Distribution Agreement

Cyrus N. Chi

Monkey See Computer Do:
Simulation of Dynamic Behavior via the Evolutionary Theory of Behavior Dynamics

By

Cyrus N. Chi
Master of Biomedical Engineering
Rutgers University and University Medicine and Dentistry of New Jersey
2005

Bachelor of Arts
Rutgers University
2001

Jack J. McDowell, Ph. D
Advisor

Elaine Walker, 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

Monkey See Computer Do:
Simulation of Dynamic Behavior via the Evolutionary Theory of Behavior Dynamics

By

Cyrus N. Chi
Master of Biomedical Engineering
Rutgers University and University Medicine and Dentistry of New Jersey
2005

Bachelor of Arts
Rutgers University
2001

Advisor: Jack J. McDowell, Ph. D

An abstract of

A thesis 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
Master of Arts
in Psychology
2019

Abstract

Monkey See Computer Do:
Simulation of Dynamic Behavior via the Evolutionary Theory of Behavior Dynamics
By Cyrus N. Chi

The Evolutionary Theory of Behavioral Dynamics (ETBD) is a computational instantiation of selection by consequences that allows the generation of simulated behavioral output in environments with known reinforcement schedules. This study extends the theory to examine its predictions within a dynamic concurrent random interval environment with unsignaled transitions between 33 unique pairs of reinforcement schedules. The results were compared with behavioral data from rhesus monkeys (n=2) that were placed in a similar environment over three different levels of analysis. At the macro-level, the generalized matching law (GLM; Baum, 1974) fit the data from the virtual organisms animated by the ETBD well and returned parameters comparable to those from GML fits to the rhesus monkeys' data. At the transition level, virtual organisms adapted more quickly at the unsignaled transitions between schedules than the rhesus monkeys. At the local level, the dynamic responses of virtual organism behavior to changes in reinforcement were comparable to that of rhesus monkeys.

Monkey See Computer Do:
Simulation of Dynamic Behavior via the Evolutionary Theory of Behavior Dynamics

By

Cyrus N. Chi
Master of Biomedical Engineering
Rutgers University and University Medicine and Dentistry of New Jersey
2005

Advisor: Jack J. McDowell, Ph. D

A thesis 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
Master of Arts
in Psychology
2019

Contents

I. Introduction

A good model of behavior is capable of reproducing patterns of behavior that are already known. Extension of the model's predictions to other contexts, timescales, and organisms help to demonstrate the generalizability and test the falsifiability of the concepts that underlie it. This study's purpose is to extend the evolutionary theory of adaptive behavioral dynamics (ETBD; McDowell, 2004) through a comparison with behavioral data generated by rhesus monkeys within a dynamic concurrent schedule paradigm (Corrado, Sugrue, Seung, & Newsome, 2005). Unlike the more typical theories of behavior, the ETBD generates sequences of behavior that are the theory's predictions. These sequences of behavior can be used to validate the model through comparisons with known animal behavior and can also be used to predict the behavior of an organism in novel untested environments. The ETBD works by placing a virtual organism in a simulated environment that it interacts with to create behavioral predictions. The ETBD will be discussed in greater detail in the next section.

The behavior of multiple species has been well modeled in static environments with continuous choice using Herrnstein's Matching Law (Herrnstein, 1961). This behavior is modeled at its equilibrium state after the organism has adapted to the environment in which it was placed. However, organisms do not exist naturally in unchanging environments. Investigation of the behavior of an organism as it adapts dynamically to changing environments may lead to insights that can support the development of more precise and functional models of behavior.

This study aims to examine the predictions of the ETBD within a dynamic environment and to compare the predictions of the theory with data collected from rhesus monkeys. Multiple levels of analysis are used in order to have a more comprehensive comparison of the theory with live data.

Evolutionary Theory of Behavioral Dynamics

The ETBD was first developed as an instantiation of selection by consequences (Skinner, 1981) using a genetic algorithm (McDowell, 2004). The ETBD is unique among behavioral models as it also qualifies as a complexity theory. Complexity theory systems have multiple components that interact with each other and the environment following simple rules. The interactions of the components create relationships or patterns that are not deducible from the basic rules that the agents follow. These relationships or patterns are considered *emergent* properties of the system. One example of an emergent property would be the shapes created by a starling murmuration (a clip of a murmuration of starlings can be found here.). While the patterns made by the starlings may seem complex, they can be generated if the starlings all follow three basic rules: (1) steer towards each other, (2) turn if any of your seven nearest neighbors turn, and (3) do not crowd other starlings (Ballerini, Cabibbo, et al., 2008; Ballerini, Calbibbo, et al., 2008). These basic rules do not make any assumptions about the starling flock as a whole, yet they still generate the flock shapes. Another example, a traffic jam (a clip of a traffic jam shockwave can be found here.), is also an emergent property of the interactions between drivers (Helbing, 2001). In order to generate the emergent property, the system's agents must be allowed to interact repeatedly. ETBD differs from traditional theories in psychology because it requires an iterative computational simulation in order to uncover the theory's predictions. This is unlike traditional theories, which derive dependent variables mathematically from independent variables. The following sections will review the elements of the ETBD: the virtual organism; the environment in which it operates; and the rules that govern the behavior of the organism.

The Virtual Organism. The virtual organism is defined by a population of one hundred behaviors that are drawn from a pool of 1024 potential behaviors. Each behavior has a phenotype and a genotype that impact the operation of the rules of the ETBD. The phenotype is a decimal integer between 0 and 1023. The genotype is a 10-digit bit string that is the binary representation of its phenotype. The ‘distance’ between two genotypes is determined by their Hamming distance (Hamming, 1950). The Hamming distance is the number of bits that must be flipped to change one genotype into another. For example, the numbers 0000000000 and 0000000001 have a Hamming distance of one since only the last of the 10 bits must be flipped to change one genotype to the other. The decimal integer values of these numbers, which are the phenotypes, are 0 and 1. In this case, the Hamming distance and the phenotypic distance are both one. In contrast, 511 and 512 are only one digit away in terms of their phenotypes but have a Hamming distance of 10 between their genotypes, which are 0111111111 and 1000000000 respectively. This is the largest Hamming distance possible using 10 digit genotypes. The distinction between phenotype and genotype is important because the two representations are used during different steps of the evolutionary process. The ability to have different phenotypic and genotypic distances allows the organism to have greater behavioral flexibility. See Figure 1 for a schematic of the relationship between the phenotypes and genotypes within the organism.

As mentioned earlier, the virtual organism has 1024 possible behaviors from which its 100-behavior population is drawn. While the total number of potential behaviors may seem low given the diversity of behaviors possible for an organism, this is reasonable in terms of environmental affordances. For example, there is a limited number of potential behaviors that a chair affords a person. A chair can be sat upon or stood upon, but it will not enable a person to fly or dig. A person may attempt to dig or fly using a chair, but it is highly unlikely to be effective and extremely unlikely to be rewarding. Due to their unrewarding nature, these behaviors are essentially interchangeable since the organism does not derive any benefit from them. Therefore, these types of behaviors may be combined with any number of other unrewarding behaviors and be represented together using a finite number of behaviors. Conversely, it is critical that the behaviors that can allow an organism to gather reward be represented within the environmental space. Within a virtual organism's environment, it is possible to designate "target classes" that allow the organism to potentially obtain a reward. If the environmental space is defined as the afforded behaviors within an operant chamber, then the behaviors within the target class would be the behaviors that cause a lever to be depressed or a button to be pushed. The reason it is a class of behaviors rather than a single behavior is that there may be multiple behaviors that can depress a lever. For example, a rat can depress a lever by pressing on it with its left paw, or its right paw.

The Rules of the ETBD. There are four rules that govern the dynamics of the ETBD's virtual organism and its interactions with its environment: Emission; Selection; Replication/Recombination; and Mutation. The sequence of the rules can be found in the flowchart in Figure 2.

The first rule in the sequence is emission. During the emission step, a behavior at random is emitted from the organism's current population.

The second rule is selection. The behavior emitted during the first step is checked to see if it is within a target class. If it is, then there is a second check to see if that behavior is scheduled to be rewarded. If the behavior emitted in the first step is rewarded, then 200 parent behaviors are chosen out of the pool of 100 behaviors. The choice is based on the behaviors' phenotypic similarity to the emitted behavior. The quantification of similarity is based on the absolute value of the phenotypic difference between the emitted behavior and the potential parent behavior. The similarity is additionally modified by linear density function in order to derive a fitness value for each behavior. The linear density function,

$$p(x) = -\frac{2}{9\mu^2}x + \frac{2}{2\mu}, \tag{1}$$

expresses the probability of a behavior becoming a parent as a function of its fitness and is used to select fitness values at random. The behaviors that correspond to the drawn fitness values are chosen to be parents. In Equation 1, $p(x)$ is the probability of the fitness value, x is the absolute value of the difference between a behavior and the emitted behavior (i.e., the behavior's fitness), and μ is the mean of the density function. It is important to note that density functions can be defined solely by their means. Small means lead to higher selectivity, while large means lead to lower selectivity. Conceptually, the mean of the density function has been theorized to capture the properties of the resource acquired by the behavior and/or the properties of the target behavior itself, like the effort required to emit that behavior (McDowell, 2013). On the other hand, if the emitted behavior is not in the target class, or if the behavior is emitted in the target class but not rewarded, then the 200 parents are chosen at random from the current population of behaviors.

The third rule in the sequence is replication/recombination. After the 200 parent behaviors are chosen, they are recombined using bitwise recombination of their genotypes to generate 100 child behaviors. In bitwise recombination, there is a 50% chance that the information in a particular bit location in a child behavior will come from the father behavior's bit in the same location or the mother's. Recombination of genotypes generally leads to child behaviors that are similar to the parents, which means that parents with high fitness values typically produce children with high fitness values.

Finally, during the fourth rule is mutation. The 100 behaviors produced by recombination undergo mutation using the bitflip-by-individual method. In the bitflip-by-individual method, a percentage of individual behaviors from the population is chosen, and then a random bit in each behavior's genotype is flipped. The mutation rate indicates the percentage of the population of behaviors that undergo mutation. The resulting population of 100 behaviors then becomes the current population of behaviors and the cycle is repeated until the end of the experiment.

Evidence Supporting the Theory. The ETBD, as described above, is a set of iterative rules for the generation of continuous behavior. It does not entail any assumptions about what behavioral characteristics may be generated from its processes. Thus, any patterns contained within the output of the theory can be considered an emergent property of the theory. To date, there have been multiple studies that compare the output of the ETBD to the behavior of live organisms. In one study by McDowell and colleagues (2004), the behavior from virtual organisms was compared to live organism behavior using the original quantitative law of effect (Herrstein, 1970),

$$B = \frac{kr}{r+r_e} \qquad (2)$$

where B represents response rate, r represents reinforcement rate, and k and r_e are parameters of the equation. Results indicated that the behavior of the virtual organism had the same hyperbolic form and violated the constant k assumptions of the classical law of effect the same way that live organisms do (McDowell, 2004). When compared to the power function version of the quantitative law of effect (McDowell, 1986),

$$B = \frac{kr^a}{r^a + \frac{r_e^a}{b}}, \tag{3}$$

behavior from the virtual organisms was found to be strictly consistent with this version and had exponents (*a*) around 0.8, consistent with the exponents generated from fits to data from live organisms (McDowell & Caron, 2007).

Behavior from virtual organisms was also compared to behavior from live organisms using the power function matching law (Baum, 1974),

$$\frac{B_1}{B_2} = b\left(\frac{r_1}{r_2}\right)^a, \tag{4}$$

where B_1 and r_1 represent the rates of behavior and acquired reinforcers respectively for one target class, B_2 and r_2 represent the rates of behavior and acquired reinforcers respectively for the alternative target class, *b* represents the bias towards one side, and *a* represents the sensitivity of the ratio of behaviors to the ratio of acquired reinforcers. The function was found to accurately describe virtual organism behavior and also generated exponents around 0.8, which were comparable to exponents obtained from fits to data from live organisms. In the same study, it was also found that asymmetry in density function means, which impacts the value of reinforcers, led to biased responding that was also consistent with the responding of live organisms (McDowell, Caron, Kulubekova, & Berg, 2008). Biased responding refers to an organism favoring one side over the other in concurrent continuous choice studies, which is also represented by a *b* value greater than or less than one in Equation 4.

In one study by McDowell and colleagues (2012), the ETBD was also evaluated using the concatenated generalized matching law (Killeen, 1972; Rachlin, 1971),

$$\frac{B_1}{B_2} = b\left(\frac{r_1}{r_2}\right)^{a_r}\left(\frac{m_1}{m_2}\right)^{a_m}, \tag{5}$$

that takes into account magnitudes, *m*, of reinforcement as well as the rates, *r*, of acquired reinforcement. From this study, it was found that Equation 5 accurately described the behavior of virtual organisms animated by the theory with exponents of 0.83 and 0.68 for a_r and a_m respectively. These exponents are nearly identical to the values obtained from live organisms (McDowell, Popa, & Calvin, 2012).

In addition, the ETBD was found to have multiple environmental interactions that mirror data collected from live organisms. The Hamming distance, which is the number of bit flips required to change one genotype into another, as mentioned earlier, has been considered analogous to a changeover delay (COD) (Popa & McDowell, 2010). CODs are used to punish switching between different target classes in operant paradigms in order to avoid continuous switching. CODs can take the form of requiring an extra action to collect a reinforcer after switching, or they can take the form of increased travel distance between levers. In the concurrent schedule paradigm, as the Hamming distance between target classes decreases, the exponent, *a*, in Equation 4 decreases rapidly and stabilizes at a value that reflects near indifference to either alternative (Popa & McDowell, 2010). This relationship between the Hamming distance and the exponent in virtual organism experiments mirrors the relationship between the COD and the exponent in live organism experiments. The changeover patterns in live organisms also show a lower rate of changeovers when the rewards rates on concurrent schedules are highly asymmetrical. The rate of changeovers increases as the reinforcement rates become more similar, and peak when the reinforcement rates are nearly equal. This changeover pattern is present in virtual organisms as well (McDowell, 2013).

The ETBD has relatively fewer comparisons with dynamic live organism data, although there are several notable findings. Virtual organisms animated by the ETBD require only a few reinforcer allocations in order to increase the number of behaviors in the target class. Once reinforcers are no longer being allocated to the target class, the amount of behavior in the target class declines slowly with irregular features. These behaviors by the virtual organism mirror behaviors seen in rats (McDowell, 2013).

One way of viewing dynamic behavior is to look at the distribution of time between responses. It was found that these distributions for virtual organisms were comparable to those of live organisms under low response cost situations (Kulubekova & McDowell, 2008). It was also found that virtual organism behavior and the way in which it changes its preference immediately following a transition between different concurrent schedules were indistinguishable from live organism behavior and preference change (Kulubekova & McDowell, 2013).

In summary, the ETBD has been found to predict a range of behavioral patterns including both steady state phenomena, like the power function form of the matching law, and dynamic phenomena, like the patterns of preference change following a transition. The theory is further able to predict the environmental interactions, such as the impact of the COD on the sensitivity exponent and the impact that asymmetrical rates of reinforcement have on changeover rates. In order to test the limits of the theory, it is useful to have additional comparisons between the theory’s predictions and live organism data. Dynamic behavior is one area with additional findings that the predictions of the theory can be compared.

Dynamic Behavior

While observations of dynamic behavior and organism learning are not new, there are only a few research groups currently engaged in the study of dynamic behavior from an operant behavioral perspective. Davison and Baum (2000) and their colleagues have focused on the behavior of organisms on dynamic variable interval (VI) schedules, using pigeons and rats. Studies that focused on preference stabilization found that preference between target classes changes rapidly, and begins approaching equilibrium quickly (Aparicio, 2008; Banna & Newland, 2009; Gallistel, Mark, King, & Latham, 2001; Hunter & Davison, 1979; Mazur, 1992). Studies that focused on molecular level dynamics, that is, changes over a small number of reinforcers, found that pigeons and rats rapidly respond to reinforcement or nonreinforcement (Aparicio, 2008; Davison & Baum, 2000), and that the relative difference between the reinforcement ratios, rather than absolute difference between ratios, determines the rate of acquisition (Mazur, 1992).

Dynamic behavior gives the opportunity to explore behavioral phenomena as part of a spectrum. Behavior can be observed from the local level all the way up to the macro-level. At the local level, it is possible to observe the moment to moment interactions of behavior with the environment. At the macro-level, phenomena such as the Matching Law relationship are observable. There are generally two types of possible relationships between the macro-level phenomena and smaller time scale behavior. They can show self-similarity i.e., the macro-level phenomena are observable at smaller time scales as well, or reduction i.e., the micro-level phenomena demonstrate smaller scale regularities that are not consistent with macro-level phenomena (Baum, 2010). In the case of reduction, the micro-level phenomena can be used to derive the macro-level phenomena, while the reverse is not true. For example, a self-similar relationship across levels would suggest that the Matching Law relationship would exist at all levels of observation, no matter how few behaviors are considered. A reductive relationship would be one where the Matching Law relationship is only present at the macro-level. At the local-level, other behavioral phenomena would be present instead. Studying dynamic behavior at multiple levels is beneficial because unique phenomena may potentially be observed at the local level that cannot be deduced from macro-level phenomena.

Monkey Experimental Paradigm

Modeling of behavior on the micro-level in monkeys on concurrent VI schedules has also been attempted using various models (Corrado et al., 2005; Sugrue, Corrado, & Newsome, 2004). Both of Corrado et al.'s studies tested models that were descriptive in nature rather than causal, but their findings suggested that previous behavior can impact present behavior. From the monkey data, it was found that local behavior allocation between two alternative VI schedules was sensitive to local fluctuations in reward within schedules and was fully adapted to new schedules within 40 trials of the transition point.

The current study is based on the paradigm used by Corrado et al. (2005). The parameters of that study follow. Two adult male rhesus monkeys were given a simulated foraging task. They had a choice between two targets, one red and one green, which they chose by gazing at one target or the other. Their gaze was continuously tracked via computer monitoring. The monkey was free to look in any direction. If it gazed at a target that was scheduled for a reward, then it would be rewarded with drops of juice. If a reward was scheduled on a target, it would be considered "setup" and the monkey would receive the reward the next time it looked at the target. If the monkey looked in any other direction during the trial, it would receive a brief two-four second timeout. The probabilities of reward for the targets were independent of each other, and the VI schedules were implemented using a Poisson process. A COD was implemented by not rewarding the first choice to a color if the previous choice had been the other color. For example, if the monkey's last choice was red and the current choice was green, then the monkey would not get a reward for choosing green, even if the reward was setup on that side. Only if the monkey chose the green target a second time immediately after choosing the green target would the setup reward be dispensed. The COD in this model can be considered weak, due to the high level of switching between targets. Using a run length distribution to show how many times a monkey would pick the same color in a row, Corrado and colleagues found that approximately 40% of the run lengths for one monkey had a length of two, and the other monkey predominantly had run lengths of two or three, signifying that the monkey would choose the same target only two or three times before switching to the other target. In other words, the monkeys continued switching back and forth for a significant portion of the time, only staying on a side long enough to satisfy the COD rule. The monkeys were exposed to blocks with a set reward schedule for 40-350 trials with approximately 0.3 rewards per trial (both sides combined). The overall reward rate was held constant while the ratio of reward was scheduled to vary between 1:8, 1:6, 1:3, 1:2, 1:1, 2:1, 3:1, 6:1, and 8:1 a majority of the time. Additional schedule ratios were also used, with the most extreme ratio being approximately 14:1. The schedule ratios used in this study can be found in Table 1. In order to have unpredictable transitions, the number of trials per schedule varied unpredictably, along with the schedule ratio. The transitions were also unsignaled.

The Present Study

Previous studies have examined dynamic behavior in multiple organisms and have been able to describe patterns of behavior that are present on differing levels of analysis (Corrado et al., 2005; Davison & Baum, 2000). However, these studies are often descriptive in nature and rarely look into causal mechanisms. The present study extends the body of literature dealing with dynamic behavior by demonstrating that the dynamic patterns of behavior can be predicted using the ETBD. The study also represents an extension of the ETBD itself. This study is the first to compare the predictions of the ETBD with the behavior of rhesus macaque monkeys. Additionally, the present study examines the predictions of the ETBD within an extended dynamic paradigm that has not yet been used with the ETBD.

The simulated environment in this study is modified to better simulate the paradigm set by Corrado et al. (2005). After generating the ETBD's behavioral prediction, the prediction was analyzed on three levels. Virtual organism behavior was examined on the macro-level, by testing the overall behavioral output data's conformance to the power function matching law (Equation 4). The virtual organism's ability to adapt to continuous, unsignaled transitions was examined using a transition-level analysis. The results of the analysis were then compared to monkey behavior. Finally, the virtual organism's ability to track acquired reinforcers at the "local-level" was examined. Observations at this level examine behavior without averaging, unlike analysis at the previous levels. Local-level analysis also maintains the basic temporal flow of the experiment. The results of the local-level analysis were then compared to monkey behavior.

II. Methods

Subjects

The experiment was conducted on 15 virtual organisms starting with a population of 100 potential behaviors from a pool of 1024 possible behaviors. The computational algorithm that animates the aforementioned organisms is described above, in the Evolutionary Theory of Behavioral Dynamics section.

Simulated Environment

The experimental environment was designed to mirror the environment used in the prior studies (Corrado et al., 2005; Sugrue et al., 2004). In these two papers, concurrent VI schedules were used to deliver rewards with differing ratios of reinforcement between the two target classes, but with a constant overall rate of reinforcement. Reinforcers were delivered approximately once every three trials for both schedules combined. Once a reinforcer was scheduled, that target would be considered “setup” and would remain set up until the reinforcer was received. The reinforcement ratio between the two VI schedules was dynamically set, switching among the 33 unique schedule ratios for one monkey and 21 unique schedule ratios for the other monkey. In the present study, the more complicated sequence of schedules, the one with 33 unique ratios, was followed. The 33 unique schedule ratios are listed in Table 1. During the experiment, the reward ratio remained constant during a ‘block’ of trials from 50 to 300 trials long. During the transitions between blocks, the monkey was given no indication that the ratio of reinforcement between the two choices was changed except through the change in the amounts of reward it acquired from either choice. A changeover delay (COD) was used to punish switching behavior. The COD rule prevented the reinforcer from being given for the choice of a target when that choice also constituted a switch between the two available targets. If the target was setup to give a reinforcer, it would additionally remain setup so a second choice on that target would allow the reinforcer to be obtained.

While the monkeys were given extensive training to familiarize them with the testing apparatus, the targets, the COD, and to reproduce reliable matching behavior, the ETBD organism did not undergo any training. The following steps were followed to generate the simulated environment. First, the exact schedules used in the Corrado et al., paradigm were followed to create a comparable environment for the virtual organisms. The transitions between schedules were also unsignaled. Second, the same COD rule was implemented. Third, reward setting was set on a trial basis, which matched the reward setting scheme used in the Corrado study.

ETBD Adaptations to fit the Experimental Paradigm

An important issue that needed to be addressed in this study was the COD. The current parameters for the virtual organisms were designed to fit behavior-analytic studies that typically use a robust COD. In the Corrado et al. (2005) study, however, the COD was found to not as robust, as discussed earlier. Adjustments to the Hamming distance of the behaviors' genotype has been shown to act similarly to changes in the COD length (Popa & McDowell, 2010). Thus, to compensate for the weaker COD, the target classes in the computational experiment were set at the phenotype ranges, 128-168 and 896-936. For these target classes, the genotype of a behavior in one target class has a minimum Hamming distance of two relative to the genotypes in the other target class. These target classes, in effect, make switching between them much more likely compared to the target classes used in previous research. See Figure 1 for a schematic of the target classes.

Time in the ETBD is typically based on the number of iterations of the rules that underlie the theory. Reinforcement setup follows the same time convention, that is, the alternatives are set up to dispense a reward based on the number of iterations that have passed. Within each iteration, there is a chance of a behavior being emitted from a target class, but it is by no means certain that it will be from the target class. However, the Corrado et al. (2005) followed a trial based system. In this system, each trial is based on the organism choosing one of the two target classes. In order to allow a more exact comparison with the Corrado et al. experiment, the ETBD computer algorithm was adapted to record data on a trial based system. Under this system, behavior being emitted from the target class would increment the trial counter while any behavior emitted outside of the target class was ignored. Between trials, there can be multiple iterations of the rules of the ETBD and the organism was allowed to evolve as usual. Only when a behavior is emitted from the target class will it be recorded. The rate of reinforcement and the length of each schedule was based on the trial number, like in the experiment by Corrado et al.

The virtual organism was set to have the following specific function parameters. Emission of behavior was at random, per usual. Selection was based on phenotypic distance using a linear continuous probability density function, with a mean of 80. This is within the normal range of density function means, although on the weaker end in order to produce a weaker selection of target behaviors (McDowell et al., 2008). This potentially allows a greater number of behaviors outside of the target classes to be parent behaviors. Replication was completed using bitwise recombination, per usual. Mutation was accomplished using bitflip-by-individual mutation, with a 20% mutation rate. This mutation rate is larger than what has been typically used in previous research. It generates a larger number of changeovers because mutation results in greater randomness in the resulting population of behaviors. Although the density function mean and mutation rate differ from values used in previous concurrent schedule studies of the ETBD organism (Kulubekova & McDowell, 2013; McDowell et al., 2008), they are not outside the typical range of parameters. These changes do not alter the operation of the basic rules of the evolutionary theory inasmuch as the methods used for emission, selection, replication, and mutation are identical to those used in previous studies.

Data Analysis

The data generated by the evolutionary theory was analyzed on three levels. First, all of the data collected was analyzed together using a macro-level analysis, as described earlier. Second, the adaptation of behavior and acquired reinforcers following a transition between schedules were analyzed using a transition level analysis. Finally, the within schedule adaption of behavior to local acquired reinforcer changes was analyzed using a local-level analysis. The methodology for each analysis depends on the specific goals and challenges for the specific level of analysis. The analysis was completed on the trial-based output data from the ETBD and monkey behavior. The monkey output data contains three major components for each trial: (1) the ratio of scheduled reinforcement, (2) the target chosen by the monkey for that trial, and (3) the acquired reinforcement if any. In the ETBD output, the monkey behavior was replaced by its ETBD analog, i.e. the target class from which the behavior was emitted. The two other output parameters were identical for both monkey and ETBD output data.

Macro-level analysis. The first level of comparison between the virtual organism and the monkey was at the macro-level. The power function matching law is an ideal macro-level comparison tool because it draws from all of the data generated in the study and describes the relationship between the ratio of acquired reinforcers and the ratio of behaviors. Fits of power function matching law to the monkey data have already been published by Corrado et al. (2005). The following data analysis steps were used to generate power function matching law fits to the data from the virtual organism. The power function matching law equation,

$$\log\left(\frac{B_1}{B_2}\right) = \log(b) + a\log\left(\frac{r_1}{r_2}\right), \tag{6}$$

was fitted to the ratios of acquired reinforcers and the ratio of behavior between the two target classes. In Equation 6, B_1 and r_1 are the numbers of behaviors and acquired reinforcers for target class one, and B_2 and r_2 are the numbers of behaviors and acquired reinforcers for target class two. Equation 6 is a logarithmic transformation of Equation 4. Equation 6 is often used instead of Equation 4 since the resulting plot is linear, which allows the function to be fitted with a linear regression. As mentioned earlier, 33 unique pairs of reinforcement schedules were used in the study. These unique scheduled reinforcement rates were alternated to generate the reinforcement rates for the 1153 schedules in the study. Each schedule in the study was sorted into a group based on its unique scheduled reinforcement rate. Each group's ratio of behaviors was averaged for each unique reinforcement rate. The same averaging step was completed on the ratio of acquired reinforcers. The logs of the average ratio of behaviors were plotted against the logs of the average ratio of acquired reinforcers. Then a linear regression was used to calculate the percentage variance accounted for, the sensitivity parameter, *a*, and the bias parameter, *b*, for that organism. All schedules that lack behavior or acquired reinforcers on at least one side for the entire schedule were not included in the fit. The percentage variance accounted for (PVAF), the sensitivity parameter, and the bias parameter were calculated for all 15 organisms. The parameters obtained were averaged to get sample means and standard errors.

Transition-level analysis. The transition between schedules was analyzed to examine how the virtual organisms adapt to unsignaled transitions. The virtual organism's adaptation can be visualized by its change in behavior proportion and acquired reinforcer proportion in response to a transition. However, in order to analyze the virtual organism's reaction to the presence of a transition, there were various challenges that needed to be addressed. First, in order to make the transitions unpredictable, the number of trials within each schedule varied between 50 and 350.

This means the number of trials available for analysis in each schedule was not constant. Second, the new schedule after the transition could be set to any of the 32 other unique reinforcement ratios. This means that to fully adapt to a new reinforcement ratio after a transition, the amount of behavior change required may vary in magnitude and direction. In order to examine the virtual organism's general adaptation response to a transition, the magnitude and the direction of behavior adaptation needs to be accounted for.

Fortunately, the magnitude of the behavior change is always relative to the absolute value of the difference between the post-transition reinforcement ratio, and the pre-transition reinforcement ratio. The direction of adaptation is considered to be positive if the post-transition reinforcement ratio is larger than the pre-transition reinforcement ratio and negative if the post-transition reinforcement ratio is smaller. The magnitude and the direction are most easily discussed in relation to one target class using a normalized scale. For the purpose of this analysis, reinforcement ratios were transformed into reinforcement proportions. The specifics of the transformation will be discussed later. The reinforcement proportions were set towards target class one, in order to be comparable to the behavior and acquired reinforcer proportions. A ratio which heavily favors target class two, e.g. 1:12, would translate to approximately 0.076, which is close to zero. A ratio which heavily favors target class one, e.g. 10:1, would translate into approximately 0.909, which is close to 1. Within this framework, behavioral adaptation across a transition from a reinforcement ratio of 1:12 to a reinforcement ratio of 10:1 has a large adaptation magnitude and a positive direction, as shown in the left panel of Figure 3. On the other hand, behavioral adaptation across a transition from schedule ratio of 5.3:1 to a schedule ratio of 5:1 would have a small adaptation magnitude and a negative direction as shown in the right panel of Figure 3.

In summary, the following steps were required to allow meaningful averaging across transitions. First, schedules with fewer trials may result in fewer data points than schedules with more trials. In order to collect the appropriate number of data points, the variation in the number of trials within a schedule needed to be accounted for. Second, reinforcement ratios are not comparable to the behavior and acquired reinforcer proportions. The reinforcement ratios needed to be transformed into a proportional scale in order to allow a direct comparison. Third, there are large and small magnitude differences between reinforcer proportions across a transition. These differences in magnitude needed to be normalized in order to make the change across transitions comparable during averaging. Finally, the direction of adaptation has two possible orientations, positive or negative. The orientation depends on which of the two reinforcer proportions across the transition is larger. The two orientations needed to be aligned in order to prevent them from canceling each other out during averaging.

In order to examine the virtual organism's behavioral response to changes in acquired reinforcers during a transition, the behavior proportion, $\frac{B_1}{B_1+B_2}$, and the acquired reinforcer proportion, $\frac{r_1}{r_1+r_2}$, for target class one were calculated from the output data. The behavior proportion and the acquired reinforcer proportion were calculated over ten trials at a time. Since previous research has found that behavior tends to adapt extremely quickly following a transition (Davison & Baum, 2000), one ten-trial set prior to the transition was used as a baseline for the change following the transition. Behavior proportions and acquired reinforcer proportions were then calculated for ten-trial sets up to the 150th trial after the transition. Since not all schedules have the same length, data was only used for each schedule up to the end of the schedule rounded down to units of ten. For example, if a schedule has 135 trials, then behavior and acquired reinforcer proportions were collected up to the 130th trial. Any ten-trial set that did not have any acquired reinforcers for either side was not used for calculation of reinforcer proportion. The previous scheduled reinforcement ratio and the current scheduled reinforcement ratio were transformed into reinforcement proportions. This allowed a direct comparison between the scheduled reinforcement and the behavior and acquired reinforcer proportions. This was accomplished by using the following formula, $\frac{tc_1}{tc_1 + tc_2}$, where tc_1 is the value for target class one and tc_2 is the value for target class two in the reinforcement ratio. For example, in the ratio 4:1, tc_1 equals 4, and tc_2 equals 1. The reinforcement proportion is calculated to be 0.8. A value closer to 1 symbolizes a ratio favoring the 128-168 target class (target class 1). Transitions that had differences of less than 0.0025 between the previous and the current reinforcement proportions were discarded, because such small transitions between schedules are likely undetectable to the organism and erroneously magnify minor fluctuations. Next, the behavior and acquired reinforcer proportions were normalized to the range between the previous scheduled reinforcement proportion and the current scheduled reinforcement proportion using the formula, $\frac{X - Rprop_{Pre}}{Rprop_{Pre} - Rprop_{Post}}$. X is the behavior or acquired reinforcer proportion before normalization, $Rprop_{Pre}$ and $Rprop_{Post}$ are the reinforcement proportions before and after the transition respectively. This normalization step also corrected the orientation alignment. The resulting values for each ten-trial set were averaged across all transitions for one organism. Finally, the results for all organisms were additionally averaged together to generate a group level plot.

Local-level analysis. The fluctuations of the behavior and acquired reinforcer proportions within schedules were examined in order to observe the dynamic response of behavior to acquired reinforcement in a within-schedule time scale. There are multiple challenges with local-level analysis that require addressing. First, since we are examining local level dynamics there is no averaging of data, unlike the previous two levels of analysis. This makes the lower quantity of acquired reinforcers a critical issue because any set without any acquired reinforcers on either side will lead to a gap in the data. In order to compensate for this, a larger number of trials per set were used to generate the proportions of behavior and acquired reinforcers. Second, since the data are not averaged, there will be 1153 schedules that potentially require individual analysis. In order to characterize the data better, the entire experiment was divided into blocks that contain 10 schedules each. Then the tracking of behavior proportion to acquired reinforcer proportion was quantified using a Pearson correlation coefficient for each block. This allows the identification of the blocks with the best, average, and worst tracking for use in analysis. Local-level analysis has an additional use, which is to look at the average behavior of multiple organisms at the local level. When the behavior and acquired reinforcer proportions from the 15 organisms are averaged, the local level variations cancel out. This allowed the examination of the behavior and acquired reinforcer proportion responses to schedule changes and transitions.

The entire behavior and acquired reinforcer output data were divided into sets of 20 trials. The proportion of behavior and acquired reinforcers were calculated for each set. Any set without any acquired reinforcers on one target was excluded from in the analysis. The reinforcement ratio for each schedule was transformed into a 0 to 1 scale using the same methodology as in the transition analysis. The transformed reinforcement ratio was plotted with the behavior and acquired reinforcer proportions. In order to quantify how well the virtual organism's behavior tracks the acquired reinforcers across the different schedules, the following steps were performed. First, the entire experiment was divided into blocks of ten schedules. The last three schedules were discarded in order to round the number of schedules to a multiple of ten. Next, the Pearson correlation coefficient between the behavior and acquired reinforcer proportions were calculated for each block. Only proportions that were fully contained within the block were used. Finally, Pearson correlation coefficients were used to compare how well the two proportions correlate per block. A histogram was used to display the distribution of the correlation coefficients for the virtual organisms and for the monkey. The blocks with the highest, average, and lowest correlation coefficients were used as exemplars of the data set. The average local level analysis was collected using the following method. The behavior and acquired reinforcer proportions sets were averaged with their corresponding sets across all virtual organisms, e.g. the first set's behavior proportion was averaged with all other first set behavior proportions, and so on. The averaged data was divided into blocks and the Pearson correlation was calculated for each block, following the procedure used in the previous local-level analysis. The blocks with the highest, average, and lowest correlation coefficients are used as exemplars of the averaged data set.

III. Results

This study aimed to examine the dynamic behavior of the ETBD and how it compares with monkey behavior. The results are presented in the following order: (1) the macro-level, (2) the transition level, and (3) the local level.

At the macro-level, data from the virtual organisms animated by the ETBD ($n = 15$) were fit well by the generalized matching law, with a PVAF of 91.3% ± 0.12% (standard error of the mean, SEM, for all of the following as well), a slope (sensitivity parameter) of 0.534 ± 0.003, and an intercept (bias parameter) of 1.056 ± 0.006. A plot of the matching law fit with the R^2 value closest to the group average is shown in Figure 4. This plot is typical for the overall sample. This demonstrated that virtual organisms, operating under the unique environments implemented here to match Corrado et al.'s (2005) experiment, can produce behavior that conforms to power-function matching. The sensitivity parameter from this fit, 0.534, is also similar to the one obtained from the generalized matching law fit to the monkey's data, 0.57.

On the transition level, virtual organism's behavior was found to adapt quickly to the new schedules and it reached an asymptote within 20 trials, as shown in Figure 5. In Figure 5, the behavior proportion (solid line) and the acquired reinforcer proportion (dashed line) are normalized to the difference between the pre-transition reinforcement proportion and the post-transition reinforcement proportion. Proportions are calculated over ten-trial sets. The error bars show ±1 SEM for each data point. The pre-transition reinforcement proportion is set at 0% and the post-transition reinforcement proportion is set at 100%. The transition is set to be at time = 0. The first data point is a baseline calculated from the ten-trial set just prior to the transition. All of the transitions within one organism are averaged.

The lower percentage shift in the proportion of behavior compared to the percentage shift in the proportion of acquired reinforcers shows undermatching at this level. Undermatching occurs when the ratio of behavior is less than what would be predicted based on the ratio of reinforcement by the matching law. On the macro-level, this is represented by having a sensitivity parameter less than one. On the transition level, this is represented by the behavior proportion being less than the acquired reinforcer proportion. A transition data comparison of virtual organisms and the monkey is shown in Figure 6. It shows that the virtual organisms and the monkey adapted quickly to transitions, although the virtual organism adapted at a slightly more rapid rate. Individual virtual organism data often looks remarkably similar to the monkey data. An example of this can be seen in Figure 6 (bottom). The closest of the 15 virtual organisms to the monkey data is plotted with the monkey's behavior proportion to show the degree of overlap. The proportion of acquired reinforcers was also found to reach its asymptote more quickly for monkey data than virtual organism data (Figure 7).

At the local level, the 1153 schedules in the experiment are divided into blocks with 10 schedules in each block. The Pearson correlation coefficients are calculated for the behavior proportion and the acquired reinforcer proportion within each block. The Pearson correlation coefficient is used as a marker of how well the behavior proportion and the acquired reinforcement proportions track each other. Good tracking of behavior to acquired reinforcers on a local level should theoretically lead to a Pearson correlation coefficient closer to one, because high levels of acquired reinforcers should also be accompanied by high levels of behavior, and vice versa. Poor tracking of behavior to acquired reinforcers should lead a Pearson correlation closer to zero.

A histogram of the Pearson correlation coefficients for virtual organism whose histogram is closest to the group average is shown in Figure 8 (left). The correlation for the same analysis is completed using monkey data and the histogram is shown in Figure 8 (right). The histograms of Pearson correlation coefficients show good tracking of behavior to acquired reinforcers for both virtual organisms and for the monkey. They both have a similar histogram form and have a mode between 0.8 and 0.9. An example of the blocks with the highest, average, and lowest Pearson correlation coefficients for the virtual organism whose histogram is closest to the group average is shown in Figure 9. The tracking of behavior to acquired reward is present in all tracking conditions at the local level. It is also worth noting that fluctuations of behavior and acquired reinforcer proportions at this level are on a comparable scale to the shifts between scheduled reinforcement proportions. Averaged data of 15 virtual organisms is shown in Figure 10. In the averaged data, the local level fluctuations are muted and the macro-level phenomenon, undermatching, is visible. As seen in the transition data, the virtual organism can quickly adapt to new schedule ratios.

A comparison of the virtual organism and the monkey at the local level is shown in Figure 11. The behavior proportions for both were found to react to the shifts caused by reinforcement ratio transitions, and also to the local level fluctuations in acquired reinforcer proportions. Both the virtual organism and the monkey exhibited local level undermatching and over matching of behavior proportion to acquired reinforcer proportion.

IV. Discussion

The results of this study suggest that organisms animated by the ETBD demonstrate similar dynamic behavioral features to those of rhesus monkeys. These results were observed at three different levels of analysis. First, at the macro-level, the organisms were capable of generating good fits to the power function matching law. This is similar to previous studies using virtual organisms (McDowell et al., 2008). The sensitivity parameter of 0.534 showed more undermatching compared to other fits of the matching law using virtual organisms but is more in line with sensitivity parameters taken from other studies that have researched dynamic phenomena (Baum & Davison, 2014; Davison & Baum, 2000). The virtual organism sensitivity parameter is also in line with the sensitivity parameter values taken from the rhesus monkey studied by Corrado et al., 2005, which was 0.57. This result also makes sense intuitively. With unsignaled, dynamic environments, undermatching in a practical sense demonstrates a greater number of choices to the less rich of the two target classes. This behavior is more adaptive since the sooner the organism discovers that the ratios have changed, the more quickly it can adapt its behavior to the new environment.

Second, virtual organisms were found to be capable of rapidly adapting to changes in reinforcement ratio during unsignaled transitions. As shown in Figure 5, rapid adaptation to the new reinforcement ratios occurred within 10 trials. This aligns with the results from investigations of pigeons (Davison & Baum, 2000) that found that preference for a side developed quickly within two to eight reinforcers. This finding differs from that of Davison and Baum's due to the continuous nature and the diversity of the schedules present. In Davison and Baum's study, the pigeons were not given any indicators as to which schedules would be present after a change, but the lights were turned off between schedules effectively signaling that a transition would occur. This was found to have an effect, as the pigeon's behavior would regress towards indifference during the blackout period (Davison & Baum, 2000). In addition, the pigeons were exposed to only five different types of reinforcement ratios, while the virtual organisms were exposed to 33 unique reinforcement ratios. This study demonstrated that rapid acquisition can still occur without a distinct period of indifference. The increase in the number of schedules further supports the idea that the rapidity of adaptation is not the result of repeated exposures to the same schedules and also provides an incremental step towards more naturalistic environment simulations.

The transition level data additionally show that undermatching continues to be present during the transition timescale. However, there are nuances in behavior at this analysis level that are not just accounted for by undermatching. The rhesus monkey data collected by Corrado et al. (2005) show rapid acquisition of the new schedule but the response proportion reaches an asymptote gradually, within approximately 40 trials. Virtual organisms animated by ETBD also show rapid acquisition of the new schedule but their behavior asymptotes quickly, within 20 trials. While the reason for this difference is uncertain, there are various potential explanations for these differences. This may be the result of temporal stimulus control, which may be acting to slow monkey behavior. This would not be present in virtual organism behavior since temporal stimulus control has not been implemented yet in the theory. Alternatively, this simply could be caused by the lack of sample monkeys. Virtual organisms can be generated indefinitely and thus better approximate average behavior. The variation in transition shape when switching between reward ratios is clearly evident when looking at dynamic pigeon data (Banna & Newland, 2009). It is not a stretch to suggest that monkey data may show a similar level of variation.

Monkey transition data also shows that the acquired reinforcer proportion for the monkey reaches its’ asymptote within the first ten trials. This differs from the acquired reinforcer proportion of virtual organisms that asymptotes at 20 trials. This difference may be caused by the use of different reinforcement set up methods. The Corrado et al. (2005), experimental paradigm uses a Poisson process to determine reinforcement setup timings which are different from the standard reinforcement methods typically used in continuous-choice experiments. A Poisson process gives a constant probability for a reinforcer to be setup on a target. In contrast, the ETBD environment uses an exponential distribution to determine the number of trials required before a target class is setup. Looking at this problem from another angle, the behavior proportions for monkey and virtual organisms show approximately 40% adaptation to the post-transition schedule during the first ten-trial set. Given the stochastic nature of reinforcement, it is highly unlikely that similar proportions of behavior would lead to drastically different proportions of acquired reinforcement. In summary, this difference between the monkey and the virtual organism results is likely an artifact.

Finally, we found that virtual organism behavior was capable of tracking acquired reinforcers at the local level, as shown in Figure 9. Local maxima or minima in acquired reinforcer proportions were almost always accompanied by or immediately followed by local maxima or minima in behavior proportions. It is also noteworthy that fluctuations in behavior and acquired reinforcer proportions within a schedule were on a similar scale as the differences between schedules. These findings suggest that, in dynamic and unsignaled conditions, adaptation to different schedules does not require any recognition of the transition between schedules. Adaptation to acquired reinforcers at the local level is sufficient to generate transition level adaptation. These observations were also present in the monkey output data, as shown in Figure 11.

Averaging of local level phenomena mutes local level phenomena and reveals macro-level phenomena, as shown in Figure 10. The averaged acquired reinforcer proportion was found to be close to the scheduled reinforcement proportion, as expected. The average behavior proportion is almost always closer to an equal proportion, 0.5, between the two sides than the acquired reinforcer proportion, showing undermatching. This is in contrast to unaveraged local level data because the local level data show some instances of overmatching and perfect matching, as well as undermatching. This contrast shows that local level tracking is a distinct phenomenon that is not explained by the matching law, although it is predicted by the ETBD.

Limitations, Strengths, and Future Directions

This study has a few limitations worth discussing as well as strengths worth reiterating. There is a limited amount of comparison monkey data. As mentioned earlier, the Corrado et al., 2005 experiment had only two monkeys. Each monkey was additionally given different sequences of reinforcement. The small sample size undermines the ability to determine the amount of natural variation inherent in rhesus monkey behavior. It may also obscure significant trends that might be present if the data could be averaged across multiple monkeys. The adaptation of ETBD parameters to accommodate a low COD environment was effective in generating the relevant phenomena as well as being aligned with the expected effect of the ETBD parameter modifications (McDowell & Popa, 2010). However, these adaptations may make it difficult to compare the results of this study with the results from other ETBD simulations. Finally, portions of the findings from this study were descriptive or qualitative comparisons which do not easily lend to further investigation. However, these descriptive phenomena were generated using the ETBD and can be considered emergent properties of the theory. In addition, a qualitative description of dynamic phenomena is also typical in the field at this level of analysis. For example, Baum described a pattern of fixing on a rich key and occasionally visiting the lean key “Fix and sample” (Aparico & Baum, 2006; Baum, 2010), likely following the pattern set by “Break and run” (Ferster & Skinner, 1957).

This study also has numerous strengths. The ETBD is found to replicate monkey dynamic behavior on all three levels of analysis, in a new paradigm. The multiple levels of analysis allow for comparisons across the different levels that would not be observable if only one level of analysis was used. The paradigm in the present study differs from previous ETBD studies in that the scheduled reinforcement ratios and the transitions themselves are unsignaled. A large number of unique reinforcement ratios allow for a greater variety of transition types which is a step closer towards more naturalistic settings.

This study lends itself to various potential future directions that may further illuminate the dynamics of behavior. The mutation rate, as mentioned earlier, was set at a higher than normal value due to the weak COD. This suggests that the mutation rate may not be static, but may be an adaptive value that varies according to environmental feedback. In the transition between reward ratios, the ETBD responded extremely rapidly to the change. This was found to be more rapid than the monkey data, though it showed some similarity to some rat transition data (Banna & Newland, 2009). This suggests that there may be a temporal stimulus component that may be a differentiating factor that is not yet part of the ETBD. Development of a temporal stimulus component may further clarify some of the transitional dynamics that currently unexplained. The ETBD was additionally modified to use a trial based system. This has interesting implications for future research as much continuous choice research done on live subjects is typically done using a trial based system. This modification may allow a more direct comparison with different types of behavioral studies.

Overall, the predictions of the ETBD aligned well with the behavior of rhesus monkeys on all three levels of analysis. It is important to note that none of the analyzed behavior patterns (e.g. the sensitivity parameter of the generalized matching law; the rates of transition adaptation; the complex overmatching, matching, and undermatching exhibited in local-level analysis) are explicitly implemented in the theory. The theory only implements the evolutionary rules of emission, selection, reproduction/recombination, and mutation in an iterative cycle on an adaptive population of behaviors. All of the behavior patterns examined were emergent properties of the theory. These results provide strong support for the ETBD account of dynamic behavior.

V. Acknowledgments

I would like to thank Dr. Jack McDowell for his guidance towards making me a better scientist and writer; my fellow lab mates for their feedback and support; my cohort mates for their support and encouragement; for the Newsome lab and the authors of Corrado et al. 2005 for their willingness to share their data for this study; and my family for their love, understanding, and tireless efforts on my behalf.

References

Aparicio, C. F. (2008). Dynamics of choice and frequency of environment change. *Behavioural Processes, 78*(2), 165-172. doi:10.1016/j.beproc.2008.03.001

Aparico, C. F., & Baum, W. M. (2006). Fix and sample with rats in the dynamics of choice. *Journal of the Experimental Analysis of Behavior, 86*(1), 43-63. doi:10.1901/jeab.2006.57-05

Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., . . . Zdravkovic, V. (2008). Empirical investigation of starling flocks: a benchmark study in collective animal behaviour. *Animal Behaviour, 76*, 201-215. doi:10.1016/j.anbehav.2008.02.004

Ballerini, M., Calbibbo, N., Candeleir, R., Cavagna, A., Cisbani, E., Giardina, I., . . . Zdravkovic, V. (2008). Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *Proceedings of the National Academy of Sciences of the United States of America, 105*(4), 1232-1237. doi:10.1073/pnas.0711437105

Banna, K. M., & Newland, M. C. (2009). WITHIN-SESSION TRANSITIONS IN CHOICE: A STRUCTURAL AND QUANTITATIVE ANALYSIS. *Journal of the Experimental Analysis of Behavior, 91*(3), 319-335. doi:10.1901/jeab.2009.91-319

Baum, W. M. (1974). 2 TYPES OF DEVIATION FROM MATCHING LAW - BIAS AND UNDERMATCHING. *Journal of the Experimental Analysis of Behavior, 22*(1), 231-242. doi:10.1901/jeab.1974.22-231

Baum, W. M. (2010). DYNAMICS OF CHOICE: A TUTORIAL. *Journal of the Experimental Analysis of Behavior, 94*(2), 161-174. doi:10.1901/jeab.2010.94-161

Baum, W. M., & Davison, M. (2014). Choice with frequently changing food rates and food ratios. *Journal of the Experimental Analysis of Behavior, 101*(2), 246-274. doi:10.1002/jeab.70

Corrado, G. S., Sugrue, L. P., Seung, H. S., & Newsome, W. T. (2005). Linear-Nonlinear-Poisson models of primate choice dynamic. *Journal of the Experimental Analysis of Behavior, 84*(3), 581-617. doi:10.1901/jeab.2005.23-05

Davison, M., & Baum, W. M. (2000). Choice in a variable environment: Every reinforcer counts. *Journal of the Experimental Analysis of Behavior, 74*(1), 1-24. doi:10.1901/jeab.2000.74-1

Ferster, C. B., & Skinner, B. F. (1957). SCHEDULES OF REINFORCEMENT. *Psychological Reports, 3*(4), 695-695.

Gallistel, C. R., Mark, T. A., King, A. P., & Latham, P. E. (2001). The rat approximates an ideal detector of changes in rates of reward: Implications for the law of effect. *Journal of Experimental Psychology-Animal Behavior Processes, 27*(4), 354-372. doi:10.1037//0097-7403.27.4.354

Hamming, R. W. (1950). ERROR DETECTING AND ERROR CORRECTING CODES. *Bell System Technical Journal, 29*(2), 147-160. doi:10.1002/j.1538-7305.1950.tb00463.x

Helbing, D. (2001). Traffic and related self-driven many-particle systems. *Reviews of Modern Physics, 73*(4), 1067-1141. doi:10.1103/RevModPhys.73.1067

Herrnstein, R. J. (1961). RELATIVE AND ABSOLUTE STRENGTH OF RESPONSE AS A FUNCTION OF FREQUENCY OF REINFORCEMENT. *Journal of the Experimental Analysis of Behavior, 4*(3), 267-272. doi:10.1901/jeab.1961.4-267

Herrstein, R. J. (1970). ON THE LAW OF EFFECT. *Journal of the Experimental Analysis of Behavior, 13*(2), 243-266. doi:10.1901/jeab.1970.13-243

Hunter, I., & Davison, M. (1979). DYNAMIC ANALYSIS OF CONCURRENT VARIABLE INTERVAL SCHEDULE PERFORMANCE. *New Zealand Psychologist, 8*(1), 45-45.

Killeen, P. (1972). MATCHING LAW. *Journal of the Experimental Analysis of Behavior, 17*(3), 489-&. doi:10.1901/jeab.1972.17-489

Kulubekova, S., & McDowell, J. J. (2008). A computational model of selection by consequences: Log survivor plots. *Behavioural Processes, 78*(2), 291-296. doi:10.1016/j.beproc.2007.12.005

Kulubekova, S., & McDowell, J. J. (2013). Computational model of selection by consequences: Patterns of preference change on concurrent schedules. *Journal of the Experimental Analysis of Behavior, 100*(2), 147-164. doi:10.1002/jeab.40

Mazur, J. E. (1992). CHOICE BEHAVIOR IN TRANSITION - DEVELOPMENT OF PREFERENCE WITH RATIO AND INTERVAL SCHEDULES. *Journal of Experimental Psychology-Animal Behavior Processes, 18*(4), 364-378. doi:10.1037/0097-7403.18.4.364

McDowell, J. J. (1986). ON THE FALSIFIABILITY OF MATCHING THEORY. *Journal of the Experimental Analysis of Behavior, 45*(1), 63-74. doi:10.1901/jeab.1986.45-63

McDowell, J. J. (2004). Computational model of selection by consequences. *Journal of the Experimental Analysis of Behavior, 81*(3), 297-317. doi:10.1901/jeab.2004.81-297

McDowell, J. J. (2013). A Quantitative Evolutionary Theory of Adaptive Behavior Dynamics. *Psychological Review, 120*(4), 731-750. doi:10.1037/a0034244

McDowell, J. J., & Caron, M. L. (2007). Undermatching is an emergent property of selection by consequences. *Behavioural Processes, 75*(2), 97-106. doi:10.1016/j.beproc.2007.02.017

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*(3), 387-403. doi:10.1901/jeab.2008.90-387

McDowell, J. J., & Popa, A. (2010). TOWARD A MECHANICS OF ADAPTIVE BEHAVIOR: EVOLUTIONARY DYNAMICS AND MATCHING THEORY STATICS. *Journal of the Experimental Analysis of Behavior, 94*(2), 241-260. doi:10.1901/jeab.2010.94-241

McDowell, J. J., Popa, A., & Calvin, N. T. (2012). SELECTION DYNAMICS IN JOINT MATCHING TO RATE AND MAGNITUDE OF REINFORCEMENT. *Journal of the Experimental Analysis of Behavior, 98*(2), 199-212. doi:10.1901/jeab.2012.98-199

Popa, A., & McDowell, J. J. (2010). The effect of Hamming distances in a computational model of selection by consequences. *Behavioural Processes, 84*(1), 428-434. doi:10.1016/j.beproc.2010.02.002

Rachlin, H. (1971). TAUTOLOGY OF MATCHING LAW. *Journal of the Experimental Analysis of Behavior, 15*(2), 249-&. doi:10.1901/jeab.1971.15-249

Skinner, B. F. (1981). SELECTION BY CONSEQUENCES. *Science, 213*(4507), 501-504. doi:10.1126/science.7244649

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science, 304*(5678), 1782-1787. doi:10.1126/science.1094765

Figure 1. The Virtual Organism's Behavioral Repertoire and Target Classes

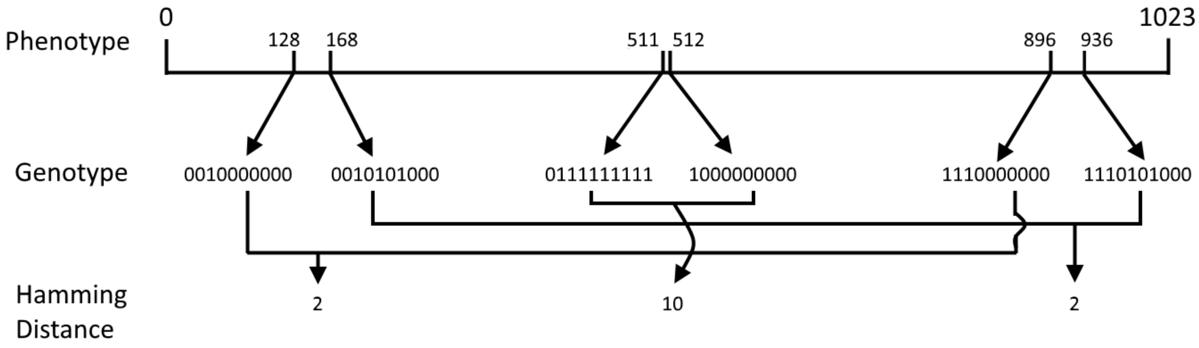


Fig 1. The 100 behaviors that constitute a virtual organism's behavioral repertoire have phenotypes between 0 and 1023, inclusively. The ranges 128-168 and 896-936 show the behaviors that fall into the two target classes used in this study. Each phenotype has a corresponding 10-bit genotype. The Hamming distance is the number of bits that must be flipped to covert one genotype into the other genotype linked to it.

Figure 2. Evolutionary Theory of Behavioral Dynamics - Flowchart

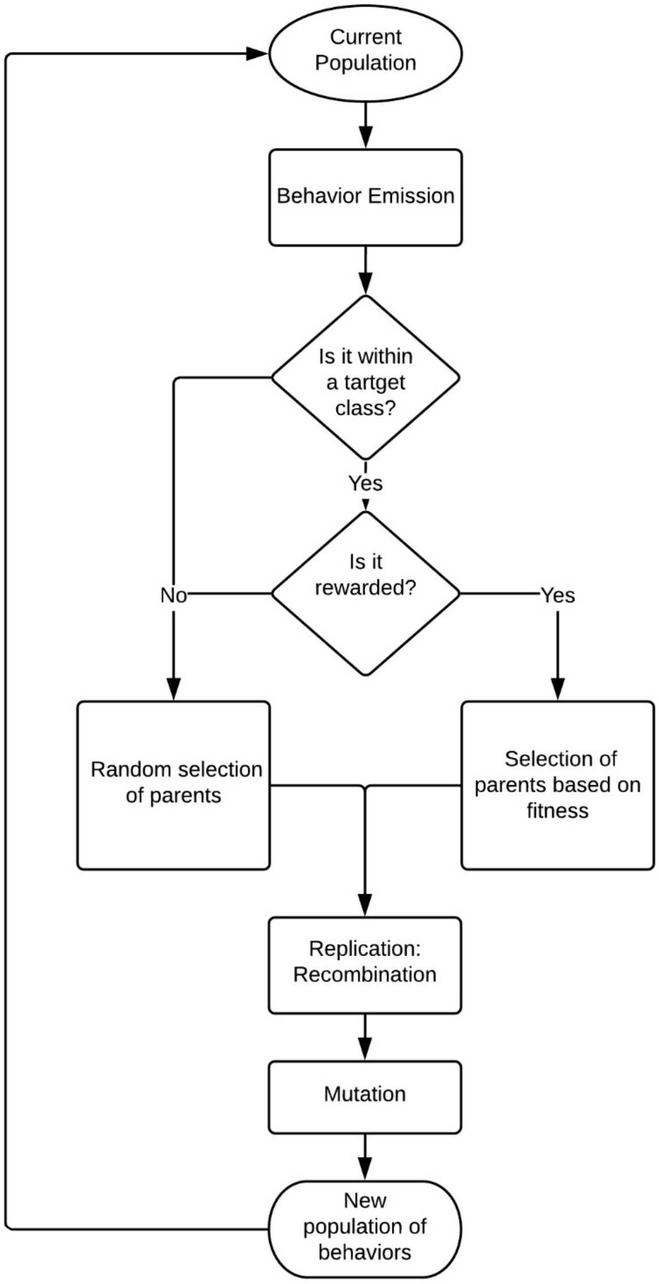


Fig 2. The flowchart demonstrates one iteration of the ETBD.

Figure 3 Transition Analysis Schematic

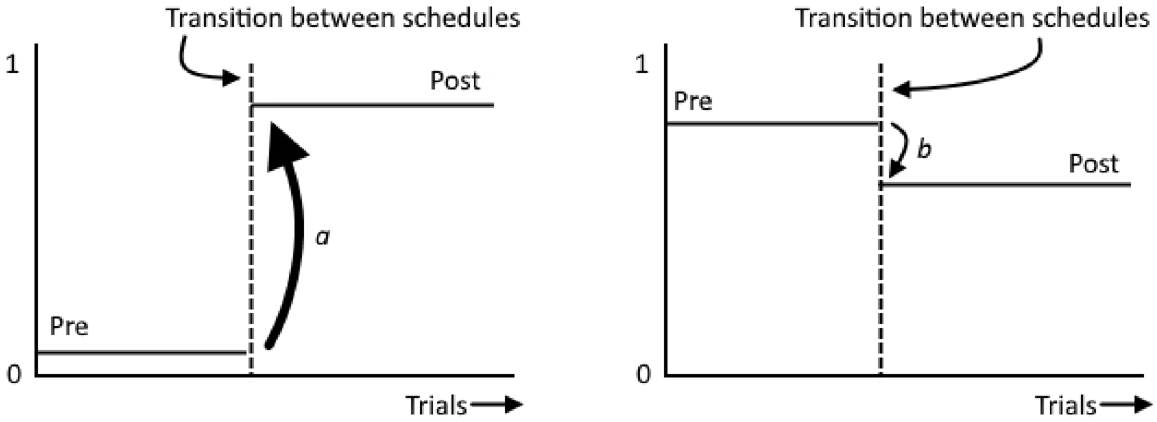


Fig 3. Transition analysis schematic. The schematic contains two example transitions to show the differences in magnitude and direction. The leftmost horizontal line, labeled Pre, represents the pre-transition reinforcement proportion and the rightmost horizontal line, labeled Post, represents the post-transition reinforcement proportion. The reinforcement proportions are based on how much they favor target class one. A reinforcement proportion that heavily favors target class one, e.g. 10:1, would have a transformed value close to one. The transition on the left shows a reinforcement proportion close to target two transitioning to a reinforcement proportion which is close to target one. The arrow, *a*, shows that this transition has a large magnitude and a positive direction within this framework. The transition on the right shows a transition between two reinforcement proportions that has a small magnitude difference and a negative direction, as shown by arrow *b*.

Figure 4. Macro Level – Power Function Matching Law Fit

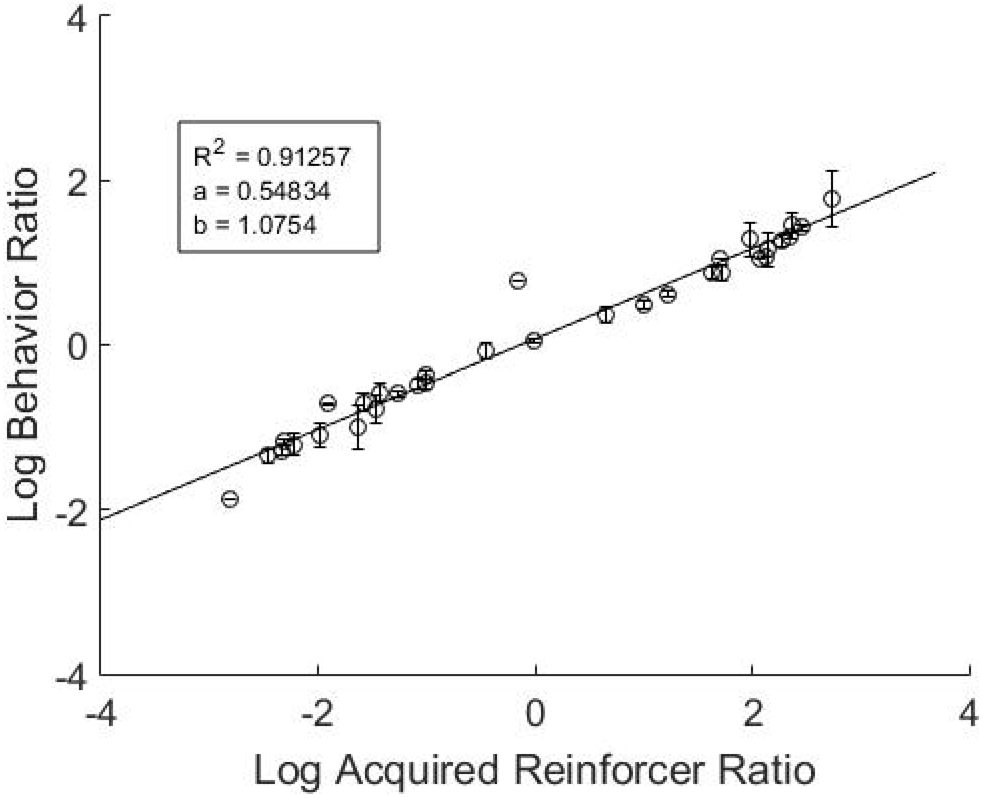


Fig 4. Power function matching law plot of the virtual organism with the closest R^2 value to the group average. Each data point represents the relationship between the log of the behavior ratio and the log of the acquired reinforcer ratio averaged across all schedules with the same scheduled reinforcement ratio. The solid line is a linear regression on the data set. The error bars show ±1 standard error. The values in the text box show the variance accounted for by the fit, R^2, the slope, *a*, and the intercept, *b*, for the fit of Equation 1.

Figure 5. Transition Level – Virtual Organism Average and Individual Plots

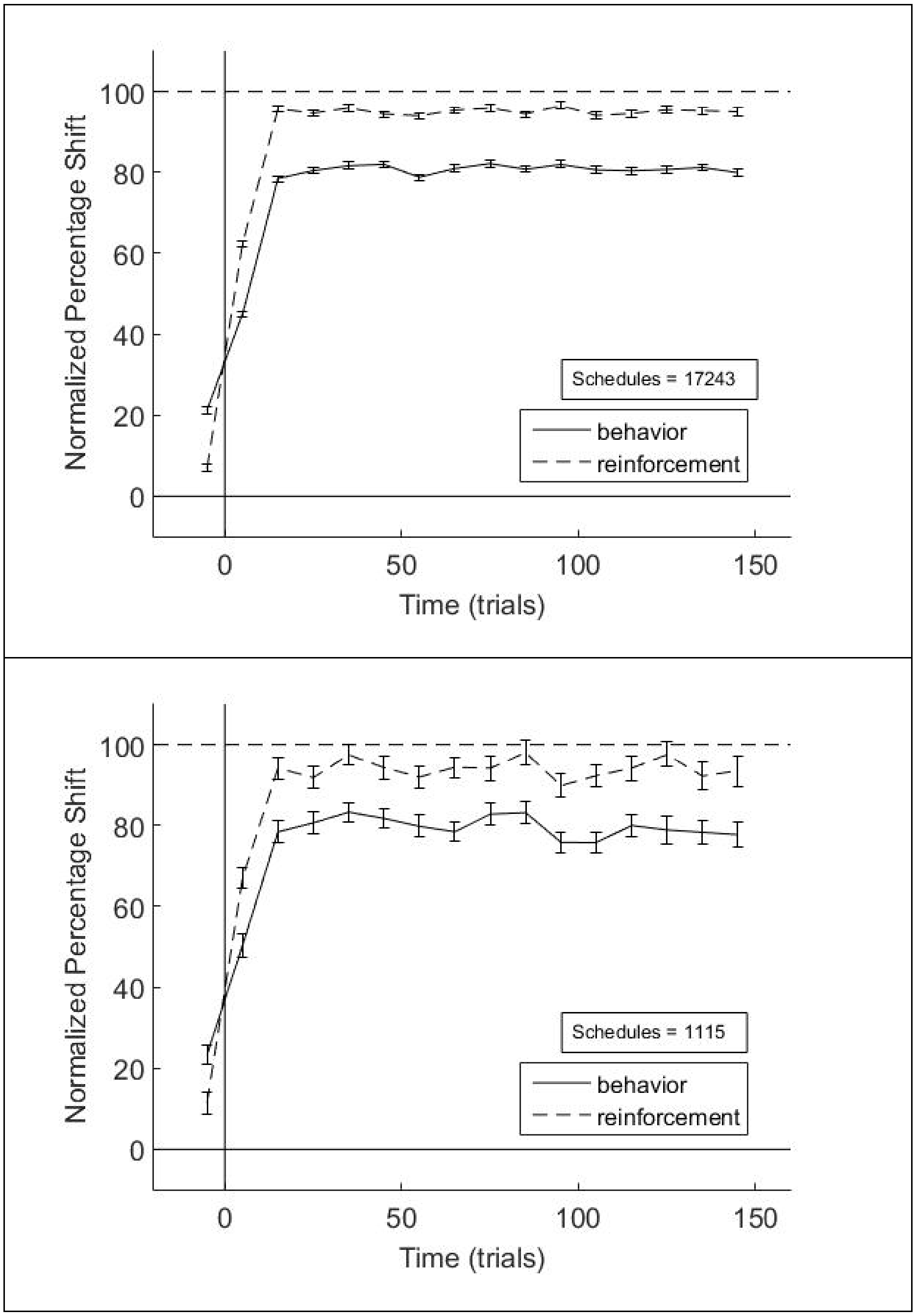


Fig 5. The averaged transition data of all 15 virtual organisms is shown at the top. The transition data for a typical single virtual organism is shown at the bottom. The behavior proportion (solid line) and the acquired reinforcer proportion (dashed line) are normalized to the difference between the pre-transition reinforcement proportion and the post-transition reinforcement proportion. The error bars show ±1 standard error for each data point. The pre-transition

reinforcement proportion is set at 0% and the post-transition reinforcement proportion is set at 100%. Proportions are calculated over ten-trial sets. The transition is set to be at time = 0. The first data point is a baseline calculated from the ten-trial set just prior to the transition. All of the transitions within one organism are averaged.

Figure 6 Transition Level – Virtual Organism and Monkey Behavior Comparisons

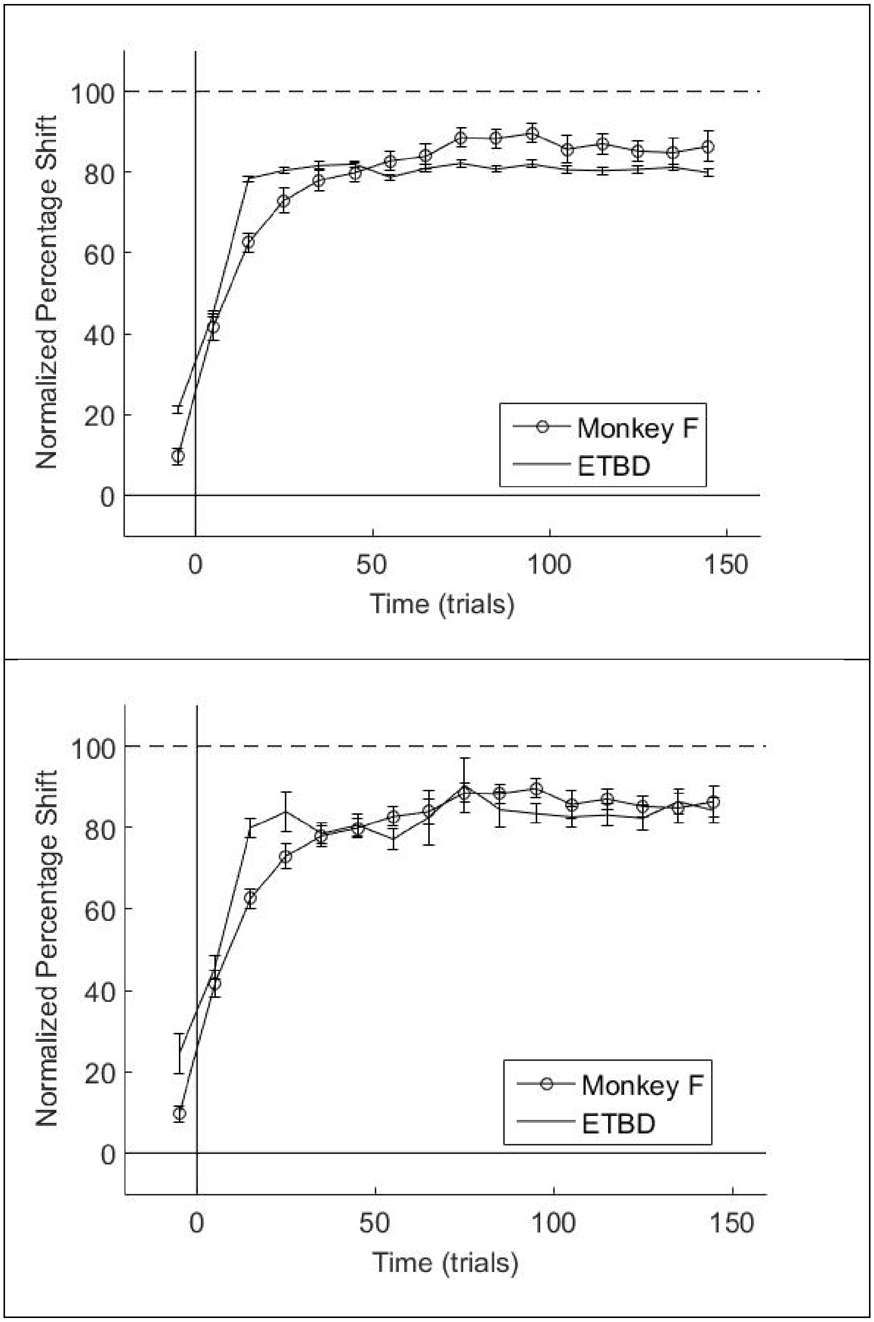
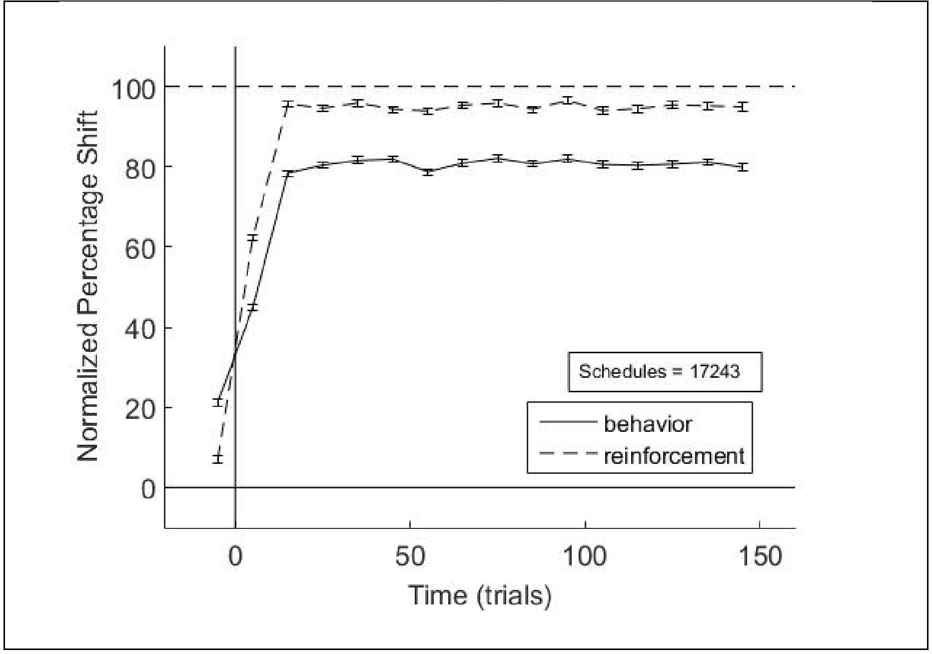


Fig 6. The averaged behavior proportion data for all 15 virtual organisms is compared with the monkey data at the top. The transition data of the virtual organism that is most similar in behavior proportion to the monkey is shown at the bottom. The behavior proportion generated by virtual organisms (solid line) and the monkey behavior proportion (circle) are normalized to the difference between the pre-transition reinforcement proportion and the post-transition

reinforcement proportion. The error bars show ±1 standard error for each data point. The pre-transition reinforcement proportion is set at 0% and the post-transition reinforcement proportion is set at 100%. Proportions are calculated over ten-trial sets. The transition is set to be at time = 0. The first data point is a baseline calculated from the ten-trial set just prior to the transition. All of the transitions within one organism are averaged.

Figure 7. Transition Level – Average Virtual Organism and Monkey Comparison

Virtual Organism



Monkey

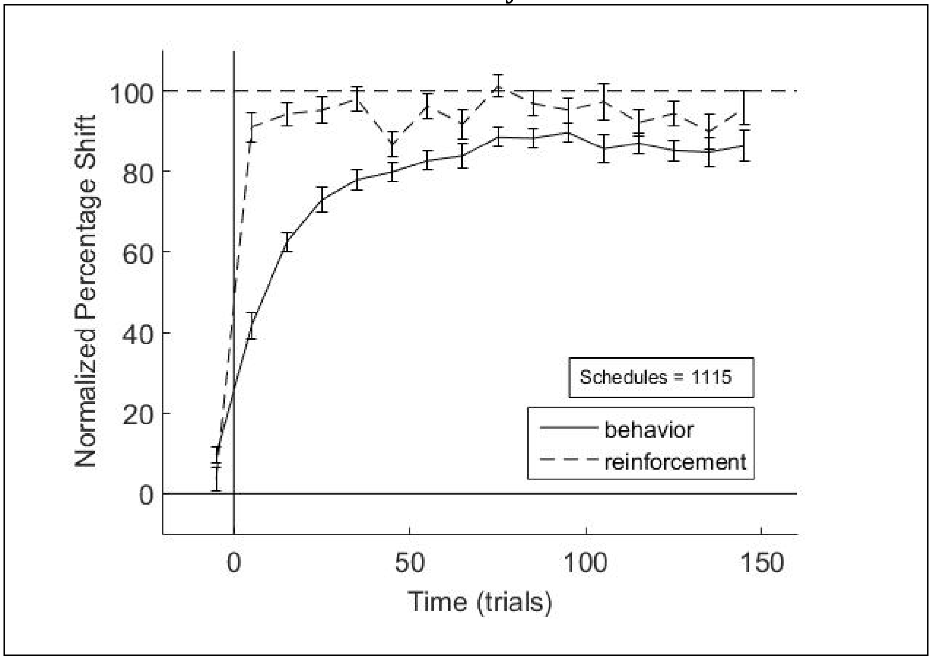
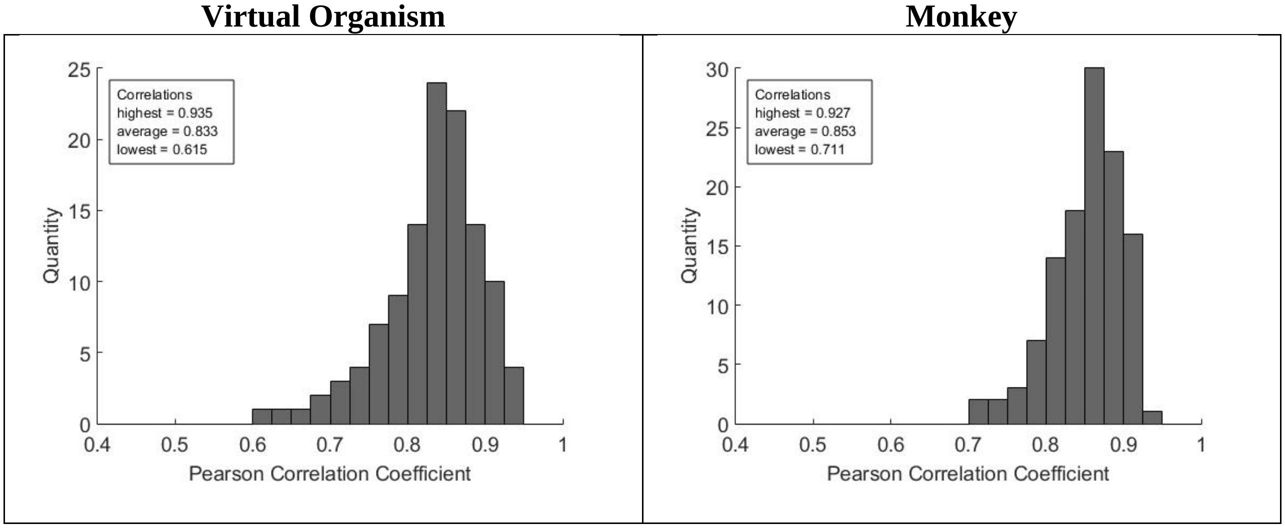


Fig 7. The average transition data of all 15 virtual organisms (top) and transition data for the monkey (bottom) are shown here. The behavior proportion (solid line) and the acquired reinforcer proportion (dashed line) are normalized to the difference between the pre-transition

reinforcement proportion and the post-transition reinforcement proportion. The pre-transition reinforcement proportion is set at 0% and the post-transition reinforcement proportion is set at 100%. Proportions are calculated over ten-trial sets. The transition is set to be at time = 0. The first data point is a baseline calculated from the ten-trial set just prior to the transition. All of the transitions within one organism are averaged. The error bars show ±1 standard error for each data point.

Figure 8. Local Level - Histograms of Block Correlation VO and Monkey Comparison


Virtual Organism

Monkey

Fig 8. Histogram of Pearson Correlation Coefficients for ten-schedule blocks. The entire sequence of dynamic schedules was broken into ten-schedule blocks. The Pearson's correlation between the behavior proportion and the acquired reinforcer proportion was calculated for each block in both the virtual organism data (left) and the monkey data (right) for the same sequence of schedules. The virtual organism shown here has the closest histogram to the group average.

Figure 9. Local Level – Virtual Organism Individual Plots

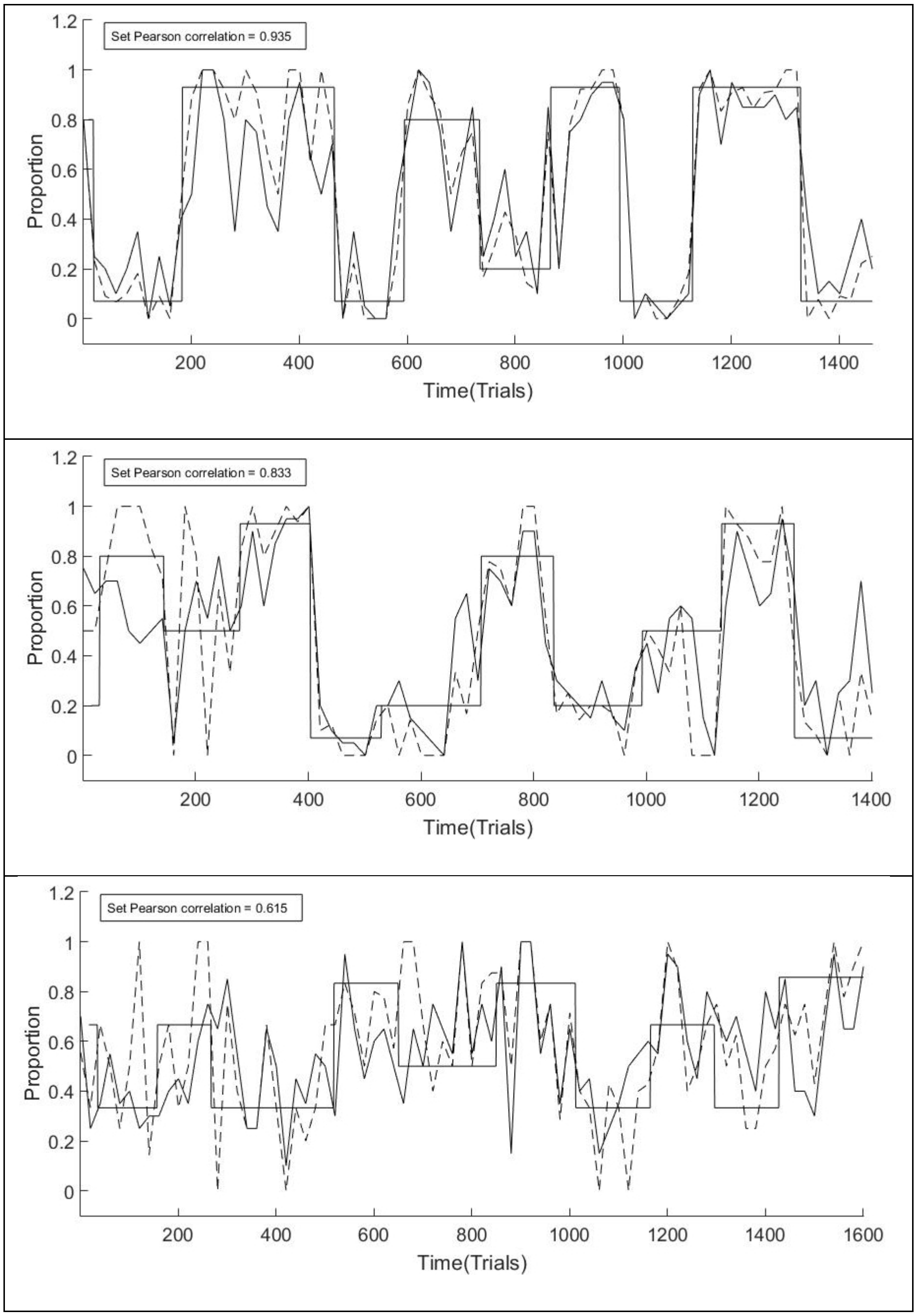


Fig 9. The blocks with the highest correlation (top), the average correlation (middle), and the lowest correlation (bottom) for a typical organism are shown here. Local level changes in reinforcer proportion (dashed line) and behavior proportion (solid line) are shown in the figure. The *y*-axis shows the proportion of behavior or acquired reinforcers on target class 1. The proportions are calculated over twenty-trial sets. The segmented horizontal line represents the

scheduled reinforcement proportions. Scheduled reinforcement proportions that favor target class 1 are near one on the *y*-axis. Scheduled reinforcement proportions that favor target class 2 are near zero on the *y*-axis. The correlation for each block is calculated between the behavior proportion and the acquired reinforcer proportion.

Figure 10. Local Level – Virtual Organism Averaged Plots

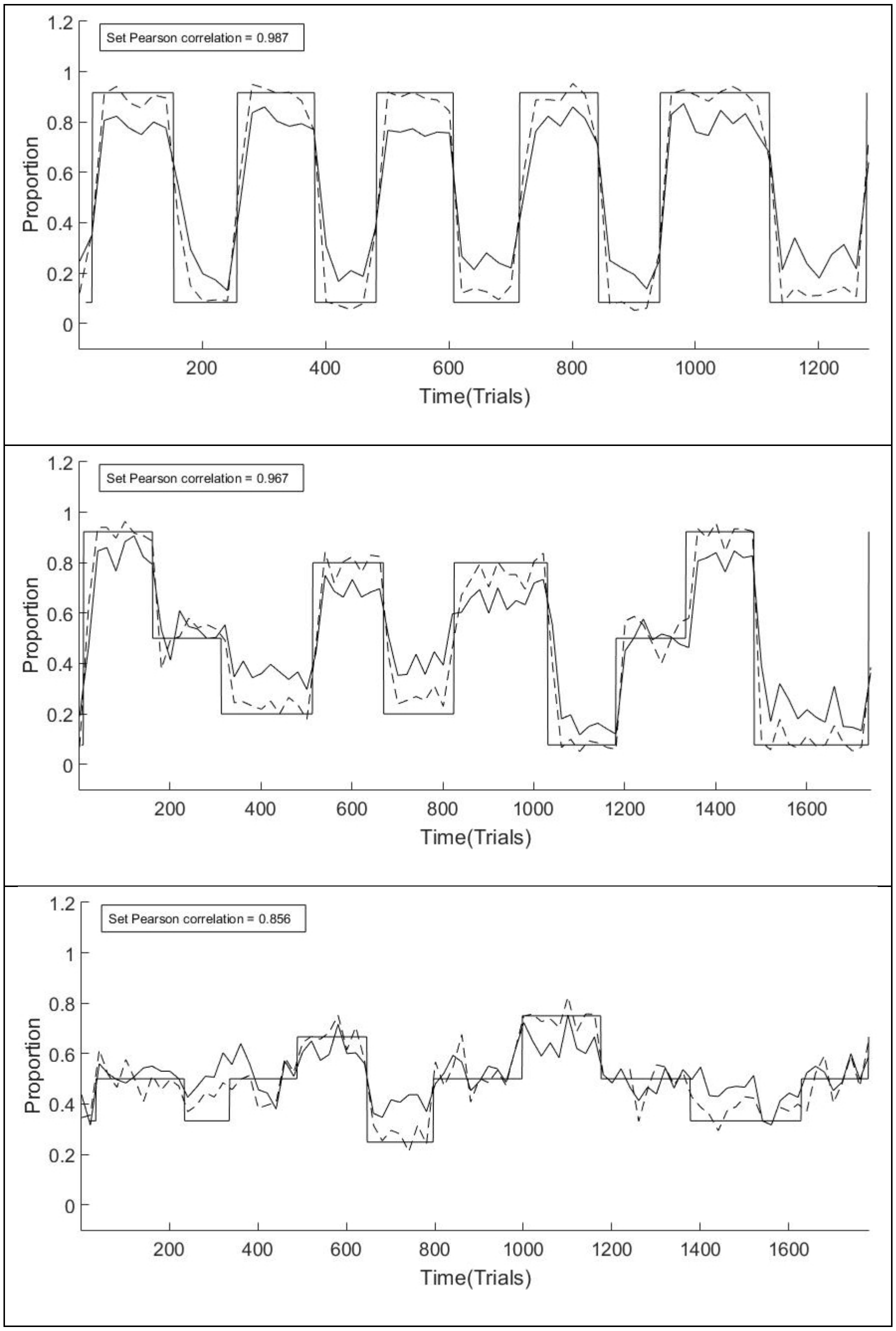
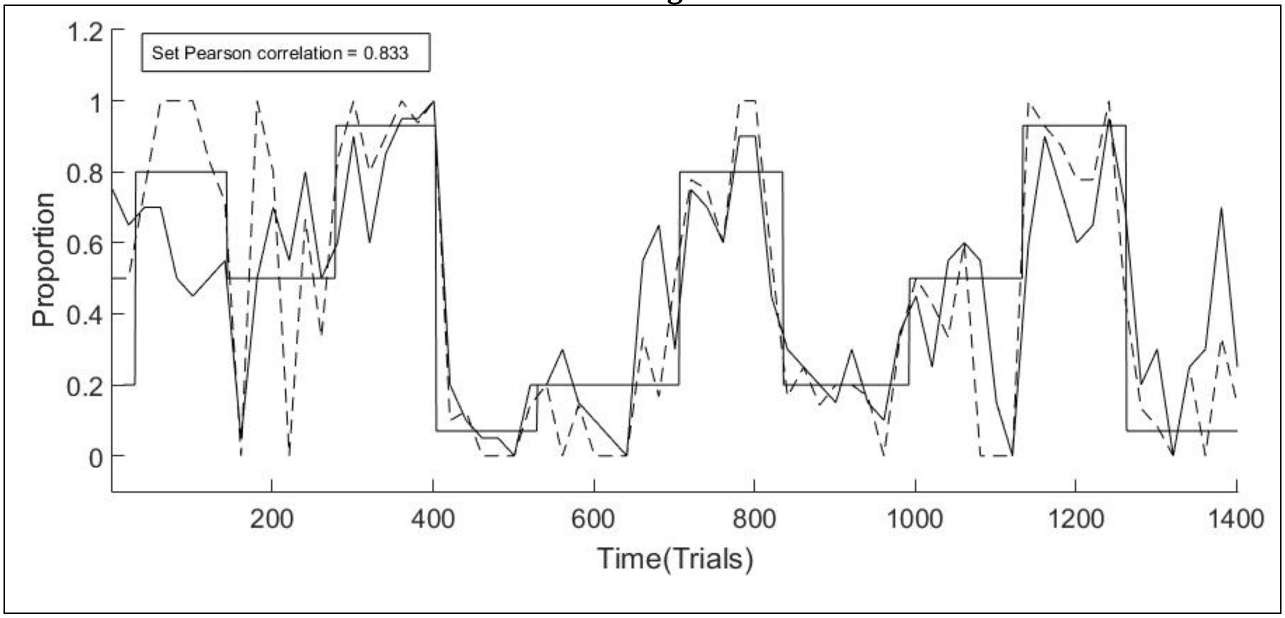


Fig 10. Local level changes are averaged over 15 virtual organisms. The blocks with the highest correlation (top), the average correlation (middle), and the lowest correlation (bottom) are shown here. The acquired reinforcer proportion (dashed line) and behavior proportion (solid line) are shown in the figure. The y-axis shows the proportion of behavior or acquired reinforcers on target class 1. The proportions are calculated over twenty-trial sets. The segmented horizontal

line represents the scheduled reinforcement proportion. Scheduled reinforcement proportions that favor target class 1 are near one on the *y*-axis. Scheduled reinforcement proportions that favor target class 2 are near zero on the *y*-axis. The correlation for each block is calculated between the behavior proportion and the acquired reinforcer proportion.

Figure 11. Local Level – Virtual Organism and Monkey Comparison

Virtual Organism



Monkey

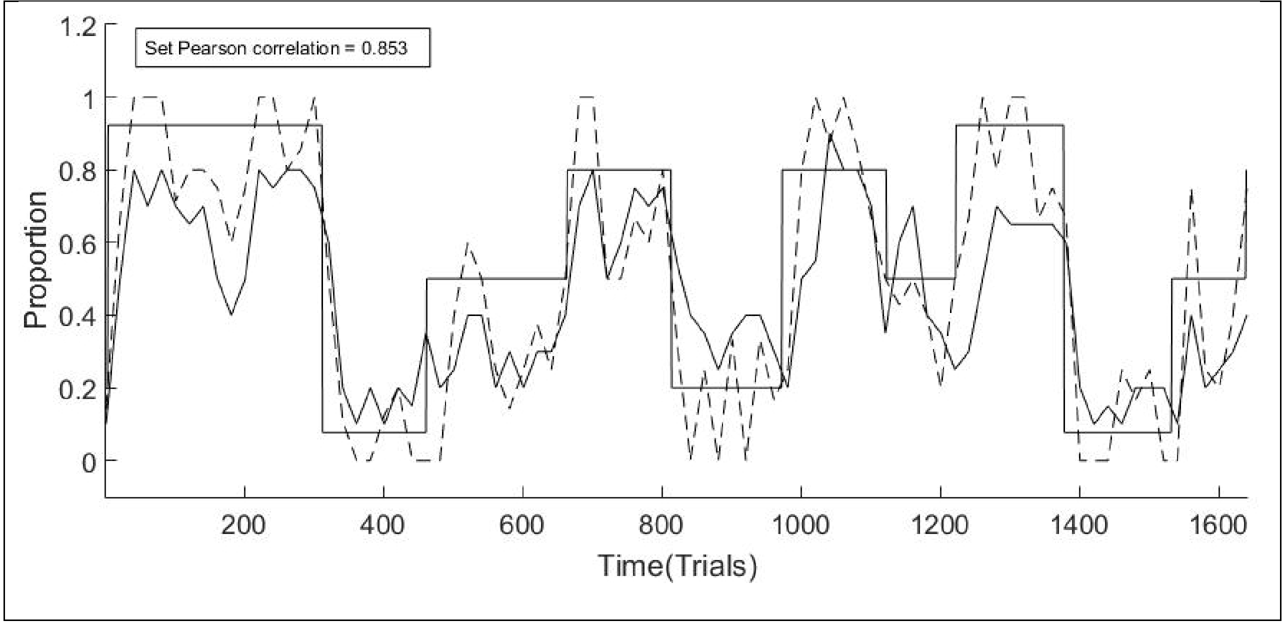


Fig 11. Blocks with an average Pearson correlation for a typical virtual organism (top) and for the monkey (bottom) are shown here. The acquired reinforcer proportion (dashed line) and behavior proportion (solid line) are shown in the figure. The *y*-axis shows the proportion of behavior or acquired reinforcers on target class 1. The proportions are calculated over twenty-

trial sets. The segmented horizontal line represents the reinforcement proportion. Reinforcement proportions that favor target class 1 are near one on the *y*-axis. Reinforcement proportions which favor target class 2 are near zero on the *y*-axis. The correlation for each block is calculated between the behavior proportion and the acquired reinforcer proportion.

Table 1: Table of schedule ratios used in the study

1:13.333	14.666:1
1:12.133	13.333:1
1:12.09	12:1
1:12	11:1
1:11	10:1
1:10	9:1
1:9	8.0769:1
1:8	8.0597:1
1:6.02	8:1
1:6	6:1
1:5	5.3:1
1:4	5:1
1:3.0189	4:1
1:3.0108	3:1
1:3	2:1
1:2.666	1:1
1:2	

Table 1. Schedule ratios between the two target classes.