Training a Sluggish Agent*

Kfir Eliaz and Ran Spiegler†

September 13, 2021

Abstract

Many biological and organizational systems need to build and maintain preparedness for external challenges. Accumulation and deterioration of such capabilities tend to be gradual. How should we design training plans for such systems to enhance their long-run preparedness? We present a simple economic model of optimal training plans for a rational, slowly adjusting agent. A "trainer" commits to a finite-state Markov process that governs the evolution of training intensity. At every subsequent period, the agent adjusts its "capability", which can only change by one unit at a time. The trainer maximizes long-run capability, subject to an upper bound on average training intensity. We consider two models of the agent's adjustment: myopic/mechanistic and forward-looking. We characterize the optimal training plan in both cases and show how stochastic, time-varying training intensity dramatically increases the agent's long-run capability. The optimal plan resembles "periodization" techniques familiar from exercise physiology.

*This is a substantial revision of a paper formerly titled "Anabolic Persuasion". We acknowledge financial support from the Foerder Institute. We thank Martin Cripps, Michael Crystal, Israel Halperin, Nathan Hancart, Ron Peled, Ariel Rubinstein, Mickey Scheinowitz and seminar audiences for helpful comments.

†Eliaz: Tel-Aviv University and University of Utah. Spiegler: Tel-Aviv University and UCL.

1 Introduction

Many real-life systems are required to develop and maintain capabilities to deal with external challenges. We want our bodies to build muscle mass in order to perform physical tasks, and our brains to preserve cognitive functions as we age. We expect organizations like the military or emergency-management agencies to have the expediency to respond to unexpected crises. To attain these objectives, we subject these systems to various kinds of training regimens. However, an effective training program needs to take into account a key feature of these systems, namely their *sluggishness* - i.e., their capabilities typically change *incrementally* over time. Even if the external stimuli that are generated by the training program change dramatically from one period to the next, capability adjusts slowly. For example, while intensity of physical exercise can fluctuate wildly between periods, the body cannot raise or lower its muscle mass instantaneously to any level. Likewise, if a well-prepared military unit faces no challenges for a while, it will take time for the unit's skills to deteriorate.

This paper explores the problem of designing training programs for slowly adapting systems from the perspective of economic theory. We construct a simple model involving two parties, an "agent" - which represents an organizational or physiological system, such as an army unit, a brain function or a muscle - and a "trainer" of this system. The agent adjusts its capability at every time period, in response to varying training intensity. We represent capability by an integer. Incremental or sluggish adjustment means that the agent can change its capability (in either direction) only by one unit at any given period. The agent's adjustment process balances two opposing forces. On the one hand, maintaining capability is costly (for example, a bigger muscle requires more energy to sustain). On the other hand, when capability falls short of the challenge that a "training session" poses, the system records this failure as a cost (for example, muscle tear or inflammation). The former cost exerts a downward force on capability, while the latter exerts an

upward force. The agent trades off these costs when choosing how to adjust its capability.

We assume that the trainer commits ex-ante to a Markov process that governs the evolution of training intensity over time. The trainer is constrained by an upper bound on the average training intensity that the Markov process induces (for example, a military organization has an allotted amount of time for drills). The trainer's objective is to maximize the agent's long-run capability - defined as the *lowest* value it gets under the stationary distribution induced by the two parties' behavior - taking into account the agent's adjustment process.

Clearly, our model is highly stylized and abstracts from specifics of the biological and organizational systems that motivate this study. Nevertheless, we believe this approach has some value. In the tradition of economic theory, it conceptualizes the problem of designing a training program for a slowly adjusting system as a constrained optimization problem that takes into account the system's optimal response. Hopefully, our simple, abstract model can generate qualitative insights that are relevant to various systems, independently of the details of their adjustment mechanisms.

To complete the economic model, we must specify the agent's planning horizon, as well as whether (and how) it forms expectations regarding future stimuli. We consider two extreme cases. The first case involves a *myopic* agent that balances the two cost components only for the current period. This effectively means that the agent's adjustment process is *mechanistic*: when capability is below (above) the required level for meeting the current challenge, it goes up (down). That is, capability always changes incrementally in the direction of current intensity. The second case involves a *forward-looking* agent that minimizes the long-run average cost, taking into account the trainer's Markov process (which the agent monitors). Unlike the first case, here the agent's behavior is not mechanistic: it involves dynamic optimization that takes into account the agent's knowledge of the stochastic

evolution of future challenges and the constraints on its own ability to adjust. These two cases fit different systems. For example, for muscle adjustment, the mechanistic case seems more appropriate; while the forward-looking version seems better suited for organizational adjustment.

Despite the differences between the two cases (which require different proof methods), we find that the trainer's optimal strategy is similar in both. The Markov process has two states: a rest state with zero training intensity and a high-intensity state. Some transitions between the two states are stochastic. For instance, in the myopic case (and for some parameter values in the forward-looking case), a high-intensity period is followed by another one with positive probability. However, the role of stochastic transitions is very different in the two cases. In the myopic/mechanistic specification of the model, it ensures that the agent's long-run capability is not sensitive to initial conditions. In the forward-looking specification, it manages the agent's dynamic incentives. Randomization keeps the agent "on its toes", deterring it from "slacking off" and losing capability during rest periods.

In both cases, the trainer's optimal plan sustains a long-run capability that is considerably higher than what the trainer could achieve with a constant training intensity - or, equivalently in our model, what the trainer could achieve if the agent's adjustment were fully flexible. In the myopic case, long-run capability is nearly *twice* as large. In the forward-looking case, the factor of increase can be *arbitrarily large* when the agent's "maintenance cost" parameter is small. Thus, our main theoretical insight is that in the presence of sluggish adjustment, *stochastic training that involves rest and high-intensity phases enhances long-run capability.*

Our optimal training plan may be viewed as a stochastic variant on "periodization" training techniques familiar from exercise physiology. Since it first began in the 1960s, this methodology has gained popularity and is currently the dominant technique used by professional athletes. Numerous studies have documented the success of periodization in terms of increased muscle

mass and athletic performance (Bompa and Buzzichelli (2018)).[1] While the physiological literature offers biological explanations for the superiority of a cyclical training design (e.g., see Issurin (2019)), our results provide a complementary perspective, by deriving the effectiveness of stochastic periodization as a logical conclusion of the assumption of sluggish adaptation (resulting from rational cost-benefit calculus) to random physical stimuli.

To our knowledge, this perspective is new: we are not aware of life-science studies that examined the hypothesis that building and maintaining long-run physiological or neurological capabilities involves forward-looking mechanisms. This theoretical conclusion does not require knowledge of specific details of the adjustment mechanism of the system in question (although it does make use of a number of simplifying assumptions). Therefore, we suggest that it might be of some relevance for a variety of biological and organizational systems that exhibit sluggish adjustment.

2 The Model

We consider a principal-agent model in which the principal is referred to as a "trainer" (she). We interpret the agent (it) as a biological or organizational system that is trained to attain and maintain high capability. The trainer commits ex-ante to a pair (P, f), where P is a discrete-time, finite-state Markov process over some finite set of states S, and $f : S \to \mathbb{N}_+$ is an output function that assigns a challenge level to every state $s \in S$. The set of states S is *endogenous*: the trainer can choose a set of any finite size. We denote by s_t and d_t the state and challenge level at period t. When there is no risk of confusion, we replace the notation $f(s)$ with $d(s)$.

We impose the following constraints on (P, f). First, P is irreducible. This ensures that it has a unique invariant distribution λ_P, and therefore

[1] See also Issurin (2010), Kiely (2012) and Kiely et al. (2019).

enables us to talk about long-run average quantities unambiguously. Second,

$$\sum_{s \in S} \lambda_P(s) f(s) \leq \mu + \varepsilon \tag{1}$$

where $\mu \geq 1$ is an integer and $\varepsilon > 0$ is arbitrarily small. That is, the long-run average challenge level cannot exceed μ by more than a negligible amount (the approximate formulation of the constraint is due to μ getting integer values).

After the trainer chooses (P, f) at period 0, the agent chooses a non-negative integer $m_t \in \{m_{t-1} - 1, m_{t-1}, m_{t-1} + 1\}$ at every $t = 1, 2, 3, ...$ The agent's choice at period t takes place *after* the realization of s_t. Henceforth, we refer to m_t as the agent's *capability* at time t. Let $m_0 \in \mathbb{N}_+$ be the agent's initial capability. The restricted choice set for m_t reflects sluggish adaptation.

Define

$$C_t = cm_t + \max(0, d_t - m_t)$$

This is the total cost that the agent incurs at period t. It consists of two terms. First, cm_t is the "maintenance cost" of the capability level. The interpretation is that higher capability carries larger maintenance costs. Second, the gap between m_t and d_t (when the latter is higher) represents a performance shortfall because the agent's capability is lower than the challenge it faces. For illustration, when m_t represents muscle mass, cm_t captures the caloric cost of maintaining it, whereas $d_t - m_t$ may represent physical damage due to excessive stress that occurs when the agent trains at an intensity that exceeds its capability.

Thus, the agent faces a trade-off whenever its current capability is not enough to meet the training intensity: increasing capability requires higher maintenance costs but lowers the cost of training beyond its capability. Our piece-wise linear cost specification implies that moving up to the next capability rung reduces net costs by $1 - c$ in the current period, regardless of the

agent's current capability (as long as it is below current training intensity). Of course, the agent may still have to take into account that increasing m today will delay its ability to lower it in response to future rest periods.

We consider two alternative specifications of the agent's intertemporal aggregation.

Myopic/mechanistic adjustment. At every period $t \geq 1$, the agent chooses m_t to minimize C_t. That is, the agent is myopic: it does not take into account future costs. Because $c \in (0, 1)$, this immediately implies the following strategy for the agent:

$$m_t = m_{t-1} + sign(d_t - m_{t-1}) \tag{2}$$

That is, capability always moves in the direction of the current challenge level. This adjustment rule can be interpreted mechanistically: it does not require the agent to know the trainer's Markov process or to monitor the evolution of its state.

Forward-looking adjustment. The agent knows the trainer's choice of (P, f). At every period t, it observes the realized state s_t before choosing m_t. The agent's objective is to minimize

$$\lim_{T\to\infty} \inf \frac{1}{T} \sum_{t=1}^{T} C_t \tag{3}$$

This is the long-run average cost that the agent incurs. The lim inf criterion reflects the assumption that the agent is not only forward-looking but also arbitrarily patient.

The mechanistic-adjustment model is probably a better fit for many physiological systems, such as muscles. The forward-looking-adjustment model is perhaps better suited for organizational systems, whose members form expectations about future challenge levels and care about their long-run cost flows.

Note that the two models could be rewritten essentially as two special cases of a single model, in which the agent minimizes the expected discounted sum of C_t, with a discount factor δ. The mechanistic-adjustment model is equivalent to the case of $\delta = 0$, whereas the forward-looking-adjustment model effectively corresponds to the case of $\delta \to 1$. Nevertheless, we treat the two models separately because they have different interpretations and their analysis is quite different.

Under both models of the agent's behavior, the agent faces an extended Markov problem, in which the state at period t is (m_{t-1}, s_t). Therefore, the agent has an optimal response that is also Markovian with respect to this extended state space. In the myopic case, this strategy is explicitly given by (2). We assume that the agent plays a Markovian best-response in the forward-looking case as well. This ensures that the extended Markov process induced by the two parties' strategies has a unique invariant distribution over (d_t, m_t), such that $m^* = \lim_{t\to\infty} \min m_t$ - namely, the lowest value that m takes beyond a sufficiently large t, also referred to as *the lowest long-run capability* - is well-defined, independently of the initial condition m_0.

The trainer's objective is to maximize m^* subject to the feasibility constraint (1). A higher m^* means that the system has greater "preparedness" - i.e., it can *consistently* meet bigger *actual* challenges - i.e. challenges that arise naturally and by surprise, independently of the training regime. Being able to meet such challenges is the whole purpose of the training program.

Note that the trainer's Markov process P does not condition on the agent's history of capability realizations. This restriction means that the training program can be implemented even when the trainer cannot monitor the agent's capability (or its performance relative to the challenge it faces) along the play path. In particular, it can neglect the initial condition m_0. We believe that all our results go through if the trainer can condition d_t on m_{t-1}.

Comment on the model's interpretation

Under the biological interpretation of our model, the trainer engages in physical or cognitive training of a biological system such as a muscle or a neurological function (not the person they belong to). The variable d represents a physical or cognitive challenge, and the system adjusts its capability m in a way that trades off the energy cost of maintaining capability against a cost of failing to meet the challenge. We can think of two interpretations of the latter cost component. First, as mentioned above, when $d_t > m_t$, the system experiences stress that can lead to injury or inflammation, hence the cost of over-stress is physical. This seems to fit the physical-training interpretation. Second, the system may record failure to meet a challenge as a negative outcome even when no physical damage is involved. This may fit the cognitive-training interpretation.

Under the organizational interpretation, it is better to think of the challenge d_t in terms of auditing or inspection. High d_t corresponds to an inspection that tests high capability. The challenge does not directly build the organization's capability, but rather *tests* it. The organizational system adapts its capability to the testing regime because of underlying incentives, which are captured by the cost function C. When the organization fails a test, its members experience a loss, which can take various forms: wage cuts, lower prospects of promotion, etc. Unlike most of the economic literature on organization design (e.g. Bolton and Dewatripont (2005)), we take these incentives as given and focus on the problem of designing a dynamic, stochastic inspection regime in the presence of sluggish adaptation.

Comment on the cost function

The key assumption embodied by the cost function C_t is that if $d_t \geq m_{t-1}$, the maintenance cost saved when the agent lowers m by one unit is more than offset by the increase in the "performance gap" cost. In the case of a myopic agent, this feature leads to the mechanistic adjustment rule (2) - namely, m_t always chases d_t. In the case of a forward-looking agent, our analysis in

Section 4 will also make use of the piece-wise linearity of the performance-gap cost component. We conjecture that our results will remain intact if we replace the term $\max(0, d_t - m_t)$ by $g(\max(0, d_t - m_t))$, where g is an increasing, convex function satisfying $g(0) = 0$ and $g(1) > c$.

Benchmark: Completely flexible adjustment

Suppose the agent could choose *any* $m_t \in \mathbb{N}_+$ at every period, regardless of m_{t-1}. In particular, it could always choose m_t to minimize C_t. In this case, it would choose $m_t = d_t$ at every t. Under this flexible-adjustment rule, the long-run average of m_t coincides with the long-run average of d_t, which by assumption cannot exceed μ (more than negligibly). Therefore, the best the trainer can do is play a constant strategy $d_t = \mu$ at every period, such that the flexible agent's mass will be μ as well. This deterministic process attains the same long-run capability of μ also when the agent is sluggish (because the agent will eventually reach this capability level and stay there indefinitely). The question is whether the trainer can outperform this benchmark with a non-degenerate Markov process.

3 Myopic/Mechanistic Adjustment

In this section we analyze the trainer's problem when the agent behaves according to the myopic/mechanistic adjustment model.

Proposition 1 *Assume the agent follows the strategy given by (2). Then:*

(i) For any trainer strategy, the lowest long-run capability is at most $2\mu - 1$.

(ii) This upper bound can be implemented by the following (P, f)*. The Markov process* P *has two states,* H *and* L*, and a transition matrix given by*

$\Pr(s_t \to s_{t+1})$	L	H
L	0	1
H	β	$1 - \beta$

where β is arbitrarily close to 1. The output function is $f(H) = 2\mu$ and $f(L) = 0$. In the $\beta \to 1$ limit, the invariant capability distribution assigns probability $\frac{1}{2}$ to $m = 2\mu$ and $m = 2\mu - 1$.

Thus, a slightly perturbed cyclic training program can dramatically increase the long-run capability of a sluggish agent, relative to the flexible-adjustment benchmark. When μ is large - corresponding to a very sluggish agent, given that we normalized the adjustment increment to 1 - the increase is by a factor of nearly 2.

The training regime approximately consists of alternating periods of high intensity ($d = 2\mu$) and rest ($d = 0$). After a period of high-intensity training, there is a small chance $1 - \beta$ that the high-intensity episode will be repeated. This stochastic perturbation ensures that the set of capability values $\{2\mu, 2\mu - 1\}$ is absorbing: the agent will reach it in finite time with probability one, regardless of m_0. The role of randomness in the trainer's strategy is thus to ensure that the agent's long-run behavior is not sensitive to the initial condition, which the trainer cannot monitor.

The intuition for the result is that changes in m depend only on the *sign* of $d - m$, whereas the trainer's "budget constraint" is expressed in terms of the *average* of d. The contrast between the ordinal adjustment rule and the cardinal constraint is the key to our result. The most economical way to get the agent's capability to go up at period t is to set $d_t = m_{t-1} + 1$; and the most economical way to bring it down is to set $d_t = 0$. In the long run, since the agent's capability moves around in increments of one unit, m goes up and down with equal frequencies. This explains the approximate factor 2 by which the trainer can increase long-run capability, relative to the flexible-benchmark μ.

Proof of part (i) of Proposition 1

Consider an arbitrary strategy for the trainer. Let $(m_{t-1}, d_t)_{t=1,2,...}$ be a possible sample path that results from the extended process. The long-run

frequency of every (m, d) in the sample path, denoted $\lambda(m, d)$, coincides with the probability of this pair according to the invariant distribution induced by the two parties' strategies. Let X be the set of recurrent pairs (m, d) in the sample path. Partition X into three classes:

$$\begin{aligned} X^+ &= \{(m,d) \in X \mid d > m\} \\ X^- &= \{(m,d) \in X \mid d < m\} \\ X^0 &= \{(m,d) \in X \mid d = m\} \end{aligned}$$

The proof now proceeds by a series of steps. Recall that we use the notation $d(s)$ as a substitute for $f(s)$.

Step 1: *λ satisfies*

$$\sum_{(m,d)\in X^+} \lambda(m,d)(m+1) = \sum_{(m,d)\in X^-} \lambda(m,d)m \tag{4}$$

Consider some period t along the sample path such that $(m_t, d_{t+1}) \in X^+$. By definition, this pair is recurrent. Therefore, m_t must be visited again in some later period. Let $t'+1$ be the *earliest* such period (while $m_{t'+1} = m_t$, we do not require $d_{t'+2} = d_{t+1}$). Since m moves only in one-unit increments, it must be the case that $(m_{t'}, d_{t'+1}) \in X^-$ and $m_{t'} = m_t + 1$. We have defined a *one-to-one* mapping from periods t for which $(m_t, d_{t+1}) \in X^+$ to periods t' for which $(m_{t'}, d_{t'+1}) \in X^-$, such that $m_{t'} = m_t + 1$. It follows that

$$\lim_{T\to\infty} \frac{\sum_{t=1}^{T} \mathbf{1}[(m_t, d_{t+1}) \in X^+] \cdot (m_t + 1)}{T} = \lim_{T\to\infty} \frac{\sum_{t=1}^{T} \mathbf{1}[(m_t, d_{t+1}) \in X^-] \cdot m_t}{T}$$

which can be rewritten as (4). □

Step 2: *The average long-run m is at most 2μ (approximately)*

The long-run average of m induced by the trainer's strategy can be written as

$$\mathbb{E}(m) = \sum_{(m,d)\in X^+} \lambda(m,d)m + \sum_{(m,d)\in X^-} \lambda(m,d)m + \sum_{(m,d)\in X^0} \lambda(m,d)m \qquad (5)$$

By the feasibility constraint,

$$\sum_{(m,d)\in X^+} \lambda(m,d)d + \sum_{(m,d)\in X^-} \lambda(m,d)d + \sum_{(m,d)\in X^0} \lambda(m,d)d \lessapprox \mu$$

By definition, $d \geq m+1$ for every $(m,d) \in X^+$, $d \geq 0$ for every $(m,d) \in X^-$, and $d = m$ for every $(m,d) \in X^0$. Therefore,

$$\sum_{(m,d)\in X^+} \lambda(m,d)(m+1) + \sum_{(m,d)\in X^-} \lambda(m,d)\cdot 0 + \sum_{(m,d)\in X^0} \lambda(m,d)m \lessapprox \mu$$

This means that

$$\sum_{(m,d)\in X^+} \lambda(m,d)m \leq \sum_{(m,d)\in X^+} \lambda(m,d)(m+1) \lessapprox \mu - \sum_{(m,d)\in X^0} \lambda(m,d)m$$

By (4), it follows that

$$\sum_{(m,d)\in X^-} \lambda(m,d)m \lessapprox \mu - \sum_{(m,d)\in X^0} \lambda(m,d)m$$

as well. Plugging the last two inequalities in (5), we obtain

$$E(m) \lessapprox 2\mu - \sum_{(m,d)\in X^0} \lambda(m,d)m \leq 2\mu$$

□

Step 3: *The minimal long-run m is at most $2\mu - 1$*

Suppose the long-run distribution over d is degenerate at some d^*. Therefore, $d^* \lessapprox \mu$. The agent's myopic best-reply implies that eventually, its capability coincides with d^*. It follows that to reach a minimal long-run capability above μ, the long-run distribution over d must assign positive probability to at least two values. This means there are infinitely many periods t in which $d_t \neq m_{t-1}$. By myopic best-replying, this precludes the possibility that the long-run distribution over m is degenerate. Since the long-run average of m cannot exceed 2μ by more than an infinitesimal amount, there must be infinitely many periods t in which $m_t \leq 2\mu - 1$. This completes the proof of part (i). □

Proof of part (ii) of Proposition 1

Consider the trainer's strategy described in part (ii) of the statement of the result. As long as $\beta \in (0,1)$, the Markov process over m that is induced by the strategy and the agent's best-reply (given by Step 1) has a unique invariant distribution, with $m = 2\mu$ and $m = 2\mu - 1$ being the only recurrent capability values. The reason is that if $m_t > 2\mu$, $m_{t+1} = m_t - 1$ with certainty; if $m_t < 2\mu - 1$, there is a positive probability that there will be a streak of realizations $d = 2\mu$ such that m will keep adjusting upward until it reaches $m = 2\mu$; and finally, if $d_t = 0$ then $d_{t+1} = 2\mu$ for sure, which means that once m hits 2μ and later goes down to $2\mu - 1$, it will return to 2μ immediately in the next period. As the exogenous upper bound on average intensity gets arbitrarily close to μ, β can be made arbitrarily close to one. In the $\beta \to 1$ limit, the invariant distribution over m assigns probability $\frac{1}{2}$ to each of the values $m = 2\mu$ and $m = 2\mu - 1$. ■

4 Forward-Looking Adjustment

In this section we characterize the solution to the trainer's problem when the agent is forward-looking. For expositional convenience, we assume μ/c is an integer.

Proposition 2 *Assume the agent evaluates cost streams by (3). Then:*

(i) The lowest long-run capability is at most $\mu/c - 1$.

(ii) This upper bound can be implemented by the following (P, f). *The Markov process* P *has two states,* H *and* L, *and a transition matrix given by*

$$\begin{array}{ccc} \Pr(s_t \to s_{t+1}) & L & H \\ L & 1-\alpha & \alpha \\ H & \beta & 1-\beta \end{array}$$

where $\alpha = 1$ *if* $c \geq \frac{1}{2}$, $\beta = 1$ *if* $c < \frac{1}{2}$, *and* $\alpha/(\alpha+\beta)$ *is arbitrarily close to* c *from above. The output function is* $f(H) = \mu/c$ *and* $f(L) = 0$. *In the* $\alpha/(\alpha+\beta) \to c$ *limit, the invariant capability distribution assigns probability* c *to* $m = \mu/c$ *and probability* $1-c$ *to* $m = \mu/c - 1$.

When $c < \frac{1}{2}$, the upper bound on the agent's lowest long-run capability is higher than in the myopic case. Moreover, it gets arbitrarily high when $c \to 0$. As c gets closer to one, the highest minimal long-run capability approaches the flexible-agent benchmark μ.[2]

The Markov process that attains the upper bound is similar to the one in Section 3, but the reasoning behind the result is different. Because the mechanistic agent of Section 3 responds only to the *current* realization of d, the role of randomization in that case is only to ensure a unique invariant capability distribution. In contrast, a forward-looking patient agent responds to trainer's *entire* continuation strategy. Randomization serves as an incentive to keep the agent "on its toes" and deter it from lowering its level of preparedness during periods of rest. In particular, when $c < \frac{1}{2}$, a rest period is followed by another one with probability approximately equal to $(1-2c)/(1-c)$. Thus, when $c < \frac{1}{2}$, the trainer's optimal program allows for a streak of $d = 0$ realizations. When this happens, the agent does not

[2]Because μ/c is an integer, we rule out the possibility that c is *arbitrarily* close to one. In that case, the trainer cannot outperform the flexible-agent benchmark of μ.

lower its capability below $\mu/c - 1$ because it takes into account the future loss $d - m$ in the event that d switches to $d = \mu/c$. The trainer designs the transition probabilities such that the agent's intertemporal trade-offs lead it to be nearly indifferent between lowering its capability and remaining at $m = \mu/c - 1$. In contrast, the mechanistic agent cannot be made indifferent when faced with a streak of $d = 0$ realizations: it repeatedly lowers its capability. This difference enables the trainer to achieve a higher minimal long-run capability when the agent is forward-looking, as long as $c < \frac{1}{2}$.

To further elucidate why randomization is necessary, consider the following example, which shows that the minimal long-run capability attained by the optimal stochastic strategy *cannot* be sustained by a particular deterministic strategy with the same long-run distribution over d. Suppose $\mu = 4$ while c is slightly below $\frac{4}{11}$. Then, the optimal training strategy of Proposition 2 induces an invariant distribution that assigns probability $\frac{4}{11}$ to $d = 11$ and probability $\frac{7}{11}$ to $d = 0$. This strategy sustains a minimal long-run capability level of $m = 10$.

Now consider a deterministic strategy that induces the same long-run frequencies of d. The strategy follows an 11-period cycle consisting of four consecutive periods of $d = 11$ and seven consecutive periods of $d = 0$. If the agent plays $m = 11$ when $d = 11$ and $m = 10$ when $d = 0$ - as it does against the strategy presented in Proposition 2 - the minimal long-run capability is $m = 10$. Moreover, this strategy is optimal for the agent among all strategies that induce this minimal long-run capability. However, given the predictable evolution of d under the cyclic deterministic strategy, a forward-looking agent can do better. Suppose that it plays the following sequence of m against the cyclic sequence of d:

d	11	11	11	11	0	0	0	0	0	0	0
m	11	11	11	10	9	8	7	7	8	9	10

Compared with the benchmark strategy of playing $m = 11$ (10) against $d = 11$ (0), the agent saves approximately

$$c \cdot (1 + 1 + 2 + 3 + 3 + 2 + 1) - 1 \approx \frac{41}{11}$$

per cycle. It follows that the agent's best-reply to the cyclic deterministic strategy leads to a minimal long-run capability below $m = 10$.

This example highlights a key role of the stochasticity of the trainer's optimal strategy. The fact that there is always a chance that the agent will face a big challenge following a rest period incentivizes the agent not to lower its capability. In contrast, the predictable nature of the cyclic deterministic strategy allows the agent to gradually lower its capability and gain it back by the time the big challenge arrives. In particular, it is profitable for the agent to lower its capability already in the final period of the high-intensity phase of the cycle, even though this involves a costly performance gap during that period, because this is more than offset by the cumulative maintenance-cost saving over the cycle.

Recall that in our discussion of the organizational interpretation of the model (see Section 2), we drew a connection between our notion of a challenge and an audit or an inspection. The idea that optimal inspection may involve random audits is very familiar in game theory and economics: when auditing is costly, making it unpredictable deters the agent from shirking (see Varas et al. (2020) and Solan and Zhao (2021) for recent examples, and the references therein). Likewise in our model, the "budget constraint" (1) represents limited resources for auditing. From this perspective, the novelty of our model is the idea that successive periods of shirking can magnify the agent's failure at an audit, and that sustained effort may be required to rebuild the ability to pass.

As to the proof of Proposition 2, in part (i) we actually prove a somewhat stronger result: to attain a strictly positive minimal long-run capability, the *average* long-run capability cannot exceed $\mu/c - 1 + c$. The Markov process

we construct in part (ii) approximates this upper bound. This means that among all trainer strategies that attain the minimal long-run capability of $\mu/c - 1$, this process cannot be outperformed in terms of average capability.

Before we give the complete proof, we provide a brief outline of its structure. The proof of part (i) proceeds in several steps. First, note that by playing a constant $d = \mu$, the trainer can attain a long-run capability of μ. Therefore, the trainer can attain a minimal long-run capability that is at least as large as μ. Hence, the invariant distribution over (m, d) - induced by an optimal trainer strategy and an agent's best-reply - satisfies $\Pr(m > 0) = 1$. Second, we establish a lower bound on the long-run frequency of positive training intensity: under the invariant distribution induced by the two parties' strategies, $\Pr(d > 0) \geq c$. To prove this, we consider the following possible deviation by the agent: pick a history in which m is at its lowest long-run value (which is positive, as we saw); move one notch below the original plan; afterwards, proceed as if the deviation never took place. The piece-wise linearity of the cost function enables a simple calculation of the net long-run profit from this deviation. This deviation saves c per period, but raises the "performance gap" cost by one unit whenever $d \geq m$ under the original strategy. When the agent is forward-looking, this deviation is unprofitable only if $\Pr(d \geq m) \geq c$. Since $\Pr(m > 0) = 1$, we have that $\Pr(d > 0) \geq \Pr(d \geq m) \geq c$. The third and final step of the proof shows that the long-run average capability cannot exceed $\mu/c - 1 + c$. If this were not true, then the average long-run cost would exceed $\mu - c(1 - c)$. But then, using the previous step, we obtain that the following deviation is profitable for the agent: descend all the way to $m = 0$ and play $m = \mathbf{1}(d > 0)$ thereafter. The upper bound on the lowest long-run capability then immediately follows.

The proof of part (ii) begins by noting that the agent has a best-reply to the trainer's strategy that induces two (and therefore adjacent) long-run values of m (this is a consequence of the fact that P has two states and

$\alpha\beta = 0$). We then show that by the piecewise linearity of the agent's cost function and the condition on α, β, c, the two long-run capability values are μ/c and $\mu/c-1$. The induced long-run average capability is then $\mu/c-1+c$.

Proof of part (i) of Proposition 2

Let p be the unique invariant distribution over (d_t, m_t) that results from the trainer's strategy and the agent's best-replying strategy. (Note the different time subscripts of d and m, compared with the proof of Proposition 1 in Section 3; our different notation highlights this difference.) We abuse notation and write $p(d)$, $p(m)$ and $p(d \mid m)$ to represent marginal and conditional distributions induced by p. As in the myopic-agent case, we first derive an upper bound on the expected capability according to p, which we use to derive the upper bound on the minimal long-run capability. Then, we show how to implement this upper bound.

In Section 2, we saw that the trainer can implement a minimal long-run capability of at least μ (by playing $d = \mu$ at every period). Therefore, we take it for granted that the minimal value of m in the support of p is at least $\mu \geq 1$.

Step 1: $p(d > 0) \geq c$

Consider the following deviation by the agent. Pick some period-t history for which $m_{t-1} \geq 1$ is at the lowest value according to p. Therefore, $m_t = m \in \{m_{t-1}, m_{t-1} + 1\}$. At this history, the agent deviates to $m'_t = m - 1$. Subsequently, the agent behaves according to its original strategy *as if the deviation did not occur.*

This deviating strategy induces an invariant distribution p' such that for every (d, m) in the support of p, $p'(d, m-1) = p(d, m)$. Therefore, the deviation saves c at every period, but raises costs by one unit per period whenever $d \geq m$ under the original strategy. In order for this deviation to be unprofitable for an arbitrarily patient agent, it must be the case that $p(d \geq m) \geq c$. Since $m > 0$ with probability one, $p(d > 0) \geq p(d \geq m)$, hence $p(d > 0) \geq c$. $\square$

Step 2: *The expectation of m according to p is at most $\mu/c - 1 + c$*
Assume the contrary. Then, the agent's average long-run cost exceeds

$$c \cdot [\frac{\mu}{c} - 1 + c] = \mu - c(1 - c)$$

Now consider a deviation to the following strategy. Descend from m_0 to $m = 0$, and then implement the following rule: $m_t = 0$ whenever $d_t = 0$, and $m_t = 1$ whenever $d_t > 0$. When the agent is arbitrarily patient, the average long-run cost from this strategy is approximately

$$\begin{aligned} p(d &= 0) \cdot 0 + p(d > 0) \cdot [c + \sum_{d>0} p(d \mid d > 0)d - 1] \\ &\lessapprox p(d > 0)(c - 1) + \mu \end{aligned}$$

Since $c < 1$, Step 1 implies that

$$p(d > 0)(c - 1) + \mu < \mu - c(1 - c)$$

such that the deviation is profitable, a contradiction. □

Step 3: *The minimal long-run capability is at most $\mu/c - 1$*
Since μ/c is an integer, $\mu/c - 1 + c$ is not an integer. Hence, in order for the average long-run cost to be weakly below $\mu/c - 1 + c$, the minimal long-run capability cannot exceed $\mu/c - 1$.[3] □

Proof of part (ii) of Proposition 2
Consider the strategy described in the statement of part (ii). Our objective is to show that given this strategy, there is a best-reply for the agent such that for every sufficiently high t, $m_t = \mu/c$ whenever $s_t = H$ and $m_t = \mu/c - 1$ whenever $s_t = L$.

[3]The proof of this step utilizes the convenient assumption that μ/c is an integer. An alternative proof that does not rely on this assumption is analogous to Step 3 in the proof of Proposition 1 (i).

Since the agent faces a Markovian decision problem with an extended state space (s, m), there exists a best-reply that is Markovian with respect to this state space. To derive such a best reply, we proceed in four steps.

Step 1: *There is no best-reply in which the invariant distribution assigns probability one to a single m.*

Proof. Assume the contrary. If $m < \mu/c$, then it is profitable for the agent to deviate to a strategy that plays $m+1$ whenever $s = H$ and m whenever $s = L$. Likewise, if $m > 0$, it is profitable for the agent to deviate to a strategy that plays m whenever $s = H$ and $m-1$ whenever $s = L$. □

Step 2: *The set of recurrent values of m (according to the unique invariant distribution induced by the two parties' strategies) is a set of consecutive numbers $\underline{m}, \underline{m}+1, ..., \overline{m}$, where $\overline{m} \leq \mu/c$.*

Proof. The agent's sluggishness implies that if the agent visits two non-adjacent capabilities m and m', then it must also visit every m'' between them. Therefore, if m and m' are recurrent, so is m''. Suppose $\overline{m} > \mu/c$. Then, there is a profitable deviation for the agent that instructs to remain at $\overline{m}-1$ whenever the original strategy instructs to switch to $\overline{m}$. □

Step 3: *There is a best-reply that induces an invariant distribution that assigns positive probability to exactly two values of m.*

Proof. Consider the invariant distribution over (d, m) induced by the trainer's strategy and the agent's best-reply. By Step 1, $\overline{m} - \underline{m} \geq 1$. If $\overline{m} - \underline{m} = 1$, we are done. Therefore, assume $\overline{m} - \underline{m} > 1$. There are two cases to consider.

First, let $\alpha = 1$ (this fits the case of $c \geq 1/2$). This means that whenever $s = L$, the state switches immediately to $s = H$ in the next period. Consider the top two values of m in the invariant distribution, namely $\overline{m}$ and $\overline{m}-1$. By Step 2, $\overline{m} \leq \mu/c$. Moreover, when $s = L$ (at which d attains its lowest value according to the trainer's strategy), the agent strictly prefers $\overline{m}-1$ to $\overline{m}$. Consider some t for which $m_t = \overline{m}$ (there are infinitely such periods because $\overline{m}$ is recurrent). If $s_{t+1} = L$, the agent necessarily switches to $m_{t+1} = \overline{m}-1$. If, on the other hand, $s_{t+1} = H$, we need to consider two possibilities.

- Suppose that when $s_{t+1} = H$, it is not optimal for the agent to play $m_{t+1} = \overline{m}$. That is, the agent switches from $m_t = \overline{m}$ to $m_{t+1} = \overline{m} - 1$ for *any* realization of s_{t+1}. But this also means that if $m_{t'} = \overline{m} - 1$ at some period t' and $s_{t'+1} = H$, it cannot be optimal for the agent to switch to $m_{t'+1} = \overline{m}$. The reason is that by revealed preference, the agent prefers being at $\overline{m} - 1$ to being at $\overline{m}$ when the state is H. And since we already saw that the agent prefers being at $\overline{m} - 1$ to being at $\overline{m}$ when the state is L, this means that the agent will *never* switch from $\overline{m} - 1$ to $\overline{m}$, contradicting the definition of $\overline{m}$ as a recurrent state.

- Suppose that when $s_{t+1} = H$, it is optimal for the agent to play $m_{t+1} = \overline{m}$. This reveals a weak preference for $\overline{m}$ over $\overline{m} - 1$ when the state is H. Therefore, there is a best-reply for the agent that prescribes $m_{t+1} = \overline{m}$ whenever the extended state (s_{t+1}, m_t) is $(H, \overline{m} - 1)$ or $(H, \overline{m})$. We already saw that when the extended state is $(L, \overline{m})$, the agent switches to $\overline{m} - 1$. Since $\alpha = 1$, this means that we have constructed a best-reply for the agent such that once it reaches $\overline{m}$, it will only visit $\overline{m}$ and $\overline{m} - 1$ from that period on, contradicting the assumption that there are additional recurrent values of m.

Thus, we have ruled out the possibility that $\overline{m} - \underline{m} > 1$ when $\alpha = 1$. Now suppose $\beta = 1$ (this fits the case of $c \leq 1/2$). An analogous argument establishes that there is a best-reply for the agent that induces an invariant distribution with only two recurrent capability values, $\underline{m}$ and $\underline{m} + 1$.

It follows that we can restrict attention to strategies of the agent that induce an invariant distribution which assigns positive probability to precisely two consecutive capability values, m and $m - 1$, where $0 < m \leq \mu/c$. $\square$

Step 4: *There is a best-reply for the agent that induces an invariant distribution on the capability values μ/c and $\mu/c - 1$.*

Proof. Given Step 3, it is clearly optimal for the agent to be at m when $s = H$ and at $m - 1$ when $s = L$. In addition, when $m > \mu/c$ ($m < \mu/c - 1$),

the agent clearly wants to move downward (upward).

The invariant distribution of the trainer's two-state Markov process assigns probability $\alpha/(\alpha+\beta)$ to state H and $\beta/(\alpha+\beta)$ to state L. Therefore, since the agent is arbitrarily patient, his long-run expected payoff is approximately

$$-\frac{\alpha}{\alpha+\beta}\cdot(cm+\frac{\mu}{c}-m)-\frac{\beta}{\alpha+\beta}\cdot c(m-1)$$

It is now easy to see that given that $\alpha/(\alpha+\beta)>c$, this expression increases with m, such that the optimal value of m is μ/c. The expected value of m according to this strategy is

$$\frac{\alpha}{\alpha+\beta}\cdot\frac{\mu}{c}+\frac{\beta}{\alpha+\beta}\cdot(\frac{\mu}{c}-1)$$

which is arbitrarily close to the upper bound. ■

5 Comment on the Trainer's Objective

In our model, the trainer's objective is to maximize the agent's minimal long-run capability. Alternatively, we could use the long-run *average* m as a criterion. However, this criterion is less attractive in our context because it does not reflect the idea of "preparedness". In particular, the average criterion allows zero to be a recurrent value for m, which means that the agent will sometimes be completely unprepared for any positive challenge.

A by-product of our analysis in Section 3 is that in the myopic case, 2μ is an upper bound on the average long-run capability that the trainer can attain. It can be shown that this upper bound can be approximated arbitrarily well, but this must come at the price of arbitrarily long recurrent stretches of $m_t=0$ realizations (which are compensated for by periods in which m_t reaches arbitrarily high values). Obviously, such paths imply that the agent's minimal long-run capability is zero. By comparison, the process we constructed in Section 3 induces an average long-run capability

of approximately $2\mu - \frac{1}{2}$ and a minimal long-run capability of $2\mu - 1$.

A similar diagnosis pertains to the forward-looking case (consider μ as a precise upper bound on average intensity, for the sake of the argument). An upper bound on the average long-run capability is μ/c. The reason is that if average m exceeds this value, it implies that the agent's average long-run cost is above μ. However, the agent can ensure an average cost of μ by always playing $m = 0$, hence a long-run capability in excess of μ/c is inconsistent with the agent's best-replying. We believe that as in the myopic case, this upper bound can be approximated arbitrarily well, at the same price of long stretches of $m = 0$ realizations. By comparison, the process we constructed in Section 4 induces an average long-run m of approximately $\mu/c - 1 + c$, and a minimal long-run m of $\mu/c - 1$. It follows that many combinations of the minimal and average criteria would lead to the same result.

References

[1] Bolton, P. and M. Dewatripont (2005), Contract theory, MIT press.

[2] Bompa, T. and C. Buzzichelli (2018), Periodization: theory and methodology of training, Human kinetics.

[3] Kiely, J. (2012), Periodization paradigms in the 21st century: evidence-led or tradition-driven? International journal of sports physiology and performance, 7(3), 242-250.

[4] Kiely, J., C. Pickering and I. Halperin (2019), Comment on "Biological background of block periodized endurance training: a review", Sports Medicine, 49(9), 1475-1477.

[5] Issurin, V.B. (2010), New horizons for the methodology and physiology of training periodization,Sports Medicine, 40(3), 189-206.

[6] Issurin, V.B. (2019), Biological background of block periodized endurance training: A review, Sports Medicine 49, 31–39.

[7] Solan, E. and C. Zhao (2021), Dynamic monitoring under resource constraints, Games and Economic Behavior, forthcoming.

[8] Varas, F., I. Marinovic and A. Skrzypacz (2020), Random inspections and periodic reviews: Optimal dynamic monitoring, Review of Economic Studies 87(6), 2893-2937.