Distribution Agreement

Signature:

____________________________ ______________
John P. Berg Date:

Implementation of Stimulus Control in a Computational Model

By

John P. Berg
Doctor of Philosophy

Psychology

Jack J McDowell, Ph.D.
Advisor

Chris Crowe, Ph.D.
Committee Member

Eugene Emory, Ph.D.
Committee Member

Joseph Manns, Ph.D.
Committee Member

Darryl Neill, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date:

Implementation of Stimulus Control in a Computational Model

By

John P. Berg
B.S., Johns Hopkins University, 2003
M.A., Emory University, 2009

Advisor: Jack J McDowell, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2011

Abstract

Implementation of Stimulus Control in a Computational Model

By John P. Berg

Reacting appropriately under different stimulus conditions is crucial for live organisms to improve their chances of survival. A computational model of behavior based on selection by consequences originally proposed by McDowell (2004) has been successful in producing a variety of behaviors consistent with live organisms (McDowell, 2004; McDowell & Caron, 2006; Kulubekova & McDowell, 2007; McDowell, Caron, Kulubekova, & Berg, 2008). However, previous versions of this model have been limited in that there is no method to change behavior in reaction to different stimulus conditions. The purpose of the current project was to modify the McDowell model to enable it to react differently under different stimulus conditions. Changes were made to the model that enabled variation in behavior across stimulus conditions using a conditioned reinforcement approach. Implementing the Rescorla-Wagner model to determine conditioned stimulus strength and therefore the degree of conditioned reinforcement, two-link, single alternative chained schedules of behavior were arranged in a series of experiments. Correspondence with live organism behavior was determined using qualitative and quantitative methods. Specifically, equilibrium values of the conditioned stimulus strength were evaluated for correspondence with the original and exponentiated versions of the Mazur hyperbolic discounting function, modified for use with variable interval schedules. Behavior on the chained schedules was evaluated qualitatively for consistency with previously published live organism data and was evaluated quantitatively in a replication of an experiment by Royalty, Williams, and Fantino (1987), which implemented a pre-reinforcement delay. Results indicated correspondence between the Rescorla-Wagner model and the exponentiated Mazur function. Chained schedules of behavior were largely consistent with live organism data. However, the Royalty et al. experiment was not successfully replicated. The results indicated that the Rescorla-Wagner model and the Mazur function provide a complete model of conditioned reinforcement. Using this model, chained schedules of behavior were successfully produced using the McDowell computational model. However, some behavioral phenomenon with pre- reinforcement delays could not be produced using the currently proposed computational methodology.

Implementation of Stimulus Control in a Computational Model

By

John P. Berg
B.S., Johns Hopkins University, 2003
M.A., Emory University, 2009

Advisor: Jack J McDowell, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2011

Acknowledgements

A dissertation, though often toiled over alone and frequently late at night, is not successfully produced in isolation. I would like to thank my advisor, Jack McDowell, for his support and guidance in completing this project; my laboratory mates Nick Cannon, Andrei Popa, Saule Kulubekova, and Marcia Caron for their advice and assistance; my parents for encouraging me to be inquisitive and skeptical; and my extended family who watched children and provided unyielding cheerleading during this process. To my wife and children, I owe the most thanks because they have seen me through this process hour after hour, day after day, and year after year. They have sacrificed more than they should, but were always there for me. To my wife, Britt Berg, I am thankful for your support, patience , and love. To my children, Zoe, Anneka, and Lena, thank you for showing me what is truly important each day.

Table of Contents

Tables

Figures

percent variance accounted for (pVAF) as polynomial order approaches the number of datapoints.

Implementation of Stimulus Control in a Computational Model

The desire to simulate realistic behavior using man-made machines is now a mainstay of modern culture. One can see in various films such as 2001: A Space Odyssey and iRobot that the idea of a computer behaving in human-like ways holds great appeal. Attempts to produce a machine with realistic behavior have often relied on a top-down approach. That is, the machine is built with the end-state behavior in mind and programmed specifically to produce that end-state behavior. An alternative approach, a bottom-up design where low-level rules generate “behaviors” with no end-state specified, has multiple advantages including the ability to test dynamic theories of behavior. In other words, these machines or programs can evaluate theories of why behavior occurs, not just how it looks in the end.

McDowell (2004) proposed a computational model of behavior that does not specify end state behavior. Instead, the model employs only low level rules that govern the workings of the model. Behavior develops dynamically from these low level rules. While this might seem risky because the end-state cannot be predicted, it allows for evaluation of basic theories of behavior because the basic rules of these theories can be programmed into the model. The behavior produced by the model can then be evaluated for consistency with live organism data. The McDowell computational model specifically tests the prediction that behaviors “evolve” in response to environmental pressures, much like the evolution of whole species in response to environmental forces (see Edelman, 1987; Skinner, 1981). The difference is in the time-scale: selection of fitter behaviors occurs within an organism’s lifetime while selection of a species’ characteristics occurs across multiple generations.

Tests of the McDowell computational model have revealed robust similarities between simulated and live organism behavior. McDowell (2004) and McDowell and Caron (2007) found that the model produced behavior on single alternative reinforcement schedules that is consistent with matching theory (Baum, 1974), which is a mathematical account of behavior at equilibrium. Kulubekova and McDowell (2007) analyzed the bouts of responding produced by the model and found that the interresponse times (IRTs), and their distributions, were consistent with those observed in live organisms. Finally, McDowell, Caron, Kulubekova, and Berg (2008) found that the model could produce behavior on concurrent schedules of reinforcement that is consistent with matching theory as well.

The current McDowell computational model does not, however, account for behavior under differing stimulus conditions. The model responds to the varying reinforcement contingencies, but there is no method to respond differentially to varied stimulus conditions. The current project’s purpose was to modify the McDowell model to enable this functionality, and to test this modified model’s behavior for agreement with live organism data.

How the McDowell Model Works

The computational model of selection by consequences proposed by McDowell (2004) consists of a virtual organism with 100 potential behaviors represented by integers between 0 and 1023. The 100 integers make up the virtual organism’s repertoire of potential behaviors. The range of 0 to 1023 integer values are split into several sub-ranges. Each sub-range represents a behavior class. During each iteration of the model, one behavior is randomly picked from the 100-behavior repertoire to be emitted. If the behavior's integer value falls within a specific behavior class, that class is considered activated. This is analogous to a rat pressing a lever, for example. Because emitted behaviors are randomly selected at each iteration of the model, the chance of a behavior class being activated at each iteration is dependent on the number of behaviors currently in that class. For example, if 40 out of the 100 behaviors in the repertoire fell within a particular class during an iteration of the model, the probability of the class being activated during that iteration would be 40/100 or 0.4.

One or more behavior classes are selected as target classes. Behaviors emitted from target classes are reinforced on random interval (RI) schedules by means of a genetic algorithm selection procedure, which has the effect of shifting the distribution of behaviors towards or into the target class. The selection procedure typically employs a linear fitness function to select "fitter" behaviors, which are defined as those closer to the target behavior class, to "mate" and produce a related "offspring" behavior. During iterations where behaviors outside a target class are emitted, or when a target-class behavior is emitted but reinforcement is not available as determined by the RI schedule, parents are chosen randomly.

Mating is accomplished by first choosing two parent behaviors to mate either by means of the selection function or randomly. The parent behavior integers are converted to ten-position binary form, which is a method of representing numbers with only 1's and 0's. For example, the integer 25 in ten-position binary form is 0000011001 while 500 is 0111110100 and 724 is 1011010100. An "offspring" behavior is created by randomly selecting which parent's bits (i.e. 1's and 0's) will be passed onto the offspring for each position in a ten-position binary number representation. The mating process is repeated until 100 new “offspring” behaviors, representing the new virtual organism behavior repertoire, are created. Overall, this method of mating produces “offspring” that resemble, but are not identical, to their parent behaviors (McDowell, unpublished data). In a final step, the new behaviors in the repertoire are subjected to mutation. Mutation is performed by randomly selecting a percentage of the behaviors from the new repertoire and changing their integer values by picking a new integer value at random from a Gaussian distribution with a mean set to the original, unmutated behavior integer.

One iteration of the McDowell computational model consists of the emission of a behavior, possible reinforcement, parental selection, mating, and mutation. In contrast to other applications of genetic algorithms, the McDowell model does not specify an end state or attempt to maximize any quantity. The model produces behavior based only on the simple rules inherent in the model.

A comprehensive account of adaptive behavior: from neurons to behavior.

Although the McDowell computational model produces behavior on a global scale (outwardly observable) based on the evolutionary principles of selection by consequences, McDowell (2010) proposed that the model fits within a larger framework of evolutionary processes at work within biological organisms. McDowell noted fundamental similarities between the selection by consequences of behavior and Gerald Edelman’s theory of neuronal group selection (TNGS), which posits that evolutionary processes are at play on a neuronal scale within biological organisms (Edelman, 1978). Like the evolutionary theory of behavior dynamics (ETBD; i.e., selection by consequences of behavior), Edelman’s TNGS proposes that evolutionary processes occur within an organism's lifetime. Instead of individual organisms being selected for fitness (traditional view of evolution) or individual behaviors being selected (behavior dynamics/selection by consequences), neurons or groups of neurons are strengthened (selected) if activation of these neurons produces adaptive behavior (Edelman, 1987). End-state neuronal activation and, thus, behavior are not specified from these rules that simply dictate neuron strengthening processes. End states emerge from underlying processes, making TNGS (like behavior dynamics) a bottom-up rather than top-down theory. This approach to brain functioning differs from other, contemporary theories of neuroscience that typically examine specific tracks of neurons. The TNGS represents a “conceptual mode [or] way of looking” (Edelman, 1992) at brain functioning that focuses less on specific circuits and more on global brain functioning (McDowell, 2010). This approach has advantages in that it attempts to explain brain functioning at a fundamental or more general level than approaches that focus on the function of anatomical structures.

The TNGS provides an explanation for how stimulus control develops via brain functioning. The TNGS proposes that stimulus control occurs when neuronal groups associated with stimulus events and neuronal groups associated with behaviors are connected. These connections are termed “reentrant signaling” and are proposed to be present in all conditioning. If the emitted behaviors produce adaptive behavior, synaptic strengthening of the neuronal groups connected by reentrant signaling occurs making behavior under the stimulus conditions more likely to occur in the future (McDowell, 2010). While the concept of reentrant signaling does provide an explanation for how stimulus control develops, it does not specify the dynamics of equilibrium conditions.

McDowell (2010) proposed that, taken together, the TNGS and ETBD constitute a framework for understanding the behavior of organisms from brain functioning to behavior. Although this comprehensive account of organism function and behavior has not been directly tested, the ETBD and TNGS have been individually evaluated and found consistent with observations in live organisms. The live organism evidence supporting the ETBD has been noted previously. The best evidence to support the TNGS would entail detailed measurements of neural activation while an organism interacts with its environment, which is not possible utilizing current technology. However, McDowell (2010) noted several indirect lines of evidence that support the TNGS. The first is the observation of various phenomena that are consistent with the TNGS, including the existence of value systems (i.e., dopaminergic and noradrenergic systems) that affect synaptic strength. The other lines of evidence come from computational and mechanical implementations of the TNGS that have shown: 1) synthetic neural networks utilizing TNGS principles function similarly to neural activity in live organisms and 2) mechanical agents operating with TNGS principles behave similarly to live organisms (See McDowell [2010] for additional details and references.).

Despite the possibilities and potential of a comprehensive, multi-level selectionist model of brain functioning and behavior represented by the ETBD and TNGS, the ETBD currently lacks methods to produce behavior that is responsive to different stimuli (as noted previously).

Chained Schedules of Reinforcement

The current project’s purpose was to give the McDowell model the ability to respond under different stimulus conditions. In other words, the purpose was to bring the model's behavior under stimulus control, which is a phenomenon consistently observed in live organisms. For example, if a rat's lever pressing only occurred when a green light was illuminated, then the lever pressing would be said to be under stimulus control.

Chained schedules of reinforcement are one method of testing stimulus control. The basic design of a chained schedule consists of two or more separate schedules of reinforcement that follow each other serially with each schedule being associated with a unique stimulus. Reinforcement in the initial links of the chained schedule consists of entry into the next link in the chain along with its associated stimulus. Reinforcement in the terminal link is the presentation of an unconditioned stimulus, which in live organisms might be something like food. Some chained designs have only one response option available at any time (simple chains) while others have two or more (concurrent chains). For the purposes of the current project, simple chained schedules were simulated with the intent that future projects will examine other, more complex, types of chained schedules.

Chained Behavior in Live Organisms and Associated Theories

In reviewing the existing literature, several phenomena are clearly seen in live organisms responding on chained schedules. First, the amount of responding in the initial link in the chain is less compared to the amount of responding in the terminal link (the one immediately preceding primary reinforcement / unconditioned stimulus) (c.f., Fantino & Romanowich, 2007; Ferster & Skinner, 1957). This makes sense given that the initial link is more removed from primary reinforcement than the terminal link. However, a number of explanations for why this occurs have been proposed.

The traditional view has held that stimuli associated with terminal links in a chained schedule gain conditioned reinforcement value as they are paired with primary reinforcement (Williams, 1994). The behavior in initial links is reinforced when a conditioned reinforcer is presented. This implies that, for example, a rat responding in a chained schedule is rewarded when a light associated with primary reinforcement/unconditioned stimulus is illuminated. In theory, a stimulus with conditioned reinforcement value could eventually beget more conditioned reinforcers in links further removed from the primary reinforcement if these stimuli were subsequently paired with an existing conditioned reinforcer stimulus (Williams, 1994). In this way, chains of behavior with a number of links should be able to be established.

An alternative approach has been proposed by Staddon and Cerutti (2003) whereby responses in initial links are proportional to the time until primary reinforcement occurs and are independent of link stimuli. Thus, behavior depends on the passage of time, not on the presentation of a stimulus with conditioned reinforcement value. This approach has been developed with, and primarily applied to, chained schedules that employ fixed interval (FI) schedules in all links of the chain. Given this restriction the relationship proposed is:

$$t_1 = aI_1 + aI_2, \tag{1}$$

where t_1 is the time to the first response in the initial link of a two link chain, a is a proportionality constant, and I_1 and I_2 are the durations of the initial and terminal links in the chain, respectively (Staddon & Cerutti, 2003).

In an experiment replicated by the current project, Royalty, Williams, and Fantino (1987) found evidence that directly contradicted the time to reinforcement approach. The experiment was designed to test whether conditioned reinforcement or time to reinforcement is the primary determinant of chained schedule behavior. Pigeons were trained to respond on a three component chain. In one condition, the reinforcement schedule (VI 33) was the same across all components. In a second condition, the initial and middle links were also VI 33 schedules. However, the transition from the middle to the terminal link was delayed by 3s after the appropriate response was emitted. In other words, once the response that would normally transition the changeover to the terminal link stimulus was registered, a 3s timer was started. After the 3s had elapsed, the terminal link stimulus was illuminated and its reinforcement schedule started. To maintain constant time-to-reinforcement across conditions, the middle link reinforcement schedule was reduced to a VI 30. In a third condition, the 3s delay occurred after the initial link. Royalty et al. found the response rates in the links immediately preceding the 3s delay were 59% lower than those where no delay was used. If time to reinforcement had been the primary factor affecting behavior, the introduction of the delay should have had no effect because the time to reinforcement was constant across all experiment conditions.

Given the wide acceptance of conditioned reinforcement as a major factor in determining behavior in chained VI schedules, the current project employed a conditioned reinforcement approach to simulate chained schedules.

The Modified McDowell Computational Model

To enable the McDowell computational model to respond realistically in a chained schedule (i.e., under differing stimulus conditions), a variation of the methodology proposed by McDowell, Soto, Dallery, and Kulubekova (2006) was used.

In the current project, each stimulus condition (i.e., link in the chain) was associated with a unique behavior repertoire. The behaviors in each stimulus condition's repertoire were selected and evolved separately from the others. However, conditioned reinforcement was introduced to allow stimuli that are paired with primary reinforcement to gain reinforcement value themselves. Therefore, behavior in an initial link repertoire was reinforced by movement into a terminal link repertoire that was associated with primary reinforcement. Behaviors that resulted in primary reinforcement were selected using the current McDowell methodology, and behaviors that resulted in conditioned reinforcement were selected using a modified methodology that employed the Rescorla-Wagner or a related function to determine the magnitude of reinforcement. This approach, by definition, implemented a conditioned reinforcement approach to chain schedules. The model was programmed for this approach and not the time-to-reinforcement approach advocated by Staddon and Cerutti (2003), which would have required an entirely different build of the computational model. Therefore, the current project had the capacity to test the conditioned reinforcement model, but not the time-to-reinforcement model.

Figure 1 illustrates the progression of reinforcement and conditioned reinforcement in a chained schedule in the modified McDowell model. In the left panel (Time Step #1), no primary or conditioned reinforcement has occurred. The behavior repertoires associated with the circle stimulus (top panel) and square stimulus (middle panel) are randomly distributed. Therefore the chance of a behavior being selected from any particular class is at its baseline level. One class of behavior (shaded) in the circle stimulus will result in the organism moving into the square stimulus. The chance of emitting a behavior from this class initially is low. The same is true for the class of behavior under the square stimulus during Time Step #1.

In Time Step #2, the chance of emitting a behavior under circle conditions from the class that will move the organism/model into square stimulus conditions is again low because it has not been reinforced. However, the chances of the model emitting a behavior in the class that resulted in primary reinforcement under square stimulus conditions during Time Step #1 are greatly increased in Time Step #2 because it was reinforced in Time Step #1. Also, the square stimulus has gained value as a conditioned reinforcer because it was paired with primary reinforcement, although this is not shown in Figure 1. This means that behavior in circle conditions that is followed by the conditioned reinforcer, the square stimulus, will receive reinforcement now. Because the value of conditioned reinforcement is less than the value of primary reinforcement, the selection event is weaker. This weaker selection effect can be observed by comparing the distributions of behaviors in the top and middles panels of Time Step #3 in Figure 1. The top panel (circle stimulus) has a more varied distribution of behaviors while the middle panel has a more concentrated distribution centered around the class of behavior that has been reinforced with primary reinforcement. Thus, the chance of emitting a behavior within the target class in the circle stimulus repertoire in Time Step #3 is lower than the chance of emitting a behavior in the square stimulus in Time Step #3 because the circle stimulus has been reinforced with conditioned reinforcement and the square with primary reinforcement. Despite having less concentrated behavior in the circle stimulus, the chances of a behavior being emitted from the class resulting in movement to the square stimulus are greatly improved compared to Time Step #2. Over these three example time steps, the model has “learned” to behave in ways that will increase the chances of obtaining primary reinforcement even across a more complicated chain of behavior.

Conditioned Reinforcement

Just as theoretical accounts exist for both static (i.e., matching theory) and dynamic (i.e., selection by consequences) properties of behavior, conditioned reinforcement can be examined on both levels of analysis. Static theories of conditioned reinforcement, simply put, describe the outcome of conditioned reinforcement after behavior has reached equilibrium, or steady state. In other words, they describe what can be observed in live organisms after behavior has developed and become stable; however, they do not provide an account of why these outcomes emerge. The questions of “why” and “how” certain phenomena develop are addressed by dynamic theories. Just as in evaluating dynamic theories of behavior, the computational environment provides an ideal method for evaluating dynamic theories of conditioned reinforcement, which can only be indirectly examined in live organisms. The computational environment can programmatically implement the simple rules of the dynamic theory and evaluate the results (or behaviors) that emerge. The current project used methodology similar to that used by McDowell and colleagues to implement dynamic theories of behavior programmatically and then evaluate them against static theory.

Dynamic theories of conditioned reinforcement.

Rescorla and Wagner (1972) proposed an iterated function system that describes how the associative strength of a stimulus (e.g., a light) develops in response to pairings with an unconditioned stimulus (e.g. food). The Rescorla-Wagner iterated function system is given by:

$$V_t = V_{t-1} + \alpha\beta(\lambda - V_{t-1}) \ , \tag{2}$$

where V_t is the associative strength of the conditioned stimulus (CS: in Pavlovian conditioning terms) at time t, V_{t-1} is the conditioned stimulus associative strength at the previous time step, α is the salience of the stimulus, β is the salience of the primary reinforcer, and λ is the ultimate level of associative learning the stimulus will support. See Table 1 for a description of parameters in this and other equations. Both β and λ take on different values when primary reinforcement occurs compared to when no reinforcement occurs. Typically, the value of λ is set to 1 during reinforcement and to 0 when no reinforcement occurs. The salience parameter, β, typically takes on values close to 1 (i.e., 0.7 to 0.9) when primary reinforcement occurs, which reflects the salience of reinforcer presence. In contrast, when no reinforcement occurs, the value of β represents the salience of reinforcer absence. Larger values of β (without reinforcement) will reduce the strength, V, quickly during unreinforced trials while smaller values will reduce the strength more gradually. Because β and λ have these alternating values, V increases when reinforcement occurs and decreases when it does not. It is important to note that Equation 2 is solely a dynamic model and does not predict an end-state. Equilibrium conditions can only be determined through iteration of the model.

Equation 2 or the Rescorla-Wagner (R-W) model, as it is commonly known, has been highly influential in the field of learning theory and in psychology in general (Seigel & Allan, 1996). The model has been successful in qualitatively accounting for a number of Pavlovian learning phenomena including acquisition and extinction of conditioned responding, generalization, discrimination, conditioned inhibition, patterning, and overshadowing, among others (Miller, Barnet, & Grahame, 1995). Although other, competing models of the associative strength of stimuli have been proposed since Rescorla and Wagner (1972) (i.e., Pearce, 1987), their model and variations of it have successfully accounted for experimental findings even in comparison to more recent models (Bahcekapili, 1998; Myers, Vogel, Shin, & Wagner, 2001).

Static theories of conditioned reinforcement.

The relationship between conditioned reinforcement and primary reinforcement has been addressed by a number of theories. Among these, hyperbolic discounting theory has received strong support and provides an account of conditioned reinforcer strength at equilibrium, that is, in the steady state (Mazur, 1997).

Although a number of researchers have proposed a hyperbolic relationship between a conditioned reinforcer's strength and the time until primary reinforcement, Mazur (1984) formalized the relationship by proposing the function:

$$V = \frac{a}{1 + bx} \tag{3}$$

where V is the value of the stimulus as a conditioned reinforcer at equilibrium, x is the latency of primary reinforcement following the onset of the stimulus, and a and b are parameters of the equation (Mazur, 1993). McDowell, Soto, Dallery, and Kulubekova (2006) showed that for random interval (RI) schedules, where there are many values of x, the equivalent form of Equation 3 is:

$$V = a\left(\frac{r}{b}\right)e^{\frac{r}{b}}\Gamma\left(0, \frac{r}{b}\right) \tag{4}$$

where *V*, *a*, and *b* are functionally the same as in Equation 3, *r* is the rate of reinforcement, and Γ is the incomplete gamma function, which is a continuous extension of the factorial. Equation 4 states that *V* increases as a function of *r* and has the general form of an exponentiated hyperbola in a reasonable parameter space (McDowell, unpublished data). In practice, Equation 4 predicts that a stimulus associated with an RI schedule will gain greater conditioned reinforcement value the greater the reinforcement rate delivered by the RI schedule (i.e., the smaller the average time between the onset of the discriminative stimulus and the delivery of primary reinforcement).

Rachlin (1989) proposed that an additional scaling factor be added to the Mazur function in the form of an exponent for the denominator, *s*, given by,

$$V = \frac{a}{(1+bx)^s}. \tag{5}$$

Although Mazur (1987) found that such an exponent did not improve fits to pigeon delay discounting data, Green, Fry, and Myerson (1994) found that the exponentiated form better accounted for human child delay discounting data. In a review article, Green and Myerson (2004) found that for human delay discounting data across a wide variety of ages, the exponentiated Mazur function provided a better fit to data with $s < 1$ in most cases. Using techniques similar to McDowell et al. (2006), an equivalent function for use with random interval (RI) schedules (i.e., *x* varies widely) was derived from Equation 5. This function is given by:

$$V = a\left(\frac{r}{b}\right)^s e^{\frac{r}{b}}\Gamma\left(1-s,\frac{r}{b}\right), \tag{6}$$

where *V*, *a*, and *b* are functionally the same as in Equation 3, *r* is the rate of reinforcement, and $\Gamma\left(1-s,\frac{r}{b}\right)$ is the incomplete gamma function, evaluated between different limits. Generally, Equation 6 gives a similar result to Equation 4. In fact, when $s = 1$, Equation 6 reduces to Equation 4. However, the plot of Equation 6 contains slight differences from Equation 4 when $s \neq 1$. It is important to note that the incomplete gamma function will return complex numbers for some values of s and r/b. Figure 2 shows a 3D plot of the exponentiated Mazur function for different values of s and r. Of note, the areas where the plot is missing or empty contain complex numbers. To obtain a real value of the function in the domains where the incomplete gamma function returns complex numbers, the imaginary parts of the returned values must be ignored. For the purposes of the current project, this method was employed to obtain a full range of exponentiated Mazur function values.

Neurobiological bases of conditioned reinforcement.

Although neither the presently discussed dynamic or static models of conditioned stimulus/reinforcement strength are based on neurological functioning, findings of brain functioning in relation to conditioned reinforcement can inform model evaluation and suggest future directions for research. A consistent finding from neurological data of conditioned reinforcement is that behavior associated with initial learning activates different brain structures than behavior under stimulus control. Learning that occurs early in stimulus-pairing conditioning has been primarily associated with activation in the ventral striatum region. As behavior comes under stronger stimulus control, the dorsal striatum region becomes more active (for a review of these findings, see Graybiel, 2008). The core of the nucleus accumbens (NAcc) also appears to have a critical role in learning and reacting to conditioned stimuli. Lesions and inhibition of neural activity in the NAcc core restricted behavior supported by conditioned reinforcement relative to pre-lesion behavior and NAcc shell-lesioned animals (Parkinson et al., 1999; Di Ciano et al., 2001). The NAcc core has also been shown to affect delay to reinforcement effects (Hutcheson et al., 2001). These findings appear to correspond with some theories (i.e., Staddon & Cerutti, 2003) that have proposed that the power of conditioned reinforcement stems from an ability to signal future primary reinforcement.

Correspondence of R-W and Mazur Models

Given that the Rescorla-Wagner function can be used to predict how the strength of a conditioned stimulus develops and the Mazur function (Equations 4 or 6) predicts the strength of a conditioned stimulus after it has fully developed (i.e., equilibrium), data produced by the Rescorla-Wagner function at equilibrium should be described by the Mazur function. In other words, Rescorla-Wagner equilibria should trace out the Mazur value function. However, when the Mazur function is fitted to dynamically generated equilibria from the Rescorla-Wagner model fit residuals are non-random (McDowell, unpublished data). Although these results represent preliminary work, they indicate that either the Mazur function does not adequately explain the Rescorla-Wagner equilibria or, alternatively, the Rescorla-Wagner function does not produce equilibria comparable to live organisms. Given the consistent success of the Mazur function in describing conditioned reinforcement at equilibrium conditions and that the majority of data supporting the R-W model have stemmed from molecular (i.e., non-equilibrium) observations (Miller, Barnet, & Grahame, 1995), it appears more likely that the latter possibility is the reason because equilibrium data is more stable. An additional possibility is that the exponentiated Mazur function adequately describes Rescorla-Wagner equilibria. To date, this possibility has not been examined, but it is supported by findings suggesting that the exponentiated Mazur function provides a better fit with delay discounting data (i.e., Myerson & Green, 2004).

The current project investigated two potential solutions to the Rescorla-Wagner equilibria and Mazur function incongruence. The first was to fit the exponentiated Mazur function to Rescorla-Wagner equilibria. The second, discussed below, was to modify the Rescorla-Wagner function, giving it a scaling exponent.

A modified R-W function.

Preliminary work conducted for the current project identified a potential modification to the Rescorla-Wagner model that may make dynamic data produced by the model consistent with the original Mazur function. This new model keeps the existing structure of the R-W function but adds an exponent to ΔV, the part of the function that increments or decrements the conditioned reinforcement strength at each iteration. The new function form is given by:

$$V_t = V_{t-1} + [\alpha\beta(\lambda - V_{t-1})]^a. \qquad (7)$$

When a>1, this new function form provides a more gradual transition to maximum conditioned reinforcement strength (i.e., equilibrium conditions). Preliminary work has shown that the exponentiated R-W function has a smoother and more gradual transition to equilibrium conditions compared to the classic R-W function and, overall, produces more stable values of *V*.

Purpose of the Current Project

The primary purpose of the current project was to extend the McDowell computational model of behavior to stimulus control, and this was done via chained schedules. To accomplish this, and test the theories underlying the extensions to the model, three series of experiments were conducted.

The first, Experiment Series I, evaluated the relationship between static and dynamic theories of conditioned reinforcement. Specifically, the Rescorla-Wagner model (a dynamic theory of conditioned reinforcement) and the currently proposed exponentiated variant of that model were programmed into the McDowell computational model. The conditioned reinforcement strengths obtained from runs with this model were then compared to the Mazur function (and its exponentiated variant) which predicts the strength of a conditioned reinforcer at equilibrium.

The second, Experiment Series II, implemented a simple two-link chain schedule in the model using the Rescorla-Wagner model to determine the strength of conditioned reinforcement. The model was evaluated across a wide range of parameters and the resulting behavior was examined for consistency with live organism behavior.

The third and final, Experiment Series III, attempted to replicate the Royalty et al. (1987) experiment, which was designed to evaluate the extent to which behavior in chained schedules is consistent with conditioned reinforcement or time-to-reinforcement theoretical explanations. However, because the modified McDowell model was explicitly programmed for a conditioned reinforcement approach, the results from the current project can only determine the degree of consistency with this approach. Within the modified McDowell model, a delay was instituted in the initial link after a target behavior was emitted and scheduled for reinforcement. After the delay lapsed, the terminal link was presented. Results of this experiment series were compared to data published by Royalty et al.

In this dissertation, the general methods used to implement the extensions to the McDowell model are presented first. This is followed by three sections reporting results from Experiment Series I, II, and III. Each experiment series section contains Methods, Results, and Discussion subsections that are relevant to that experiment series. Finally, a General Discussion synthesizes the entirety of the current project's results and discusses their implications.

General Methods

Unless noted otherwise, the following methods were used in all project experiments.

Subject and Environment

The subject was a digital organism with behavior governed by principles of selection by consequences as specified in McDowell (2004). The organism was operated in a simulated environment consisting of two stimulus conditions chained together. The reinforced behaviors emitted from the target class in the initial link in the chain resulted in presentation of the terminal link stimulus, and transferred the action to the behavior repertoire associated with that stimulus. Behaviors emitted from the terminal link target class resulted in primary reinforcement using procedures identical to those used by McDowell (2004) followed by presentation of the initial link stimulus and action transfer to the repertoire associated with the initial link stimulus. The terminal link stimulus gained conditioned reinforcement value via the Rescorla-Wagner model or a variant thereof (i.e., Equation 2 or 7). Behaviors emitted in the initial link were reinforced using a modification of the McDowell (2004) procedures as described below.

Upon action being transferred to the terminal link, the behaviors in the initial link underwent modified reproduction rules as follows. Once a behavior in the target class was emitted, the fitness function was centered at the midpoint of the class and used to select behaviors for reproduction. However, to reduce the strength of reinforcement due to conditioned reinforcement, a "weakened" selection event was used where only a proportion of the next generation's (i.e., next time step) behaviors were produced by using the fitness function. The remaining behaviors were produced by randomly selecting parent behaviors. For example, of the 100 behaviors in the repertoire, 90 might be produced by selecting parents using the fitness function with the remaining produced by randomly selecting parents. In evolutionary terms, this is equivalent to a weaker selection event: only a subset of the population is affected by the selection. The percentage of behaviors undergoing selection by the fitness function during each conditioned reinforcement event was determined by using the conditioned reinforcer strength, V, to determine the proportion of behaviors produced using the fitness function. Thus, if the conditioned reinforcer strength was $V = 0.9$, then 90 behaviors for the next generation would be produced using a fitness function and the remaining 10 would be produced by selecting parent behaviors randomly.

Apparatus and Materials

Software to implement the digital organism and conduct the experiments was written in VB.net and run on computers using Windows XP, Vista, or Windows 7 operating systems. Computers had at least 1.7-Ghz processors with 1 Gb of RAM and 5 GB of hard disk space.

Computational Procedures for the Modified McDowell Model

The original McDowell (2004) selection by consequences algorithm was modified to allow for chained schedule experiments to be arranged as previously specified. The following steps were implemented programmatically to animate the model:

1. At the start of an experiment, initial model parameters were collected and the corresponding model variables were set accordingly.
2. At the beginning of an experiment, the initial link behavior repertoire was active.
3. A behavior was selected at random from the repertoire to be emitted.
4. If the emitted behavior came from the target class (the class of behavior allowing action to be transferred to the terminal link), the initial link schedule of reinforcement was consulted for the availability of reinforcement.

a. If reinforcement was available, then the conditioned stimulus strength (*V*) associated with the terminal link was evaluated. If $V = 0$, then parents were selected randomly (i.e., no selection occurred). If $V > 0$, a proportion of behaviors equal to *V* was produced using a fitness function as previously detailed. Recall that *V* varies between 0 and 1. The remainder of the behaviors were produced by randomly selected parents.

b. If reinforcement was not available, all parents were selected randomly.

5. Two parent behaviors were selected (using the method described in Step 4) and “mated” using bitwise reproduction as specified in McDowell (2004). This process was repeated until a new repertoire was produced.
6. The new repertoire underwent Gaussian mutation (standard deviation = 25) as specified by McDowell (2004).

7. If reinforcement occurred, the action was transferred to the terminal link repertoire. If not, the program returned to Step 3.
8. Once in the terminal link, a behavior was selected at random to be emitted.
9. If the behavior came from the target class designated to produce reinforcement, the schedule of reinforcement for the terminal link was consulted for the availability of reinforcement.

a. If reinforcement was available, parent behaviors were selected using the fitness function.

b. If reinforcement was not available, then parent behaviors were selected at random.

10. Two parent behaviors were selected using the method specified in Step 9 and “mated” using “bitwise” reproduction as specified in McDowell (2004). This process was repeated until a new repertoire is produced.
11. The new repertoire underwent Gaussian mutation (standard deviation = 25) as specified by McDowell (2004).
12. The strength of the conditioned stimulus, V, was calculated using Equation 4 or 6 (depending on the model being evaluated).
13. If reinforcement occurred, the action was transferred back to the initial link. If not, the program returned to Step 8.

At each iteration of the model, relevant behavior data and model parameters were stored for later analysis. Specifically, the behavior that was emitted, whether reinforcement occurred, and the conditioned reinforcement strength for the terminal link were collected and stored in a file for later analysis.

Simulations employed random interval (RI) schedules of reinforcement. Consistent with previous work by McDowell and colleagues, a range of RI schedules was employed for reinforcement in the links of the chain. Specifically, RI 1, 2, 3, 5, 8, 10, 18, 25, 68, 112, and 200 schedules were employed in each condition where relevant. Schedule timing was specific to each link meaning that the RI schedule timer stopped when action was transferred out of each link.

Mutation rate was fixed at 10% for all conditions in both chain links. A linear parental selection function as specified by McDowell (2004) was employed with a mean of 40. The target class of behavior for both the initial and terminal links was the range 0 to 40, inclusive, out of the 0 to 1023 integer range. Baseline rates of unreinforced target behavior emission (i.e., operant level of behavior) are approximately 4% given this range. McDowell and colleagues (McDowell, 2004; McDowell, Caron, Kulubekova, & Berg, 2008) and Popa (2009) have found that these values produce behavior that is consistent with the behavior of live organisms responding on single and concurrent schedules of reinforcement. For each combination of factors, the model was run for 20,000 cycles to obtain sufficient data for analysis.

Project Structure

Three series of experiments were conducted to examine the different components of the modified McDowell model and to test the model for correspondence with theory and previously published live organism data.

Experiment Series I.

In this series, correspondence between dynamically produced values of the conditioned stimulus strength, *V*, produced by the Rescorla-Wagner function (or the exponentiated variant) and Mazur's delay discounting theory (original and exponentiated variant) was examined.

Experiment Series II.

This series examined two-link chained schedules of behavior produced by the modified McDowell model. In particular, inter-link dependent behavior was examined across a variety of model parameters. Additionally, the new technique of only producing a proportion of behaviors using a fitness function (with the others being produced by randomly selecting parent behaviors), was evaluated to ensure that this technique was consistent with matching theory.

Experiment Series III.

This series attempted to replicate the Royalty et al. (1987) experiments. Data from experiments implementing a pre-reinforcement delay and those without such a delay was compared to the live organism data published by Royalty et al.

Experiment Series I: Rescorla-Wagner Model and Mazur Function Correspondence

Methods

The subject, environment, apparatus, materials, and computation procedures detailed in the General Methods were used for Experiment Series I.

To evaluate the correspondence between the Rescorla-Wagner model (or its exponentiated variant) and the Mazur function, the average values of conditioned reinforcer strength, V, produced by the modified McDowell were obtained for 11 different RI schedules (RI 1, 2, 3, 5, 8, 10, 18, 25, 68, 112, and 200) by calculating the average value of V across all time-points. This allowed the relationship between r, the reinforcement rate generated by the RI schedule, and the average V to be evaluated using the Mazur function.

The r-V relationship was examined across multiple model conditions by systematically varying the Rescorla-Wagner parameters (α, β_1, β_0, and a). While each parameter of interest was varied, the other parameters were held constant. To evaluate a wide variety of model conditions, several combinations of parameters were used across experiments. Table 2 shows the Rescorla-Wagner parameters that were varied, the range of values tested, and the values of the other parameters.

For each set of r-V (reinforcement rate – conditioned reinforcer strength) datapoints, plots were created, and the Mazur and exponentiated Mazur functions, Equations 4 and 6, were fitted to the data. Equation 4, the Mazur function, was fitted using the Excel Solver add-in, and Mathematica Link for Excel was used to resolve the value of the gamma function, $\Gamma\left(0, \frac{r}{b}\right)$. Equation 6, the exponentiated Mazur function, was fitted using the FindFit function in Mathmematica, because complex numbers were generated by some solutions due to the incomplete gamma function, $\Gamma\left(1 - s, \frac{r}{b}\right)$. Goodness of fit was determined by the percentage of variance accounted for (pVAF) by Equation 4 or 6, and by residual analyses, which are described in detail in the Appendix.

Results

Effects of traditional Rescorla-Wagner parameters on conditioned stimulus strength.

Figure 3 shows the relationship between the reinforcement rate and the conditioned stimulus strength for 3 different values of α. As can be seen in Figure 3, there was no virtually no effect of the stimulus salience, α, on the relationship between the reinforcement rate and conditioned stimulus strength (r-V) at equilibrium. In other words, each value of α produced the same r-V profile. Although Figure 3 only shows the results from one set of model conditions, this invariance in the equilibrium values of V for different values of α was observed in all model conditions (i.e., different values of β_0, β_1, and, a). However, Figure 3 only shows the asymptotic values of V. It does not illustrate the molecular level or moment-by-moment values of V. Figure 4 shows values of V on the molecular time-scale: the first 500 time-ticks for three different values of α. Although the time at which the value of V begins to increase varies between each plot, this was likely due to normal variability in the model (i.e., the time to the first reinforcement event varied across the experiments). A close examination of the plots indicated that the value of V decreases faster at higher values of α. This, in turn, resulted in more variability in the values of V. To further examine this effect, the mean and standard deviation for V were calculated for the same three experiments. Figure 5 shows these values. Although the mean of V was unchanged across the three values of α, the standard deviation increased as α increased. These observations confirm that while α did not affect the equilibrium or mean value of V, it did affect the variability of V, with higher values of α producing more variability in V.

The effect of β_0, the salience of reinforcer absence, on the relationship between reinforcement rate and the conditioned reinforcement strength, V, can be seen in Figure 6. Although Figure 6 only shows the results from varying β_0 under one set of model conditions, similar results were obtained for all other model conditions. Compared to α, β_0 had a marked effect on V with lower values of β_0 producing steeper r-V (higher levels of *V* with lower reinforcement rates). Additionally, the *r*-*V* relationship varied greatly depending on the value of β_0. A wide range of *r*-*V* profiles was produced by varying the value of β_0. Although not shown in the figure, the relationship between *V* and β_0 was consistent across differing values of β_1 in that a wide range of *r*-*V* profiles were obtained from varying β_0 across differing values of β_1.

The effect of β_1, the salience of reinforcer presence, on the conditioned reinforcement value *V* can be seen in Figure 7. Although Figure 7 only shows the results from varying β_1 under one set of model conditions, similar results were obtained for all other model conditions. Similar to β_0, differing values of β_1 produced marked changes in the *r*-*V* relationship. Higher values of β_1, produced steeper *r*-*V* curves (i.e., higher levels of *V* were observed for lower rates of reinforcement). Notably, a wide range of *r*-*V* curves was obtained indicating that β_1 had a significant effect on the conditioned reinforcement strength. The range of *r*-*V* profiles obtained by varying β_1 depended on the value of β_0, the salience of reinforcer absence. For lower values of β_0 (i.e., 0.01), only a restricted range of *r*-*V* profiles were obtained, as shown in Figure 8, while for higher values of β_0 (i.e., 0.05), wider ranges of values were obtained, as shown in Figure 7. This is likely because β_0 and β_1 produce opposing effects in the R-W equation with β_0 being responsible for the amount of decrease in *V* following no reinforcement and $\beta_1$being responsible for the amount of increase in *V* following reinforcement.

To summarize: the findings from modifying the parameters of the traditional Rescorla-Wagner model showed that the stimulus salience, α, had little effect on asymptotic or equilibrium values of the conditioned stimulus strength, *V*. However, α did have an effect on how much the conditioned stimulus strength varied, with higher values of α producing more variability. In contrast, both β_0 and β_1 (the salience of reinforcer absence or presence, respectively) produced wide ranges of conditioned stimulus strengths. Lower values of reinforcer-absence salience produced higher asymptotic values of conditioned reinforcer strengths, whereas lower values of reinforcer-presence salience produced the opposite effect: lower asymptotic values of conditioned reinforcer strength. Finally, there was an interaction between the salience of reinforcer absence and presence with lower reinforcer-absence values restricting the effect of the reinforcer presence. Taken together, these results indicate that 1) the Rescorla-Wagner model was successfully implemented in the McDowell model, and 2) a range of conditioned reinforcer strengths could be produced within the model.

Effect of the currently proposed Rescorla-Wagner exponent, *a*, on conditioned reinforcer strength.

The effect of varying the exponent, *a*, in the modified R-W equation can be seen in Figure 9 under the model conditions $\beta_0 = 0.05$, $\beta_1 = 1.0$, and $\alpha = 0.7$. Increases in *a* produced higher values of *V*, but this effect was more pronounced at lower rates of reinforcement. At the highest rates of reinforcement, the values of *V* obtained under different values of *a* were similar. This effectively increased the curvature of the *r*-*V* profile without changing the asymptote. Finally, the minimum value of *V* for each *r*-*V* profile increased with *a* indicating that higher values of *a* restricted the range of *V* to larger values.

The molecular view of the moment-by-moment values of *V* for differing values of *a* are shown in Figure 10, which shows plots of *V* for the first 500 time-ticks for three values of *a* (1, 1.5, 2). The plots indicate that exponent values greater than 1 decreased the variability in V, the conditioned reinforcer strength, markedly, and hence created a more stable value of conditioned reinforcement.

Correspondence of Rescorla-Wagner and Mazur functions.

The Mazur and exponentiated Mazur functions (Equations 4 and 6, respectively) were fitted to the average conditioned stimulus strength values generated by the Rescorla-Wagner model (Equation 2) for the parameter combinations listed in Table 2. The best-fit parameter values are shown in Table 3 (Effect of varying β_1), Table 4 (Effect of varying β_0), Table 5 (Effect of varying α), and Table 6 (Effect of varying a). Each table shows the Rescorla-Wagner parameters in the left hand column. The classic Mazur function fits are shown in the middle section, with the exponentiated Mazur fits shown in the far right section. For both classic and exponentiated fits, the values of R^2, which are the proportions of variance accounted for by the function, are given. The other indication of goodness-of-fit is the result of the residual analysis for each fit, namely whether a cubic polynomial trend was present in the residuals.

The best-fit parameter values for the experiments varying β_1 are shown in Table 3. A comparison of these parameter values indicated that a wider range of values for both a and b parameters was obtained for fits of the classic Mazur function than for fits of the exponentiated function. The values of a varied between 1.06 and 11.4 for fits of the Mazur function, while that range was much more restricted (1.00-1.72) for fits of the exponentiated Mazur function. The highest value of b for the exponentiated function was 19.6 while the classic Mazur function required a greater range of values (8.85-13,370). The values of the exponentiated Mazur scaling parameter, s, ranged from 4.06 to 34.1. These large values may explain why a large proportion of the exponentiated fits did not converge, which means that the iterated step size in the fit algorithm was larger than the precision of the fit algorithm itself. Although this precision was increased to a very high number of significant digits (a range of 40-80 digits was explored), these algorithm errors continued to occur. This was likely because large values of s produced very small values of the incomplete gamma function, $\Gamma\left(1 - s, \frac{r}{b}\right)$, and very high values of $\left(\frac{r}{b}\right)^s$ in the exponentiated Mazur function (Equation 6). It is also possible that the wide range of values of s resulted from an incomplete fitting process (i.e., the fitting algorithm did not identify a stable, best-fit solution) and, hence, there may be inconsistency across fit attempts.

Most importantly for the purposes of the current project, residual analyses of the classic Mazur function fits indicated the presence of non-random residuals for all R-W parameter combinations tested despite R^2 values all greater than 0.99. Figure 11 shows the residuals for all fits. Clear evidence of cubic trends was visible in all plots. In contrast, non-random residuals were only evident in 2 out of the 15 fits of the exponentiated Mazur function to the R-W generated data. Figure 12 shows the residual plots for all exponentiated Mazur function fits. Despite problems fitting the exponentiated Mazur function to the conditioned stimulus strength data noted previously, the fits that did not converge still had excellent goodness-of-fit statistics and no evidence of non-random patterns in the residuals.

Table 4 shows the parameters for fits to the experiments varying β_0. Similar to the results of β_1, the residuals from fits of the Mazur function showed evidence of non-random patterns. In comparison, the fits of the exponentiated Mazur function showed evidence of non-random residual patterns in only 1 of the 17 fits. Figures 13 and 14 show the residual plots for fits of the Mazur and exponentiated Mazur function, respectively. Clear cubic trends are visible in the Mazur fits, while exponentiated Mazur fits show a good scatter of residual values. Across both functions, increasing values of β_0 produced higher values of *b* and *a.* Similar to the series of experiments varying β_1, values of *b* were larger for the Mazur function fits. The values of *s* varied widely from 0.40 to 71.0 across the experiments that varied β_0. Two fits did not fully converge. It is notable that these fits produced the only fit with non-random residuals; the other fit produced the highest value of *s* of all experiments in the current project. For the exponentiated Mazur function fits, there were several cases where the FindFit function of Mathematica was unable to resolve a fit without throwing an error. These “indeterminate fits” were typically caused by problems with complex number calculations.

Table 5 shows the parameters for fits to the experiments varying α. The quality of Mazur and exponentiated Mazur function fits was similar to other results with the exponentiated Mazur function, showing no evidence of non-random residuals (see Figures 15 and 16 for residual plots). There were again, some problems fitting the exponentiated Mazur function to several datasets. The fits for the β_0: 0.05, β_1: 0.5 series illustrate some of the difficulties encountered when fitting the exponentiated Mazur function. The experiment with $\alpha = 0.5$ was fitted successfully, whereas the experiment with $\alpha = 0.7$ produced errors, and finally the experiment with $\alpha = 0.9$ produced values of *b* and *s* that were markedly divergent from the $\alpha = 0.5$ experiment. Considering that the *r-V* profile was almost identical for these three experiments, this variability may indicate that fits using the exponentiated Mazur function may be unstable for some datasets.

Finally, Table 6 shows the parameters for fits to the experiments varying a, the currently proposed Rescorla-Wagner exponent. (Note that the Mazur functions and the exponentiated R-W model contain a parameters, but these parameters are not the same. For the purposes of clarity in this section, the subscripts *"rw"* and *"m"* will be used to differentiate between the two.) Non-random residuals were present in all the classic Mazur function fits with the exception of $a_{rw} = 1.25$. The residuals were random for exponentiated Mazur function fits with $a_{rw} = 1$ and 1.25, but they were not random for higher values of a_{rw}. Increasing a_{rw} appeared to have several effects on the model parameters. First, the Mazur model parameters a and s decreased with increases in the R-W parameter, a_{rw}. Notably, values of $a_m < 1$were observed when no other fit in the current project produced values of $a_m < 1$. A similar result with s was obtained with increasing values of the R-W parameter, a_{rw}, producing values of $s < 1$, which was not observed in any other fits. Another unexpected result emerged when $a_{rw} = 1.25$. At this value, and under these conditions, the Mazur and exponentiated Mazur functions were markedly similar. Figure 17 shows the residuals for both the Mazur and exponentiated Mazur function fits. The pattern of residuals is almost identical between the two function fits to data produced with $a_{rw} = 1.25$. An examination of the fit parameters revealed that the value of s was very close to 1 ($s = 0.97$), indicating that the classic and exponentiated Mazur functions were nearly equivalent for this case. (Recall that the exponentiated Mazur function reduces to the classic Mazur function when $s = 1$.)

In summary, the current results showed that dynamically produced values of V were described well by the exponentiated Mazur function (Equation 6), but not by the classic Mazur function. Fit residuals from classic Mazur function fits to average values of

V produced by the Rescorla-Wagner function contained non-random patterns for all R-W parameter values except for $a_{rw} = 1.25$. Although non-random residuals were only present in 3 out of 35 fits using the exponentiated Mazur function, the function required sophisticated techniques to find the best fit, did not always converge despite these techniques, and required large values of s to provide the best fit, which has not been observed in live organism data to date.

Discussion

Effects of Rescorla-Wagner parameters on conditioned stimulus strength, *V*.

The results of varying R-W parameters demonstrated that the conditioned stimulus strength is affected in several ways by these parameters. The range of R-W parameters tested provided substantial detail about how each parameter affects V. While changes in α, the salience of the conditioned stimulus, produced little difference in equilibrium levels of V, it did affect the instantaneous values of V, with larger values of α corresponding to greater variability in V across the time series. In contrast, both β_0 and β_1 affected the asymptotic values of V. Larger values of β_0, the salience of reinforcer absence, resulted in shallower r-V curves or, in other words, decreased the sensitivity to reinforcement rates. Larger values of β_1, the salience of reinforcer presence, had the opposite effect: they generated steeper r-V curves or greater sensitivity to reinforcement. Additionally, the salience of reinforcer absence (β_0) restricted the range of r-V profiles obtained when varying the salience of reinforcer strength (β_1). This indicates that these two parameters interact to some extent. The final R-W parameter, a, newly proposed in the current project, affected the curvature of the r-V profile without changing the asymptote. It also affected the variability of V.

The association between the Rescorla-Wagner and Mazur models.

The current project constitutes the first effort to fully examine the relationship between the Rescorla-Wagner model and the Mazur function. Two dynamic models of conditioned stimulus strength and two static models of conditioned reinforcer strength were examined. The dynamic models consisted of the Rescorla-Wagner equation and an exponentiated variant, which are given by Equations 2 and 7, respectively. The static model consisted of the Mazur function and an exponentiated variant, which are given by Equations 4 and 6, respectively. Although the exponentiated R-W model was developed for the current project, the exponentiated Mazur function was proposed previously (Rachlin, 1989; Myerson, Fry, & Green, 1994).

The results of the current project consistently indicated problematic fits between the values of Rescorla-Wagner generated conditioned stimulus strength, V, and the Mazur function for random interval schedules, Equation 4. Although the Mazur function fit the r-V data relatively well, as evidenced by percent variance accounted for (pVAF) of 99% and higher, the fits consistently left non-random residuals. In contrast, the exponentiated variant of the Mazur function fit the data very well, with pVAF values consistently nearing unity. Analyses of the residuals revealed non-random trends in a small minority of the fits. Comparing the classic Mazur and exponentiated Mazur function residual plots showed clear evidence that the exponentiated variant provided a greatly improved fit to the data. Overall, the results of the current project indicated that the exponentiated version provided a very good fit to the conditioned stimulus strength values obtained via the Rescorla-Wagner function implemented in the McDowell computational model.

Some fits of the Mazur functions to average conditioned stimulus strength values using the currently proposed exponentiated Rescorla-Wagner model (Equation 7) showed evidence of good correspondence while others did not. For $a_{rw} > 1.25$, neither the exponentiated nor the classic Mazur function fit the conditioned reinforcement data produced using the exponentiated Rescorla-Wagner function. Fits of both functions to the data contained non-random patterns in the residuals. Additionally, increases in the exponent, a, resulted in decreased goodness-of-fit. Clear patterns in the residuals were visible at higher values of a_{rw} and the pVAF decreased as well. While these patterns were not unexpected for the classic Mazur function given the presence of non-random residuals in all other fits using this function, the clear non-random patterns in the residuals were not observed for any other condition fit by the exponentiated Mazur function. However, for the case where $a_{rw} = 1.25$, both Mazur functions appeared to fit the exponentiated Rescorla-Wagner data well. Both functions had high pVAFs and neither fit showed evidence of non-random residuals. It is notable that this was the only experiment where the classic Mazur function fit did not contain non-random residuals. Additionally, the best-fit parameters indicated that the two Mazur function forms were essentially equivalent. The value of the exponentiated Mazur scaling parameter, s, was near unity. Under these conditions, the exponentiated Mazur function reduces to the classic Mazur function. These results indicated that for most values the exponentiated Rescorla-Wagner model is not consistent with existing theory. The unique result when $a_{rw} = 1.25$ indicates that certain values of a_{rw} may represent special cases where data produced by the model is consistent with findings from live organisms.

It is important to note that the exponentiated version of the Mazur function that is adapted for random interval schedules contains the incomplete gamma function, which generates complex numbers for many parameter ranges. This resulted in many of the predicted values of *V* in the current project having both real and imaginary parts because of the complex nature of the incomplete gamma function. This in turn produced residuals with imaginary parts as well. For the purposes of the current project, imaginary parts of the predicted values of *V* were ignored in the analyses. For example: if $V_{\text{predicted}} = 0.8925 + .0000542i$, where *i* is the non-real representation of $\sqrt{-1}$, the entire value 0.0000542*i* was ignored. Only the real parts (i.e., 0.8925) were used in calculating pVAF and determining whether non-random patterns existed in the residuals. While it is not desirable to ignore parts of a function, the current results suggest that this approach produces a function that accurately predicts values of conditioned stimulus strength.

The results of the conditioned reinforcement analyses have important theoretical consequences. Although the general consensus in the discounting literature is that the relationship between reinforcer strength and delay to reinforcement is likely a hyperbolic one (Rachlin, 2006), the exact form has been debated extensively. Mazur (1987) originally proposed the relationship between conditioned reinforcer strength and time to primary reinforcement describe by Equation 3. Later, Rachlin (1989) proposed that exponentiating the denominator of Equation 3 would provide a scaling factor to better fit data under conditions of delay. The results from the current project clearly support utilizing the exponentiated version of the Mazur function based on residual analysis and pVAF. Interestingly, if pVAF were solely used to determine which function provided a better fit, the comparison between a function that accounts for 98.9% of variance and one with 99.9% would be difficult because both values indicate very high correspondence with data. However, this was the methodology used by previous research (e.g., Green, Myerson, & Ostaszewski, 1999; Rachlin, 2006) to determine goodness-of-fit. Given the reliance on pVAF to evaluate models in previous work, it is difficult to make definitive conclusions about the correspondence of the current model to live organism data. For example, Jones and Rachlin (2006) concluded that the original Mazur function provided a robust fit to human subject data while Green and Myerson (2004) concluded that the exponentiated version provided the best fit to a wide variety of human subjects data including children, and college aged and elderly individuals. However, the predictions resulting from theoretically based computational models such as the current one illustrate the benefit of a computational model. The current results indicate that the fits of the original Mazur function will contain non-random residuals while the exponentiated version will not. To date, this analysis has not been conducted with data from live organisms, but it is a prediction based on the current results.

An additional prediction based on the current results is that the exponentiated Mazur function for random interval schedules (i.e., Equation 5) corresponds to conditioned stimulus strength values and, by extension, to the response rates dependent on conditioned reinforcement. To date, this analysis has not been conducted with live organisms. Future studies could test this relationship in live organisms by varying the reinforcement rate of a behavior that is supported by primary reinforcement and paired with a discriminative stimulus. Then the response rates of a behavior supported by conditioned reinforcement via this stimulus could be compared with the reinforcement rates supported by the primary reinforcement. The r_{terminal}-$B_{\text{conditioned}}$ profile should correspond to the exponentiated Mazur function, Equation 6.

Despite the success in describing conditioned stimulus strengths seen in the current project, the results of this project also indicated some problems with the exponentiated Mazur function. First, the fitting algorithm did not converge on a best-fit solution in many cases. This was likely due to the high number of decimal places required for fitting values of the incomplete gamma function. The presence of imaginary numbers also caused some fits to fail to converge. Despite these problems, most fits that did not converge still provided an excellent fit to the R-W data, indicating that the function provides a good account of the data. However, these problems do raise the possibility that the function may be difficult to use, and this may restrict its practical application.

The second problem with the exponentiated Mazur function was the difference in parameter values obtained between the Mazur and exponentiated Mazur function. Across model conditions, the parameters a and b were consistently different for these two function forms despite being derived from similar functions. Despite these functions being the same when $s = 1$, it is possible that the addition of this scaling parameter may fundamentally alter the function.

Finally, the third problem with the exponentiated Mazur function concerns the estimates of s. Many of the best-fit values of s were relatively large and often were much greater than one. The reason for this is apparent when considering the very small values produced by the incomplete gamma function. Although the offset between incomplete gamma function and the $(r/b)^s$ factor within Equation 6 produces a function that describes the data from RI schedules well, the large values of *s* are incompatible with the general delay discounting function, Equation 5. Values of $s > 1$ in Equation 5 produce very low values of conditioned reinforcer strength. Additionally, previous work has shown that fits of Equation 5 to delay discounting data in humans yields best-fit values of *s* that are typically less than 1 (Green & Myerson, 2004). Therefore, the current results suggest that the two forms of the exponentiated Mazur function (Equation 5 for delay discounting and Equation 6 for RI schedules) are not comparable.

The incompatibility between the two forms of the exponentiated Mazur function illustrates problems stemming from their development. Each form was developed to describe different behavioral phenomena. The delay discounting function form has primarily been used to describe the effect that delay to reinforcement has on behavior (see Mazur, 1997; Green & Myerson, 2004). In contrast, the derivations for use with variable interval schedules implemented in the current project were designed to describe the relationship between conditioned stimulus strength and reinforcement rate. To date, this relationship has not been investigated in live organisms. Given the fundamental difference in how these two forms were developed and intended to be used, it is not surprising that there would be some differences in the parameters between them. Future research could help clarify whether the RI function form is useful in practical settings by fitting this form to live organism behavior that is supported by conditioned reinforcement.

Experiment Series II: Chained Schedules in the Computational Environment

Methods

The subject, environment, apparatus, materials, and computation procedures detailed in the General Methods were used for Experiment Series II.

Varying the proportion of behaviors produced with a fitness function.

In preparation for using the conditioned stimulus strength, *V*, to determine the strength of selection events (i.e., reinforcement), the effect of producing fewer than 100% of behaviors after a reinforcement event using a fitness function was examined. Previous versions of the McDowell model produced all behaviors for the next time-step using a fitness function after reinforcement. The currently proposed modified model used the conditioned stimulus strength values produced by the Rescorla-Wagner model to determine the proportion of behaviors produced using the fitness function after a reinforcement event in the initial link. The remaining behaviors were produced by randomly selecting parent behaviors. The primary purpose of the current experiments was to ensure that changing these proportions would not modify the correspondence of model behavior to matching theory (and by extension to live organism behavior).

A series of experiments was conducted that varied the proportion of behaviors produced using the fitness function from 0.2 to 1.0. For each condition, data from 11 single alternative RI schedules (RI 1, 2, 3, 5, 8, 10, 18, 25, 68, 112, and 200) were used to evaluate the relationship between reinforcement rate, *r*, and response rate, *B.* The average response and reinforcement rates for each RI schedule were calculated by counting the number of responses for a particular link during the experiment and dividing this result by 500. Although this method of determining the rates differed from previous methods used by McDowell and colleagues, this method allowed a standardized rate calculation without discarding data from incomplete 500 time tick windows that result when running chained schedules (the exact number of time ticks within a particular link cannot be determined a priori due to the emergent nature of movement from initial to terminal link). The average *r*-*B* data was plotted and the classic and modern matching theory equations were fitted to this data using the Solver add-in component in Microsoft Excel. McDowell (2005) has discussed the details of, and the differences between, the classic and modern theories of matching. The residuals from the fits were evaluated for non-random trends. Goodness of fit was determined by percent variance accounted for (pVAF), and by residual analysis.

Initial link correspondence with matching theory.

The modified McDowell model was run on chained schedules to determine whether behavior in the initial link remained consistent with matching theory under these conditions. Although this set of experiments was similar to the experiments presented above that varied the proportion of behaviors produced using randomly selected parents versus parents selected using a fitness function those experiments had static/constant proportions. The present set of experiments examined behavior where the proportion of behaviors produced using the fitness function varied based on the conditioned reinforcement strength, *V*, of the terminal link stimulus. It should be noted that the terminal link behavior was not evaluated because the terminal link was not different from an ordinary single alternative schedule, which has already been shown to produce results consistent with matching theory (McDowell, 2004; McDowell & Caron, 2007). These analyses were conducted primarily to ensure that initial link behavior, which depended on the terminal link R-W parameters of the modified McDowell model, conformed to matching theory, and by extension to live-organism data. For each of the conditions listed in Table 7, 11 RI schedules were run in the initial link to obtain an *r-B* profile so that the matching theory equations could be fitted to the data.

Simple behavior chains.

A variety of methods were used to qualitatively evaluate the behavior of the modified McDowell model running on chained schedules of reinforcement. Using a set of Rescorla-Wagner parameters found to produce a stable reinforcement rate-conditioned stimulus strength relationship in Experiment Series I (β_0: 0.05, β_1: 0.5, α: 0.7, a: 1.0), a variety of experiments were conducted to examine the chained schedule behavior across different conditions. First, four chained schedules were run at different RI values (RI1-RI1, RI5-RI5, RI25-RI25, and RI112-RI112). Cumulative records of initial and terminal link behavior were constructed for the first 1000 time steps. These were examined to ensure adequate progression through the chain and general consistency with live organism behavior. Next, the relationship between initial and terminal link response and reinforcement rates was examined by plotting response-reinforcement rate data for the terminal link/initial link schedule combinations noted above. Finally, the response rate data were examined for consistency with previously published two-link chained schedule data from live organism. Based on limited published data from chained VI VI schedules (Ferster & Skinner, 1957, Figures 859, 860, and 861), a 60% reduction in initial link response rate was expected (compared to terminal link response rate).

Results

Effect of varying the proportion of behaviors produced using a fitness function after reinforcement.

Figure 18 shows the *r-B* (reinforcement-response) plot for single alternative schedule experiments that varied the proportion of behaviors produced using a fitness function from 0.2 to 1.0 (the remaining proportion of behaviors were produced by selecting parent behaviors randomly). The plot demonstrates that reducing the percent produced using a fitness function reduces both the response and reinforcement rates relative to the traditional method of replacing all behaviors. This reduction in both response and reinforcement rates resulted in a shallower *r-B* curve thus reducing the strength of each reinforcement event, as expected. Notably, the same effect has been observed when increasing the extraneous reinforcement rate, which is represented by the r_e parameter in the matching theory equations. Table 8 shows the fit parameters for the percent replaced analyses. As shown in Table 8, both k and r_e (or $\frac{{r_e}^a}{b}$ for the modern matching equation) vary with the percent of behaviors replaced during selection events. Additionally, the exponent, a, obtained from the modern matching equation ranged from 0.82 to 0.94 indicating a degree of undermatching. This was consistent with previously published results (McDowell & Caron, 2007). Residual analysis indicated the presence of cubic polynomial trends in 6 of 9 fits using the classic matching function, whereas only one modern matching fit contained significant non-random residuals. Figure 19 shows the residuals resulting from fits of the classic and modern matching equations to the *r-B* data. Some classic matching equation residual plots appear to have non-random patterns while others appear randomly distributed. Generally, non-random patterns were not apparent in the modern matching equation residual plots. This result is consistent with previously published reports (McDowell & Caron). Consequently, it can be concluded that (1) varying the percent of behavior produced using the parental fitness function does not disrupt the model's correspondence with matching theory, and (2) selection events are weakened proportionally to the percent of behaviors randomly replaced during selection events.

Initial link correspondence with matching theory.

The analyses of the initial link correspondence with matching theory indicated that both the classic and modern matching theory functions adequately described model behavior. However, less evidence of non-random residuals was observed in the modern matching equation fits to model data compared to classic matching fits (Equation 2 versus Equation 6), as indicated by the asterisks in Table 8. The account of model data by matching theory was consistent across R-W model parameter combinations and terminal link rates of reinforcement, as seen in Table 7. As expected, the value of the initial link r_e increased with decreasing rates of reinforcement in the terminal link, which depress the value, V, of the terminal link conditioned reinforcement and therefore the percentage replaced using a fitness function after reinforcement. There was no clear relationship between the modern matching exponent, a, and model conditions, which ranged from 0.70 to 0.92 indicating that undermatching was present in all conditions. These results indicate that conditioned reinforcement as implemented in the computational model did not affect the phenomena of undermatching that has been shown to be an emergent property of the model (McDowell & Caron, 2007).

Simple behavior chains.

Cumulative records.

The cumulative records from four chained schedule experiments with RI1-RI1, RI5-RI5, RI25-RI25, and RI112-RI112 schedules in the intial and terminal links are shown in Figures 20 through 23, respectively. These figures show cumulative records, which provide a method of viewing the relationship between responses, reinforcement, and the passage of time. For each time-step, the plot makes one step horizontally to the right. For each response, the plot makes one positive vertical step. For each reinforcement, a horizontal dash is placed next to the response plot. Once the number of responses reaches 100, the vertical location is reset to 0 (the time axis location is not reset). High rates of response result in steep plot slopes. No responding, such as occurs after extinction, results in flat plot slopes.

As can be seen in the RI1-RI1 cumulative record plot (Figure 20), the chained behavior takes time to develop with the first target behavior in the initial link being emitted after approximately 100 time steps (denoted by the 1 in Figure 20). Following this behavior emission, which was reinforced (as denoted by the small dash), the terminal link repertoire became active. Within the terminal link, approximately 100 time steps elapsed before a target behavior was emitted and reinforced (as denoted by the 2). These two behaviors completed the chain and the initial link repertoire was again presented. Note that again, approximately 100 time steps elapsed before another target behavior in the initial link was emitted and reinforced (as denoted by the 3). After this behavior was reinforced and the terminal link repertoire was presented for the second time, the latency to terminal link completion was markedly smaller. This quick increase in behavior rate was due to the previous target behavior being reinforced with primary reinforcement whereas the first behavior emitted in the initial link was reinforced with presentation of the terminal link stimulus that had yet to be paired with primary reinforcement. However, once the terminal link stimulus was paired with reinforcement and therefore gained conditioned reinforcing strength (as determined by the Rescorla-Wagner function), it quickly produced high rates of responding in the initial link. This effect can be seen in the rapid acquisition of high response rates to the right of the “3” in Figure 20.

A qualitative comparison of the response rates in the initial and terminal links revealed similar rates across both links. As can be seen in Figure 20, the slopes of both cumulative records were similar. Additionally, the response rates were approximately constant throughout the remainder of the 1000 time steps shown in the cumulative record. Taken together, these results indicate that chained schedules of behavior were established in the computational model and, once the chained behavior was acquired, it continued at a constant rate.

An examination of the RI5-RI5 cumulative record revealed similar patterns of behavior: the behavior chain took time to develop, but once acquired, continued at a relatively constant rate as can be seen in Figure 21. However, the rates of behavior were notably lower in the initial link compared to the terminal link as can be observed by noting the number of times the cumulative records reached 100 behaviors or, alternatively, by observing the relatively lower slope in the top panel of Figure 21. An additional phenomenon observed was the presence of significant pauses in responding that occurred before some reinforced responses. These pauses were also present in the cumulative records of the RI25-RI25 and RI112-RI112 schedules as can be seen in

Figures 22 and 23. As would be expected, the response rates in these two chained schedules were appropriately lower. However, the difference between initial and terminal link response rates became more apparent. In the RI25-RI25 and RI112-RI112 experiments, the response rate in the initial link was approximately half that in the terminal link as can be seen in Figure 22 and 23. In comparison, the difference was much smaller in the RI5-RI5 experiment and largely unnoticeable in the RI1-RI1 experiment.

Initial and terminal response-reinforcement rate dependency.

Plots of the response-reinforcement relationship for the initial link at different constant values of the terminal link are shown in Figure 24. As would be expected from the previous results showing shallower *r-B* curves for lower proportions of behavior produced with fitness functions, leaner schedules in the terminal link produced shallower curves.

Initial versus terminal link response rates.

The difference between response rates for the initial and terminal links varied across reinforcement rates. Figure 25 shows the ratio of initial to terminal link response rates for RI1-RI1, RI5-RI-5, RI25-RI25, and RI112-RI112 reinforcement rates with Rescorla-Wagner parameters (β_0: 0.05, β_1: 0.5, α: 0.7, a: 1.0). Richer schedules resulted in higher response rates in the initial link relative to the terminal link. The response rate ratios ranged from 0.97(RI1-RI1) to 0.59 (RI25-RI25). Although the ratio increases for the RI112-RI112 schedule (0.64), it is likely that this pattern was unique to the individual experiment: this pattern was not observed across other conditions not specifically reported here. Unfortunately, there is little published data to compare with these results. Overall, the results demonstrate that the initial link response rates were lower than the terminal link response rates, as would be expected in chained schedules. These results also show that the ratio of initial to terminal link response rates varies with reinforcement rate.

Discussion

Effect of varying the proportion of behaviors produced using a fitness function after reinforcement.

This set of experiments examined a new technique for reducing the strength of a selection event to reflect the reduced strength of conditioned reinforcement. This technique produced only a proportion of new behaviors for the next time-step using a fitness function following reinforcement. By randomly picking parent behaviors to produce the remaining behaviors, the strength of a reinforcement event can be reduced. This random parent selection places no selection pressure on these behaviors and thus reduces the likelihood of new behaviors being in the target class. Additionally, the current results demonstrated that producing less than 100% of the behaviors using the fitness function resulted in behavior consistent with matching theory and, by extension, with the behavior of live organisms. Furthermore, reductions in the proportion of behaviors produced using the fitness function (i.e., reinforcer strength) corresponded with an increase in extraneous reinforcement, r_e, (or reinforcement not contingent with a target behavior) and a decrease in the matching parameter k, the maximum response rate. This effect on the matching function parameters is consistent with previously published live organism data (e.g.,Bradshaw, Szabadi, & Bevan, 1978). Put more simply, reductions in the reinforcer strength reduced the response rate, a result that is consistent with previously published reports using live organisms (e.g., Reed & Wright, 1988).

Matching analyses.

The initial link behavior was evaluated for correspondence with matching theory. The primary reason to conduct these analyses was to ensure that making initial link behavior dependent on terminal link behavior did not alter the model in a way that made it unrealistic. The current results confirmed that initial link behavior was well described by matching theory. Although there was some evidence of non-random residuals in the fits of the classic matching function to model data, the evidence suggested that modern matching function fits fully accounted for the variance in model data. This result has been observed before by McDowell and colleagues and, as such, was expected (McDowell & Caron, 2007).

Simple behavior chains.

Several different techniques utilized in the current project demonstrated that 1) simple chained schedules of behavior could be established and that 2) this behavior was largely consistent with previously published data on simple chained schedules (i.e., two-link, single alternative chains). The cumulative records from the current project demonstrated the presence of several relevant phenomena. First, the cumulative records showed how the chains are established. Although the chains take time to establish, once the model (or digital organism) has acquired the behavior, response rates were generally consistent over time. Additionally, the model was sensitive to reinforcement rates with higher rates of reinforcement producing chains of behavior that move quickly from link to link. The cumulative records also showed evidence of pre-reinforcement pauses. This occurred more often and was more marked in leaner schedules. These pauses were likely caused by instances when very few behaviors were in the target class. Close examination of these instances showed that immediately preceding the pauses there were relatively steady rates of behavior.

Finally, the simple comparison of initial and terminal response rates revealed findings that were generally consistent with live organism behavior. Although there is little previously published data on two link chains, the studies reviewed for this project indicated that 1) response rates in the initial link should be lower than in the terminal link and that 2) the response rate in the initial link should be approximately 60% of the terminal link rate. However, this latter assertion was derived from only one set of data published by Ferster and Skinner (1957). The results from the current experiments clearly showed that the response rates in the initial link were lower than the response rates in the terminal link. However, the amount of rate decrement seen in the initial link depended on the terminal link reinforcement rate. To further verify these results, live organism data with varied reinforcement rates in chained links would be necessary.

Experiment Series III: Replication of the Royalty et al. Experiments

The subject, environment, apparatus, materials, and computation procedures detailed in the General Methods were used for Experiment Series III.

While Experiment Series I evaluated how well the behavior of a computationally based dynamic theory of conditioned reinforcement corresponded to live organism behavior, and Experiment Series II examined the qualitative properties of a two-link chained schedule, Experiment Series III attempted to replicate a previous experiment that was originally designed to examine the role that conditioned reinforcement and time-to-reinforcement play in chained schedules. The Royalty et al. (1987) experiment demonstrated that in a 3-link chained schedule, a short, 3 second delay immediately prior to presenting the conditioned stimulus reduced the response rates by 60%. In Experiment Series III, the Royalty et al. experiment was replicated by delaying the presentation of the conditioned stimulus (terminal link) by a single time-step in a two-link chain.

Methods

Two experiments using methods similar to Royalty et al. (1987) were conducted and the results were compared. In the first experiment, a chain RI 31 RI 31 schedule was run for 20,000 cycles. In the second experiment, the RI 31 initial link was followed by a 1 cycle delay before the terminal link stimulus is presented. In other words, once a behavior in the target class was emitted and reinforcement was available as determined by the RI 31 schedule, an additional cycle of the model was run (this is the delay) before the terminal link stimulus was presented. Once the terminal link stimulus was presented, the normal selection and reproduction rules as specified in Step 3a of the program procedures described in the General Methods section were executed. To maintain the same time (i.e., cycle) to reinforcement, the terminal link schedule was changed to RI 30. Again, the model was run for 20,000 cycles.

Additional experiments were run to determine if the Royalty et al. experiment was better replicated by the model in richer or leaner schedules. In a series of experiments the value of the initial and terminal link RI schedules was varied from RI1RI1 to RI120RI120. Results from these experiments were analyzed to determine their correspondence with the Royalty et al. findings.

Results

The results from experiments replicating the Royalty et al. (1987) study were partially consistent with previously published live organism data. The response rates from experiments with RI31-RI31 schedules did produce higher rates than those where a delay preceded the presentation of the terminal link (i.e., reinforcement). Consistent with the Royalty et al. data, the response rate for the initial link was 68% lower when a 1 time-tick delay was added before presentation of the terminal stimulus. However, follow-up analyses indicated that this result was likely an artifact of the computational model. Figure 26 shows the results of implementing a 1 time-tick delay in a variety of RI schedules (RI1-RI1 to RI22-RI22). As can be seen, there was no relationship between the schedule value and the response rate for the initial link when a 1 time-tick delay occurred before presentation of the terminal link. Additionally, the actual response rate value was consistent with the operant level of the model. The stable rate of response seen in Figure 26 and its value (i.e., the operant level) indicated that adding a 1 time-tick delay essentially removed all reinforcement contingencies. Thus, the 68% reduction observed in the Royalty et al. replication experiment response rate was merely a coincidence caused by reaching a response rate floor.

Discussion

The failure to replicate the Royalty et al. (1987) findings demonstrated that the model in its current form cannot fully simulate chained schedule behavior. While the results from the current experiment initially appeared to be consistent with the Royalty et al. study, they were, in fact, due to the removal of reinforcement contingency. By instituting a delay-to-reinforcement paradigm for a wide variety of schedules and observing that the response rates for the initial link were all the same, and were equal to the operant level, it was determined that in the McDowell model a delay of reinforcement removes all reinforcement contingencies. Rates of responding return to operant levels with even the shortest delay (i.e., a single time-step). However, live organism data suggest that discounting functions like Equations 3 or 5 describe the relationship between delay of reinforcement and response strength (Rachlin, 2006), not a complete dissociation of reinforcement contingencies as seen in the current data.

Because the time-to-reinforcement theory was not implemented in the current model, the current results do little to resolve the theoretical debate on whether it is conditioned reinforcement or time-to-reinforcement responsible for behavior in chained schedules. However, because the current project did implement a strictly conditioned reinforcement approach to chained schedules and fully realistic behavior could not be produced, a possible interpretation of the current results is that conditioned reinforcement is not exclusively responsible for the behavior observed in chained schedules or other situations under stimulus control. It is possible that a combination of the two approaches may be necessary. Given that Royalty et al. (1987) showed that conditioned reinforcement effects explained their results, it is possible that, while delay-to-reinforcement is involved in some way, it plays a lesser role compared to conditioned reinforcement. An alternative explanation is that the conditioned reinforcement approach is correct, but without adequate modeling of delay to reinforcement effects, the Royalty et al. experiment cannot be replicated. Thus, by correcting the McDowell model's inadequate simulation of delay to reinforcement effects, the current conditioned reinforcement modeling approach might work. Efforts to resolve these alternative explanations are a clear direction for future research.

General Discussion

The current project’s purpose was to implement stimulus control within the McDowell computational model of behavior. To this end, two-link chained schedules of behavior were established, maintained, and were found to be qualitatively similar to live organism behavior. However, the failed attempt to replicate the Royalty et al. (1987) experiment indicated that certain phenomena could not be simulated in the current model configuration. Despite some problems, the current project produced other findings, namely, the correspondence between the Rescorla-Wagner and Mazur models of conditioned reinforcement, that have important theoretical implications for the learning theory field.

Problems with Delay of Reinforcement: Failure to Replicate Royalty et al. (1987)

Despite the model’s success in producing realistic “looking” behavior, it could not replicate the theoretically important Royalty et al. (1987) experiment, which found that pre-reinforcement delays in initial links decreased response rates by 60%. Although the current results did indicate a decrease in response rates with the addition of a one time-step delay, the rates dropped to operant levels. This indicated that the delay effectively removed all reinforcement contingencies. In other words, the same result could have been obtained in a setup where no reinforcement existed at all. Given that hyperbolic discounting theory (i.e., Mazur, 1984) has repeatedly demonstrated that delays to reinforcement should follow a hyperbolic decrease in reinforcement effectiveness, it is clear that the current computational model setup lacks this type of decay.

The model is not completely devoid of memory for previous states since the behaviors in one generation are related to the next by means of parental mating. (Note that the term, “memory,” as used here refers to a basic process where previous events have some affect on events occurring later.) There is also a form of memory when the parental fitness function is used to select parent behaviors for mating. The selection tends to concentrate behaviors in and near the target class. When all of the behaviors are clumped together, it takes some time steps before the clumping dissipates. Anecdotally, the number of time ticks necessary to dissipate behaviors after reinforcement is on the order of 1-3 ticks depending on the model parameters being used. Clearly, three ticks is a very short memory and would only represent the simplest of organisms.

Although the clumping after reinforcement is some form of memory, there is relatively little, if any, memory for behaviors that undergo random mating processes (i.e., non-reinforced behaviors). As noted above, there are some similarities between behavior generations because the new generations are built from behaviors in previous generations. However, there is virtually no meaningful memory effect from this property especially in regard to non-reinforced behaviors. Behaviors after a non-reinforced response are generally as randomly distributed as those before the response. Because the emitted behavior is randomly drawn from the existing behavior population, there is little continuity between responses. For example, if a behavior from Target 1 is emitted in Time-Step 1 and not reinforced, the next emitted behavior has equal likelihood of being selected from any of the other targets (proportional to the target sizes). The real world analogy of this might be the following: at Time 1, a rat is pressing a lever, then at Time 2, the rat is sniffing the corner of the cage completely opposite to the lever location. In reality, the rat would have to dismount the lever, turn around, walk to the other side of the cage, then sniff the corner. This chain of behavior is not represented in the current model.

In other words, there is no requirement for series of emitted behaviors in the current model to be related to each other.

There are several potential solutions to this problem of discontinuous behavior. One solution would be to require the next generation's emitted behavior to be close to the previous generations. However, a top-down approach to modeling such as this would essentially mandate the desired solution. In other words, the end result would be determined a priori.

An alternative approach to creating more continuity between behaviors would be to implement a simple low level rule and determine whether the desired consequences emerge from the behavior produced by this low-level rule. Taking inspiration from brain function, the model could implement a requirement that behavior emission would not be determined by just one phenotype randomly drawn from the population. Instead, the emission would be determined by the location of a group of behaviors. This is an attempt to be analogous with neurobiological functioning where groups of neuron activations are responsible for behavior, not single, isolated neuron firings. The new emission rule would be as follows:

1) To determine which behavior will be emitted, the repertoire will be searched for where the greatest concentration of behaviors is located.
2) The local mean of that clump of behaviors will be calculated and that mean will represent the emitted behavior.
3) The target class where the emitted behavior resides will represent the emitted target class.
4) Selection, mating, and mutation will occur using existing methods.

Naturally, without reinforcement for a target class behavior, more behaviors would be emitted from non-target classes, but this is the case in the current model version. To ensure that sufficient behaviors exist to determine groupings of behavior, the total number of behavior phenotypes in the population should be increased (e.g., by using 1000 individual behavior phenotypes instead of 100). These methods would likely increase the contiguity between series of emitted behaviors. In the case of delayed reinforcement like the Royalty et al. (1987) study, behaviors emitted in a delay condition will likely more closely resemble the behavior that triggered the delay to reinforcement.

Correspondence of the Rescorla-Wagner and Mazur Models

Although the current project's purpose was to make a computational model that produced realistic chained schedule behavior, other results were obtained that have practical and theoretical implications. The most salient of these findings was the correspondence of a dynamic theory of conditioned stimulus strength (i.e., Rescorla-Wagner) and a static theory of conditioned reinforcement (i.e., the Mazur function). This correspondence was first proposed by McDowell et al. (2006), but had not been formally evaluated. The current results did not support the original proposition that the Rescorla-Wagner model would produce conditioned stimulus values that corresponded to the RI variant of the classic Mazur function. However, correspondence between the Rescorla-Wagner model and an exponentiated variant of the Mazur function was found. This finding was important for two reasons: 1) a complete model (dynamic and static) of conditioned reinforcement was supported and 2) the exponentiated variant of the Mazur function provides a better description of conditioned reinforcement at equilibrium conditions than the original function. Both findings contribute significantly to the conditioned reinforcement and hyperbolic discounting literature, respectively.

To date, a complete model of conditioned reinforcement accounting for both dynamic and static conditions has not been fully evaluated. Having a complete model of a behavioral phenomenon that describes both what can be observed and what factors create the phenomenon is a crucial step in fully understanding behavior. The Rescorla-Wagner and Mazur models together account for not only the outcomes (conditioned reinforcement), but also the processes that generated these outcomes (an iterated model of conditioned stimulus strength).

It is notable that within this complete model, it is an exponentiated variant of the Mazur hyperbolic discounting function that best describes dynamically produced conditioned stimulus strength. As noted previously, there has been significant debate regarding which form, the original Mazur or exponentiated Mazur, better describes live organism behavior data (Green & Myerson, 2004). The current results support the exponentiated variant over the original function. The fact that this finding originated from a computational model rather than live organism data might lead some to dismiss its significance. However, computational data offers certain advantages over live organism data. Specifically, computational data is highly controlled and plentiful. Both of these factors markedly increase the statistical strength of the analyses. Therefore, phenomena are less likely to be lost within noise that is inherent to live organism data. Despite these advantages, the true scientific value of a computational model is in generating findings that can be tested in live organisms. In regard to the current results, the findings suggest that the reinforcement rate-conditioned reinforcement strength relationship is best accounted for by an exponentiated hyperbolic discounting function. However, existing live-organism data do not directly address this relationship. Therefore, future research could focus on evaluating this relationship in live organisms.

Neurobiological Correlates with the Computational Model

Two elements of the model design most relevant to neurobiological findings appear to be the functioning of the Rescorla-Wagner model and the separate behavior repertoires that were used to represent separate conditioned stimuli.

Although the dynamic Rescorla-Wagner model provides an account of how a stimulus develops conditioned reinforcement through pairings with unconditioned stimuli, it can only be grossly compared to neurobiological functioning. The brain structure most closely related to this functioning would be the ventral striatum. However, as discussed above, this region has been primarily associated with the initial learning of conditioned stimuli associations. As learning develops further, the dorsal striatum appears to become more relevant (Graybiel, 2008). The Rescorla-Wagner model does not account for the separation of the initial learning and behavior maintenance processes. Despite the lack of this functioning, it is possible that the R-W model provides an estimation of overall/combined striatum functioning. A review of the literature did not reveal any specific studies that have explored these connections so this remains conjecture at the current time. Future research could address this connection with collaborative neurobiologist and learning theorist project. Although neurobiologists and learning theorists come from different training backgrounds, this cross-disciplinary approach could be a productive area for future research (c.f., McDowell, 2010).

The second aspect of the computational model design that may benefit from some comparison to neurobiological findings is the proscribed separation of the stimulus-associated repertoires (i.e., the initial and terminal links). This separation was programmed into the model giving the separated repertoires “built-in” stimulus control. Although the resulting behavior was consistent with live-organism behavior in several ways, the functioning of the model was restricted to behavior that has already developed stimulus associations. Neurobiological findings have found consistent evidence that separate brain regions are involved in the learning and maintenance of stimulus associated behavior (Graybiel, 2008; Everitt & Robbins, 2005). The modified model does not address how stimulus associations develop beyond the conditioned stimulus strength values produced by the Rescorla-Wagner model. Because development of stimulus associations is an important part of realistic behavior, future research would likely benefit from modeling these phenomena.

The TNGS (theory of neuronal group selection) proposed by Gerald Edelman does specify how stimulus associations develop: by connecting separate neuronal groups associated with a stimulus and a behavior, respectively, by means of reentrant signaling. It may be possible to use the concept of reentrant signaling within the current computational model. If this implementation of reentrant signaling were technically feasible, the behavior produced by the model using reentrant signaling and Rescorla-Wagner methods could be compared. An experiment like this would have important theoretical implications because a neurobioloical theory (TNGS) could be tested alongside a traditional learning theory (Rescorla-Wagner) that has a large amount of research supporting its usefulness.

Future Directions for Research

There are a variety of possible future research directions based on the current project's results. Some of these directions have been discussed previously. One of these possibilities is the change in behavior emission methodology. The new methodology would entail behavior emission by groupings of behaviors rather than by randomly selected behaviors. Although this change would represent a major modification to the McDowell model, it has the potential to improve the continuity between behaviors emitted at each time-step, which could improve model realism when pre-reinforcement delays are instituted. A second possibility would be to examine model behavior on concurrent chained schedules of behavior. Although concurrent chains are more complex, there are a variety of quantitative theories that have been successful in describing live organism behavior on them (e.g., delay reduction theory - Fantino, 1969), which would improve the validity of any evaluation of model behavior.

Several additional directions for future research stem from examining the correspondence of live organism behavior and the current model results. Although the current model implemented relatively simple experiments (two link chained schedules), there was a paucity of existing live organism data to compare to the current results. Future work could coordinate computational model and live organism experiments to compare the results directly. A first series of these coordinated experiments could systematically vary the reinforcement rate in chained schedules and examine the changes in behavior rates in the links as a result. Comparison of the computational model data and live-organism data would be very informative in determining the realism of the computational data. A second series of experiments could examine the relationship between conditioned stimulus strength and reinforcement rate to determine whether the original or exponentiated Mazur functions best account for live organism data variance. Given the current project’s results, it is expected that the exponentiated Mazur function would best account for the data.

Conclusion

The current project represents a major and important advance in computational modeling of behavior. Although previous versions of the McDowell computational model could produce behavior on single and concurrent alternative schedules that was consistent with live organisms, there was no account for changes in behavior due to different stimulus conditions. The current project’s modifications to the model allow for the production of behavior on chained schedules. However, the range and type of behavior capable of being produced was limited. In fact, the current project’s results highlighted some areas of the model that need improvement (i.e., the effect of delay to reinforcement).

In the process of developing the modified model, the project also found evidence for a complete model of conditioned reinforcement that accounts for both dynamic and static effects. The finding that the Rescorla-Wagner model’s dynamically produced conditioned stimulus strength are accounted for by the exponentiated variant of the Mazur hyperbolic discounting function has important theoretical implications, as noted previously.

Overall, the project has produced predictions about phenomena that should be observable in live-organisms. Although some have predicted that analytical techniques give rise to a “flight from the laboratory” (Skinner, 1959), the current results clearly demonstrate the co-productive nature of computational modeling and live organism data: results from each discipline can suggest future directions of research in the other. It is this co-evolution of the two disciplines that holds great potential for spurring a greater understanding of behavior.

References

Bahcekapili, H.G. (1998). An evaluation of Rescorla and Wagner's elementistic model versus Pearce's configural model in discrimination learning. *Dissertation Abstracts International*. *58(11-B)*, 6262.

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

Bradshaw, C. M., Szabadi, E., & Bevan, P. (1978). Relationship between response rate and reinforcement frequency in variable-interval schedules. The effect of the concentration of sucrose reinforcement. *Journal of the Experimental Analysis of Behavior*, 29, 447-452.

Edelman, G.M., 1978. Group selection and phasic re-entrant signaling: a theory of higher brain function. In: Mountcastle, V. (Ed.), The Mindful Brain. MIT Press, Cambridge, MA.

Edelman, G.M. (1987). Neural Darwinism: The Theory of Neuronal Group Selection. New York: Basic Books.

Edelman, G.M., 1992. Bright Air, Brilliant Fire: On the Matter of the Mind. Basic Books, New York.

Everitt, B.J., & Robbins, T.W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature Neuroscience*, 8, 1481-1489.

Fantino, E., & Romanowich, P. (2007). The effect of conditioned reinforcement rate on choice: a review. *Journal of the Experimental Analysis of Behavior.* 87, 409-421.

Fantino, E. (1969). Choice and the rate of reinforcement. Journal of the Experimental Analysis of Behavior, 12, 723-730.

Ferster, C.B., & Skinner, B.F. (1957). *Schedules of reinforcement*. New York: Appleton-Century-Crofts

Graybiel, A. (2008). Habits, rituals, and the evaluative brain. *Annual Review of Neuroscience*, 31, 359-387.

Green, L., Myerson, J., & Ostaszewski, P. (1999). Amount of reward has opposite effects on the discounting of delayed and probabilistic outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 418-427.

Green, L., Fry, A. F., & Myerson, J. (1994). Discounting of delayed rewards: A life-span comparison. *Psychological Science*, 5, 33-36.

Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin*, 130, 769-792.

Hutcheson, D., Parkinson, J., Robbins, T., & Everitt, B. (2001). The effects of nucleus accumbens core and shell lesions on intravenous heroin self-administration and the acquisition of drug-seeking behavior under a second-order schedule of heroin reinforcement. *Psychopharmacology*, 153, 464-472.

Jones B., & Rachlin H. (2006), Social discounting. Psychological Science, 17, 283-6.

Kulubekova, S., & McDowell, J. J (2007). A computational model of selection by consequences: log survivor plots. *Behavioural Processes, 78,* 291-296.

Mazur, J.E. (1984). Tests of an equivalence rule for fixed and variable reinforcer delays. *Journal of Experimental Psychology: Animal Behavior Processes, 10,* 426-436.

Mazur, J.E. (1993). Predicting the strength of a conditioned reinforcer: effects of delay and uncertainty. *Current Directions in Psychological Science, 2,* 70-74.

Mazur, J.E. (1997). Choice, delay, probability, and conditioned reinforcement. *Animal Learning & Behavior, 25*, 131-147.

McDowell, J. J (2004). A computational model of selection by consequences. *Journal of the Experimental Analysis of Behavior. 81*, 297-317.

McDowell, J. J (2005). On the classic and modern theories of matching. *Journal of the Experimental Analysis of Behavior, 84*, 111-127.

McDowell, J. J (2010). Behavioral and neural Darwinism: Selectionist function and mechanism in adaptive behavior dynamics. *Behavioural Processes*, 84, 358-365.

McDowell, J. J & Caron, M. L. (2007). Undermatching is an emergent property of selection by consequences. *Behavioural Processes, 75*, 97-106.

McDowell, J. J, Caron, M. L., Kulubekova, S., & Berg, J. P. (2008). A computational theory of selection by consequences applied to concurrent schedules. *Journal of the Experimental Analysis of Behavior, 90*, 387-403.

McDowell, J. J, Soto, P. L., Dallery, J., & Kulubekova, S. (2006). A computational theory of adaptive behavior based on an evolutionary reinforcement mechanism. In M. Keijzer (Ed.), *Proceedings of the 2006 Conference on Genetic and Evolutionary Computation (GECCO-2006)* (pp. 175-182). New York: ACM Press.

Miller, R., Barnet, R., & Grahame, N. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin, 117,* 363-386.

Myers, K., Vogel, E., Shin, J., Wagner, A. (2001). A comparison of the Rescorla-Wagner and Pearce models in a negative patterning and a summation problem. *Animal Learning & Behavior, 29,* 36-45.

Parkinson, J., Olmstead, M., Burns, L., Robbins, T., & Everitt, B. (1999). Dissociation in effects of lesions of the nucleus accumbens core and shell in response to cocaine cues and during cocaine-seeking behavior in rats. *Journal of Neuroscience*, 20, 7489-7495.

Pearce J.M. (1987). A model of stimulus generalization for Pavlovian conditioning. *Psychology Review*, 94, 61–73.

Popa, A. (2009). *The effect of Hamming distances in a computational model of selection by consequences.* Unpublished master's thesis, Emory University, Atlanta, Georgia.

Rachlin, H. (1989) *Judgement, decision, and choice.* New York: Freeman.

Rachlin H. (2006), Notes on discounting. *Journal of the Experimental Analysis of Behavior*, 85, 425-35.

Reed, P., & Wright, J. (1988). Effects of magnitude of food reinforcement on free-operant response rates. *Journal of the Experimental Analysis of Behavior*, 49, 75-85.

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

Royalty, P., Williams, B., & Fantino, E. (1987). Effects of delayed conditioned reinforcement in chain schedules. *Journal of the Experimental Analysis of Behavior*, *47*, 41-56.

Siegel, S., & Allan, L. (1996). The widespread influence of the Rescorla-Wagner model. *Psychonomic Bulletin and Review*, *3*, 314-321.

Skinner, B. F. (1981). Selection by consequences. *Science*, *213*, 501–504.

Skinner, B.F. (1959). The flight from the laboratory. In B. F. Skinner, Cumulative record. New York: Appleton-Century-Crofts.

Staddon, J., & Cerutti, D. (2003). Operant conditioning. *Annual Review of Psychology, 54,* 115-144.

Williams, B.A. (1994). Conditioned reinforcement: neglected or outmoded explanatory construct? *Psychonomic Bulletin and Review, 1,* 457-475.

Appendix

Residual Analysis

The residuals resulting from fits of various functions to model data were analyzed in several steps. First, the standardized residuals were calculated after fitting a function to model data. Second, these residuals were plotted against the predicted value of the function being used. Third, these plots were qualitatively examined for non-random patterns. Previous work by Berg (unpublished masters, 2008) and Berg and McDowell (2010) found that lower order polynomial fits could miss significant non-random residuals. Although visual inspection of plots does not provide statistical confidence regarding ones' conclusions, the method can identify whether patterns might exist but are not being identified using appropriate statistical means. Finally, after visual inspection of the residuals, cubic polynomials were fitted to the predicted value-standardized residual data. Cubic polynomials were chosen (as opposed to quadratic, quartic, or quinitic) in the hopes of balancing Type I and II error. Figure A1 illustrates how the pVAF and chance of Type I error increases as polynomial order approaches the number of datapoints. Although the datapoints are random, a 6th order polynomial accounts for 100% of the variance. Thus, in determining what order of polynomial to fit to 11 residual datapoints, a cubic polynomial was chosen to minimize the chance of Type I error.

Table 1

Parameter descriptions

Model/Function	Parameter	Description
Rescorla-Wagner		
	V	Conditioned stimulus strength
	α	Stimulus salience
	β_0	Salience of reinforcer absence
	β_1	Salience of reinforcer presence
	λ	Maximum stimulus strength
	a	Currently-proposed scaling exponent
Mazur Function		
	a,b	Non-specific equation parameters
	x	Latency of primary reinforcement
Mazur Function(s) for Variable Intervals		
	a,b	(same as Mazur Function)
	r	Reinforcement rate
	s	Scaling parameter for exponentiated variant

Table 2

Model parameters varied in Experiment Series I

Parameter of interest	Parameter values tested	Other model conditions*			
		α	β_0	β_1	a
α	0.5, 0.7, 0.9	-	0.01	0.5	1
		-	0.1	0.5	1
		-	0.05	0.5	1
		-	0.05	1	1
		-	0.05	2	1
β_0	0.001,0 .005, 0.01, 0.05, 0.1	0.7	-	0.25	1
		0.7	-	0.5	1
		0.7	-	1	1
		0.7	-	2	1
β_1	0.25,0 .5,0 .75, 1, 2	0.7	0.01	-	1
		0.7	0.1	-	1
		0.7	0.05	-	1
a	1, 1.25, 1.5, 1.75, 2	0.7	0.05	1	-

* for each value of the parameter of interest

Table 3

Best-fit parameter values and goodness-of-fit statistics for fits of classic and exponentiated Mazur functions to conditioned stimulus strength values, V, for different values of β_1 under several Rescorla-Wagner parameter conditions.

	Classic Mazur			Exponentiated Mazur			
β_1	a	b	R^2	a	b	s	R^2
β_0: 0.01, α: 0.7							
0.25	1.18	51.5	0.99*	*1.05*	*2.62*	*8.95*	*1.00*
0.5	1.09	21.4	0.99*	*1.03*	*1.37*	*8.33*	*1.00*
0.75	1.06	12.6	0.99*	1.01	1.42	5.34	1.00
1	1.04	8.85	0.99*	*1.01*	*1.39*	*4.13*	*1.00**
2	1.09	21.4	0.99*	*1.00*	*0.62*	*4.52*	*1.00**
β_0: 0.1, α: 0.7							
0.25	11.4	13370	1.00*	1.72	19.6	18.7	1.00
0.5	2.42	859	1.00*	*1.25*	*3.81*	*34.1*	*1.00*
0.75	1.74	345	1.00*	*1.16*	*3.54*	*22.9*	*1.00*
1	1.49	198	1.00*	1.14	3.54	22.9	1.00
2	1.22	70	0.99*	1.06	4.01	7.60	1.00
β_0: 0.5, α: 0.7							
0.25	2.49	899	1.00*	*1.26*	*3.86*	*34.0*	*1.00*
0.5	1.51	207	1.00*	1.14	10.4	6.39	1.00
0.75	1.30	105	0.99*	1.08	4.58	8.84	1.00
1	1.21	67.3	0.99*	*1.06*	*3.49*	*8.40*	*1.00*
2	1.11	29.1	0.99*	1.03	3.97	4.06	1.00

* statistically significant ($p < .05$) cubic polynomial pattern in the residuals

Italics indicate that the fit did not converge

Table 4

Best-fit parameter values and goodness-of-fit statistics for fits of classic and exponentiated Mazur functions to conditioned stimulus strength values, V, for different β_0 Rescorla-Wagner parameter values.

	Classic Mazur			Exponentiated Mazur			
β_0	*a*	*b*	R^2	*a*	*b*	*s*	R^2
β_1: 0.25, α: 0.7							
0.001	1.02	3.12	1.00*	*1.00*	*0.65*	*3.62*	*1.00**
0.005	1.09	20.9	0.99*	1.02	1.40	8.01	1.00
0.01	1.18	50.3	0.99*	1.04	2.62	8.82	1.00
0.05	2.44	873	1.00*	*1.25*	*1.81*	*71.0*	*1.00*
0.1	11.4	13270	1.00*	1.69	10.8	32.6	1.00
β_1: 0.5, α: 0.7							
0.001	1.00	1.38	1.00*	-	-	-	-
0.005	1.05	9.31	0.99*	1.02	1.75	3.53	1.00
0.01	1.09	21.2	0.99*	1.03	1.95	6.00	1.00
0.05	1.51	206	1.00*	1.14	8.15	7.90	1.00
0.1	2.42	860	1.00*	1.27	12.3	11.3	1.00
β_1: 1, α: 0.7							
0.001	1.00	0.50	0.99*	1.00	1.67	0.40	1.00
0.005	1.02	3.99	1.00*	1.00	0.64	4.41	1.00
0.01	1.04	8.84	0.99*	1.03	2.97	5.23	1.00
0.05	1.22	68.3	0.99*	1.06	4.85	6.25	1.00
0.1	1.47	194	0.99*	1.11	4.5	13.0	1.00
β_1: 2, α: 0.7							
0.001	*0.99*	*0.21*	*0.86**	0.99	0.20	1.63	1.00
0.005	1.01	1.71	0.99*	-	-	-	-
0.01	1.01	3.75	0.99*	-	-	-	-
0.05	1.11	28.9	0.99*	1.04	5.64	3.02	1.00
0.1	1.22	70.1	0.99*	1.06	7.59	4.31	1.00

* statistically significant ($p < .05$) cubic polynomial pattern in the residuals

\- indeterminate fit

Italics indicate that the fit did not converge

Table 5

Best-fit parameter values and goodness-of-fit statistics for fits of classic and exponentiated Mazur functions to conditioned stimulus strength values, V, for different values of α under several Rescorla-Wagner parameter conditions.

	Classic Mazur			Exponentiated Mazur			
α	*a*	*b*	R^2	*a*	*b*	*s*	R^2
β_0: 0.01, β_1: 0.5							
0.5	1.09	21.0	0.99*	1.03	2.70	5.65	1.00
0.7	1.09	21.2	0.99*	1.03	2.97	5.23	1.00
0.9	1.09	21.0	0.99*	1.03	3.83	4.00	1.00
β_0: 0.1, β_1: 0.5							
0.5	2.37	834	1.00*	-	-	-	-
0.7	2.46	886	1.00*	1.25	4.10	31.8	1.00
0.9	2.44	881	1.00*	*1.25*	*2.28*	*60.9*	*1.00*
β_0: 0.05, β_1: 0.5							
0.5	1.49	201	1.00*	1.13	7.91	8.02	1.00
0.7	1.46	193	0.99*	-	-	-	-
0.9	1.50	205	1.00*	1.11	2.30	25.5	1.00
β_0: 0.05, β_1: 1							
0.5	1.22	68.9	0.99*	1.06	3.13	9.40	1.00
0.7	1.22	69.6	0.99*	1.07	5.49	5.70	1.00
0.9	1.21	68.4	0.99*	1.05	2.67	10.9	1.00
β_0: 0.05, β_1: 2							
0.5	1.11	28.6	0.99*	1.03	2.70	5.64	1.00
0.7	1.11	28.9	0.99*	1.03	2.97	5.23	1.00
0.9	1.10	27.1	0.99*	1.03	3.83	4.00	1.00

* statistically significant ($p < .05$) cubic polynomial pattern in the residuals

\- indeterminate fit

Italics indicate that the fit did not converge

Table 6

Best-fit parameter values and goodness-of-fit statistics for fits of classic and exponentiated Mazur functions to conditioned stimulus strength values, V, for different a.

	Classic Mazur			Exponentiated Mazur			
a	*a*	*b*	R^2	*a*	*b*	*s*	R^2
β_0: 0.05, β_1: 1, α: 0.7							
1	1.22	70.0	0.99*	1.09	7.60	4.51	1.00
1.25	1.03	23.0	0.99	1.04	24.4	0.97	1.00
1.5	0.95	8.52	0.99*	0.99	23.8	0.57	1.00*
1.75	0.92	4.00	0.98*	0.98	37.9	0.31	1.00*
2	0.92	2.32	0.96*	0.97	42.3	0.22	1.00*

* statistically significant ($p < .05$) cubic polynomial pattern in the residuals

Table 7

Rescorla-Wagner parameters, terminal link RI schedule values, parameters of the best-fitting hyperbola, and fit statistics (R^2) for fits of the classic and modern matching functions to initial link response-reinforcement data.

	Classic hyperbola Parameters			Modern hyperbola parameters			
Model Parameters and Terminal Link Schedules	k	r_e	R^2	k	a	$\frac{r_e^a}{b}$	R^2
β_0: 0.01, β_1: 0.25							
RI 1	250	34	1.00*	306	0.81	26	1.00
RI 5	242	37	1.00	303	0.85	35	1.00*
RI 25	222	52	0.99	258	0.89	46	0.99
RI 112	223	103	1.00	-	-	-	-
β_0: 0.01, β_1: 1.0							
RI 1	252	31	1.00*	283	0.86	25	1.00
RI 5	239	28	1.00	260	0.89	24	1.00
RI 25	214	25	0.99	233	0.90	21	0.99
RI 112	238	35	0.98	266	0.91	34	0.98
β_0: 0.005, β_1:0.5							
RI 1	253	32	1.00*	294	0.84	25	1.00
RI 5	250	32	1.00*	301	0.85	28	1.00
RI 25	239	33	1.00	257	0.92	29	1.00
RI 112	251	57	0.98*	527	0.70	57	0.99*
β_0: 0.05, β_1: 0.5							
RI 1	237	32	1.00*	268	0.86	26	1.00
RI 5	241	42	0.99	320	0.79	33	1.00
RI 25	239	59	1.00	296	0.87	52	1.00
RI 112	252	89	0.99	444	0.80	92	0.99
β_0:0.05,β_1:0.5, a:1.5							
RI 1	249	38	1.00	309	0.81	29	1.00
RI 5	242	38	1.00	303	0.80	29	1.00
RI 25	235	53	1.00	261	0.89	43	1.00
RI 112	159	44	0.94	202	0.86	41	0.94

* statistically significant ($p < .05$) cubic polynomial pattern in the residuals

\- indeterminate fit

Table 8

Best-fit parameter values from fits of classic and modern matching functions to model response-reinforcement rate data for different values of proportion of behaviors produced using a fitness function following reinforcement.

Proportion Produced Using Fitness Function	Classic hyperbola Parameters			Modern hyperbola parameters			
	k	r_e	R^2	k	a	$\frac{r_e^a}{b}$	R^2
20	151	83	99.0	225	0.85	84	0.99
30	186	69	1.00	211	0.92	63	1.00
40	231	72	1.00*	304	0.83	59	1.00
50	226	55	1.00	241	0.94	49	1.00
60	234	48	1.00*	267	0.87	37	1.00
70	232	39	1.00*	262	0.85	30	1.00
80	239	37	1.00*	266	0.86	29	1.00*
90	245	35	1.00*	279	0.83	26	1.00
100	244	31	1.00*	278	0.82	22	1.00

* statistically significant ($p < .05$) cubic polynomial pattern in the residuals

Figure Captions

Figure 1. Schematic illustrating development of primary and conditioned reinforcement in modified McDowell computational model implemented in a chained schedule.

Figure 2. Plot of the exponentiated Mazur function for values of s and r with parameters a and b set to 1.1 and 5, respectively.

Figure 3. The effect of reinforcement rate (r) on conditioned stimulus strength (V) for 3 different values of α under the model conditions β_0: 0.01, β_1: 0.5, a: 1.0.

Figure 4. Instantaneous values of conditioned stimulus strength, V, for the first 500 time ticks. Top panel shows results for $\alpha = 0.5$, middle panel for $\alpha = 0.7$, and bottom panel for $\alpha = 0.9$.

Figure 5. Mean and standard deviation of the conditioned stimulus strength, V, for three experiments using $\alpha = 0.5$, $\alpha = 0.7$, and $\alpha = 0.9$.

Figure 6. The effect of reinforcement rate on conditioned stimulus strength (V) for 5 different values of β_0 under the model conditions α: 0.7, β_1: 0.25, a: 1.0.

Figure 7. The effect of reinforcement rate on conditioned stimulus strength (V) for 5 different values of β_1 under the model conditions α: 0.7, β_0: 0.05, a: 1.0.

Figure 8. The effect of reinforcement rate on conditioned stimulus strength (V) for 5 different values of β_1 under the model conditions α: 0.7, β_0: 0.01, a: 1.0.

Figure 9. The effect of reinforcement rate on conditioned stimulus strength (V) for 5 different values of a under the model conditions β_0: 0.05, β_1: 1.0, α: 0.7.

Figure 10. Instantaneous values of conditioned stimulus strength, V, for the first 500 time ticks. Top panel shows results for $a = 1$, middle panel for $a = 1.5$, and bottom panel for $a = 2$ (β_0: 0.05, β_1: 1.0, α: 0.7).

Figure 11. Standardized residuals (vertical axis) versus predicted values of *V* (horizontal axis) from fits of the classic Mazur function for random interval schedules. Model conditions are detailed below each plot.

Figure 12. Standardized residuals (vertical axis) versus predicted values of *V* (horizontal axis) from fits of the exponentiated Mazur function for random interval schedules. Model conditions are detailed below each plot.

Figure 13. Standardized residuals (vertical axis) versus predicted values of *V* (horizontal axis) from fits of the Mazur function for random interval schedules. Model conditions are detailed below each plot.

Figure 14. Standardized residuals (vertical axis) versus predicted values of *V* (horizontal axis) from fits of the exponentiated Mazur function for random interval schedules. Model conditions are detailed below each plot.

Figure 15. Standardized residuals (vertical axis) versus predicted values of *V* (horizontal axis) from fits of the Mazur function for random interval schedules. Model conditions are detailed below each plot.

Figure 16. Standardized residuals (vertical axis) versus predicted values of *V* (horizontal axis) from fits of the exponentiated Mazur function for random interval schedules. Model conditions are detailed below each plot.

Figure 17. Standardized residuals (vertical axis) versus predicted values of *V* (horizontal axis) from fits of the Mazur and exponentiated Mazur functions for random interval schedules under 4 different values of the exponentiated Rescorla-Wagner exponent, a. For all plots the Rescorla-Wagner parameters used were: β_0: 0.05, β_1: 1, α: 0.7.

Figure 18. Response and reinforcement rate plots for different proportions of behaviors produced for the next time-step using the parental fitness function after reinforcement (as opposed to production of all new behaviors using the fitness function to select parent behaviors). The percentage of behavior replaced using the linear fitness function is denoted by the legend on the right.

Figure 19. Standardized residuals (vertical axis) versus predicted values of B (horizontal axis) from fits of the classic (top two rows) and modern (bottom two rows) matching functions to reinforcement-response rate data.

Figure 20. Cumulative distribution plots showing responses and reinforcements for the first 1000 time-steps of both the initial (top) and terminal (bottom) links for a chain RI1-RI1 schedule.

Figure 21. Cumulative distribution plot showing responses and reinforcements for the first 1000 time-steps of both the initial (top) and terminal (bottom) links for a chain RI5-RI5 schedule.

Figure 22. Cumulative distribution plots showing responses and reinforcements for the first 1000 time-steps of both the initial (top) and terminal (bottom) links for a chain RI25-RI25 schedule.

Figure 23. Cumulative distribution plots showing responses and reinforcements for the first 1000 time-steps of both the initial (top) and terminal (bottom) links for a chain RI112-RI112 schedule.

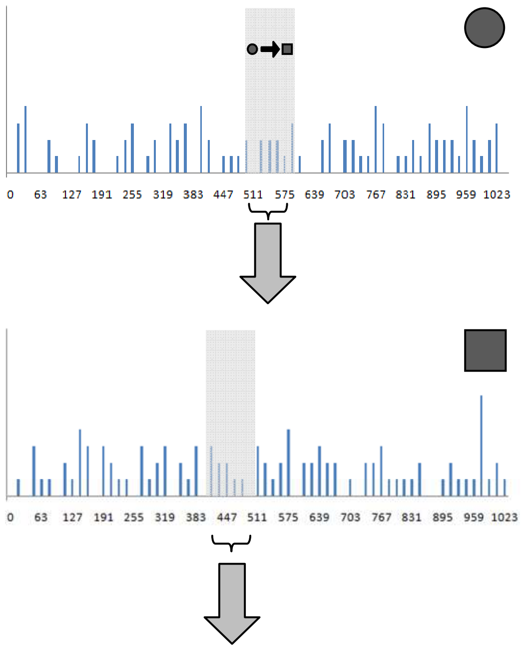
Figure 24. The initial link reinforcement and response rates for 4 constant terminal link reinforcement rates under the Rescorla-Wagner conditions β_0: 0.05, β_1: 0.5, α: 0.7, a: 1.0.

Figure 25. Initial to terminal link response rate ratios. Model conditions: β_0: 0.05, β_1: 0.5, α: 0.7, a: 1.0.

Figure 26. Initial link response rates (B) for 22 experiments varying the initial and terminal link schedules from RI1-RI1 to RI22-RI22 while maintaining a 1 time-tick delay to reinforcement in the initial link. Operant level of response rate for the computational model is denoted for comparison.

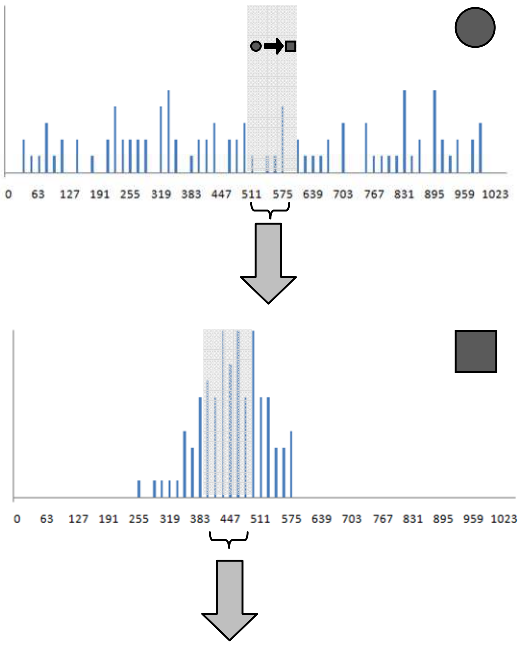
Figure A1. Random distribution of 6 datapoints fit with quadratic (top panel), cubic (middle panel), and a 6th order polynomial (bottom panel) illustrating the increasing percent variance accounted for (pVAF) as polynomial order approaches the number of datapoints.

Time Step #1



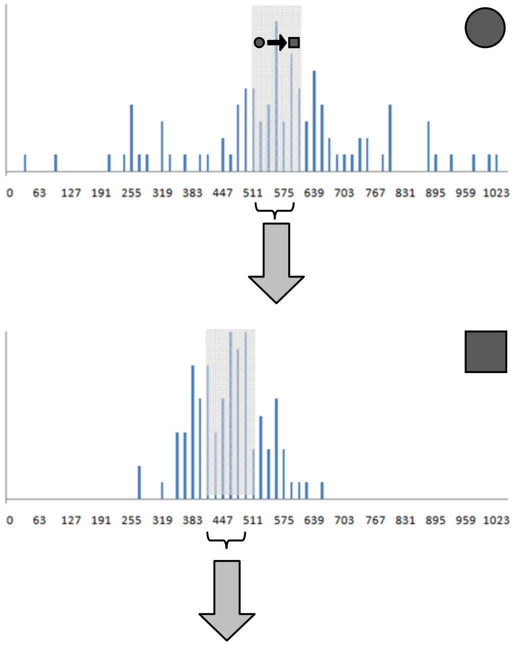
S^{r+}

Time Step #2

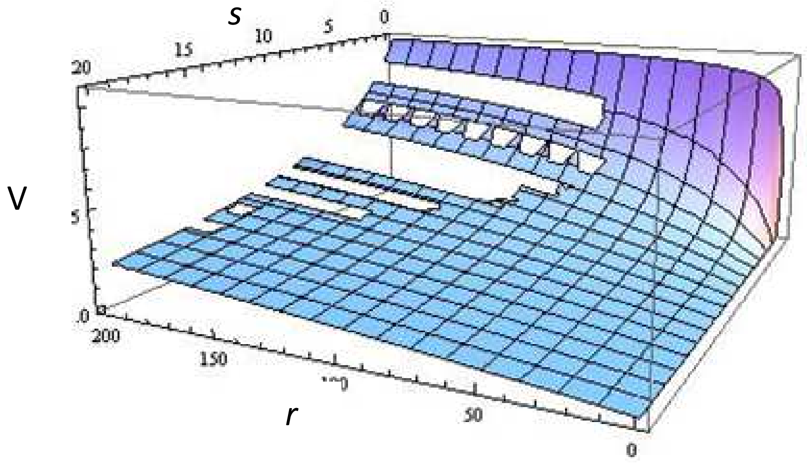


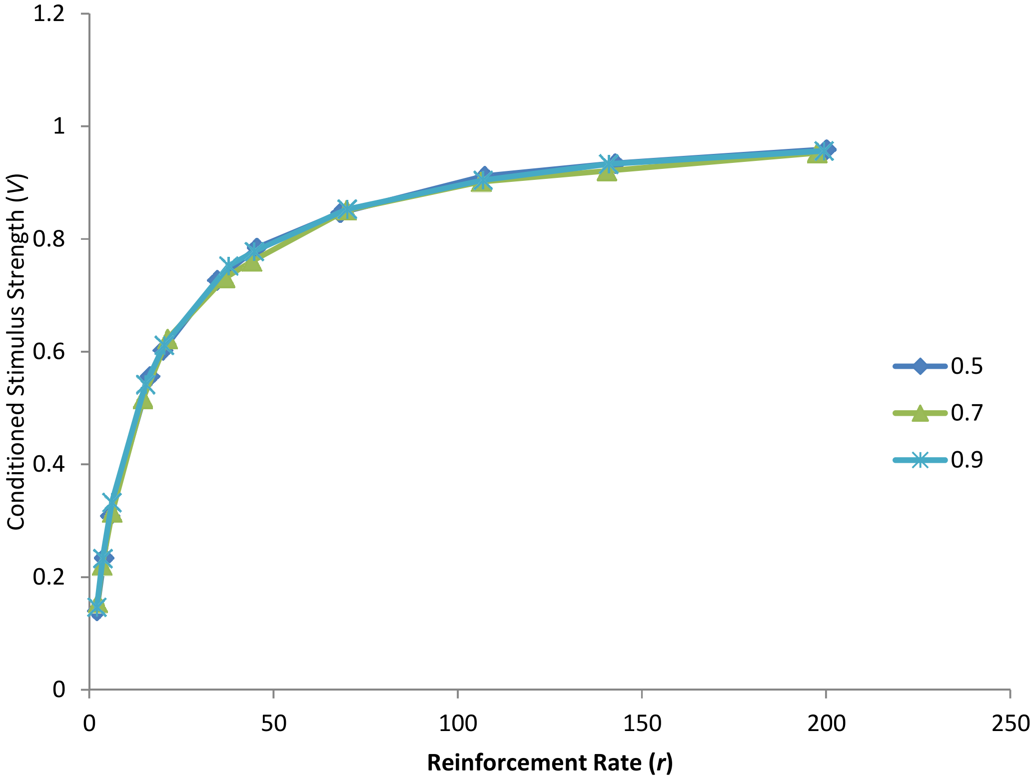
S^{r+}

Time Step #3

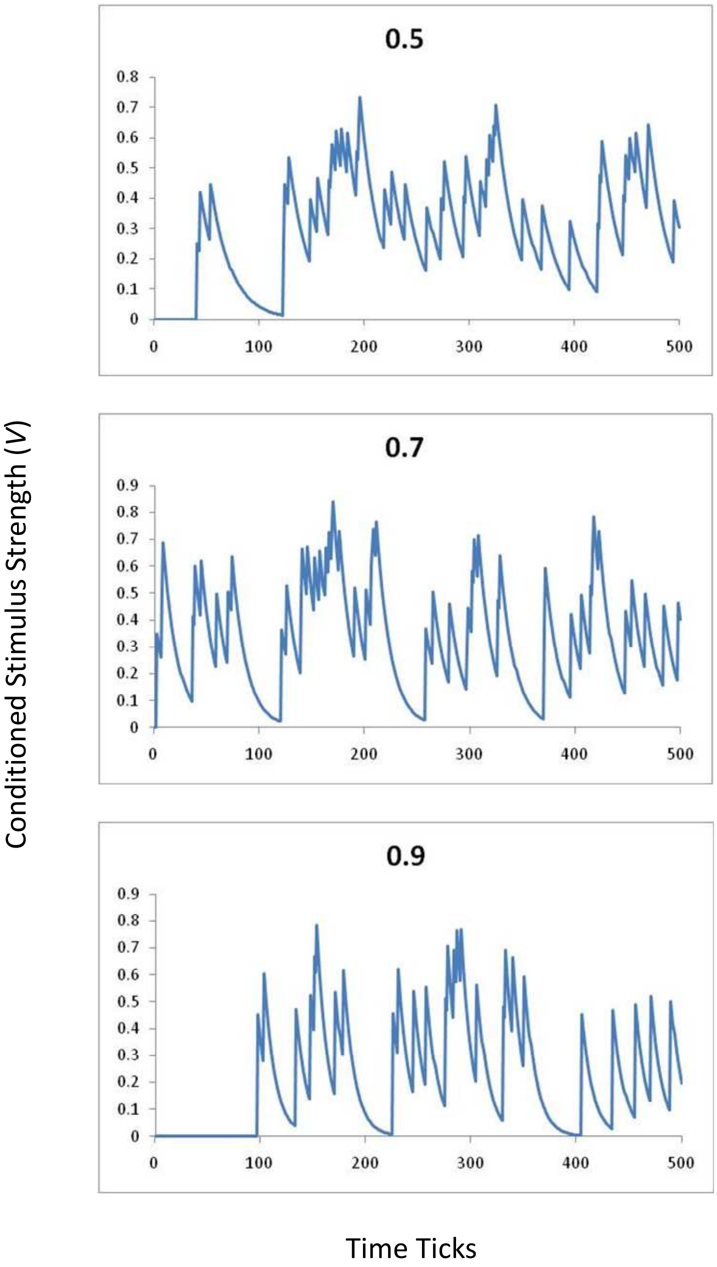


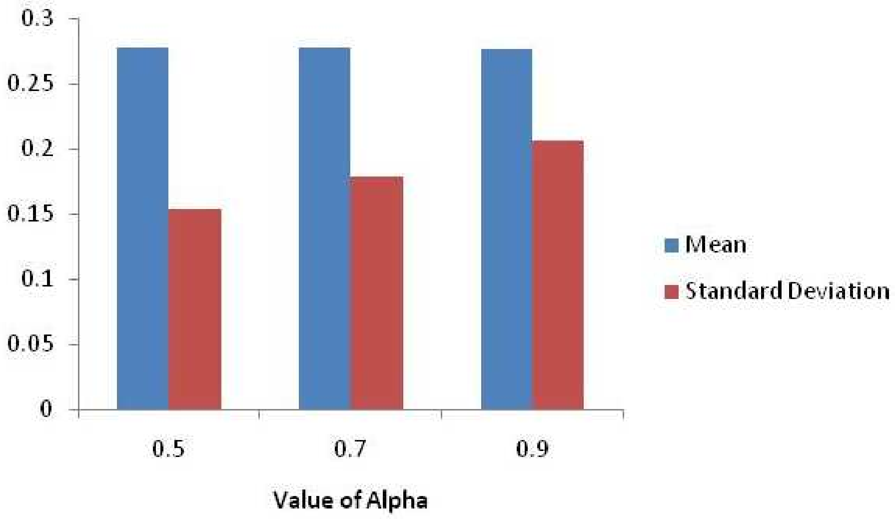
S^{r+}

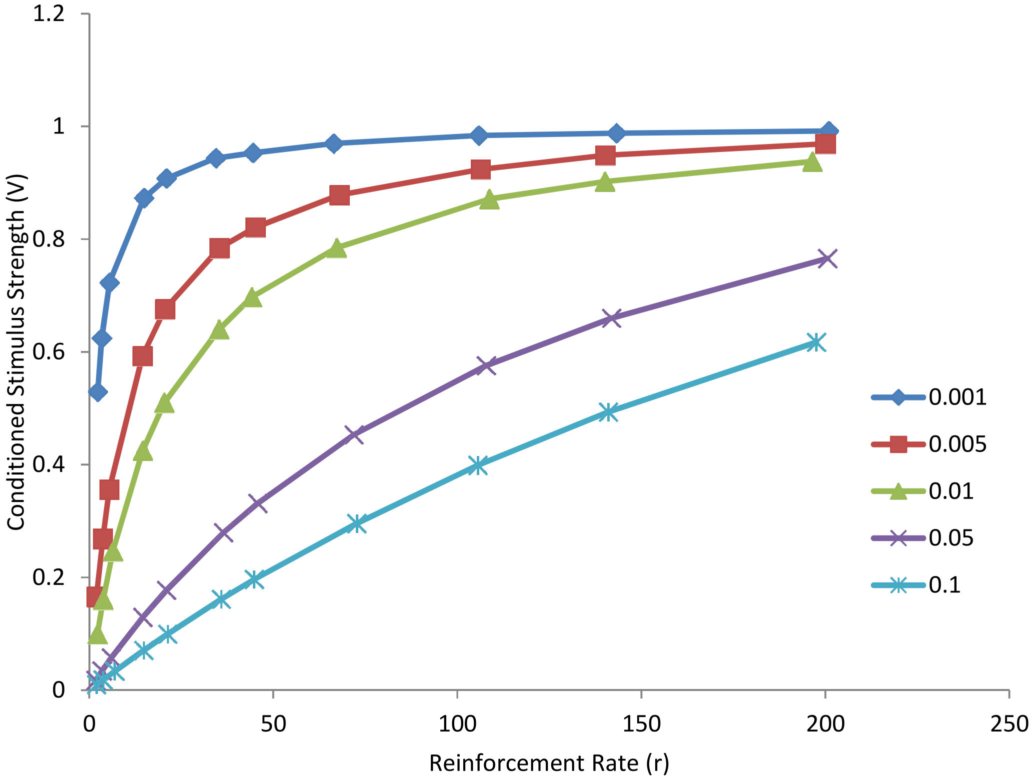

s
0
5
10
15
20
V
5
0
200
150
100
50
0
r

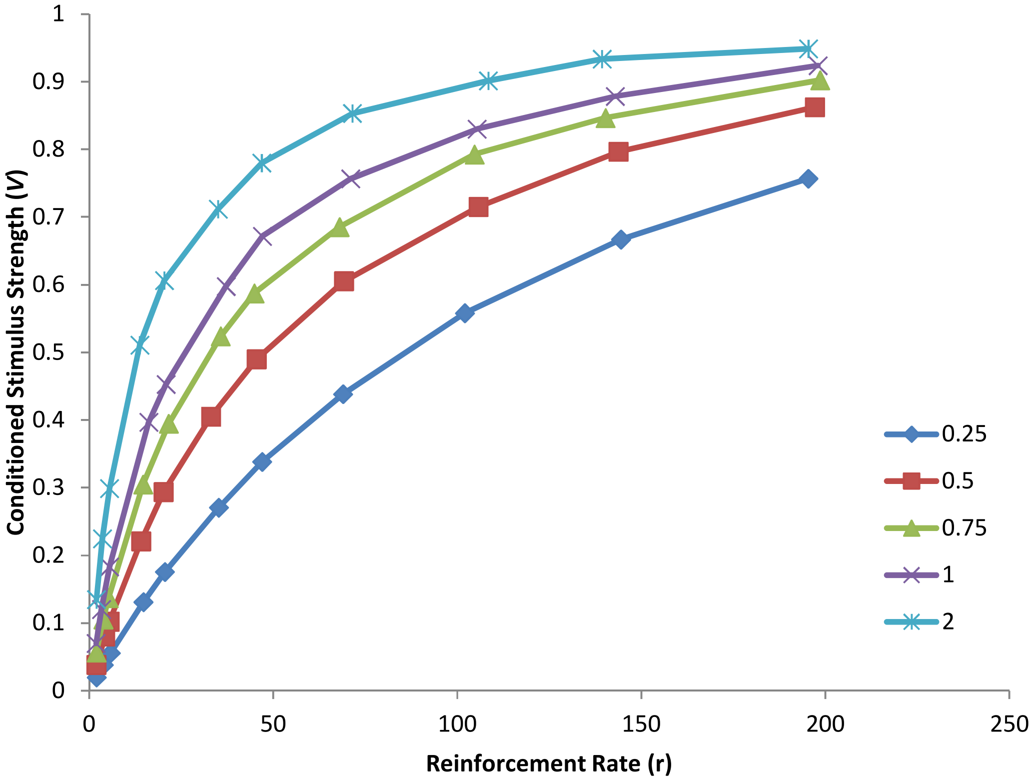


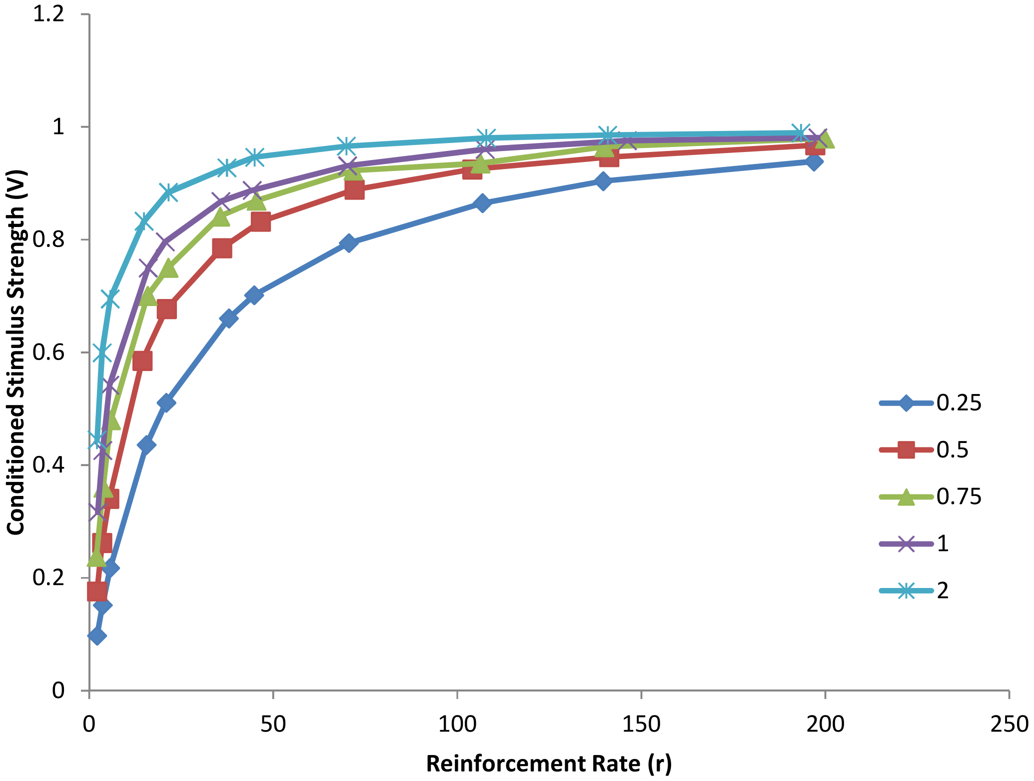
1.2
1
0.8
0.6
0.4
0.2
0
Conditioned Stimulus Strength (V)
0
50
100
150
200
250
Reinforcement Rate (r)
0.5
0.7
0.9

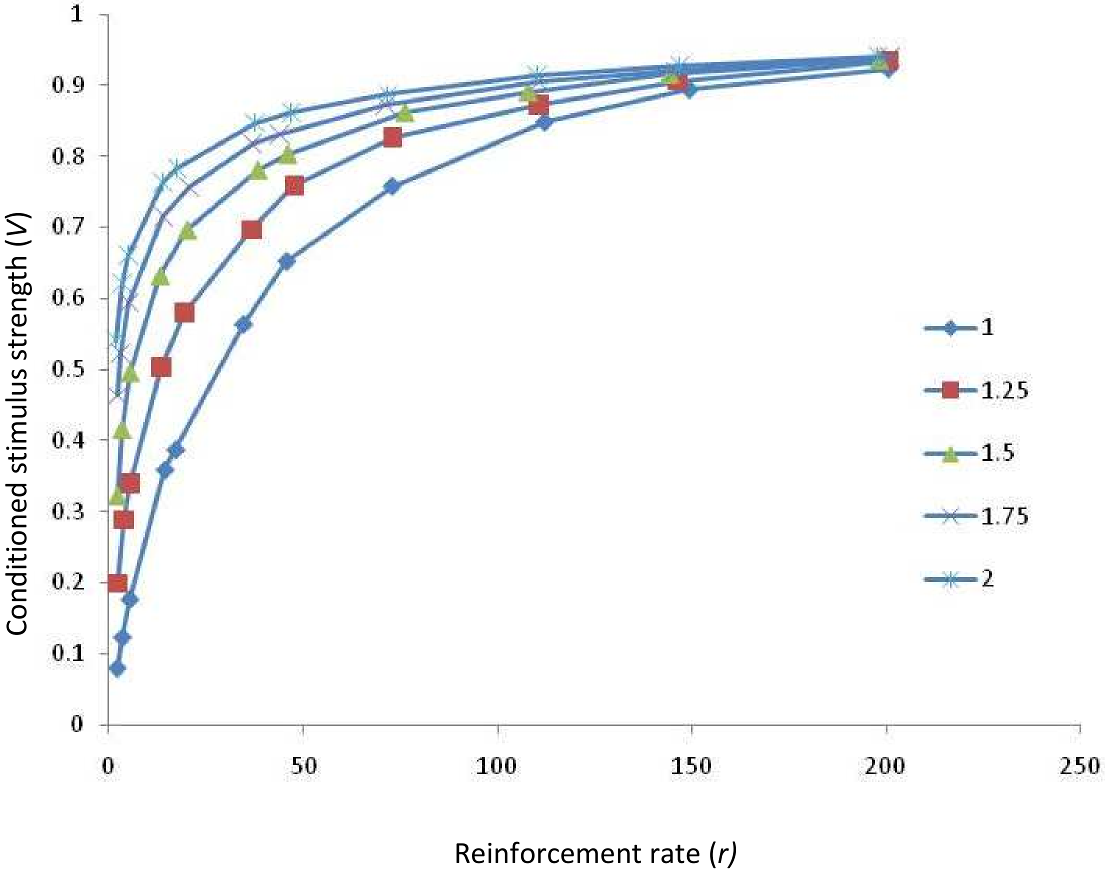

0.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0
100
200
300
400
500
0.7
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0
100
200
300
400
500
0.9
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0
100
200
300
400
500
Conditioned Stimulus Strength (V)
Time Ticks

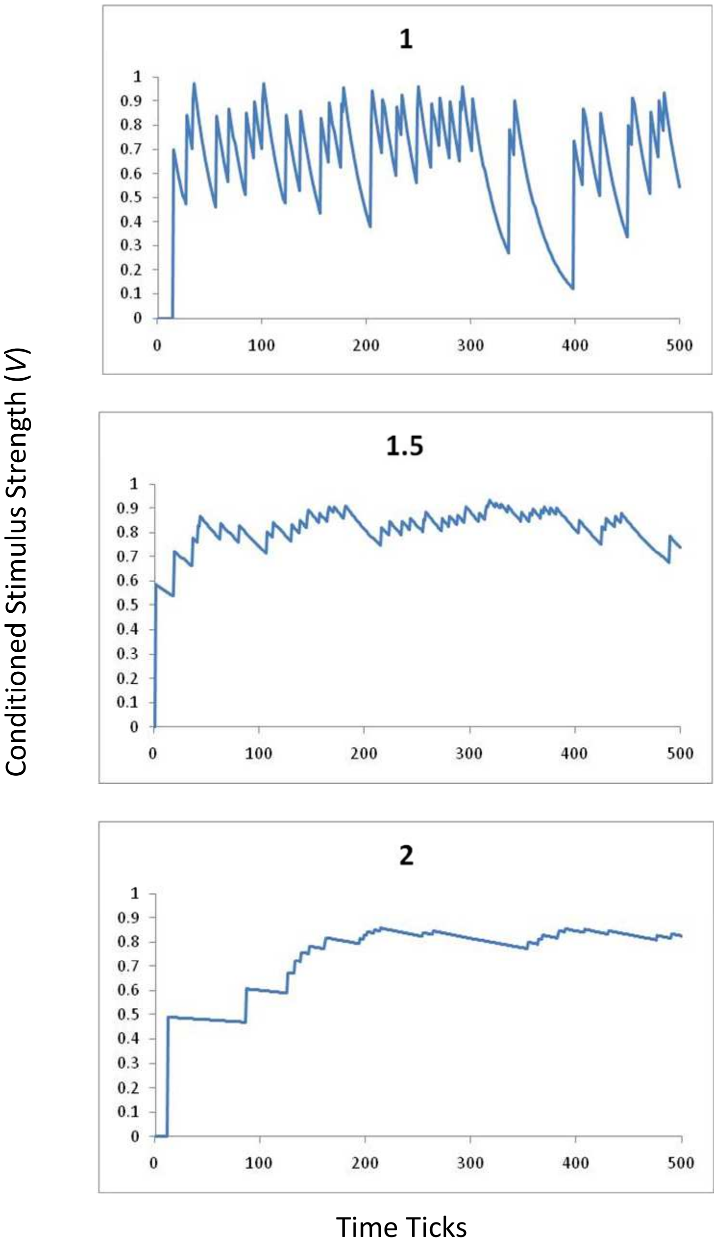

0.3
0.25
0.2
0.15
0.1
0.05
0
0.5
0.7
0.9
Value of Alpha
Mean
Standard Deviation

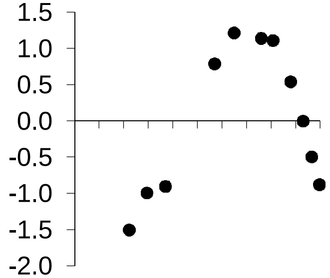

1.2
1
0.8
0.6
0.4
0.2
0
Conditioned Stimulus Strength (V)
0
50
100
150
200
250
Reinforcement Rate (r)
0.001
0.005
0.01
0.05
0.1


Conditioned Stimulus Strength (*V*)
Reinforcement Rate (r)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
50
100
150
200
250
0.25
0.5
0.75
1
2

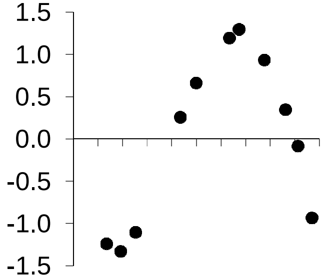

1.2
1
0.8
0.6
0.4
0.2
0
Conditioned Stimulus Strength (V)
0
50
100
150
200
250
Reinforcement Rate (r)
0.25
0.5
0.75
1
2


Conditioned stimulus strength (V)
Reinforcement rate (r)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
50
100
150
200
250
1
1.25
1.5
1.75
2

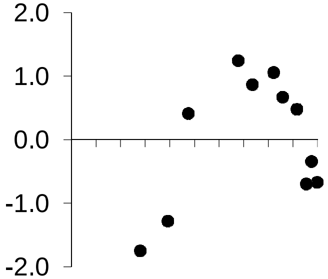

1
1.5
2
Conditioned Stimulus Strength (*V*)
Time Ticks
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100
200
300
400
500



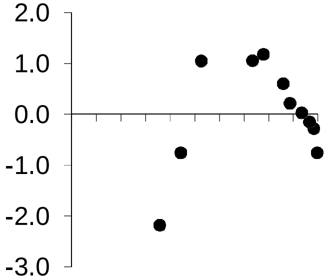
β_0: 0.01, β_1: 0.5



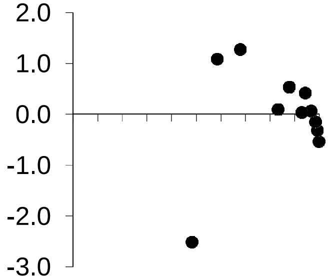
β_0: 0.01, β_1: 0.25



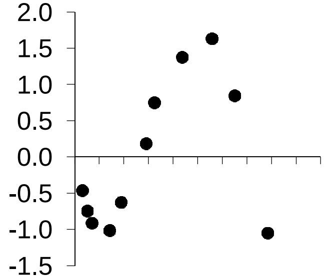
β_0: 0.01, β_1: 0.75



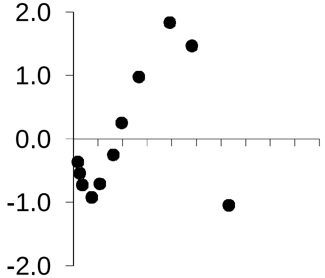
β_0: 0.01, β_1: 1



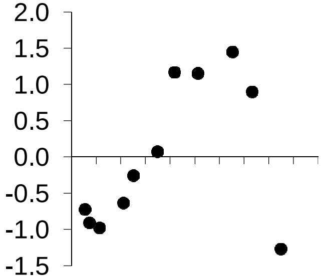
β_0: 0.01, β_1: 2



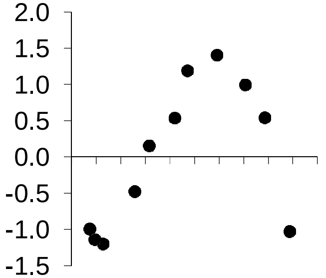
β_0: 0.1, β_1: 0.5



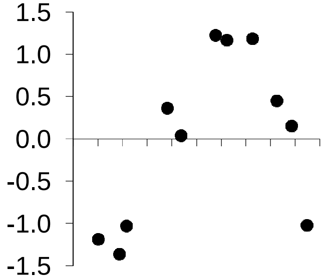
β_0: 0.1, β_1: 0.25



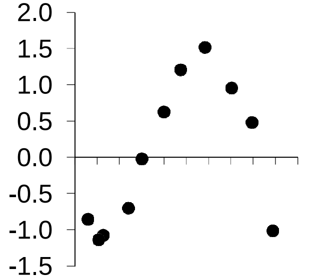
β_0: 0.1, β_1: 0.75



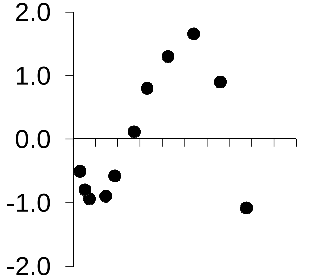
β_0: 0.1, β_1: 1



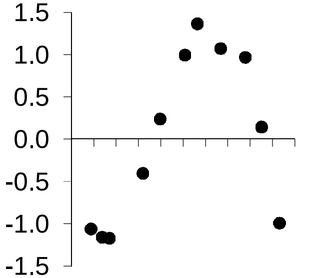
β_0: 0.1, β_1: 2



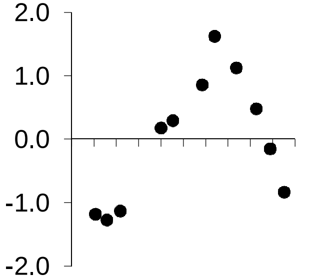
β_0: 0.05, β_1: 0.5



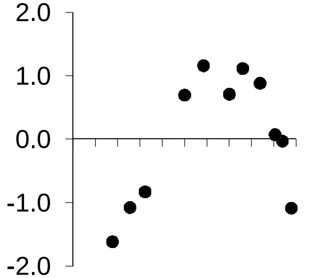
β_0: 0.05, β_1: 0.25



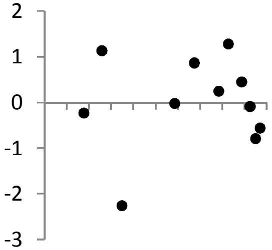
β_0: 0.05, β_1: 0.75



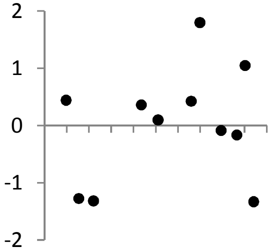
β_0: 0.05, β_1: 1



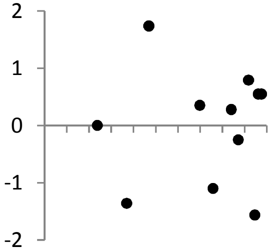
β_0: 0.05, β_1: 2



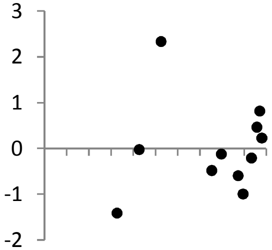
β_0: 0.01, β_1: 0.5



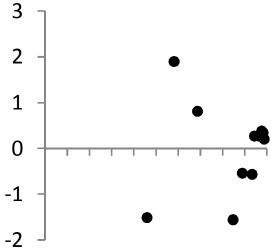
β_0: 0.01, β_1: 0.25



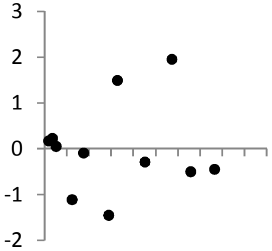
β_0: 0.01, β_1: 0.75



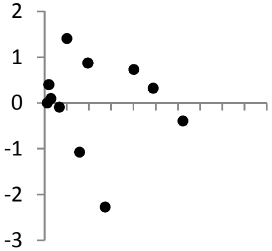
β_0: 0.01, β_1: 1



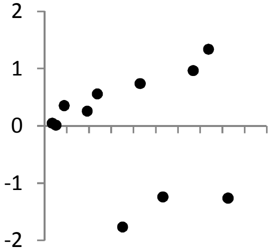
β_0: 0.01, β_1: 2



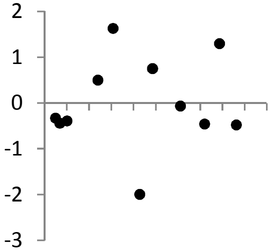
β_0: 0.1, β_1: 0.5



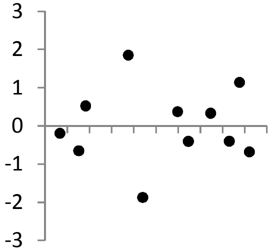
β_0: 0.1, β_1: 0.25



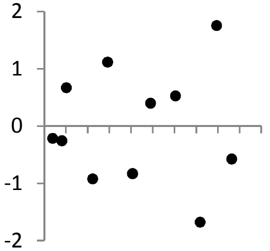
β_0: 0.1, β_1: 0.75



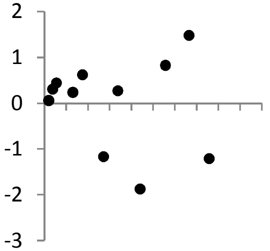
β_0: 0.1, β_1: 1



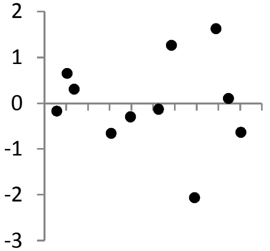
β_0: 0.1, β_1: 2



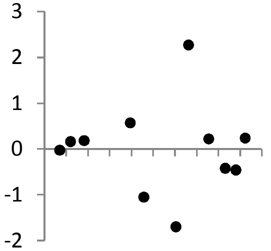
β_0: 0.05, β_1: 0.5



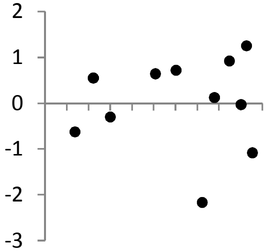
β_0: 0.05, β_1: 0.25



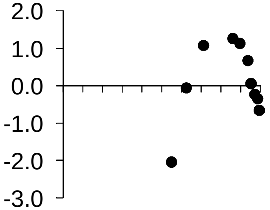
β_0: 0.05, β_1: 0.75



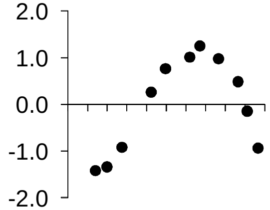
β_0: 0.05, β_1: 1



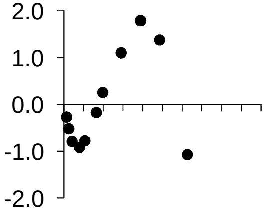
β_0: 0.05, β_1: 2



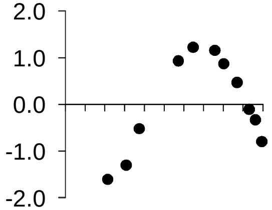
$?_0$: 0.001, $?_1$: 0.25



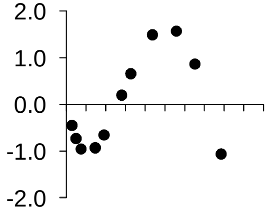
$?_0$: 0.01, $?_1$: 0.25



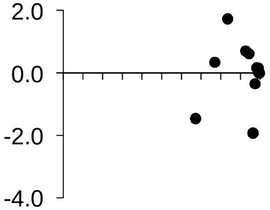
$?_0$: 0.1, $?_1$: 0.25



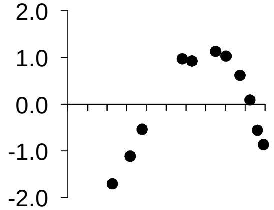
$?_0$: 0.005, $?_1$: 0.25



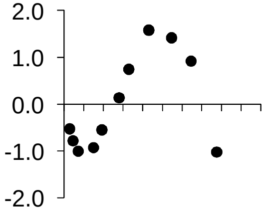
$?_0$: 0.05, $?_1$: 0.25



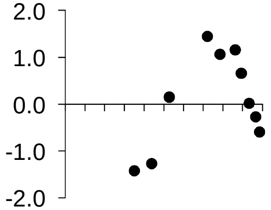
$?_0$: 0.001, $?_1$: 0.5



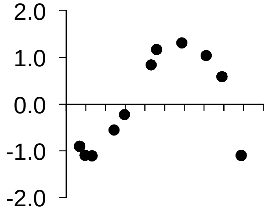
$?_0$: 0.01, $?_1$: 0.5



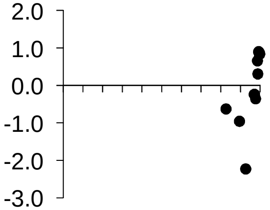
$?_0$: 0.1, $?_1$: 0.5



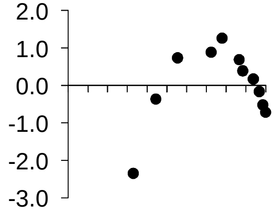
$?_0$: 0.005, $?_1$: 0.5



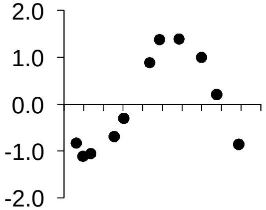
$?_0$: 0.05, $?_1$: 0.5



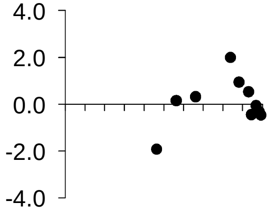
$?_0$: 0.001, $?_1$: 1



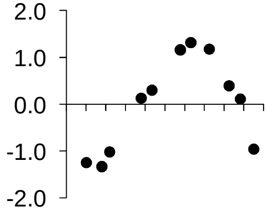
$?_0$: 0.01, $?_1$: 1



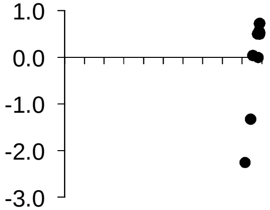
$?_0$: 0.1, $?_1$: 1



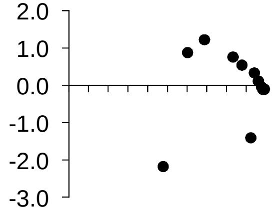
$?_0$: 0.005, $?_1$: 1



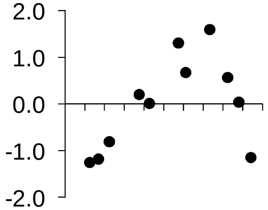
$?_0$: 0.05, $?_1$: 1



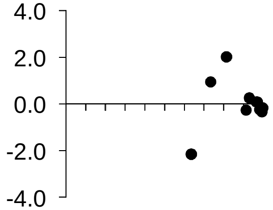
$?_0$: 0.001, $?_1$: 2



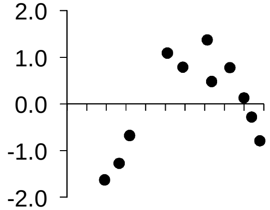
$?_0$: 0.01, $?_1$: 2



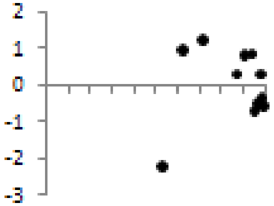
$?_0$: 0.1, $?_1$: 2



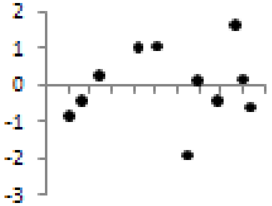
$?_0$: 0.005, $?_1$: 2



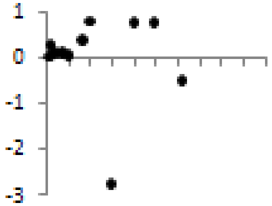
$?_0$: 0.05, $?_1$: 2



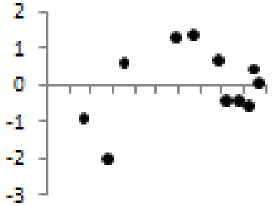
B_0: 0.001, B_1: 0.25



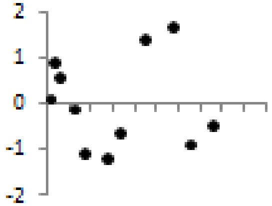
B_0: 0.01, B_1: 0.25



B_0: 0.1, B_1: 0.25



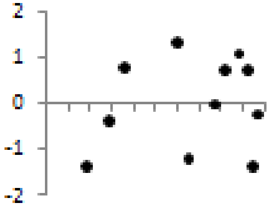
B_0: 0.005, B_1: 0.25



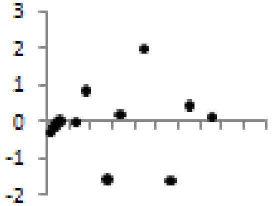
B_0: 0.05, B_1: 0.25

Indeterminate

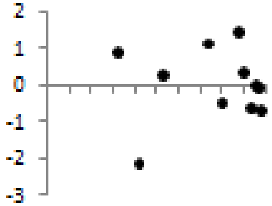
B_0: 0.001, B_1: 0.5



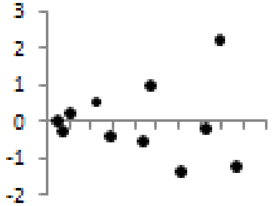
B_0: 0.01, B_1: 0.5



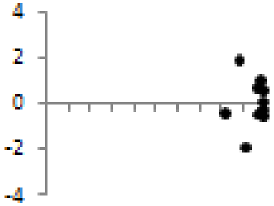
B_0: 0.1, B_1: 0.5



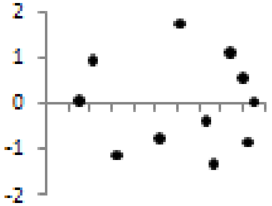
B_0: 0.005, B_1: 0.5



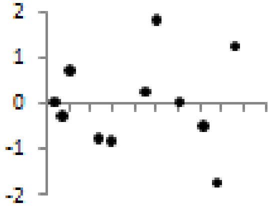
B_0: 0.05, B_1: 0.5



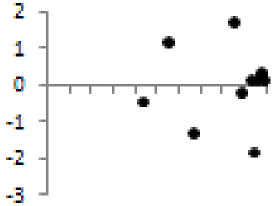
B_0: 0.001, B_1: 1



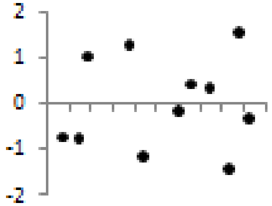
B_0: 0.01, B_1: 1



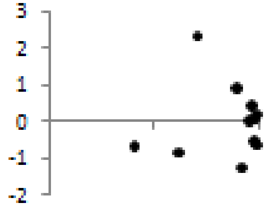
B_0: 0.1, B_1: 1



B_0: 0.005, B_1: 1



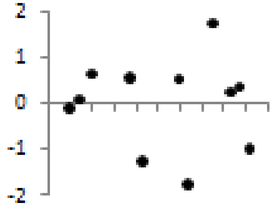
B_0: 0.05, B_1: 1



B_0: 0.001, B_1: 2

Indeterminate

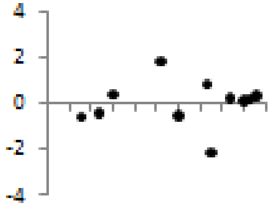
B_0: 0.01, B_1: 2



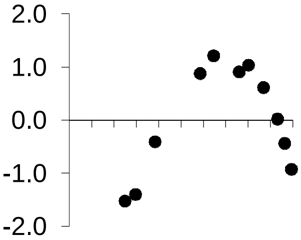
B_0: 0.1, B_1: 2

Indeterminate

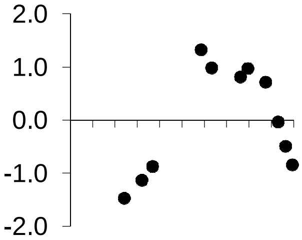
B_0: 0.005, B_1: 2



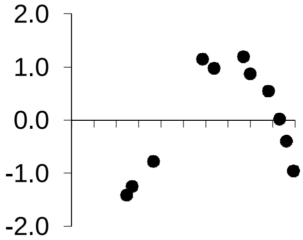
B_0: 0.05, B_1: 2



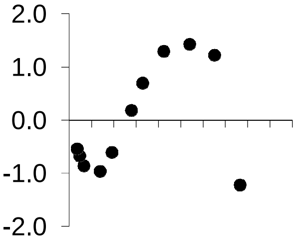
β_0: 0.01, β_1: 0.5,α: 0.5



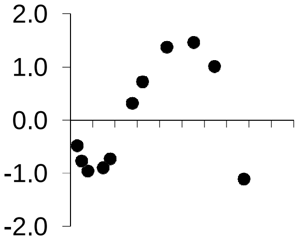
β_0: 0.01, β_1: 0.5, α: 0.7



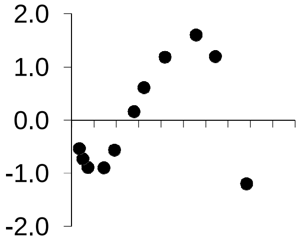
β_0: 0.01, β_1: 0.5, α: 0.9



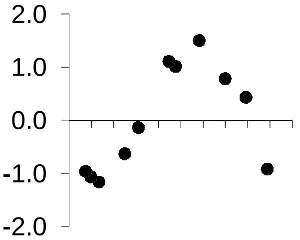
β_0: 0.1, β_1: 0.5, α: 0.5



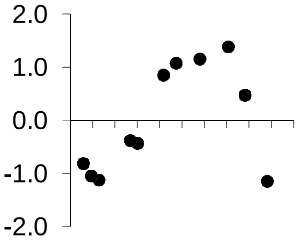
β_0: 0.1, β_1: 0.5, α: 0.7



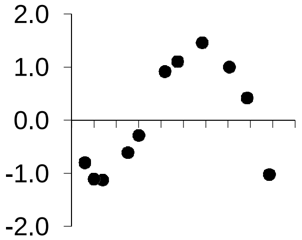
β_0: 0.1, β_1: 0.5, α: 0.9



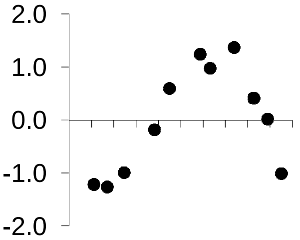
β_0: 0.05, β_1: 0.5, α: 0.5



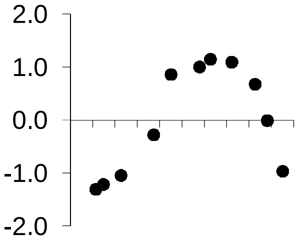
β_0: 0.05, β_1: 0.5, α: 0.7



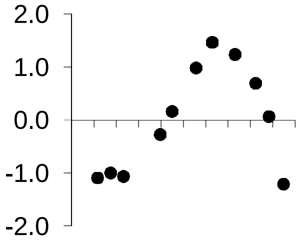
β_0: 0.05, β_1: 0.5, α: 0.9



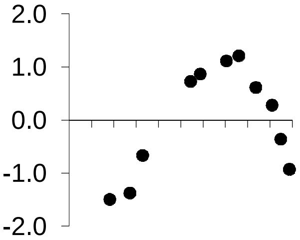
β_0: 0.05, β_1: 1, α: 0.5



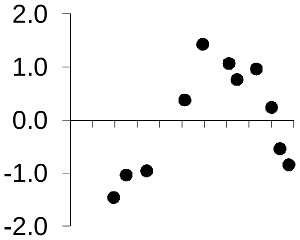
β_0: 0.05, β_1: 1, α: 0.7



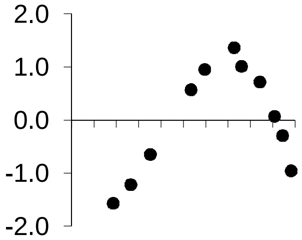
β_0: 0.05, β_1: 1, α: 0.9



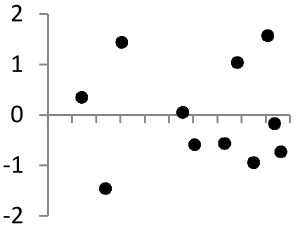
β_0: 0.05, β_1: 2, α: 0.5



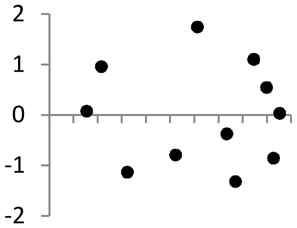
β_0: 0.05, β_1: 2, α: 0.7



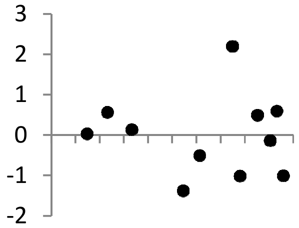
β_0: 0.05, β_1: 2, α: 0.9



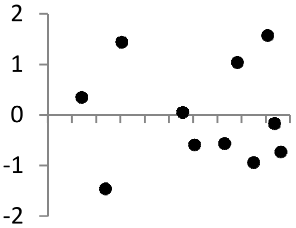
β_0: 0.01, β_1: 0.5,α: 0.5



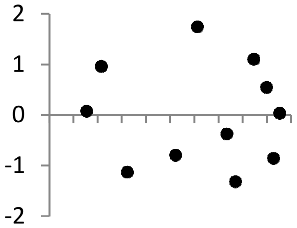
β_0: 0.01, β_1: 0.5, α: 0.7



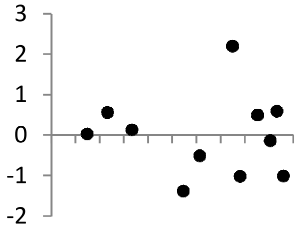
β_0: 0.01, β_1: 0.5, α: 0.9



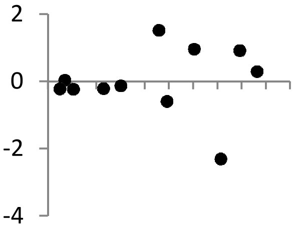
β_0: 0.1, β_1: 0.5, α: 0.5



β_0: 0.1, β_1: 0.5, α: 0.7



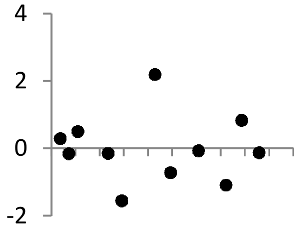
β_0: 0.1, β_1: 0.5, α: 0.9



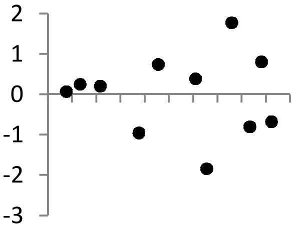
β_0: 0.05, β_1: 0.5, α: 0.5

Indeterminate

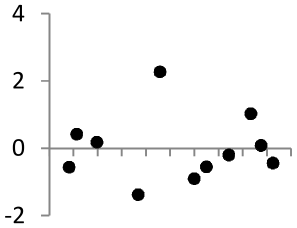
β_0: 0.05, β_1: 0.5, α: 0.7



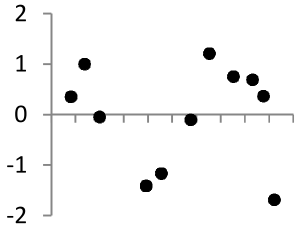
β_0: 0.05, β_1: 0.5, α: 0.9



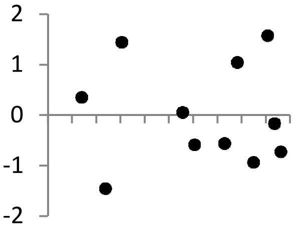
β_0: 0.05, β_1: 1, α: 0.5



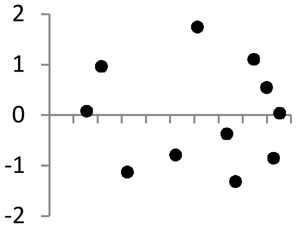
β_0: 0.05, β_1: 1, α: 0.7



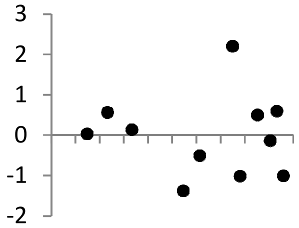
β_0: 0.05, β_1: 1, α: 0.9



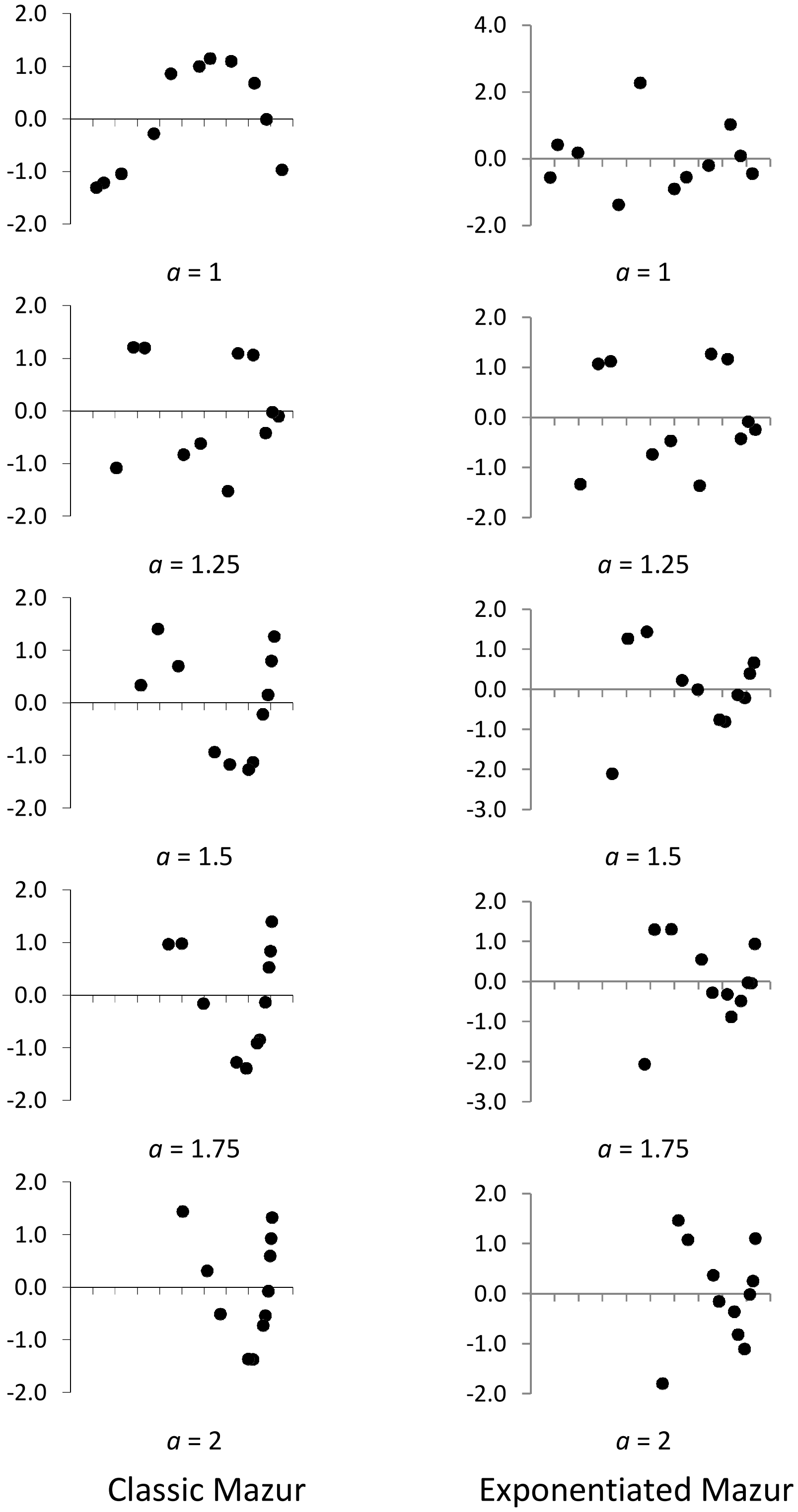
β_0: 0.05, β_1: 2, α: 0.5

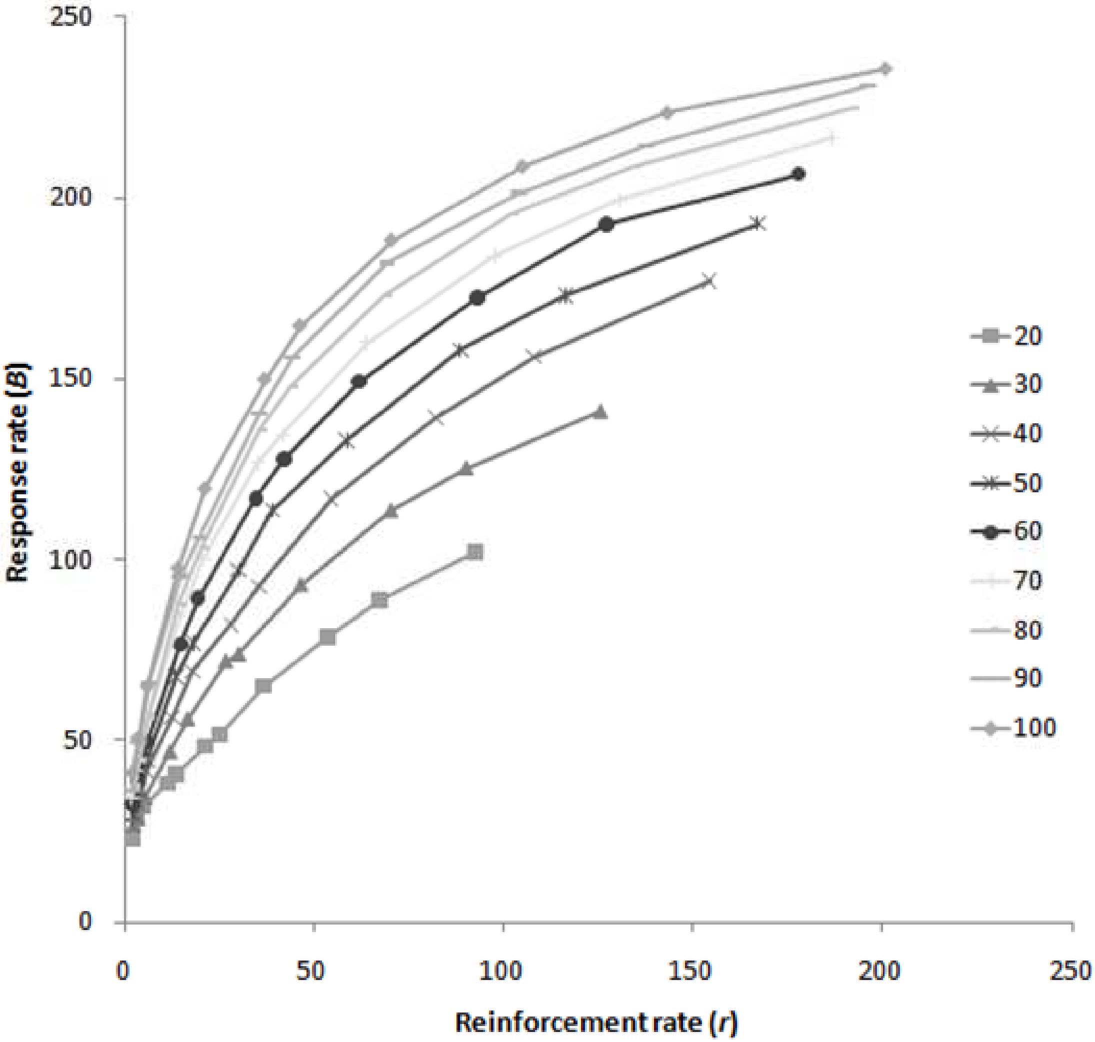


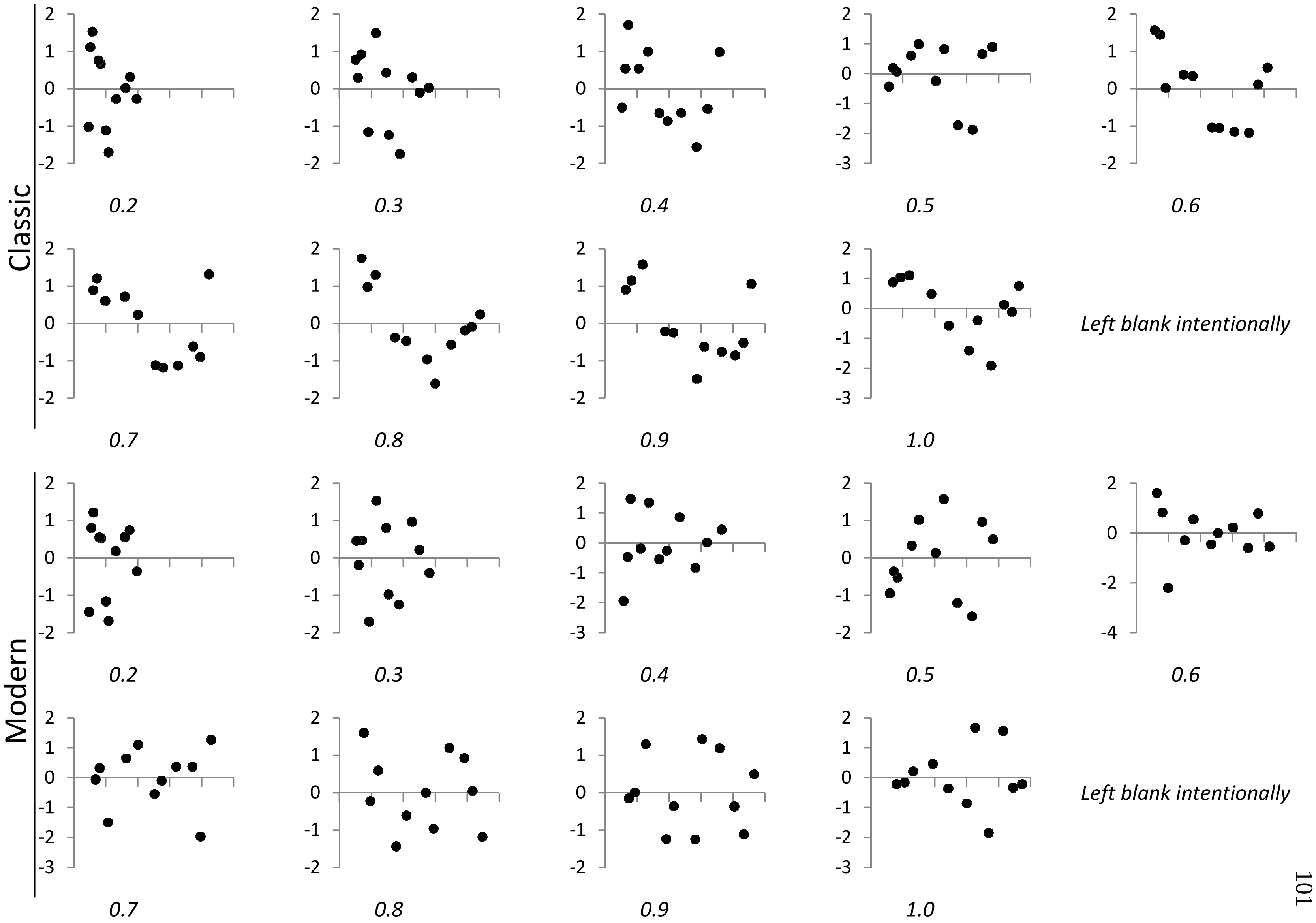
β_0: 0.05, β_1: 2, α: 0.7

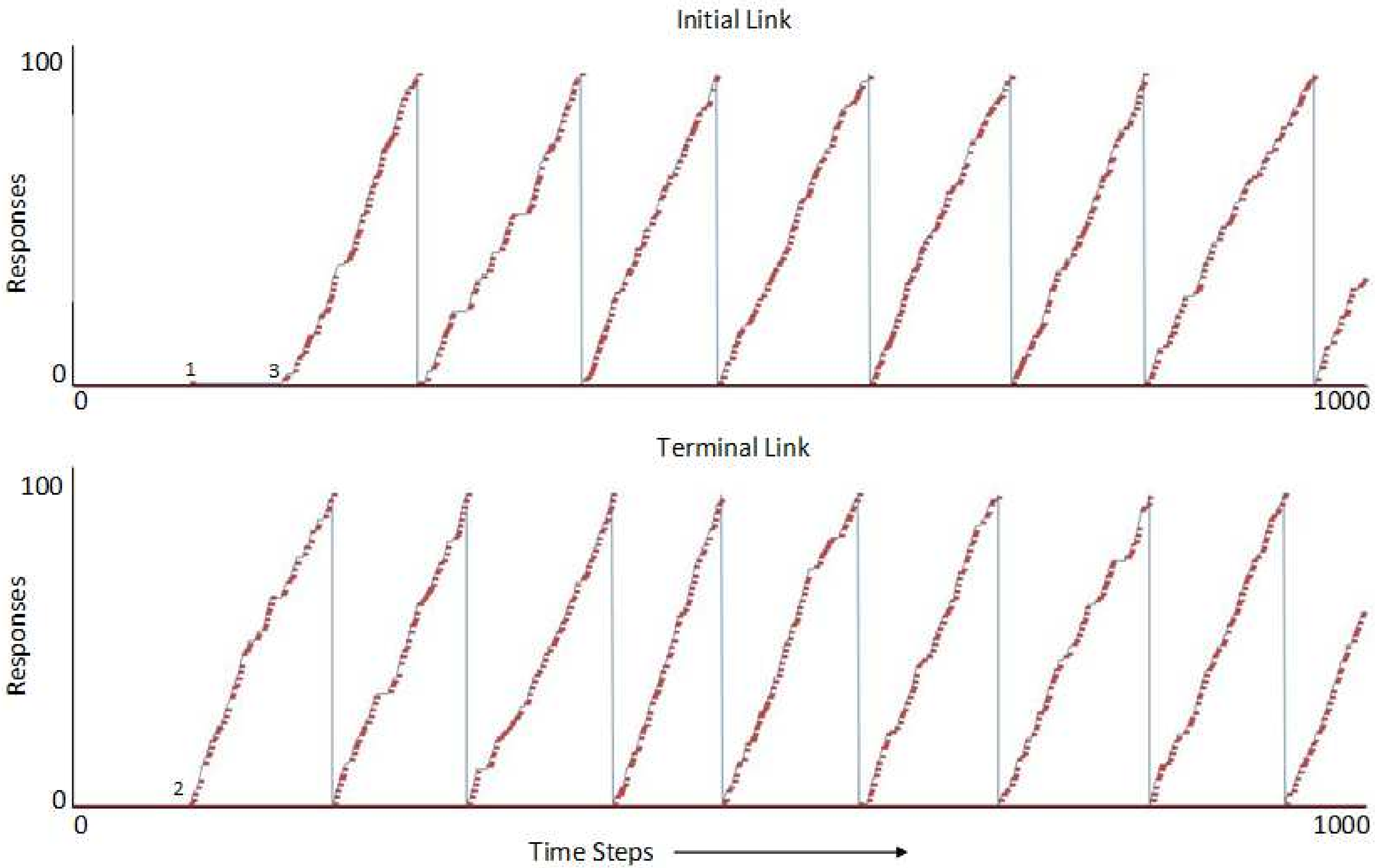


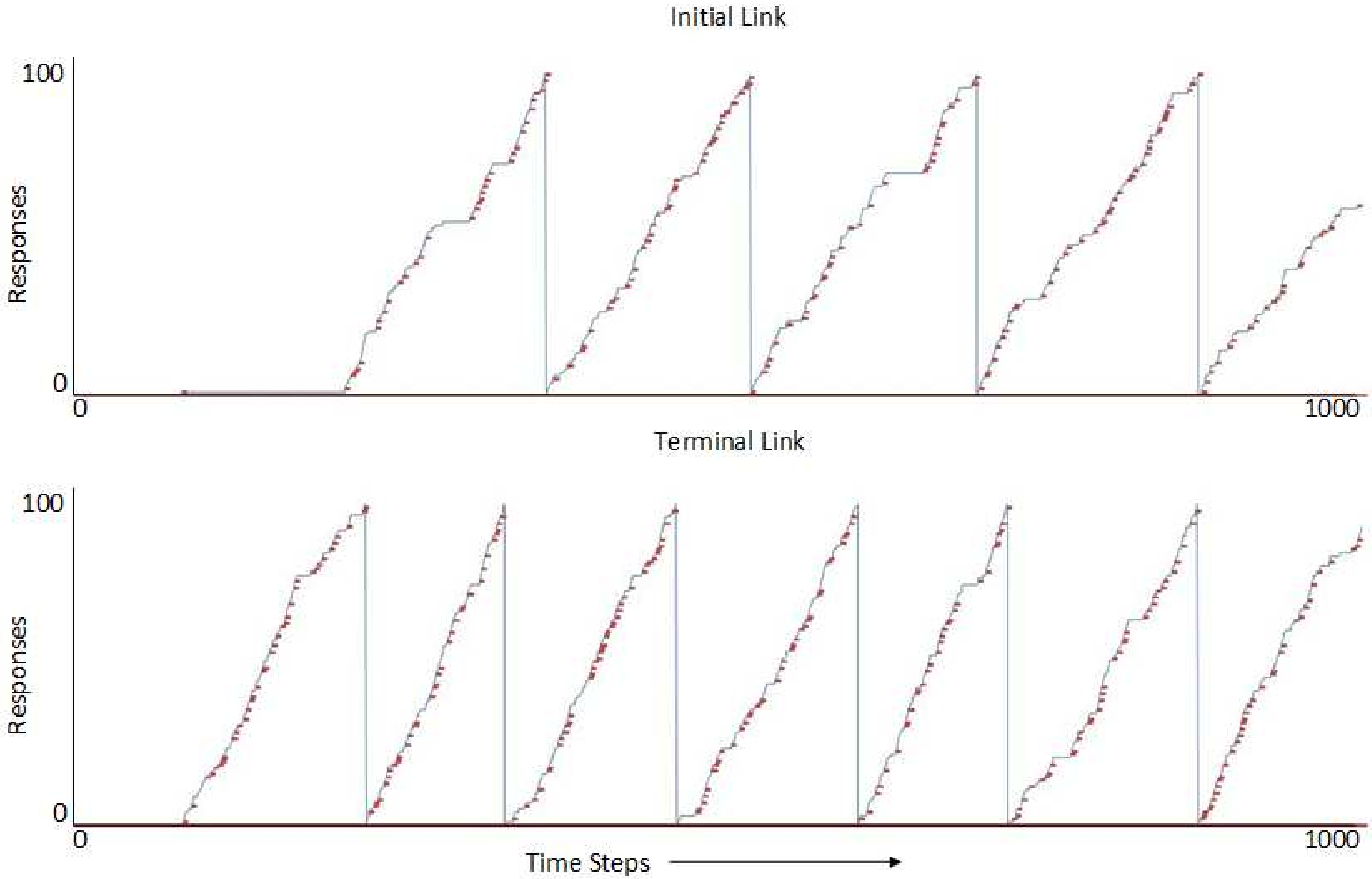
β_0: 0.05, β_1: 2, α: 0.9

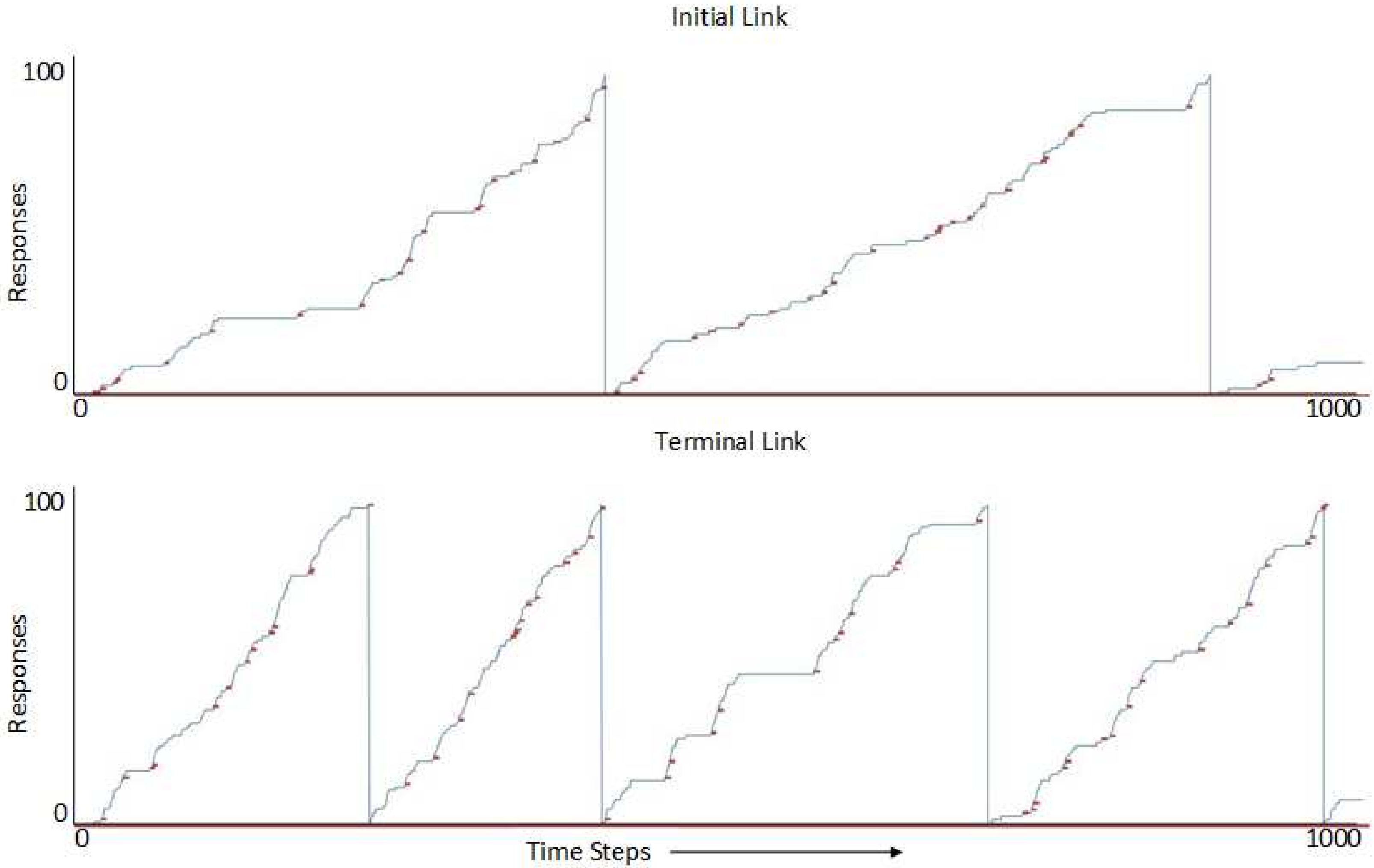

2.0
1.0
0.0
-1.0
-2.0
a = 1
4.0
2.0
0.0
-2.0
a = 1
2.0
1.0
0.0
-1.0
-2.0
a = 1.25
2.0
1.0
0.0
-1.0
-2.0
a = 1.25
2.0
1.0
0.0
-1.0
-2.0
a = 1.5
2.0
1.0
0.0
-1.0
-2.0
-3.0
a = 1.5
2.0
1.0
0.0
-1.0
-2.0
a = 1.75
2.0
1.0
0.0
-1.0
-2.0
-3.0
a = 1.75
2.0
1.0
0.0
-1.0
-2.0
a = 2
2.0
1.0
0.0
-1.0
-2.0
a = 2
Classic Mazur
Exponentiated Mazur

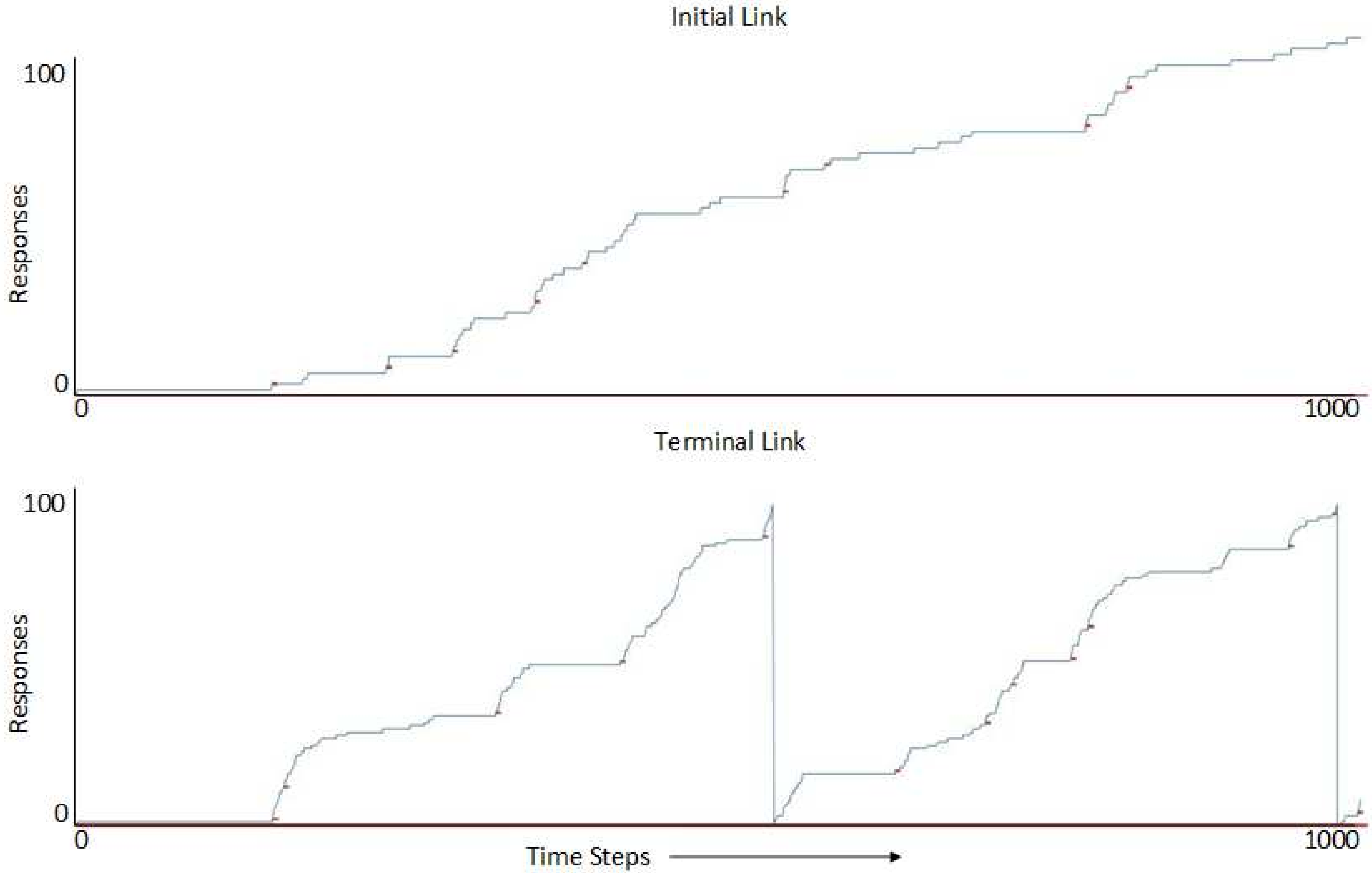

250
200
150
100
50
0
Response rate (*B*)
0
50
100
150
200
250
Reinforcement rate (*r*)
20
30
40
50
60
70
80
90
100

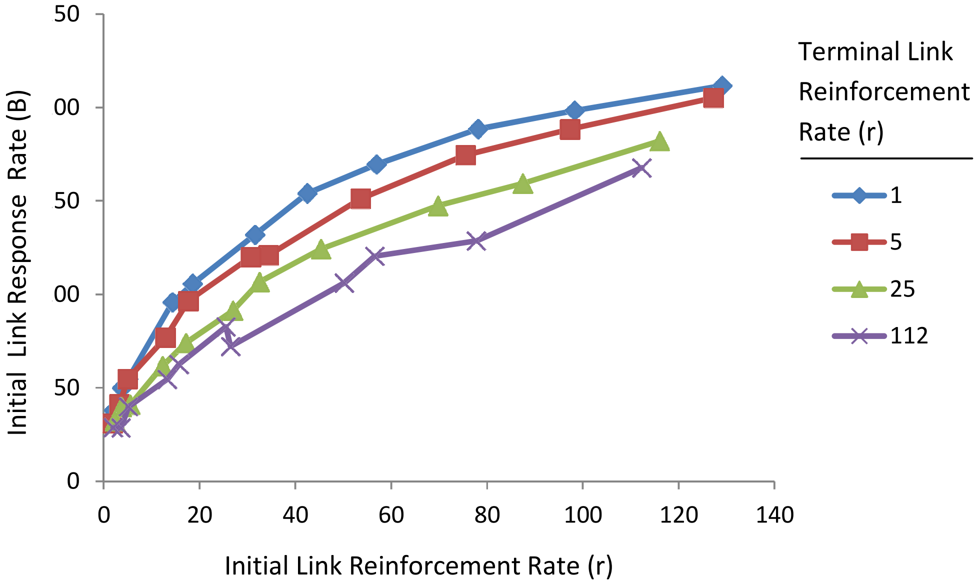

Classic
Modern
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Left blank intentionally
2
1
0
-1
-2
-3
-4

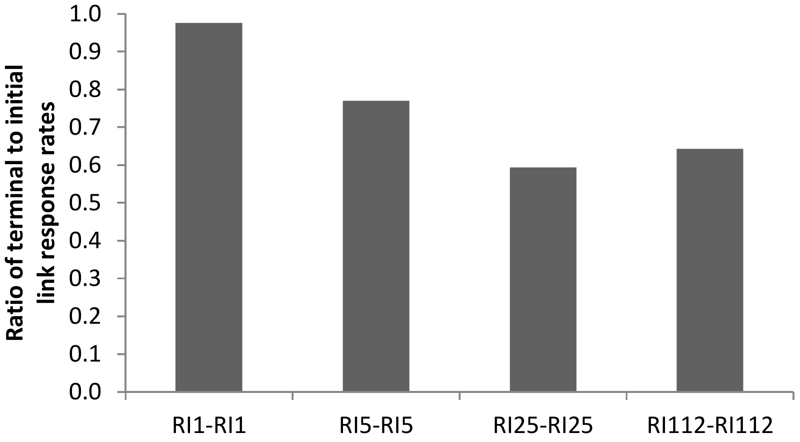

Initial Link
100
Responses
0
1
3
0
1000
Terminal Link
100
Responses
2
0
0
1000
Time Steps


Initial Link
100
Responses
0
0
1000
Terminal Link
100
Responses
0
0
1000
Time Steps

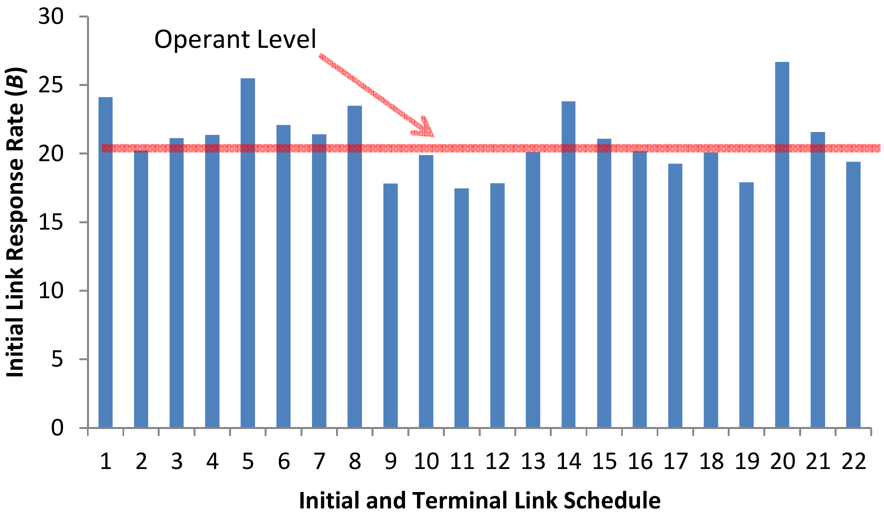

Initial Link
100
0
Responses
0
1000
Terminal Link
100
0
Responses
0
1000
Time Steps

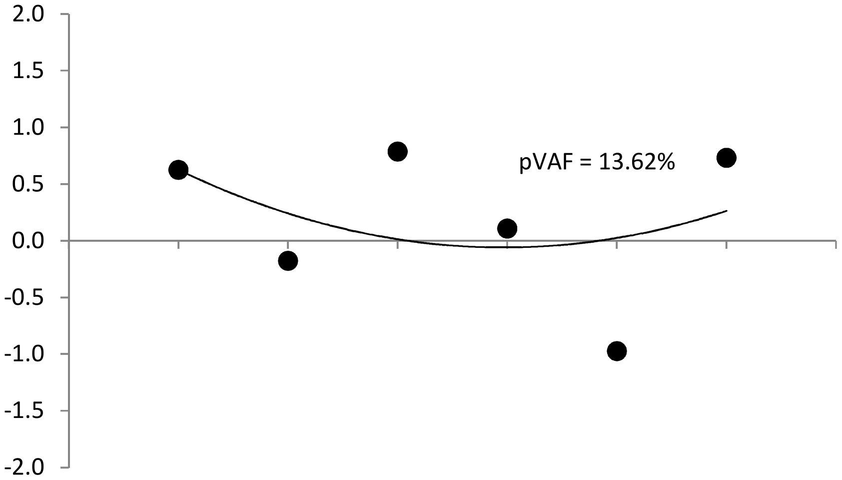

Initial Link
100
Responses
0
0
1000
Terminal Link
100
Responses
0
0
1000
Time Steps

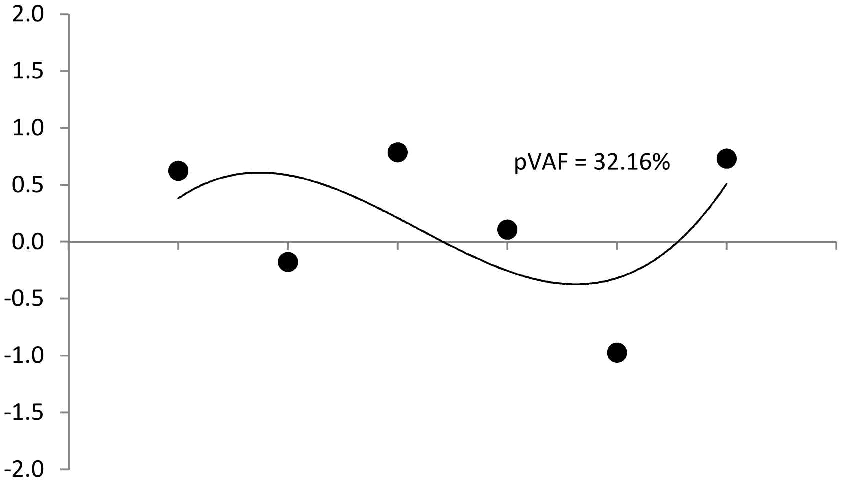

50
00
50
00
50
0
Initial Link Response Rate (B)
0
20
40
60
80
100
120
140
Initial Link Reinforcement Rate (r)
Terminal Link Reinforcement Rate (r)
1
5
25
112

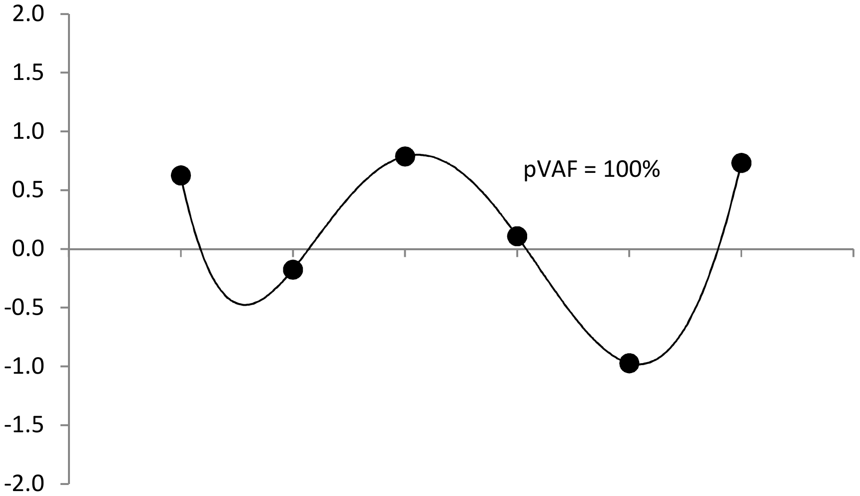


$\beta 0$: 0.05, $\beta 1$: 0.5


Operant Level
Initial Link Response Rate (B)
Initial and Terminal Link Schedule
0
5
10
15
20
25
30
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22


2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
pVAF = 13.62%


2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
pVAF = 32.16%


2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
pVAF = 100%