper provides the formal foundation of such a unified theory. The model is based on the most general description of natural selection as given by the Price equation. We extend the Price equation to cover reinforcement learning as the result of a behavioral selection process within individuals and relate it to the principle of natural selection via the concept of statistical fitness predictors by means of a multilevel model of behavioral selection.

The main result is the *covariance-based law of effect*, which describes reinforcement learning on a molar level by means of the covariance between behavioral allocation and a statistical fitness predictor. We further demonstrate how this abstract principle can be applied to derive theoretical explanations of various empirical findings, like conditioned reinforcement, blocking, matching and response deprivation.

Our model is the first to apply the abstract principle of selection to derive a unified description of reinforcement learning and natural selection within a single model. It provides a general analytical tool for behavioral psychology in a similar way that the theory of natural selection does for evolutionary biology. We thus lay the formal foundation of a general theory of reinforcement as the result of behavioral selection on multiple levels.

Keywords: *selection by consequences, behavioral selection, natural selection, reinforcement learning, Price equation, multilevel model of behavioral selection*

This manuscript has been accepted for publication in Behavioural Processes. The final publication can be retrieved from: https://doi.org/10.1016/j.beproc.2021.104370

1 Introduction

It is a long held belief that reinforcement learning can be characterized by analogy to evolution by natural selection (e.g., Broadbent, 1961; D. T. Campbell, 1956; Gilbert, 1970; Herrnstein, 1964; Pringle, 1951; Skinner, 1966; Thorndike, 1900). For example, Staddon and Simmelhag (1971), state that the 'Law of Effect […] can best be understood by analogy with evolution by means of natural selection' (p. 40). Skinner (1981) even claims that natural selection and reinforcement learning are two instances of the same underlying causal principle: *selection by consequences*. The selectionist account of reinforcement has also found its way into textbooks on behavioral psychology as the *Darwinian metaphor* (Baum, 2005; Staddon, 2016) and is now a popular theme in theoretical work on behavior analysis (e.g., Baum, 2017, 2018; Becker, 2019; Borgstede, 2020; Donahoe, 2011; Donahoe, Burgos, & Palmer, 1993; Hull, Langman, & Glenn, 2001; Richerson, 2019; Simon & Hessen, 2019).

The appeal of the Darwinian metaphor conceivably stems from its generality: given selection by consequences is a fundamental principle of behavior, it might constitute the foundation of a unified theory of behavior. Such a theory would provide a general analytical framework for behavioral psychology in a similar way that the theory of natural selection does for evolutionary biology. A theoretical description of fundamental behavioral principles that goes 'beyond the collection of uniform relationships' (Skinner, 1950, p.215) could offer a theory-driven explanation of the basic laws of learning. This would help to understand why the regularities in empirical findings of behavioral psychology – for instance, the matching law – are to be expected and might even generate new testable hypotheses.

However, the theoretical status of 'selection by consequences' has been subject to criticism (e.g., the open peer commentaries to Skinner, 1984; also Burgos, 2019; Pennypacker, 1992; Tonneau & Sokolowski, 2000 for more recent accounts). Apart from an ongoing debate about the ontological status of the Darwinian metaphor, there are three recurring questions concerning the adequacy of the analogy between natural selection and reinforcement learning: the first one addresses the

hereditary particles of behavioral selection[1] (i.e., a gene analogue), the second one tackles the problem of a fitness equivalent on the behavioral level (i.e., some kind of reinforcer value), and the third one is about the units of selection (i.e., the question of what is selected). While the question of a gene analogue for behavioral selection is surely interesting (cf. Donahoe et al., 1993), it is not essential for selectionist theory, because it is possible to describe evolutionary change on a phenotypic level without loss of generality (Frank, 1997, 1998; Grafen, 2014). The second and third question, however, are crucial for the Darwinian metaphor to make sense – if there is no fitness equivalent on a behavioral level, there is no criterion for selection, and if the units of selection are unclear, we do not know what is selected.

The question of a behavioral fitness equivalent has been clarified from a maximization perspective (Borgstede, 2020): if there is a behavioral maximand ('reinforcer value') that reinforcement selects for, and if maximization of this value leads to maximization of evolutionary fitness, reinforcer value must be proportional to marginal fitness (i.e., fitness change per unit change in behavioral allocation). However, Borgstede (2020) does not link the maximization principle to the dynamics of change. Hence, whilst providing a valid mathematical definition of reinforcer value, the implications for reinforcement learning as a process of selection by consequences remain open.

The issue of the units of selection in the context of reinforcement learning has been addressed explicitly by McDowell (2013), who models reinforcement learning by means of an evolutionary algorithm that is applied to a population of 'behaviors'. In this view, behaviors relate to learning in the same way that individuals relate to evolution. However, in the context of learning, behavior is often treated as the *target* of selection, as well (Hull et al., 2001). Hence the analogy between learning and evolution is at least vague in this respect: do we conceptualize learning as selection *of* behaviors or as selection *for* behaviors (cf. Sober, 1984)?

[1] In this paper, we use the term 'behavioral selection' exclusively for behavioral adaptations by means of reinforcement learning. We do *not* refer to behavior being the target of evolution by natural selection.

In this paper, we aim to resolve the conceptual ambiguities of the analogy between learning and evolution by formally integrating reinforcement learning with natural selection in a unifying model that captures both levels of selection simultaneously. We build our model around the most abstract description of selection by means of the Price equation (Price, 1970, 1972). The main result is a molar account of reinforcement learning in terms of selection that applies to different levels on different time scales. In this view, reinforcement learning can be described as a Darwinian process where the units of selection are individuals showing behavioral variability and the target of selection is the relative allocation of behavior over time within a specified context. This process is universal in that it does not depend on the specific (molecular) mechanisms involved in learning but constitutes a general invariance principle which we call the *covariance based law of effect*. The dynamics of reinforcement are thus re-conceptualized in a molar way, shifting the focus from contiguity between single behavioral instances (Thorndike, 2010/1911) to the correlation between behavior and reinforcement (Baum, 1973). When applied to different experimental paradigms, the covariance based law of effect explains why *conditioned reinforcers* work (Skinner, 1969), why conditioning is sometimes *blocked* by previous reinforcement (Kamin, 1969), why *response deprivation* can establish reinforcers (Timberlake & Allison, 1974), and why individuals tend to *match* relative behavioral allocation to relative reinforcement in concurrent variable interval schedules (Herrnstein, 1961),

Our model integrates these empirically well-established regularities in a mathematically rigorous way by means of a single theoretical principle that is derived from the theory of evolution by natural selection. We thus lay the foundation for a general theory of behavior that is of 'greater generality than any assemblage of facts' (Skinner, 1950, p.216).

This manuscript has been accepted for publication in Behavioural Processes. The final publication can be retrieved from:
https://doi.org/10.1016/j.beproc.2021.104370

2 Natural selection and the Price equation

The Price equation provides a mathematical description of evolutionary processes on the most general level by partitioning the change in mean character value from one generation to the next generation into a covariance term and an expectation term (Price, 1970)[2]:

$$\bar{w}\Delta\bar{z} = \mathrm{Cov}(w_i, z_i) + \mathrm{E}(w_i \Delta z_i) \quad (1.)$$

Here, z_i refers to an arbitrary character (usually an allele frequency) and $\bar{z}$ designates the arithmetic mean of z over all individuals i. There is nothing special about gene frequencies here. z may be any quantitative character (e.g., body size or parental investment). Also, i does not necessarily refer to individuals, but can designate the members of an arbitrary set (e.g., groups). In order to define $\Delta\bar{z}$ one needs to relate the first set (the 'parent population') to a second set (the 'offspring population'): $\Delta\bar{z}$ is defined as $\sum q_i' z_i' - \sum q_i z_i$, where q_i refers to the frequency of the value z_i in the parent population and q_i' refers to the frequency of value z_i' in the offspring population. The index i does not refer to the individuals in the offspring population but to the parent population. This means that q_i' is the number of elements in the offspring population that originate from parents of type i and z_i' is their corresponding character value.

Mathematically, the Price equation builds on the existence of a *right-total relation* (i.e., a *mapping*) between two sets. Using this relation it is possible to define the fitness w_i as the contribution of a type i parent to the offspring population, resulting in $q_i' = q_i w_i / \bar{w}$, where $\bar{w}$ is the mean fitness of the parent population. z_i' is also defined with respect to the parent population, which means that it refers to the average character value z of descendants from parent type i. The change in z from parent to offspring is defined accordingly: $\Delta z_i = z_i' - z_i$. The Price equation is valid given these definitions (mathematical proofs can be found in Frank, 1998, Gardner, 2020 and elsewhere). The change in mean character value can always be partitioned into one covariance term and one

[2] Price used a different notation in his original paper. However, the notation adapted in this paper has become more common in the literature. Compare Luque (2017) for a review of the many different versions of the Price equation.

expectation term. In biological models, the covariance term captures the change in character value due to natural selection, whereas the expectation term refers to changes from parent to offspring due to imperfect transmission or environmental factors.

It is possible to extend the formalism to capture different genetic architectures, class structured populations, stochasticity and inclusive fitness (e.g., Frank, 1998; Grafen, 2000; Taylor, 1990). Moreover, the second term can be further partitioned by inserting the Price equation recursively into the expectation term. Taking individuals to be nested within groups and adding a new index g for these groups, one may partition the change in mean character value within groups, z_g, into a within group covariance term and a within group expectation term:

$$w_g \Delta z_g = \mathrm{Cov}(w_{gi}, z_{gi}) + \mathrm{E}_i(w_{gi} \Delta z_{gi}) \quad (2.)$$

Here, z_{gi} and w_{gi} stand for the individual character value and fitness of individual i in group g. z_g and w_g refer to mean character value and mean fitness in group g. Taking the expectation of mean change $w_g \Delta z_g$ over groups, one can extend the Price equation to capture selection within groups and selection between groups at the same time:

$$\overline{w} \Delta \bar{z} = \mathrm{Cov}(w_g, z_g) + \mathrm{E}_g \left(\mathrm{Cov}(w_{gi}, z_{gi}) + \mathrm{E}_i(w_{gi} \Delta z_{gi}) \right) \quad (3.)$$

This multilevel Price equation is useful to model fitness trade-offs between the individual and the group, thereby explaining, how a trait that is harmful to the individual can spread in a population by positively affecting average fitness on a group level (Price, 1972).

3 The Price equation and behavioral selection

Price himself noted that his formal account of selection was not restricted to natural selection acting on gene frequencies but might well be applied to other areas such as operant learning (Price, 1995, written ca. 1971). The Price equation has been applied to such diverse fields as probability theory, particle physics and information theory (Frank, 2017, 2018, 2020). Especially the application of a selectionist framework to the field of information theory seems to imply that there might be an

intricate relationship between selection and learning. Nevertheless, whilst there have been various attempts to apply the Price equation to cultural evolution (see Nettle, 2020 for an overview), its potential for behavioral psychology and reinforcement learning in particular remained largely unnoticed until a recent publication by Baum (2017).

Baum (2017) identifies the objects in the 'parent population' with operant behavior in a fixed time interval, and the objects in the 'offspring population' with operant behavior in a later time interval of equal length. These time intervals correspond to different trials in a behavioral experiment, where the individual is repeatedly confronted with the same reinforcement contingencies. In order to construct the necessary 'parent-offspring' relation between the two sets he argues that behaviors *recur* in the sense that a behavior emitted in interval one may be emitted in interval two, as well. Following this rationale, recurrence in behavioral selection means 'to occur again'. However, this understanding of recurrence departs from the conceptual foundation of the Price equation: 'to occur again' is by no means sufficient to establish a right-total relation between two sets of behavioral episodes. Baum further elaborates his position by linking reinforced behavior to Phylogenetically Important Events (PIE[3]). Following Baum, a PIE like, e.g., the availability of food induces PIE-related behavior like feeding and foraging. Thus, behaviors in interval one recur because they co-vary with a PIE, which in turn induces the same behaviors in interval two (Baum, 2012, 2017).

Even though induction possibly plays an important role for the allocation of behavior, it does not account for recurrence in the sense of a 'parent-offspring' relation as required by the Price equation. The indices in the Price equation always refer to the parent population – this means that z_i designates the mean character value of objects *descending from parent type* i. Therefore, in order to apply the mathematical apparatus of the Price equation it is necessary that every object in the offspring population can be *individually linked* to an object in the parent population. Applying Baum's claim that behaviors recur due to induction, we need to identify which behavioral instance in interval

[3] A PIE is an event that directly affects an individual's evolutionary fitness, e.g., the availability of food or the presence of a physical threat (Baum, 2012).

This manuscript has been accepted for publication in Behavioural Processes. The final publication can be retrieved from:
https://doi.org/10.1016/j.beproc.2021.104370

one induces which behavioral instance(s) in interval two. But this is not possible. Therefore, the presented account of behavioral recurrence by means of induction remains unclear.

Even though Baum's model is formally consistent for some special cases (namely, when there are no sources of behavioral change apart from selection), it is impossible to retrieve the original meaning of 'fitness' as the contribution from one generation to the next generation because the necessary set mapping (a right-total relation) does not exist. Moreover, formally treating behavioral selection as analogous to natural selection does not provide a functional relation between both levels of selection, thereby missing the opportunity to integrate learning and evolution within the same model.

4 A multilevel model of behavioral selection

As shown in the previous section, behaviors do not recur in the sense that one could establish a 'parent-offspring' relation between behaviors occurring in one time interval and behaviors occurring in a future time interval. Therefore, it is not possible to derive a coherent definition of fitness from the recurrence of behaviors themselves. We solve this issue by not defining fitness on the level of single behaviors, but on the level of the whole organism (i.e., in the standard way the Price equation is applied in biological models of natural selection). Thus, we do not conceptualize behavioral selection as a process *similar* to natural selection, but as *a part* of natural selection itself. In this view, behavioral selection is not a 'Darwinian metaphor' but a theoretically derived scientific fact: behavioral selection is *literally* a part of evolution by natural selection as described by the Price equation.

If the objects used in the Price equation are whole organisms instead of behaviors, several challenges arise. First, natural selection and behavioral selection are usually taken to act on different time scales. This poses a formal problem since the Price equation only deals with the change from one generation to the next. Second, natural selection and behavioral selection are mediated by different mechanisms of inheritance (or recurrence). Evolving characters may be influenced by both,

genetically transmitted components, as well as learned components (for which the mechanisms of transmission are yet to be understood). Third, since fitness is defined on the level of the individual (i.e., fitness refers to the contribution of an individual to the future population), it needs to be clarified how evolutionary fitness relates to reinforcement in order to make sense of assigning values of evolutionary fitness to *behavior*. We will provide solutions to each of the above problems by integrating natural selection and behavioral selection using a multilevel extension of the Price equation and linking evolutionary fitness to reinforcement via the concept of statistical fitness predictors. The general idea is that reinforcement learning can only be effective to the degree that the average outcome of reinforcement (i.e., the learned behavior) contributes to expected evolutionary fitness (Borgstede & Simon, under review).

The formal aspects of our model are very general in that they apply to any kind of behavior and that they are independent of specific learning algorithms. In fact, we do not attempt to model molecular mechanisms at all. Instead, we adopt a molar view that treats behavior as being extended over time and over contexts (Rachlin, 1978). Hence, when we speak of behavior, we mean the relative allocation of competing behavioral options over time within a certain context, which in itself is defined by a certain structure of contingencies. In the course of reinforcement learning, behavior becomes controlled by context-specific discriminative stimuli. This molar view implies that we analyze behavior in terms of allocated time per interval while in a certain context. Therefore, we do not need to worry about the measurement units of specific behaviors, which may very well vary between behaviors. We call these context-specific intervals 'behavioral episodes' to stress the fact that behavior is not measured at a point in time but extended over time within a specified context. A behavioral episode is thus defined with respect to the structure of contingencies that are effective in a specified environmental context. From the perspective of behavioral psychology, one may think of behavioral episodes as the trials in a reinforcement experiment, exposing the individual to the same contingencies again and again. However, the concept can also be applied to settings outside the laboratory. Here, we may identify behavioral episodes with recurring contexts the animal

encounters, for example, different food patches, whose contingency structures are signaled by the presence of certain discriminative stimuli.

To keep the mathematical derivation as simple as possible, we restrict our analysis to learning within only one type of behavioral episode. In other words, we treat learning separately in different contextual settings. Thus, when we calculate the mean behavioral allocation over several episodes, these need not be adjacent in time, but instead are recurring instances of the same contingency structure. Similarly, when we speak of statistical fitness predictors, we refer to predictors that are valid within the current class of behavioral episodes (i.e., they are context-dependent). This is in line with reinforcement learning as well as extinction being context-specific[4].

4.1 The covariance based law of effect

The first step to integrate natural selection and behavioral selection is to partition the expectation term of the Price equation using recursive expansion (cf. section 2). However, here the two levels of selection are not individuals nested within groups but *behavioral episodes nested within individuals*. Like in the simple Price equation, we designate individuals by the index i. Behavioral episodes are designated by the index j. We use the letter b to refer to the behavioral allocation within a behavioral episode. Behavioral allocation is measured as relative time spent at a certain behavior within a given behavioral episode[5]. For reasons of simplicity, we only deal with one type of behavior here. Hence, b_{ij} refers to the behavioral allocation of individual i in episode j.

To capture the rather small time scale of learning, we formally treat surviving individuals as if they were their own offspring. Even though this treatment of survival may be counter-intuitive, it is not uncommon in biological models, since it provides a mathematically simple way to capture the survival part of evolutionary fitness (cf. Taylor, 1990). Moreover, since we focus only on the learned

[4] It would be possible to include several contexts into the model by introducing a class structured version of the Price equation (cf. Taylor, 1990). However, this would inflate the mathematical formalism and distract the reader from the general import of the model.

[5] Actually, the model allows for different measures of behavior, as well, e.g., relative response rate, running speed, spatial position, or even neural activity. However, for reasons of consistency, we restrict ourselves to cases where behavior is measured as the relative allocation over time.

248 components of behavior, we restrict the analysis to the survival part of fitness, thereby excluding the
249 actual offspring of the individuals[6]. This means that, strictly speaking, b does not refer to behavioral
250 allocation per se, but to the part of behavioral allocation that is 'transmitted' within individuals from
251 one set of behavioral episodes to a second set of behavioral episodes . The mechanisms of
252 transmission certainly involve some kind of neural processing and might be referred to as 'memory'.
253 We do not imply the notion of a cognitive process of 'storing' and 'retrieving' here, but use the term
254 'memory' for any mechanism that integrates past experiences with present behavior. However, since
255 the model does not depend on these supposed mechanisms, we prefer to speak of 'behavior' b.
256 Formally, this definition of b is an analogue to the biological concept of 'breeding value' or 'additive
257 genetic value', which is defined as the component of the evolving character that is genetically
258 transmitted. Using breeding value instead of the actual character comes without loss of generality
259 when using the Price equation to describe natural selection (Frank, 1998). Effectively, treating
260 surviving individuals as their own offspring provides a way to bridge the time scales between natural
261 selection and reinforcement learning because no matter how short the interval between the two sets
262 of behavioral episodes, if fitness is defined by surviving individuals, the Price equation still describes
263 (a part of) natural selection. At the same time, restricting the analysis to the survival component of
264 fitness circumvents the problem of different mechanisms of transmission.

265 We can thus model several time steps on an evolutionary scale within the life span of a single
266 individual. There are no formal constraints on the choice of the time scale, hence we can choose the
267 level of analysis to fit our experimental requirements. Conceptually, this corresponds to a multiscale
268 view of behavior analysis as advocated by Baum (2018). Note, however, that no matter how small we
269 choose our evolutionary time scale to be, behavior is still understood as extended over time with
270 several behavioral episodes occurring in each (evolutionary) time step.

[6] Including transmission by reproduction would require a separation of hereditary mechanisms on the different levels of the Price equation. Although this might help to understand how the levels of selection interact and influence one another, it is not necessary in order to derive the general structure of behavioral selection.

This manuscript has been accepted for publication in Behavioural Processes. The final publication can be retrieved from:
https://doi.org/10.1016/j.beproc.2021.104370

Within this conceptual framework, we can now describe average behavioral change simultaneously for individuals and for the whole population by means of a multilevel Price equation:

$$\bar{w}\Delta\bar{b} = \mathrm{Cov}(w_i, b_i) + \mathrm{E}_i\left(\mathrm{Cov}(w_{ij}, b_{ij}) + \mathrm{E}_j(w_{ij}\Delta b_{ij})\right) \quad (4.)$$

Using behavioral allocation b as an evolving character and focusing on a time scale small enough to capture behavioral adaptations on an individual level, the two terms on the right hand side can be interpreted as follows: the covariance term over individuals i designates the survival part of natural selection; the expectation term over individuals i designates behavioral changes within the individuals. The latter term is further partitioned into one selection part on the level of behavioral episodes j and one expectation part referring to changes within behavioral episodes. Individual fitness w_i is defined in the usual way as the contribution of individuals to the future population (because the only 'offspring' in this model are the individuals themselves, this is essentially survival). Since we are mainly interested in the selection part, we can simplify our notation by treating the within individual expectation term as a residual term δ, defining $\mathrm{E}_j(w_{ij}\Delta b_{ij}) = \delta$. Due to the recursive extension of the Price equation, change in behavioral allocation within each individual can now be expressed as:

$$w_i\Delta b_i = \mathrm{Cov}(w_{ij}, b_{ij}) + \delta \quad (5.)$$

Although fitness can be explicitly defined on an individual level, it is difficult to make sense of the 'fitness' ascribed to behavioral episodes w_{ij}. Since it is always the whole organism that dies or survives (or, more formally, has a certain survival probability), it is not reasonable to attribute evolutionary fitness to a behavioral episode (note that we are talking about fitness in a literal way here – thus, it would be inconsistent to invoke a metaphorical interpretation). Moreover, it is unlikely that individuals adapt their behavior with regard to their *actual* fitness, since this would require information about their future survival. Therefore, it is reasonable to assume that individuals adapt their behavior with regard to *fitness proxies* p. These fitness proxies are essentially statistical predictors of evolutionary fitness. In some cases, they may coincide with the aforementioned PIEs,

but this is not necessarily the case. Formally, we predict fitness w by a context-dependent linear regression of the form $w = \beta_0 + \beta_{wp}p + \varepsilon$. The regression is calculated over all individuals that are exposed to the contextual factors that constitute the class of behavioral episodes. This means that we calculate separate regressions for each context in accordance with the definition of context by its contingency structure.

We now use these regression coefficients to obtain statistical estimates of the expected fitness for each behavioral episode. Formally, we substitute the fitness of each behavioral episode w_{ij} with the corresponding predicted value from the regression model:

$$w_i \Delta b_i = \mathrm{Cov}\left(\beta_0 + \beta_{wp} p_{ij}, b_{ij}\right) + \delta \tag{6.}$$

Since the regression coefficients are calculated on a between individuals level, they can be treated as constants within individuals. Therefore we can simplify to get[7]:

$$w_i \Delta b_i = \beta_{wp} \mathrm{Cov}\left(p_{ij}, b_{ij}\right) + \delta \tag{7.}$$

This means that the change in behavior for an individual i equals the covariance between behavioral allocation b and a linear fitness predictor p, weighted by the statistical regression effect of the fitness proxy p on evolutionary fitness, with δ being a residual term capturing all changes in behavior that are not caused by selection. We call this the *covariance based law of effect*, since it provides the conditions under which behavior is changed by reinforcement learning: *the change in mean behavioral allocation due to selection is proportional to the covariance between the behavior and a reinforcer*. Furthermore, the coefficient β_{wp} acts as a weighting factor to scale the covariance term. This means that behavioral change is also proportional to the degree to which a reinforcer is predictive for evolutionary fitness. Therefore, β_{wp} is called the *reinforcing power* of a p (cf. Borgstede, 2020).

[7] See Appendix 1 for the complete derivation.

This manuscript has been accepted for publication in Behavioural Processes. The final publication can be retrieved from:
https://doi.org/10.1016/j.beproc.2021.104370

5 Application of the model

The covariance based law of effect is a theoretical description of reinforcement learning on the most abstract level. It provides a quantitative account of behavioral selection that formally links the level of individual learning to the level of natural selection. We will now demonstrate how this abstract principle can be applied to various experimental paradigms. Using the method of path analysis, we will partition the reinforcing effects of behavioral selection into different components[8]. The first application deals with the implications of our model for steady state behavior. The second application is concerned with the factors that constitute a reinforcer.

5.1 Path analysis

The covariance based law of effect describes reinforcement by means of behavioral selection. The covariance term stresses the understanding of learning as a selection process as described by the Price equation. However, it is useful to translate the law of effect into a form that is more easily tractable in practical applications. Therefore, we introduce an alternative formulation of the same law by means of a statistical path model.

Because, by definition, $\beta_{pb} = \frac{\mathrm{Cov}(p_{ij},b_{ij})}{\mathrm{Var}(b_{ij})}$, the covariance based law of effect can be equivalently stated as:

$$w_i \Delta b_i = \beta_{wp}\beta_{pb}\mathrm{Var}(b_{ij}) + \delta \qquad (8.)$$

Here, β_{pb} and β_{wp} are the partial linear effects from a path model where the effect of behavior b on evolutionary fitness w is fully mediated by a fitness proxy p (see Figure 1). The parameter β_{wp} refers to the reinforcing power of the fitness proxy. It corresponds to the expected change in evolutionary fitness per unit change in reinforcement. The parameter β_{pb} is the slope of the feedback function of

[8] Path analysis is routinely applied to partition different fitness effects on evolutionary fitness in biological models of selection (cf. Okasha and Otsuka, 2020; Scheiner and Gurevitch, 2001). The difference here is that, apart from the fitness effects of the fitness proxies, we use within individual regressions to calculate the contextual effects that are effective in the current class of behavioral episodes.

This manuscript has been accepted for publication in Behavioural Processes. The final publication can be retrieved from:
https://doi.org/10.1016/j.beproc.2021.104370

the schedule of reinforcement. It thus captures the expected gain in reinforcement per unit change in behavior at the current point of behavioral allocation.

Using the technique of statistical path analysis, we can calculate the total effect β_{wb} of behavior b on fitness w by multiplication of the two partial regression slopes, which gives $\beta_{wb} = \beta_{wp}\beta_{pb}$. This total effect corresponds to the marginal value of a change in behavioral allocation in terms of evolutionary fitness. According to Borgstede (2020), this may be regarded as marginal *reinforcer value* $r(b)$ of behavior b. We may thus state the covariance based law of effect in terms of reinforcer value[9]:

behavioral selection equals the product of reinforcer value and behavioral variance.

If we calculate the path model using standardized variables, behavioral variance will be one, leaving us with the standardized total effect of behavior on fitness. We can thus investigate different components of reinforcement by partitioning the reinforcer value of a behavior into partial regression effects. Similarly, we can start with a given partitioning and retrieve the corresponding expression of $r(b)$ by summing up the products of the partial regression coefficients for each path (Shipley, 2016). This method can be applied to analyze the structure of different contexts from the perspective of behavioral selection.

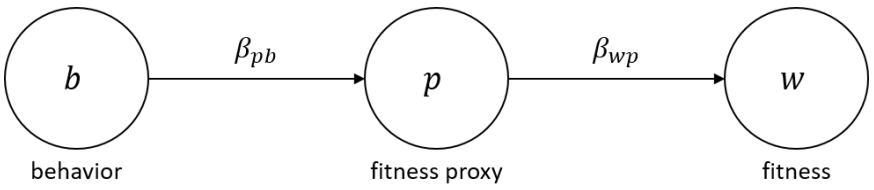

Figure 1: Path diagram depicting the statistical relation between behavior b, fitness proxy p and evolutionary fitness w. β_{pb} and β_{wp} designate the slopes of linear regression models that predict p from behavior b, and fitness w from p controlling for b, respectively. Statistically, the total effect of b on w equals the product of β_{pb} and β_{wp}.

[9] Note that in Borgstede (2020), $r(b)$ refers to the *partial* effect of behavior on fitness, whereas here, we refer to the *total* effect, which is the sum of the partial effects over all paths from b to w.

This manuscript has been accepted for publication in Behavioural Processes. The final publication can be retrieved from:
https://doi.org/10.1016/j.beproc.2021.104370

5.2 Steady state behavior

Our first application deals with several mutually exclusive behaviors that compete for time within behavioral episodes. Formally, behavioral allocation of behaviors $b_1, b_2, \dots, b_n$ is subject to a fixed time budget constraint such that the relative time spent at each behavior sums up to one. From this it follows that any increase in mean behavioral allocation towards one behavior will result in an equal amount of decrease in the sum over all other behaviors.

In the following, we will focus on the dynamics of only one behavior b. Like before, the amount of behavioral selection is determined by the corresponding reinforcer value. In contrast to the simple mediation model of the previous section, we now have to incorporate the time constraint when calculating the total effect of behavior b on fitness w. Figure 2 depicts the simplest case with only one additional behavior (we call this second behavior b')[10]. In the path model, we account for the time constraint by adding a path from behavior b to b' (dashed line in Figure 2). The corresponding partial regression coefficient $\beta_{b'b}$ expresses the expected change in behavior b' per unit change in b. Due to the above budget constraint, it holds that $\beta_{b'b} = -1$. Consequently, we can calculate the marginal reinforcer value of behavior b as:

$$r(b) = \beta_{wp}\beta_{pb} - \beta_{wp'}\beta_{p'b'} \tag{9.}$$

This means that we have to subtract the partial fitness effect of behavior b' from the partial fitness effect of behavior b to predict behavioral selection on b.

We can use this result to derive a general equilibrium condition for behavioral selection. Behavioral equilibrium (or 'steady state behavior') is characterized by a constant behavioral allocation (i.e., the absence of change due to reinforcement). Formally, this means that behavior b is at a steady state, if

[10] We restrict ourselves to this minimal example to avoid unnecessary complexity of the path model. This does not restrict the generality of the analysis, however, because we only consider the point of behavioral equilibrium here. This means that, even if behaviors b and b' do not take up the whole time interval at the beginning of the experiment, in the absence of selection on other behaviors, they will eventually compete for the whole available time.

This manuscript has been accepted for publication in Behavioural Processes. The final publication can be retrieved from:
https://doi.org/10.1016/j.beproc.2021.104370

and only if behavioral selection on b equals zero (i.e., $w_i \Delta b_i = 0$). The equilibrium condition thus becomes:

$$\beta_{wp}\beta_{pb} = \beta_{wp\prime}\beta_{p\prime b\prime} \quad (10.)$$

Hence, behavior is stable if and only if the marginal fitness effects of competing behaviors are equal. This result is equivalent to the one derived in Borgstede (2020). However, we did not assume a tendency to maximize reinforcer value, but derived the same condition for steady state behavior from the covariance based law of effect.

It has been shown that in concurrent variable interval schedules of reinforcement, the condition of equal marginal reinforcer value coincides with the *matching law* (Baum, 1981). However, this is not the case for all schedules of reinforcement, since the resulting behavioral allocation depends on the shape of the corresponding feedback functions. Hence, the matching law is best understood as a special case of the above equilibrium condition and therefore follows from the covariance based law of effect.

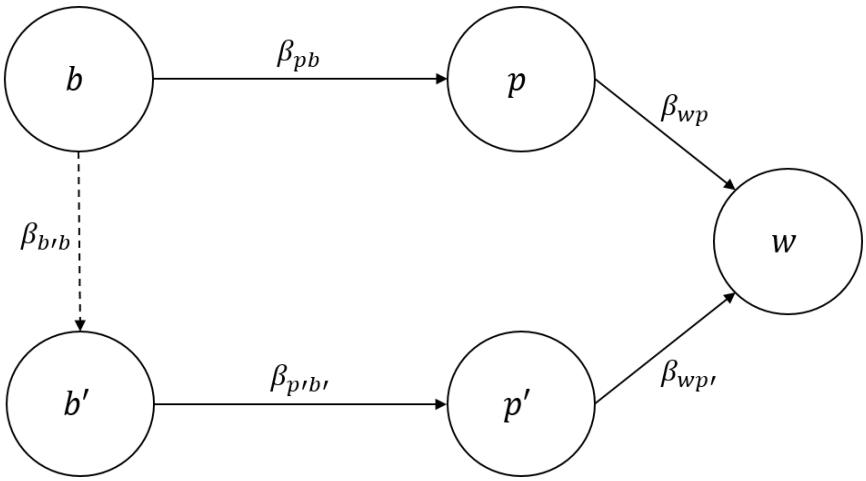

Figure 2: Path diagram for behavioral selection on a behavior b that is constrained by a temporal budget. In addition to the direct path from behavior b to fitness w, the time constraint induces a negative correlation with behavior b'. At the point of behavioral equilibrium, this results in a second path from b to w that is mediated by b'.

This manuscript has been accepted for publication in Behavioural Processes. The final publication can be retrieved from: https://doi.org/10.1016/j.beproc.2021.104370

5.3 The nature of reinforcement

Let us now consider a case where behavioral allocation of behavior b has reached an equilibrium state (we may call this the 'baseline condition'). Effectively, this means that there is no behavioral change due to selection. Given behavioral variance does not equal zero, this is equivalent to a marginal reinforcer value of zero, yielding $r(b) = 0$. In other words, there is no benefit in terms of evolutionary fitness when behavioral allocation is changed.

We now imagine a second behavior b' that we try to reinforce by pairing it with b, thereby establishing a positive covariance between b and b' (this is indicated by the dashed line in Figure 3). We can easily see that this will not affect b' because, under baseline conditions, the marginal reinforcer value of b' is:

$$r(b') = \beta_{bb'} r(b) = 0 \quad (11.)$$

We can now constraint the behavior b, thereby disturbing the equilibrium state we observed under baseline conditions. Following the principle of diminishing returns (i.e., assuming that the marginal effect of a change in behavioral allocation becomes smaller for higher values of b), this will result in a positive reinforcer value for behavior b. Therefore, if we make b contingent on a second behavior b', the reinforcer value of b' becomes positive, as well, which in turn results in behavioral selection. This explains why *response deprivation* can establish constrained behaviors as reinforcers (Timberlake & Allison, 1974).

The covariance based law of effect provides us with a general definition of a reinforcer as a context-dependent fitness predictor. This implies that behaviors may become reinforcers, given they predict an expected gain in fitness, which will be the case for most constrained behaviors. Under the assumption that high-probability behaviors are often constrained by the environment (i.e., individuals would engage even more in high-probability behaviors, if they could), this also explains Premack's principle, which states that usually, behaviors that occur at a high probability under

This manuscript has been accepted for publication in Behavioural Processes. The final publication can be retrieved from:
https://doi.org/10.1016/j.beproc.2021.104370

baseline conditions, function as reinforcers for behaviors that occur with a lower probability (Premack & Premack, 1963).

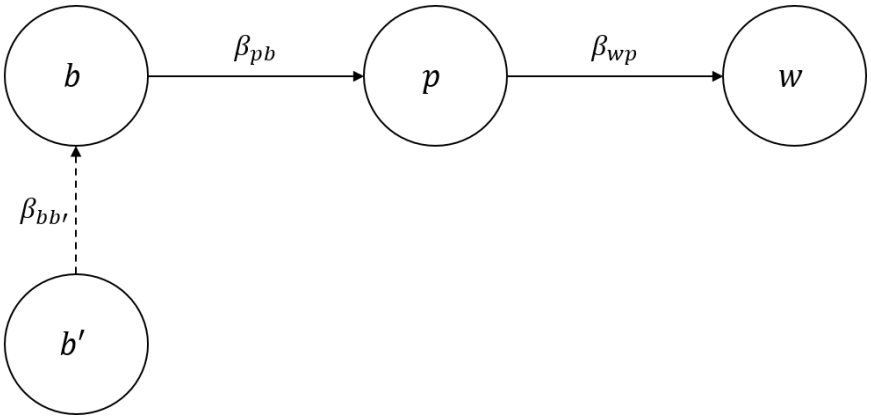

Figure 3: Path diagram for the reinforcing effects of a behavior b for a different behavior b'. Assuming diminishing returns of fitness proxy p for increasing values of b, constraining behavior b below baseline will raise the value of β_{pb}. This, in turn, will establish the constrained behavior as a reinforcer for b'.

We can extend the method of path analysis to even more complex scenarios, such as the establishment of conditioned reinforcers. In this paradigm, a formerly neutral stimulus (e.g., a flashing light) is repeatedly paired with the availability of food (or any other reinforcer). The pairing of a stimulus with a reinforcer corresponds to establishing an empirical covariance between the stimulus s and a fitness proxy p. Given there is variation in s, we can thus calculate a linear regression of p on s. Since the stimulus is made contingent on behavior b, this results in a mediation model with s being a mediating variable between b and p. The total fitness effect of behavior b (i.e., its reinforcer value) thus becomes:

$$r_0(b) = \beta_{wp}\beta_{ps}\beta_{sb} \quad (12.)$$

When a second discriminative stimulus s' is added to the experiment, we establish a new path in the predictive model (this is indicated by the dashed lines in Figure 4). Given s has already been established as a discriminative stimulus, the statistical effect of behavior b on the new stimulus s' is completely mediated by s. Therefore, the total fitness effect of the behavior can be partitioned as follows:

This manuscript has been accepted for publication in Behavioural Processes. The final publication can be retrieved from:
https://doi.org/10.1016/j.beproc.2021.104370

$$r_1(b) = \beta_{wp}\beta_{ps}\beta_{sb} + \beta_{wp}\beta_{ps\prime}\beta_{s\prime s}\beta_{sb} \quad (13.)$$

$\beta_{ps\prime}$ corresponds to the partial effect of s' on p when controlled for the effect of s. As long as s' always occurs together with s, there is no additional variance in p explained when s' is added to the regression. Consequently, $\beta_{ps\prime}$ will be zero. If we now stop presenting the first stimulus, the expected fitness gain of behavior b reduces to the second path:

$$r_2(b) = \beta_{wp}\beta_{ps\prime}\beta_{s\prime s}\beta_{sb} \quad (14.)$$

Since $\beta_{ps\prime} = 0$, the reinforcer value of b becomes zero. Therefore, we do not expect the second stimulus to affect behavior in the absence of the first one.

This phenomenon has been demonstrated repeatedly in classical and operant conditioning experiments and is known as the *blocking effect* (Kamin, 1969). Note that applying the covariance based law of effect to this special case does not only explain the blocking effect, but also makes a quantitative prediction about the amount of reinforcer value when s and s' are not perfectly correlated (i.e., if . $\beta_{ps\prime} \neq 0$). By partitioning the predictive effects of a behavior in a given context, we can thus explain how conditioned reinforcers acquire control over an individual's behavior. The general rule is that new stimuli affect the current reinforcer value (and thus the amount of reinforcement) to the degree that they provide additional information about the availability of fitness proxies (cf. Rescorla & Wagner, 1972). Thus, we can understand the well-known Rescorla-Wagner-Model as a molecular model that follows the general principle of reinforcement as described by the covariance based law of effect.

This manuscript has been accepted for publication in Behavioural Processes. The final publication can be retrieved from:
https://doi.org/10.1016/j.beproc.2021.104370

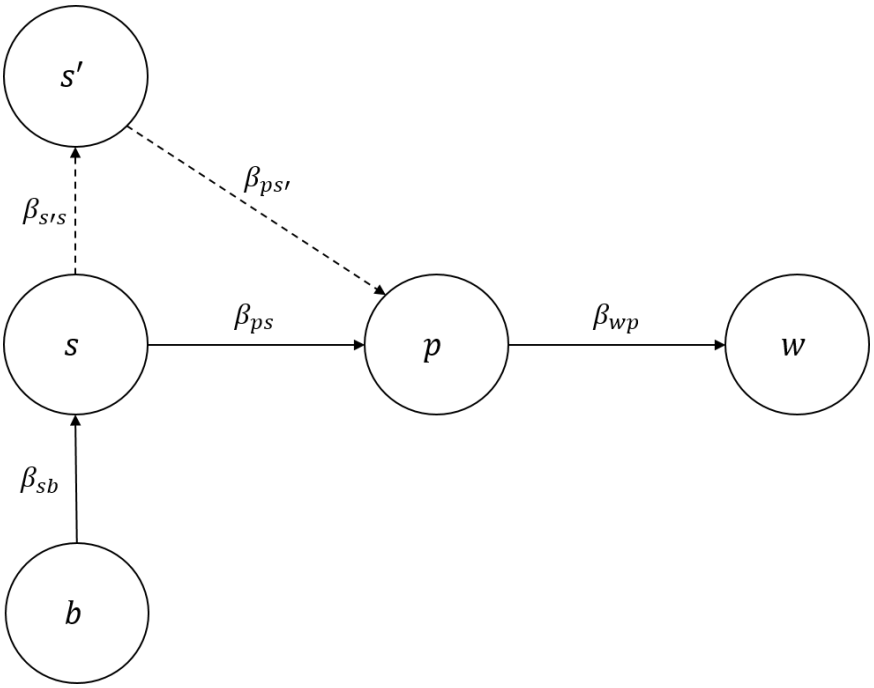

Figure 4: Path diagram to illustrate the principle of conditioned reinforcement and the blocking effect. If a stimulus s is paired with a fitness proxy p, it will become predictive of fitness w and can then act as a reinforcer itself. However, if we pair an additional stimulus s' with the fitness proxy p, there will be no gain in predictive power, as long as s' is presented alongside with the previously established conditioned reinforcer s (i.e., the second stimulus is 'blocked' by the first one).

6 Discussion

This paper deals with the question how reinforcement learning can be formally described as a Darwinian process. We present an evolutionary model of behavioral selection using the formalism of the Price equation, thereby clarifying some common conceptual ambiguities. The model treats behavioral selection (by means of reinforcement learning) as a part of natural selection, rather than a process that is merely analogous to natural selection. Hence, the 'Darwinian metaphor' advocated by many behavioral scientists is replaced by the view that behavioral selection literally *is* a Darwinian process (i.e., a part of evolution by natural selection). Therefore, in contrast to existing approaches to behavioral selection (e.g., Baum, 2017; Donahoe et al., 1993; McDowell, 2004), the presented multilevel model of behavioral selection is more than a re-statement of standard learning theories using a selectionist vocabulary – it is a true integration of behavioral psychology and evolutionary biology.

The model is used to derive the *covariance based law of effect*, which describes reinforcement learning as a selection process that is proportional to the covariance between behavioral allocation and a fitness proxy, weighted by the corresponding statistical fitness effect. We further showed how this abstract principle of behavioral selection can be applied to various experimental paradigms, thereby integrating such diverse empirical regularities as conditioned reinforcement, response deprivation, the blocking effect and the matching law. We thus lay the formal foundation for a general theory of reinforcement that is grounded in the theory of evolution by natural selection.

The covariance based law of effect also provides a formal definition of a 'reinforcer': a reinforcer is anything that is statistically predictive of evolutionary fitness. Due to the statistical nature of reinforcers, it does not matter whether we conceive them as external events (e.g., the availability of food), activities of the individual (e.g., eating), or perceptions (e.g., taste). Hence, we might just as well attribute the reinforcing power of a p to the feeding behavior that it induces, or to the resulting taste perception. A reinforcer is not a 'thing' that somehow changes the individual , it is essentially a (context-dependent) statistical fitness predictor.

Since the multilevel model of behavioral selection translates behavioral change into statistical terms like covariances, variances and regression coefficients, behavioral selection is inherently linked to the concept of *prediction*. In fact, it has been shown that natural selection as described by the Price equation maximizes Fisher Information[11] (Frank, 2009). This means that evolutionary change can be understood as a mechanism that maximizes predictive power with regard to the environment. Adopting this view, we interpret behavioral selection as a mechanism to optimize the predictions of an organism about the environment. This is in line with the theory of *predictive coding* (Helmholtz, 1909) and the *Bayesian brain* hypothesis (Clark, 2013). Although the formal integration of the 'predictive brain' hypothesis and the principle of behavioral selection has yet to be accomplished, the link between the Price equation and information theory provides a promising approach towards a

[11] Fisher information is a statistical measure of how much information observations provide about an unknown parameter of a probability distribution.

general theory of learning that explains the structure of learning processes by means of a universal selection principle.

Of course, the high level of generality of the presented model is limited by the underlying assumptions. First, we did not model the hereditary mechanisms that mediate behavioral selection. Treating surviving individuals as their own offspring and leaving aside the contribution to the population by reproduction, it was formally possible to leave the question of hereditary mechanisms open. However, this means that the 'offspring population' only consists of surviving individuals, thereby ignoring all behavioral changes that stem from (genetical) transmission and reproduction. This means that newborn individuals are not counted to calculate evolutionary fitness. Therefore, strictly speaking, the model only refers to the survival component of evolutionary fitness and thus does not capture the whole effect of natural selection. However, since reinforcement occurs only within individuals, omitting the reproduction part of evolutionary fitness does not pose a major problem. Nevertheless, it would be interesting to investigate the effects of different hereditary mechanisms on different levels of selection. Apart from the genetic part, in social species, cultural mechanisms (like imitation, model learning, or verbal instruction) partly mediate the transmission of behavior from parent (in the usual meaning) to offspring. Moreover, individual learning invokes at least one more mechanism of transmission (some kind of 'memory'). Disentangling these effects requires a careful mathematical treatment and will be targeted in subsequent work.

Finally, it should be mentioned that the Price equation in its original form implies a homogeneous population, thereby assuming inter-individual variation negligible. This assumption has been made here to keep the mathematical notation as simple as possible. Inter-individual variation can be added using a class structured version of the Price equation. This introduces additional weighting factors for the classes, where each class refers to a 'type' of individual, defined by a certain combination of characteristics (Taylor, 1990). In population biology, these weighting factors are the *reproductive values* of the different types of individuals and depend on the demography and long-term dynamics of the population (cf. Caswell, 2001). If the aim is to make specific predictions about adaptive

behavior, at least some kind of population structure has to be modelled. However, as long as one is mainly concerned with the fundamental principles of selection, class structure can be ignored. Therefore, here we stick to the most general form of the Price equation, resulting in a most general account of behavioral selection. Incorporating inter-individual variation will be the objective of subsequent work.

Despite these limitations, this paper provides a consistent quantitative account of reinforcement learning on a molar level. It integrates behavioral selection with natural selection and provides new insights into the quantitative relation between reinforcement, behavioral allocation and evolutionary fitness. This is an important step towards a general account of learning and of behavior in general, based on the theory of evolution by natural selection.

7 Author contributions

XX conceived the original idea, developed the theoretical formalism, performed the mathematical derivations and wrote the original draft. XY verified the analytical methods, provided additional ideas and critical feedback and helped shape the research, analysis and interpretation. Both authors contributed to the final version of the manuscript.

8 Competing interests

The authors declare no competing interests.

9 Appendix

9.1 Derivation of the covariance based law of effect

We start with the elementary Price equation for mean individual behavioral allocation b_i:

$$\bar{w}\Delta\bar{b} = \mathrm{Cov}(w_i, b_i) + \mathrm{E}_i(w_i \Delta b_i)$$

Applying the logic of the multilevel Price equation, the expectation term can be separated into a within individual covariance term and a within individual expectation term:

This manuscript has been accepted for publication in Behavioural Processes. The final publication can be retrieved from:
https://doi.org/10.1016/j.beproc.2021.104370

$$w_i \Delta b_i = \mathrm{Cov}(w_{ij}, b_{ij}) + \mathrm{E}_j(w_{ij} \Delta b_{ij})$$

By substitution, we arrive at the multilevel Price equation for behavioral selection:

$$\bar{w} \Delta \bar{b} = \mathrm{Cov}(w_i, b_i) + \mathrm{E}_i \left(\mathrm{Cov}(w_{ij}, b_{ij}) + \mathrm{E}_j(w_{ij} \Delta b_{ij}) \right)$$

We now assume a statistical fitness predictor p of the form $w = \beta_0 + \beta_{wp} p + \varepsilon$ and substitute the w_{ij} with the predicted values from this regression:

$$\bar{w} \Delta \bar{b} = \mathrm{Cov}(w_i, b_i) + \mathrm{E}_i \left(\mathrm{Cov}(\beta_0 + \beta_{wp} p_{ij}, b_{ij}) + \mathrm{E}_j((\beta_0 + \beta_{wp} p_{ij}) \Delta b_{ij}) \right)$$

This can be rearranged to:

$$\bar{w} \Delta \bar{b} = \mathrm{Cov}(w_i, b_i) + \mathrm{E}_i \left(\mathrm{Cov}(\beta_0, b_{ij}) + \mathrm{Cov}(\beta_{wp} p_{ij}, b_{ij}) + \mathrm{E}_j((\beta_0 + \beta_{wp} p_{ij}) \Delta b_{ij}) \right)$$

Since β_0 is a constant, $\mathrm{Cov}(\beta_0, b_{ij})$ equals 0. This results in:

$$\bar{w} \Delta \bar{b} = \mathrm{Cov}(w_i, b_i) + \mathrm{E}_i \left(\mathrm{Cov}(\beta_{wp} p_{ij}, b_{ij}) + \mathrm{E}_j((\beta_0 + \beta_{wp} p_{ij}) \Delta b_{ij}) \right)$$

Rearrangement yields:

$$\bar{w} \Delta \bar{b} = \mathrm{Cov}(w_i, b_i) + \mathrm{E}_i \left(\beta_{wp} \mathrm{Cov}(p_{ij}, b_{ij}) + \mathrm{E}_j((\beta_0 + \beta_{wp} p_{ij}) \Delta b_{ij}) \right)$$

Because the expectation term in the Price equation is taken over the fitness weighted changes in behavioral allocation $w_i \Delta b_i$, which has been separated into $\mathrm{Cov}(w_{ij}, b_{ij}) + \mathrm{E}_j(w_{ij} \Delta b_{ij})$ above, all rearrangements of the multilevel Price equation within this expectation term equally apply to $w_i \Delta b_i$. Therefore, we can write the change in behavioral allocation within individuals as:

$$w_i \Delta b_i = \beta_{wp} \mathrm{Cov}(p_{ij}, b_{ij}) + \mathrm{E}_j((\beta_0 + \beta_{wp} p_{ij}) \Delta b_{ij})$$

Defining $\delta = \mathrm{E}_j((\beta_0 + \beta_{wp} p_{ij}) \Delta b_{ij})$ we arrive at the covariance based law of effect:

$$w_i \Delta b_i = \beta_{wp} \mathrm{Cov}(p_{ij}, b_{ij}) + \delta$$

10 References

Baum, W. M. (1973). The correlation-based law of effect. *Journal of the Experimental Analysis of Behavior*, *20*, 137–153.

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. (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. (2018). 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

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

Borgstede, M. (2020). An evolutionary model of reinforcer value. *Behavioural Processes*, 104109.

Borgstede, M., & Simon, C. (under review). A conceptual synthesis of learning and evolution.

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.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Donahoe, J. W. (2011). Selectionism. In K. A. Lattal & P. A. Chase (Eds.), *Behavior theory and philosophy* (Vol. 33, pp. 103–128). New York, London: Springer. https://doi.org/10.1007/978-1-4757-4590-0_6

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

Frank, S. A. (1997). The Price equation, Fisher's fundamental theorem, kin selection, and causal analysis. *Evolution; International Journal of Organic Evolution*, *51*(6), 1712–1729. https://doi.org/10.1111/j.1558-5646.1997.tb05096.x

Frank, S. A. (1998). *Foundations of social evolution*. *Monographs in behavior and ecology*. Princeton, NJ: Princeton Univ. Press.

Frank, S. A. (2009). Natural selection maximizes Fisher information. *Journal of Evolutionary Biology*, *22*(2), 231–244. https://doi.org/10.1111/j.1420-9101.2008.01647.x

599 Frank, S. A. (2017). Universal expressions of population change by the Price equation: Natural
600 selection, information, and maximum entropy production. *Ecology and Evolution*, *7*(10), 3381–
601 3396. https://doi.org/10.1002/ece3.2922

602 Frank, S. A. (2018). The Price Equation Program: Simple Invariances Unify Population Dynamics,
603 Thermodynamics, Probability, Information and Inference. *Entropy*, *20*(12), 978.
604 https://doi.org/10.3390/e20120978

605 Frank, S. A. (2020). Simple unity among the fundamental equations of science. *Philosophical*
606 *Transactions of the Royal Society of London. Series B, Biological Sciences*, *375*(1797), 20190351.
607 https://doi.org/10.1098/rstb.2019.0351

608 Gardner, A. (2020). Price's equation made clear. *Philosophical Transactions of the Royal Society of*
609 *London. Series B, Biological Sciences*, *375*(1797), 20190361.
610 https://doi.org/10.1098/rstb.2019.0361

611 Gilbert, R. M. (1970). Psychology and biology. *Canadian Psychologist/Psychologie Canadienne*, *11*(3),
612 221–238. https://doi.org/10.1037/h0082574

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

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

617 Helmholtz, H. von (1909). *Treatise on physiological optics. Vol. III 3rd edition*. Hamburg: Voss.

618 Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of
619 reinforcement. *Journal of the Experimental Analysis of Behavior*, *4*, 267–272.
620 https://doi.org/10.1901/jeab.1961.4-267

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

622 Hull, D. L., Langman, R. E., & Glenn, S. S. (2001). A general account of selection: Biology, immunology,
623 and behavior. *The Behavioral and Brain Sciences*, *24*(3), 511–528.
624 https://doi.org/10.1017/S0140525X01004162

625 Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In B. A. Campbell & R. M.
626 Church (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York.

627 Luque, V. J. (2017). One equation to rule them all: a philosophical analysis of the Price equation.
628 *Biology & Philosophy*, *32*(1), 97–125. https://doi.org/10.1007/s10539-016-9538-y

629 McDowell, J. J. (2004). A computational model of selection by consequences. *Journal of the*
630 *Experimental Analysis of Behavior*, *81*(3), 297–317. https://doi.org/10.1901/jeab.2004.81-297

631 McDowell, J. J. (2013). A quantitative evolutionary theory of adaptive behavior dynamics.
632 *Psychological Review*, *120*(4), 731–750. https://doi.org/10.1037/a0034244

633 Nettle, D. (2020). Selection, adaptation, inheritance and design in human culture: The view from the
634 Price equation. *Philosophical Transactions of the Royal Society of London. Series B, Biological*
635 *Sciences*, *375*(1797), 20190358. https://doi.org/10.1098/rstb.2019.0358

636 Okasha, S., & Otsuka, J. (2020). The Price equation and the causal analysis of evolutionary change.
637 *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *375*(1797),
638 20190365. https://doi.org/10.1098/rstb.2019.0365

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

641 Premack, D., & Premack, A. J. (1963). Increased eating in rats deprived of running. *Journal of the*
642 *Experimental Analysis of Behavior*, *6*, 209–212. https://doi.org/10.1901/jeab.1963.6-209

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

Price, G. R. (1995, written ca. 1971). The nature of selection. (Written circa 1971, published posthumously). *Journal of Theoretical Biology*, *175*(3), 389–396. https://doi.org/10.1006/jtbi.1995.0149

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

Rachlin, H. (1978). A molar theory of reinforcement schedules. *Journal of the Experimental Analysis of Behavior*, *30*(3), 345–360. https://doi.org/10.1901/jeab.1978.30-345

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

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

Scheiner, S. M., & Gurevitch, J. (2001). *Design and analysis of ecological experiments* (2nd ed.). Oxford, New York: Oxford University Press.

Shipley, B. (2016). *Cause and correlation in biology: A user's guide to path analysis, structural equations and causal inference with R* (2nd edition). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139979573

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

Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review*, *57*(4), 193–216.

Skinner, B. F. (1966). The phylogeny and ontogeny of behavior. Contingencies of reinforcement throw light on contingencies of survival in the evolution of behavior. *Science (New York, N.Y.)*, *153*(3741), 1205–1213. https://doi.org/10.1126/science.153.3741.1205

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

Sober, E. (1984). *The nature of selection: Evolutionary theory in philosophical focus*. Cambridge, Mass.: Bradford Books/MIT Press. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=909783

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

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

688 Thorndike, E. L. (1900). The associative processes in animals. *Biological Lectures from the Marine*
689 *Biological Laboratory of Woods Holl, 1899*, 69–91.

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

692 Timberlake, W., & Allison, J. (1974). Response deprivation: An empirical approach to instrumental
693 performance. *Psychological Review*, *81*(2), 146–164. https://doi.org/10.1037/h0036101

694 Tonneau, F., & Sokolowski, M. B. C. (2000). Pitfalls of Behavioral Selectionism. In F. Tonneau & N. S.
695 Thompson (Eds.), *Perspectives in Ethology: Vol. 13. Perspectives in Ethology: Evolution, Culture,*
696 *and Behavior* (Vol. 13, pp. 155–180). Boston, MA: Springer. https://doi.org/10.1007/978-1-4615-
697 1221-9_6