2018

Computational Model for Behavior Shaping as an Adaptive Health Intervention Strategy

Vincent Berardi
Chapman University, berardi@chapman.edu

Ricardo Carretero-González
San Diego State University

Neil E. Klepeis
San Diego State University

Sahar Ghanipoor Machiani
San Diego State University

Arash Jahangiri
San Diego State University

See next page for additional authors

Computational Model for Behavior Shaping as an Adaptive Health Intervention Strategy

Authors

Vincent Berardi, Ricardo Carretero-González, Neil E. Klepeis, Sahar Ghanipoor Machiani, Arash Jahangiri, John Bellettiere, and Melbourne Hovell

Computational Model for Behavior Shaping as an Adaptive Health Intervention Strategy

V. Berardi[1], R. Carretero-González[2], N.E. Klepeis[3], S. Ghanipoor Machiani[4], A. Jahangiri[4], J. Bellettiere[3], M. Hovell[3]

[1]Chapman University, Department of Psychology

[2]San Diego State University, Department of Mathematics and Statistics

[3]San Diego State University, Center for Behavioral Epidemiology and Community Health

[4]San Diego State University, Department of Civil, Construction, and Environmental Engineering

Corresponding Author: Vincent Berardi, Chapman University, Psychology Department, 1 University Drive, Orange, CA 92866, USA. Email: berardi@chapman.edu

Computational Model for Behavior Shaping as an Adaptive Health Intervention Strategy

Introduction

Behavior Shaping in Health Interventions

The leading causes of death and disease in the United States are modifiable behavioral factors such as tobacco use, poor diet, physical inactivity, alcohol consumption, and avoidable injuries [1, 2]. Studies have indicated that preventable, extrinsic factors contribute >70-90% of lifetime cancer risks [3]. Consequently, enormous gains to public health are achievable through behavior-altering interventions. Most behavioral interventions, though, generate limited, short-lived effects [4], partly due to the reliance on episodic assessments of behavior captured by tools such as counseling sessions, ecological momentary assessments, surveys, and discrete direct observations. In contrast, recent advances in mobile technology have enabled the development of just-in-time, adaptive interventions (JITAIs) that have the potential to improve upon the shortcomings associated with traditional trials [5, 6]. JITAIS typically use assessment technology that is capable of observing and recording behavior in a natural environment on a near-continuous basis over a long period of time. By pairing intensive data collection with analytic systems capable of real-time decision making, JITAIs enable interventions to be provided on an ongoing basis and automatically adapt in response to participants' varying behaviors, environmental contexts, and past history. This process is hypothesized as an ongoing two-way conversation between patients and providers. While still in the preliminary stages, adaptive interventions have been implemented to, for example, encourage physical activity [7], assist with in-home living for older adults [8], and manage HIV medication adherence [9].

The implementation of JITAIs can be enhanced by consideration of the precise mechanisms by which interventions should adapt in response to participants' behavior. Often, adaptation strategies are developed in an ad-hoc fashion with little consideration of theoretical underpinnings, despite findings indicating that adherence to established theory

can increase the efficacy of behavioral interventions [10]. Current behavioral theories provide little insight in this regard since they rarely consider behavior as a dynamic entity [11]. Introducing theoretically-sound, responsive adaptation strategies will likely lead to more effective interventions and, furthermore, can generate results that will allow the underlying theories to be refined. One possible adaptation strategy with a theoretical foundation is *behavior shaping*, defined as the process whereby a targeted behavior is gradually cultivated via the differential reinforcement of successive approximations to the target [12]. When implementing this procedure, the range of behaviors that are reinforced narrows with time. [See Fig. 1i)]. The shaping process can lead to complex behaviors that would otherwise not be emitted as quickly or at all. In a traditional shaping scenario, a practitioner must discriminate which behaviors are sufficiently similar to the target behavior in order to receive reinforcement and determine the optimal time to discontinue the reinforcement associated with crude approximations. The withholding of reinforcement produces a temporary extinction condition that might occasion novel or differential rates of behavior that are appropriate for shaping. This process typically occurs in a controlled environment such as classroom or training center. Proficiency in performing these tasks arbitrate the ultimate effectiveness of shaping routines and specialists such as teachers and coaches do this as an art. JITAIs, though, offer the opportunity to precisely gauge behavior on a nearly constant basis and to continually assess its similarity to a target behavior. This enables shaping procedures to be automatically implemented in a much wider variety of contexts than has typically been possible [13].   Many behaviors are reliably shaped throughout society. For instance, the entire educational system can be viewed as a high-level shaping procedure where successively close approximations to proficiency in certain disciplines are reinforced as individuals proceed through each grade. Subsequent to these formal shaping protocols, the environment continues to shape behavior, albeit under less predictable schedules. For example, education is reinforced by the admission to college and subsequent employment. In contrast, behavior shaping

routines that would be deployed in JITAIs, at least during this preliminary stage, are likely to be rudimentary and not have the benefit of a host of strong, supporting contingencies. In some cases, such as when attempting to shape tobacco smoking cessation, the opposite may even be true and environmental factors could discourage the target behavior. The extent to which simple shaping procedures deployed within a natural environment would be successful in producing elevated levels of healthy human behavior is an open question that this manuscript begins the process of addressing.

Outside of the JITAI domain, the effectiveness of behavior shaping in humans has been reliably demonstrated in areas ranging from promoting motor activity in patients recovering from strokes [14] to improving dental treatment acceptance among children [15] to the management of cellular service consumption [16]. For individuals with autism, shaping has been used to promote socio-emotional functioning [17], to aid in toilet training [18], to increase the duration of sustained attention [19], and to develop social-cueing skills [20]. The latter of these examples utilized automatic behavior assessment features similar to those that are required for JITAIs.

Behavior Shaping and Computational Models

Using computational models to assess the effectiveness of behavior shaping routines in JITAIs is an attractive preliminary approach since it allows for the components of complex systems to be efficiently isolated and manipulated. Methodologies can be explored, tweaked, and sometimes abandoned without the complications associated with their real-world counterparts. The ultimate aim of the procedure presented herein is to leverage the insight gained within simplified, digital domains to develop increasingly more realistic controlled laboratory experiments and subsequent real-world trials, which will all share a common theoretical underpinning. This will allow behavior shaping programs to be designed with a degree of theoretical fidelity [10] that has been absent in this area thus far.

There is a rich history of implementing digital shaping programs within reinforcing learning models that are popular in the realm of artificial intelligence (AI) research. For

example, behavior shaping routines have been implemented in reinforcement learning schemes within computational models of bicycle riding [21] and navigating a rod around obstacles [22]. In addition to experiments occurring in a virtual environment, shaping has also been included within the reinforcement learning protocol for robots learning to simulate foraging and other survival behaviors [23, 24]. The shaping protocols in these reinforcement learning models typically consist of having the agent preliminarily complete a simplified version of the full target task and demonstrating that this priming increases the rate at which the target behavior is acquired. In contrast to the hypothesized implementation of shaping within JITAIs, these shaping routines have only a rudimentary temporal component and do not adapt over time, which limits their generalizability to the JITAI domain.

Due to the shortcomings of the existing computational shaping routines discussed above, this paper aims to develop computational models that are suitable for a JITAI framework. This is accomplished by modifying McDowell's evolutionary model of behavior dynamics [25] by incorporating behavior shaping. In accordance with Darwinist principles, the McDowell model emits a stream of behaviors chosen from a population via a system of selection, reproduction, and mutation, a process that may be equivalent to reinforcement learning [26]. As will be detailed below, this is an abstract model that considers a digital organism emitting generic behaviors. The absence of specificity regarding behaviors and targets is an attractive feature since, as opposed to the behavior-specific AI reinforcement learning tasks described above, it enables findings to be generalized to many different JITAIs. As described in Table 1, the McDowell computational model has consistently produced results that agree with many material world experimental findings. In the case of temporally-adaptive behavior shaping routines, the appropriate real-world experiments required for model comparison have not yet been performed. The results outlined within this paper lay the groundwork for the development of such experiments that will allow the consistency of computational and material-world findings to be assessed in order to inform

behavior shaping JITAIs.

Behavior Shaping in McDowell Model

Summary of Previous Work

Mcdowell's model [25] considers a hypothetical digital organism whose behavior evolves over time according to low-level rules informed by principles of behavior. This foundation defines the relationship between the emission of a behavior and its consequence, as specified by the probability of this behavior being emitted in the future. In a process similar to agent-based modeling, the interaction of these rules for various behaviors are simulated via computational experiments that produce emergent, higher-order results that cannot be extrapolated by solely examining the structure of the rules. The system is entirely decentralized without explicit considerations of global outcome and has stochasticity built into it. Drawing on the parallels between operant behavior and natural selection, model components are presented in evolutionary terms.

The behavior of the digital organism evolves over time according to the algorithm illustrated in Fig. 1ii) and now described. Each behavior is associated with a unique integer within the interval [0,1000] and at each time step a repertoire of 100 behaviors is active. The integers are stratified into behavior classes, one or more of which represents a targeted class which contains behaviors that are eligible to receive reinforcement when emitted. The experiments detailed herein include three classes: Class I: [0,494], Class II: [495,505],and Class III: [506,1000], the second of which is the target class. The range of integers defining behaviors, the size of the active repertoire and the specification of the reinforced class can be freely chosen. At each time step, the probability that a particular class of behavior is emitted is given by the proportion of total behaviors within the repertoire that correspond with this class. Based on the probabilities calculated for each class, one class is selected at random for emission. The specific behavior from within the class that is emitted is randomly selected from the behaviors in the repertoire associated

with this class. To create the behavioral repertoire at the next time step so this procedure can be iterated, "parent" behaviors are selected from the current behavioral repertoire and "cloned" and "mutated" to generate a new set of 100 behaviors comprising the new repertoire. The full details of this process are detailed in Ref. [25] and in the Appendix A, which describes deviations from the methodology outlined in Ref. [25].

The computational model includes a reinforcement component, which allows behavior shaping to be introduced. Reinforcement is defined as the delivery of a stimulus contingent upon performance of a behavior which results in the increased probability of future occurrences of this behavior and *others similar to it* [27]. This construct is included in the model as follows. If the emitted behavior is from the target class, a reinforcement schedule is consulted to determine whether this behavior should be reinforced. If reinforcement should occur, the emitted behavior is characterized as "fit" and the fitnesses of the other behaviors in the repertoire are based on their similarity (i.e., distance) to the reinforced behavior. A fitness function, fully detailed in Ref. [25], is then used to select the parent behaviors for the next repertoire that are similar to the emitted, reinforced behavior with preference given to the most similar behaviors. As a result, after cloning and mutation, the behaviors comprising the repertoire at the next time step will be similar to emitted behavior, satisfying the definition for reinforcement. If the emitted behavior is not reinforced, the parent behaviors are selected at random.

Operationalizing Behavior Shaping

The primary focus of the research summarized in this manuscript is to operationalize the construct of behavior shaping within the McDowell computational model so that digital experiments concerning its optimal implementation can be performed. To simulate the reinforcing of successive approximations to a target behavior, the McDowell model was modified so that a class of behaviors wider than the target behavior class was reinforced. Since shaping requires, as time progresses, behaviors to be increasingly similar to the desired behavior in order to receive reinforcement, the width of this reinforcement class

was gradually tightened according to some non-increasing function $w(t)$. The reinforced class at any given time is defined as $[500 - W, 500 + W]$, where $W = \frac{w(t)-1}{2}$. It follows that $w(t) = 11$ defines the reinforcement of only the targeted behavioral class [495,505]. To ensure that the reinforced class is defined by integers, all values of $w(t)$ are rounded to the nearest odd integer. As the behaviors defining the reinforced class are updated according to $w(t)$, the other classes must be updated as well. Class I is thus defined as $[0, 500 - W - 1]$ and Class III is defined as $[500 + W + 1, 1000]$. This study aims to identify $w^*(t)$, the function that *optimally* narrows the reinforcement class towards the target class.

Discrete Shaping Procedure

The simplest shaping procedures start with a reinforcement window that is wider than the target window and tighten it at discrete time point(s), which can be summarized by treating $w(t)$ as a step function. Several example step functions chosen for exploration and denoted as $w_n(t)$ for $n = 1 \ldots 6$ are shown in Fig. 2. Panel i) in this figure represents the baseline case where only the target behavior class $(w(t) = 11)$ is reinforced at all times. Figures 2ii) and iii), are denoted as 1-step shaping procedures. In this case, the class [489-511] $(w(t) = 22)$ was initially reinforced and then the reinforcement window was reduced to the target class at $t = 60$ and $t = 120$, respectively. Figures 2iv) and v), illustrates 2-step shaping procedures and Fig. 2vi) illustrates a 3-step shaping procedure.

Figure 2vii) depicts the results generated by these shaping functions. The metric shown is the percentage of target behaviors in the behavior repertoire at each time step. To account for stochasticity in the system, this value is averaged over 5,000 simulations. After approximately 50 time steps, each of the shaping procedures produces higher levels of the target behavior than does $w_1(t)$, where only the targeted class was reinforced. This demonstrates that the operationalization of shaping within the computational algorithm is functioning as expected since higher levels of behavior were eventually produced relative to the scenario where only the target behavior class was reinforced. $w_2(t)$ and $w_3(t)$ are both 1-step functions that utilize the same two reinforcement windows, but spend a different

amount of time in each window. The different results for these functions demonstrate the effect of this temporal feature on results. The 3-step function $w_6(t)$ produced a higher level of target behavior than the 2-step functions $w_4(t)$ and $w_5(t)$, which in turn produced a higher level of behavior than the 1-step functions. Taken together, these results indicate that continuously narrowing the reinforcement with time, as opposed to the discrete contraction used in this section, might lead to more pronounced behavior change.

Continuous Shaping Procedure

The first step in the continuous shaping procedure is to develop functions that can be used to guide the reinforced class width. These functions are analogous to those illustrated in Fig. 2, but are continuous and nonlinear. The following, generic piecewise function is used:

$$w(t) = \begin{cases} A(1 - e^{bt}) + w_0 & t \le t_f \\ w_c & t > t_f, \end{cases} \tag{1}$$

where w_0 is the initial width of the reinforcement class, b is an exponential loss/gain parameter, A is the distance between the horizontal asymptote and w_0, and t_f is the time at which the target reinforcement width is met. For $t > t_f$, only behaviors within the target behavior class are reinforced. To ensure that $w(t)$ is continuous, the restriction $w(t_f) = w_c$ is established, which leads to the condition $A \equiv \frac{w_c - w_0}{1 - e^{bt_f}}$.

In Eq. (1), negative values of b correspond to concave-up functions, which represent an initial rapid decrease in the reinforced class. Positive values of b correspond to concave-down functions, representative of a gradual initial decrease in the reinforced class along with a rapid narrowing of the reinforced class later in time. For $b = 0$, the non-constant component of $w(t)$ is undefined. However, a linear Taylor expansion about $b = 0$ shows that, in the limit, this function can be approximated by a straight line crossing through the two points $(0, w_0)$ and (t_f, w_c), i.e. the slope is $\frac{w_c - w_0}{t_f}$ and the y-intercept is w_0. Therefore, when $b = 0$ the reinforced class is narrowed at a constant rate. Figure 3i) illustrates the qualitative shape of Eq. (1) for different values of b.

In order to fully define Eq. (1), values for t_f, w_c, b, and w_0 are required. The chosen target behavior class of [495,505] corresponds to $w_c = 11$. t_f was set equal to 100 and the simulations were run until $t = 250$. These three values are free parameters of the system. Exploratory analyses outside the scope of this manuscript indicated that different values for these parameters did not affect the qualitative nature of the findings detailed below. The computational experiments performed in this work explore the effects of varying w_0, the initial value of $w(t)$ that can be interpreted as the maximal deviation from the target behavior that will result in reinforcement and b, the reinforcement window narrowing rate. The experiments were conducted for values of $(b, w_0) \in B \times W_0$, where $B = [-0.2, -0.19, ..., 0.19, 0.2]$ and $W_0 = [80, 120, ..., 200, 210]$.

Results of the Continuous Shaping Procedure

Figure 4i) illustrates the target behavior level (measured as in Fig. 2vii) as the percentage of target behaviors in the behavioral repertoire at each time step (averaged over 5,000 simulations) versus time for selected values of $(b, w) \in B \times W_0$. For all cases, including those not shown in Fig. 4i), the evolution of target behavior proceeds in, roughly, three different ways. For certain sets of parameters, the percentage of target behavior increases rapidly at the onset of the simulations and then asymptotes. For other sets of parameters, the trajectory is sigmoid-like with a nearly constant low level of behavior followed by a large, rapid jump to a higher level of behavior that approaches an asymptote. The last class of behavior does not begin to increase until t_f, when the target class is reached, and then increases gradually before asymptoting. To characterize the trajectory for the entire set of parameters, the following two metrics were calculated: 1.) the time at which the trajectory begins to increase, denoted by t_c and approximated by time at which a target behavior value equal to 15 is first breached [see horizontal dashed line in Fig. 4i)] and 2.) the asymptote as time approaches ∞, denoted by h_M and approximated by the maximum value a given trajectory realizes over the course of a simulation. The highest h_M values summarized in Figs. 4iii) and v) are higher than the those produced by the discrete

shaping functions in Figure 2, demonstrating the superiority of continuous shaping procedure.

Figures 4ii) and iii) illustrate t_c and h_M over $B \times W_0$. The variation in target behavior associated with parameter b, the curvature of the function, is much greater than the variation associated with w_0 for all metrics. For example, in Fig. 4ii), if b is fixed at zero and w_0 is varied, the range of the t_c values is approximately 50 to 90 while if w_0 is fixed at 150 and b is varied, t_c ranges from approximately 20 to 120.

As summarized in Fig. 4ii), the smallest value for t_c was 16, generated by six (b, w_0) combinations, all located in the bottom-left corner of the figure. These are concave up shaping functions characterized by a rapid initial decrease in reinforcement class width followed by long time intervals with a width relatively close to w_c. High values for t_c, on the other hand, are produced by parameter combinations in the upper-right corner of Fig. 4ii), which represent $w(t)$ functions that start relatively far from w_c and decrease very slowly until t is near t_f. Both of these results indicate that $w(t)$'s proximity to w_c is associated with a jump in the level of target behavior.

h_M (the asymptotic height) changes dynamically and only a snapshot at a particular moment can be illustrated. For instance, Fig. 4iii) presents the results at $t = t_f$, where the maximum value is 51.2% of target behaviors in the repertoire, which is associated with the parameter set $(b, w_0) = (-0.01, 130)$. In this snapshot, the largest values of h_M are associated with nearly linear functions that have negative b values near 0. But t_c, the time at which the trajectory jumps, is larger for these values of b than for large negative values of b. This represents competing effects where the target behavior trajectories that jump most quickly are not associated with highest levels of target behavior. To account for this competition between t_c and h_M in determining the overall levels of target behavior, the area under the trajectory curve (AUC) was also considered as a metric, as illustrated in Fig. 4iv). The maximum AUC is associated with $(b, w_0) = (-0.06, 100)$ and in general, larger area values are associated with concave-up functions as opposed to concave-down

functions.

The analyses described above were calculated at time $t = t_f$. As Fig. 4i) illustrates, many trajectories have not reached their maximum height at this time. At $t = 250$, which represents the end of the simulation, the trajectories have developed further and Fig. 4 v) and vi) illustrate h_M and AUC at this time. The concave-down functions associated with positive b values now have the highest h_M levels, although the largest values are still associated with essentially linear functions. The highest values are also associated with larger w_0 values. This rightward shift in the graph is mirrored when considering the AUC, as shown in Fig. 4vi). In this case the maximum values is for $(b, w_0) = (-0.01, 130)$. The functions with negative b values, i.e. the concave-up shaping functions, jumped to elevated target behavior levels very quickly. While the concave-down functions take longer to jump, their h_M values are higher. As longer time frames are considered, this elevated level of behavior outweighs the initial gains made by the concave-up functions and the AUC increases, as shown in Fig. 3ii). The maximum h_M values, though, remain centered around $b = 0$, indicating that a linear function will ultimately result in the most target behavior.

To summarize, concave-up shaping functions $(b < 0)$ produce rapid increases to asymptotic target behavior levels, but this asymptote is lower than that of linear $(b = 0)$ and concave-down functions $(b > 0)$, particularly as the simulation runs for longer periods of time. The linear and concave-down shaping functions result in extended periods of a low rate of target behavior before jumping up to higher levels.

Effects of Model Parameter Variation

Fixed Initial Behavioral Repertoire

The results summarized in Fig. 4i) point to an upper limit for the horizontal asymptote of the behavior trajectories. The generic $w(t)$ functions used in the continuous shaping procedure were selected to capture a range of concavity characteristics. It is possible that more complex shaping functions could ultimately lead to higher levels of

targeted behavior. This section explores the level of targeted behavior supported by the computational model with the specific parametrization described in Appendix A. Simulations were performed where some portion of the *initial* behavioral repertoire was required from the outset to be from the target behavior class. This stands in contrast to the standard procedure where the initial behavior class is chosen at random. Furthermore, setting a fraction of the initial behavior class to the target behavior simulates a previous learning history for the organism, a scenario which more accurately reflects the real-world conditions to which this model can be applied.

Once a fraction of the initial behavioral repertoire was fixed, only behaviors within the targeted class were reinforced, i.e. there was no shaping. The behavior trajectories, averaged over 5,000 simulations, for various proportions are illustrated in Fig. 5. There appears to be an upper limit of approximately 65% for the asymptotic level of targeted behavior, even for the ideal case when *all* behaviors initially in the repertoire are target behaviors and each of these target behaviors are reinforced when emitted. This feature is due to the shaping and cloning procedures (see. Ref. [25] and Appendix A) which result in sufficient variance in the next generation of behaviors to ensure that non-target behaviors are included in the repertoire. Interestingly, beginning with 60% and 70% of the initial behavioral repertoire in the target class results in an overshoot of this asymptotic level, but eventually the rate of behavior decreases to the asymptote. In the simulations detailed thus far, asymptotes as high as approximately 55% have been observed. Given an upper limit of 65% and the fact that the standard shaping procedure begins with a random initial behavioral repertoire, it is not expected that alternate shaping functions would drastically improve the ultimate levels of behavior generated.

Parameter Variation within the Computational Model

In all of the previously-detailed findings, a fixed-ratio 1 (FR1) reinforcement schedule was implemented, meaning that every time a behavior from the reinforcement class was emitted, it was reinforced. The effects of utilizing an FR3 schedule, where every third

reinforcable behavior that is emitted is reinforced, was also explored. Figure 6 illustrates the effects of this schedule, where differences compared with the previously-described results can be seen. As is expected with less reinforcement, the overall target behavior levels are lower. This effect is particularly pronounced for parameter sets pairing large, negative values of b with small values of w_0 (i.e. the lower left corner of the figure), where t_c values are much larger. It appears that the infrequent reinforcement coupled with a sharp initial reduction in reinforcement window does not result in sufficient reinforcement for shaping to be effective. This reinforcement schedule also resulted in the highest AUC values being more localized around $b = 0$ than was the case for previous analyses. As was the case for the FR1 schedule, at $t = 250$, the maximum h_M values are associated with linear functions, but, in general, the values are relatively higher for concave-down shaping functions.

The effects of varying the reinforcement strength (how similar the next repertoire is to a reinforced behavior), target class size, and time to target class were also explored. A full accounting of these results is beyond the scope of this article, but the results of these analyses were in accordance with the findings above, namely they contained a tradeoff between trajectories that quickly jump to elevated levels of target behavior versus trajectories that took longer to jump to ultimately higher levels of target behavior. Approximately linear shaping functions produced the highest levels of behavior. This indicates that the results detailed within are not a function of the choice of modeling parameters.

Discussion

This work demonstrated the viability of using computational models to investigate behavior shaping routines, a process that may be valuable in developing an alternative to the ad-hoc modifications often incorporated into adaptive, just-in-time, health behavioral interventions. The results indicate that shaping was effective at engendering higher levels of target behavior than when only the target behavior class is reinforced. When shaping

target behavior, narrowing the scope of behaviors that are reinforced on a continuous basis rather than at discrete time points, is more effective in producing the target behavior. Within this continuous framework, computational experiments were performed to explore the role of both w_0, the initial size of the reinforced behavior class, and b, which determines the concavity of reinforced class narrowing, on the effectiveness of shaping routines. The b values were more crucial in arbitrating the ultimate effectiveness of shaping. When considering the total amount of target behavior produced, there were two competing effects to consider: concave-up shaping functions that resulted in the percentage of target behavior quickly jumping to a relatively low asymptotic value **versus** concave-down shaping functions where the percentage of target behavior took a longer period of time to jump to higher asymptotic levels. Approximately linear functions did the best job of managing these two effects and led to the highest levels of target behavior.

There are practical conclusions to be drawn from the results outlined in the previous paragraph. If high levels of target behavior are the chief concern of the shaping system, then either a linear or concave-down shaping function should be used to guide the evolution of reinforcement windows with the latter being appropriate if there it is only possible to reinforce behaviors that are relatively similar to the target behavior, i.e. small w_0. An example of this scenario is someone training for an athletic competition, where the highest levels of target behavior are desired regardless of the time it takes to reach this goal. As an alternative scenario, consider a target behavior that is defined as the absence of some deleterious behavior, such as cigarette smoking. Shaping routines could be used within a gradual cessation program by informing the scheduling of increasingly longer intervals between prompts designed to support cessation. If a linear or concave-down function is used to guide the shaping, considerable harm could be done during the extended period of low-level target behavior (high levels of smoking). In scenarios like this where it is critical to change the behavior as soon as possible, a concave-up shaping procedure might be preferred. Once the individual has reached a constant level of target behavior that has a

sufficiently low risk, the shaping schedule can be transitioned to a linear/concave function that may increase the level (or improve the topography) of targeted behavior even further.

Behavior shaping can be implemented in nearly any domain, but the argument for formalized, automated shaping procedures presented herein requires the use of real-time sensing technology such as accelerometers, particle sensors, and smart outlets. These technologies provide the ability to continually assess an emitted behavior's similarity to a target behavior in order to determine the ideal moment for reinforcement. As an example, our research team recently completed an intervention that used real-time air monitors in homes to discourage secondhand smoke exposure by providing reinforcement when air particle levels stay below a threshold for some extended period of time [28]. These low-particle time intervals represent approximations to the desired behavior of no particle exposure at any time. Shaping would proceed by requiring increasingly large time intervals in order to receive reinforcement. The computational shaping platform in this manuscript represents a tool to investigate the optimal way to expand the duration of intervals required for reinforcement. The intervention model in this example can be replicated in many fields, including the use of accelerometers to shape shorter bouts of sedentary behavior or the use of screen tracking devices to promote less screen time. Each of these interventions can be informed by the computational shaping model, but as discussed in the next paragraph, the translation to real-world scenarios presents complications to be addressed.

The shaping procedures outlined within this report proceed by reinforcing a class of behaviors that are wider than a target class. In the computational model, the similarity between non-target and the target behavior is clearly defined as the difference between two integers. This feature is not easily translated into most real-world scenarios. For instance, while a large w_0 value in $w(t)$ indicates that behaviors that are quite dissimilar from the target will be reinforced at the outset of the shaping procedure, there is no interpretation as to whether this difference is based on function or topography or how the model would handle very rare, distal approximation to the target. It is not likely that it will be possible

to associate a one-to-one correspondence between the model constituents and the components of any material-world models. It has been argued in Ref. [29] though, that this lack of accordance is a feature of many successful models. For example, consider quantum theory where underlying model components do not conform to standard descriptions of space and time and, therefore, do not have an analogue in experimental model components. Rather, consistency across model predictions and experimental findings are sufficient to declare the two systems to be computationally equivalent, making the computational model a suitable platform for investigation. While agreement between McDowell computational model findings and several real-world experiments have been demonstrated (see Table 1), to the best of our knowledge, behavior shaping experiments with continuously-adapted reinforcement criteria have not yet been performed. Laboratory experiments including this feature are currently under development by the authors of this paper and, when performed, they will allow the computational equivalency of the model and real-world findings to be assessed. This will likely increase the interpretability and generalizability of computational results and refine the ability to automatically programming generalizable shaping procedures.

In addition to being compared to real world experiments, the computational modeling can be made more robust by exploring more complex modeling scenarios, such as shaping the extinction rather than the establishment of behavior. More complicated reinforcement schedules can be implemented and the consequences of errors within the shaping routine (e.g., incorrectly reinforcing a behavior that is not eligible for reinforcement) can also be explored. In particular, the use of variable ratio (and interval) reinforcement schedules should be explored since the ultimate goal of public health interventions is to sustain healthy behavior on a long-term basis and variable schedules are known to lead to longer maintenance effects. Whether it is better to implement the shaping routine with a variable schedule or to shape with a continuous schedule and then transition to a variable schedule is an open question to be investigated.

A focus on single simulations rather than aggregate results would be more appropriate for comparison to JITAIs. As these features are added and the computational shaping model becomes more rigorous, it will have an increasing potential to serve as a beneficial tool to be used in the design and refinement of JITAIs.

Implications

Practice: A computational model of behavior shaping has been developed in order to aid with the formalization and automation of behavior shaping procedures within adaptive behavioral interventions that utilize real-time technology.

Policy: Policymakers interested in leveraging real-time sensing technology to automatically personalize behavioral interventions should consider theoretically-rooted intervention adaptation strategies.

Research: Further research should be aimed at verifying computational model outcomes with real-world experiments.

References

1   Danaei, G, Ding, E L, Mozaffarian, D, et al. The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. PLoS Med. 2009; 6: e1000058.

2   Mokdad, A H, Marks, J S, Stroup, D F, and Gerberding, J L. Actual causes of death in the United States, 2000. Jama. 2004; 291: 1238–1245.

3   Wu, S, Powers, S, Zhu, W, and Hannun, Y A. Substantial contribution of extrinsic risk factors to cancer development. Nature. 2016; 529: 43–47.

4   Glanz, K and Bishop, D B. The role of behavioral science theory in development and implementation of public health interventions. Annual review of public health. 2010; 31: 399–418.

5   Nahum-Shani, I, Hekler, E B, and Spruijt-Metz, D. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. Health Psychology. 2015; 34: 1209.

6   Spruijt-Metz, D, Hekler, E, Saranummi, N, et al. Building new computational models to support health behavior change and maintenance: new opportunities in behavioral research. Translational behavioral medicine. 2015; 5: 335–346.

7   Adams, M A, Sallis, J F, Norman, G J, Hovell, M F, Hekler, E B, and Perata, E. An adaptive physical activity intervention for overweight adults: a randomized controlled trial. PloS one. 2013; 8.

8   Reeder, B, Meyer, E, Lazar, A, Chaudhuri, S, Thompson, H J, and Demiris, G. Framing the evidence for health smart homes and home-based consumer health technologies as a public health intervention for independent aging: A systematic review. International journal of medical informatics. 2013; 82: 565–579.

9   Pellowski, J A, Kalichman, S C, White, D, Amaral, C M, Hoyt, G, and Kalichman, M O. Real-time medication adherence monitoring intervention: test of concept in people living with HIV infection. The Journal of the Association of Nurses in AIDS Care: JANAC. 2013; 25: 646–651.

10   Rovniak, L S, Hovell, M F, Wojcik, J R, Winett, R A, and Martinez-Donate, A P. Enhancing theoretical fidelity: an e-mail-based walking program demonstration. American Journal of Health Promotion. 2005; 20: 85–95.

11   Riley, W T, Rivera, D E, Atienza, A A, Nilsen, W, Allison, S M, and Mermelstein, R. Health behavior models in the age of mobile interventions: are our theories up to the task? Translational behavioral medicine. 2011; 1: 53–71.

12   Cooper, J, Heron, T, and Heward, W. Applied Behavior Analysis. Always learning. Pearson Education, Limited, 2013.

13   Nakajima, T, Lehdonvirta, V, Tokunaga, E, and Kimura, H. Reflecting human behavior to motivate desirable lifestyle. Proceedings of the 7th ACM conference on Designing interactive systems. ACM. 2008: 405–414.

14   Taub, E, Crago, J E, Burgio, L D, et al. An operant approach to rehabilitation medicine: overcoming learned nonuse by shaping. Journal of the experimental analysis of behavior. 1994; 61: 281–293.

15   Hoist, A and Ek, L. Effect of systematized 'behavior shaping' on acceptance of dental treatment in children. Community dentistry and oral epidemiology. 1988; 16: 349–355.

16   Dawson, C, Rick, A, Seaman, J, and Waters, T. Traffic shaping of cellular service consumption through modification of consumer behavior encouraged by cell-based pricing advantages. US Patent 7,328,001. Feb. 2008.

17   Corbett, B A, Gunther, J R, Comins, D, et al. Brief report: theatre as therapy for children with autism spectrum disorder. Journal of autism and developmental disorders. 2011; 41: 505–511.

18   Kroeger, K and Sorensen-Burnworth, R. Toilet training individuals with autism and
      other developmental disabilities: A critical review. Research in Autism Spectrum
      Disorders. 2009; 3: 607–618.

19   Gutbrod, T. "Evaluating the Efficacy of Shaping with a Percentile Schedule of
      Reinforcement to Increase Duration of Sustained interaction in Children Diagnosed
      with Autism." PhD thesis. University of South Florida, 2014.

20   Greczek, J, Kaszubski, E, Atrash, A, and Mataric, M. Graded cueing feedback in
      robot-mediated imitation practice for children with autism spectrum disorders. Robot
      and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International
      Symposium on. IEEE. 2014: 561–566.

21   Randlov, J and Alstrom, P. Learning to Drive a Bicycle Using Reinforcement
      Learning and Shaping. ICML. Vol. 98. Citeseer. 1998: 463–471.

22   Konidaris, G and Barto, A. Autonomous shaping: Knowledge transfer in reinforcement
      learning. Proceedings of the 23rd international conference on Machine learning. ACM.
      2006: 489–496.

23   Dorigo, M and Colombetti, M. Robot shaping: Developing autonomous agents
      through learning. Artificial intelligence. 1994; 71: 321–370.

24   Mataric, M J. Reward functions for accelerated learning. Machine Learning:
      Proceedings of the Eleventh international conference. 1994: 181–189.

25   McDowell, J. A computational model of selection by consequences. Journal of the
      Experimental Analysis of Behavior. 2004; 81: 297–317.

26   Walsh, M M and Anderson, J R. Navigating complex decision spaces: Problems and
      paradigms in sequential choice. Psychological bulletin. 2014; 140: 466.

27   Skinner, B F. The behavior of organisms: an experimental analysis.
      Appleton-Century, 1938.

28   Hovell, M F. NIH/NHLBI Grant: Innovation for Smoke-Free Homes: Real-Time Feedback (R01 HL103684-05). 2011.

29   McDowell, J. Representations of complexity: how nature appears in our theories. The Behavior Analyst. 2013; 36: 345–359.

30   Moler, C. Matrix computation on distributed memory multiprocessors. Hypercube Multiprocessors. 1986; 86: 181–195.

31   McDowell, J, Caron, M L, Kulubekova, S, and Berg, J P. A computational theory of selection by consequences applied to concurrent schedules. Journal of the Experimental Analysis of Behavior. 2008; 90: 387–403.

32   McDowell, J, Popa, A, and Calvin, N T. Selection dynamics in joint matching to rate and magnitude of reinforcement. Journal of the experimental analysis of behavior. 2012; 98: 199–212.

33   Popa, A and McDowell, J. The effect of Hamming distances in a computational model of selection by consequences. Behavioural processes. 2010; 84: 428–434.

34   Kulubekova, S and McDowell, J J. Computational model of selection by consequences: patterns of preference change on concurrent schedules. Journal of the experimental analysis of behavior. 2013; 100: 147–164.

35   Kulubekova, S and McDowell, J. A computational model of selection by consequences: Log survivor plots. Behavioural Processes. 2008; 78: 291–296.

Appendix A: Details of the McDowell Computational Model

The computational model begins by randomly selecting a repertoire of 100 behaviors from the 1,001 possibilities and grouping them by their respective classes. The probability that a particular class of behaviors is emitted is defined as the proportion of total behaviors within the repertoire that correspond to this class. Based on these probabilities, a class of behavior is randomly chosen to be emitted and a specific behavior in the repertoire from this class is randomly selected. If the selected behavior is from the target class, a reinforcement schedule is consulted to determine whether this behavior is reinforced. Any behavior that is to be reinforced is characterized as "fit" and the fitness of the other behaviors in the repertoire is based on their similarity (distance) to the reinforced behavior. The fitness metric is used to select 100 "parent" behaviors that will generate the behavior repertoire at the next time step with preferential selection given to behaviors that are most similar to the behavior that is being reinforced. The relationship between a behavior's fitness and the likelihood of its selection as a parent behavior is governed by a *parental fitness function*, which can be any function that assigns higher probabilities to smaller distances. The analyses in this report used the exponential parental fitness function developed in Ref. [25] as follows. Let $p(x) = re^{-rx}$ be the probability of reinforcing a fitness value of x for $x \in [0, \infty)$. The mean is calculated as $\mu \equiv \int_0^\infty xp(x) \, dx = \frac{1}{r}$. This relationship implies $r = \frac{1}{\mu}$ making the cumulative density $P(x) = 1 - e^{-\frac{1}{\mu}x}$. Using this function, inverse transform sampling is used to select a value at random from this distribution. The fitnesses of the current behavioral repertoire are searched for a match and if one does not exist, a new random value is generated. This process continues until 100 parent behaviors have been selected. μ, the mean, is the only value required to parameterize this procedure. If the reinforcement schedule indicates that reinforcement is not available or if the emitted behavior is not from the target class, then rather than using the procedure described above, 100 parent behaviors are selected at random from the repertoire.

Each of selected parents is "cloned" to generate the behavior repertoire at the next time step. In the original McDowell model, parent behaviors were "mated" via a bitwise, binary procedure rather than cloned, but diagnostic analyses revealed biases within this procedure that were not present with cloning. Cloning proceeds by considering a collection of Gaussian distributions with a mean set equal to each of the parent behaviors. The standard deviation of these distributions, a free parameter of the system, was set equal to 2 and one new behavior for the next repertoire was then selected from each one of these distributions. A certain percentage of this new generation of behaviors are selected at random for mutation, which is also done by considering a Gaussian distribution, with mean set equal to integer representation of the behavior to be mutated and some standard deviation as a parameter of the system. For the analyses presented herein, a proportion of 0.01 behaviors in the new repertoire were selected, at random, for mutation and the standard deviation was 2.5. For both of the cloning and mutating steps, all calculations are performed using modulo-1,001 arithmetic so that all behaviors are guaranteed to fall within [0,1000]. Once the mutation has occurred, the behavior repertoire at the next time-step has been completely determined and the probability that a particular class will be emitted is updated accordingly as the proportion of total behaviors in this new repertoire that belong to each class. The process then repeats. Table 2 details each of the parameters in the system and default values, if appropriate.

All model simulations were performed in MATLAB R2012b. The model has a stochastic feature to it so when assessing results it is important to consider outcomes that are averaged over many runs. Due to the dimensionality of the parameter space that was explored, this was a computationally demanding requirement. However, because the runs are not dependent on each other, the computations are "embarrassingly parallel" [30]. To take advantage of this feature, the runs were executed using Matlab's built-in parallel `parfor` loops. A batch script was created that allowed multiple, parallelized simulations to be run simultaneously on different nodes within a cluster.

Study	Finding
Ref. [25]	Consistency with the law of effect.
Ref. [31]	Consistency with the power-function matching equation.
Ref. [32]	Consistency with an extension of the power-function matching equation that considers reinforcement magnitude.
Ref. [33]	Demonstrated the effect of changeover delays when switching between reinforcement schedules.
Ref. [34]	Changing response preference based on concurrent reinforcement schedules.
Ref. [35]	Consistency with known inter-response time distributions.

Table 1

Summary of previous findings from computational model.

Variable	Symbol	Value
Computational Model		
Range of Behaviors	-	[0,1000]
Target Behavior Class	-	[495-505]
Number of Behaviors in Repertoire	-	100
Fitness Function Mean (Reinforcement Strength)	μ	5
Cloning Std. Dev.	-	2
Proportion Mutated	-	0.01
Mutation Std. Dev.	-	2.5

Table 2

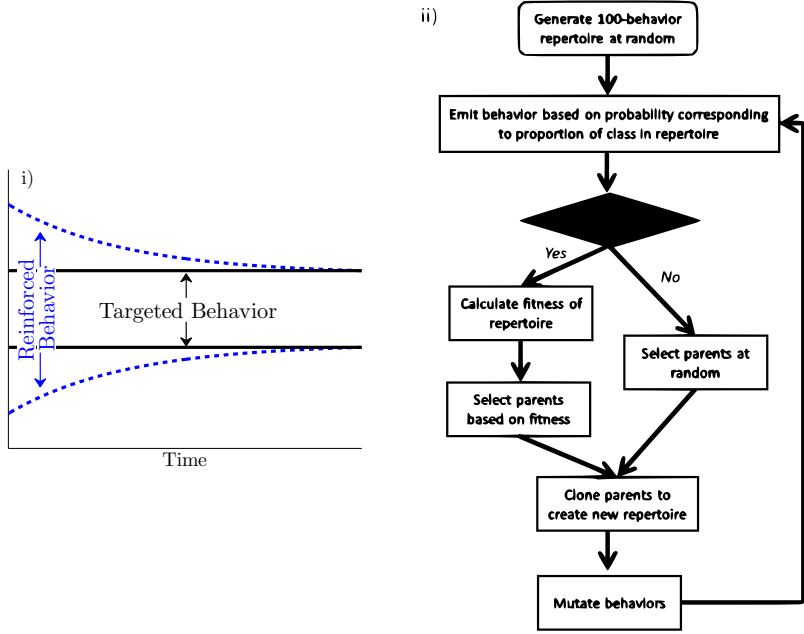*Variable definitions and default values used for all simulations, unless otherwise noted.*

Figure 1. Panel i): Behavior shaping schematic. Panel ii): McDowell computational model flowchart.
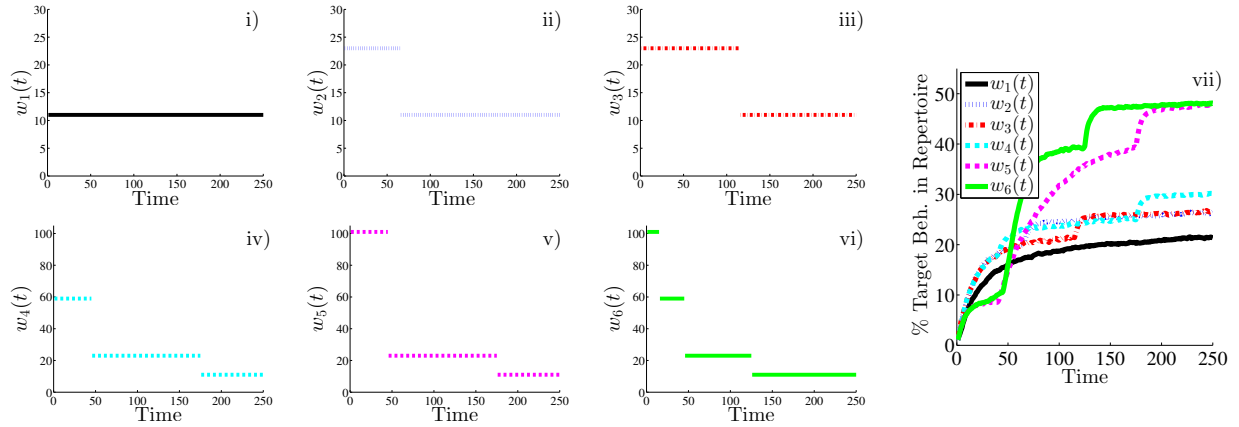
Figure 2. Panels i)-vi) illustrate step $w(t)$ functions for discrete shaping and panel vii) illustrates and analytic results for each function.
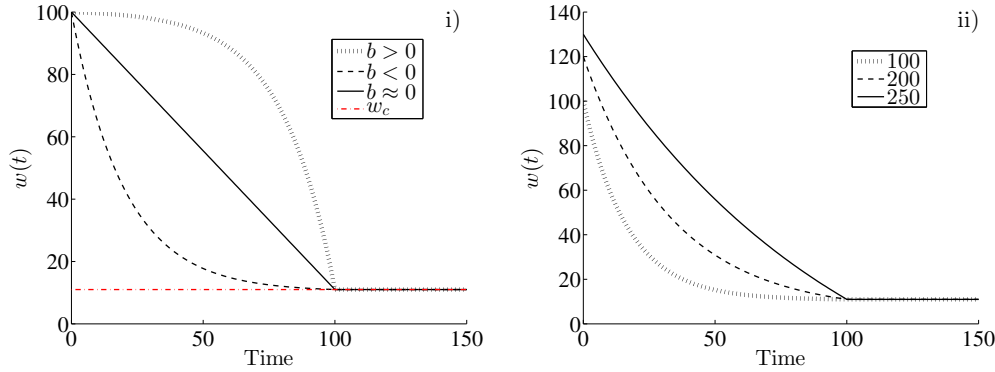

Figure 3. Panel i): Qualitative shapes for the shaping function described by Eq. (1) for different values of b. w_c, depicted as the horizontal line represents the width of the target class. Panel ii): The $w(t)$ functions with optimal parameters at time $t = 100$, 200 and 250 for the FR1 simulations outlined in the text. The largest area under the curve of the target behavior trajectories (see next section) was used to determine which parameters were optimal.
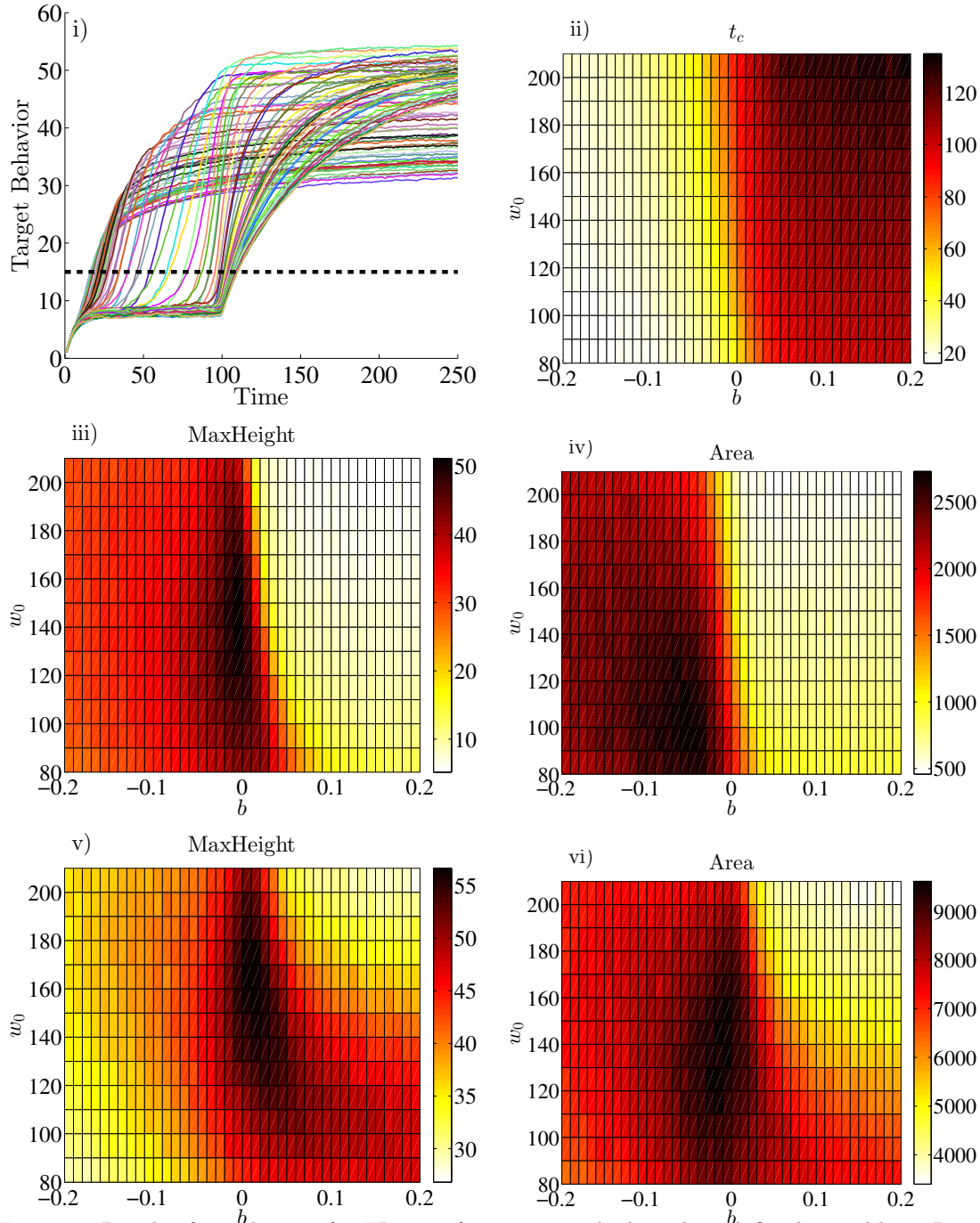
Figure 4. Result of simulations for FR1 reinforcement with the values defined in Table 2. Panel i)
illustrates the level of target behavior with respect to time and panel ii) illustrates t_c, the time required for
the targeted reinforcement to reach a level of 15, for all $(b, w_0) \in B \times W$, where b is the concavity and w_0
is the initial value of the shaping function $w(t)$. Panels iii) and iv) illustrate the maximum height (h_M)
and the area under the curve (AUC) for all combinations of parameters calculated at $t = 100$. Panels v)
and vi) are a recalculation of the results in Panels iii) and iv), but with the metrics calculated at $t = 250$
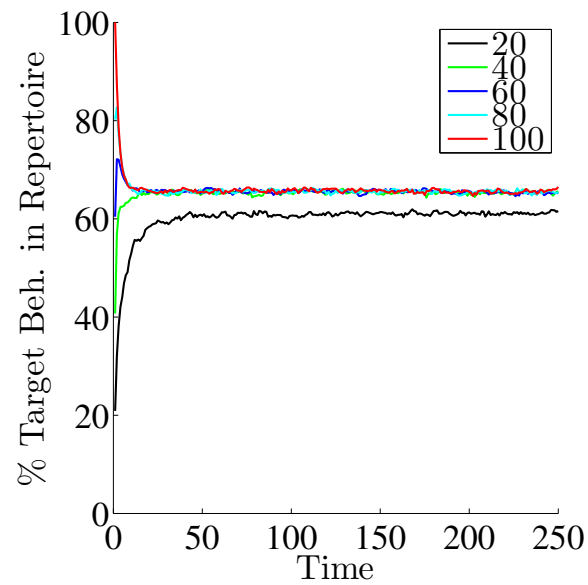rather than $t = t_f$. All results are averaged over 5,000 simulations.

Figure 5. Target behavior trajectories for simulations with parameters in Table 2 with no shaping routine. Each line represents fixing a certain proportion of the initial repertoire with behaviors from the target class, rather than choosing the initial repertoire randomly as is the case with other simulations.
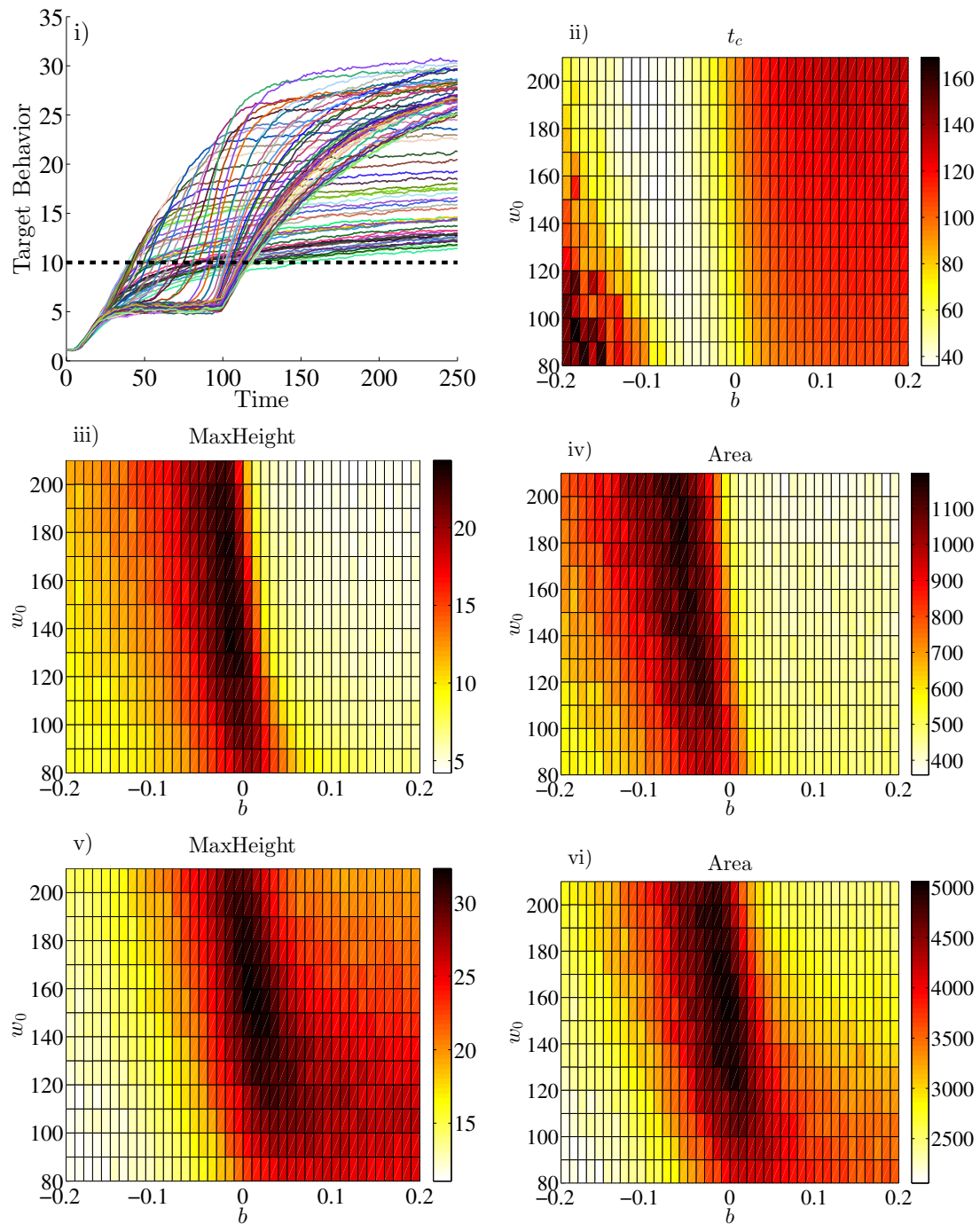
Figure 6. Result of simulations with the default values defined in Table 2 but with an FR3 reinforcement schedule. The metrics illustrated in each panel are the same as in Fig. 4 with the exception that panel ii) illustrates the time required for the targeted reinforcement to reach a level of 10 rather than 15.