It Is Time to Get Some Rest

Manel Baucells
University of Virginia Darden School of Business, P.O. Box 6550, Charlottesville, Virginia 22906.
BaucellsM@darden.virginia.edu.

Lin Zhao
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190.
zhaolin@iss.ac.cn.

We introduce the fatigue disutility model. Set in continuous-time, the model captures in a precise way the notion that fatigue accumulates with effort, decays with rest, and produces an increase in marginal instant disutility. The model is meant to be prescriptive, although it rationalizes several anomalies observed in time preference. Under very general conditions the most efficient temporal profile of effort exhibits a high-low-high pattern. We also consider a variant whereby marginal productivity decreases with fatigue, and examine the optimal management of fatigue and productivity under various scenarios. We bring the model to the data and provide a calibration that predicts well the speed profiles observed in swimming competitions.

Key words: Disutility, Effort, Fatigue, Time preferences, Scheduling, Athletic time trial.

1. Introduction

Fatigue has been recognized as an important factor in determining the cost of providing effort, as well as in reducing the productivity. Numerous studies have noted the hazards associated with work overload, pointing to work-related fatigue as a substantial cause of cognitive errors, reduced job performance, and increased safety risks (Veasey et al. 2002, Kanai 2009, Brachet et al. 2012). In a study of cardiothoracic surgery, Kc and Terwiesch (2009) find that an increase in overwork by 10 percent is associated with a two percent increase in the likelihood of mortality. Similarly, paramedics' performance deteriorates toward the end of long shifts, resulting in a 76 percent increase in 30-day mortality Brachet et al. (2012, p. 219).

The notion that past effort reduces psychophysical resources and lowers productivity, and that rest restores resources, is well accepted in psychology via the effort-recovery and willpower paradigms (Meijman et al. 1998, Muraven et al. 1998). Evidence supporting this directional effect of recent effort is abundant (Hunter and Wu 2016, Hagger et al. 2010). In biomechanics, it is well understood that muscle force degrades with fatigue (Hawkins and Hull 1993).

In the management sciences, Bechtold et al. (1984) and Bechtold and Sumners (1988) propose a continuous-time model in which work and rest affect productivity, and show that nonzero breaks during a working day increase productivity. Their formulation is tailor made for the work-rest problem, and not necessarily derived from a more fundamental model to evaluate effort profiles.

In economics, fatigue is largely ignored by assuming inter-temporal separable preferences. After introducing the separable model, Samuelson (1937, p. 159) admits "the serious limitations of the previous kind of analysis, which almost certainly vitiate it even from a theoretical point of view." Indeed, the marginal rate of substitution between today and next week does not depend on what is happening in between. More critically, the time separable model violates local substitution, namely, that effort supplied at two nearby points in time should be close substitutes (Hindy et al. 1992).

A natural way to relax inter-temporal separability is to introduce a state variable, say fatigue, that depends on past effort and influences current preferences. The few formal models of fatigue seem to be motivated by tractability, rather than by the more fundamental question of how any such model should look like. For example, Fehr and Goette (2007, p. 305) propose a discrete-time model where effort in the previous period increases the marginal cost in the current period. Their model also violates the local substitution property. Moreover, discrete-time models cannot be directly applied to practical management questions such as the optimal duration of breaks or the optimal sequencing of tasks of variable duration. Dragone (2009) and Ozdenoren et al. (2012) consider continuous-time models where a state variable called fatigue and willpower, respectively, must be kept positive. Their state variable does not directly affect the instantaneous costs or benefits of effort, and hence does not reflect our intuitive notion of fatigue.

Hence, the literature is missing a continuous-time preference model over temporal profiles of effort that takes fatigue into account, that satisfies the local substitution property, and that is relatively simple (i.e., not very different from discounted utility).

In line with Kahneman et al. (1997), we assume total disutility is the time integral of an instant disutility profile. Following Hindy et al. (1992), instant utility is a general function of fatigue, effort, and time; and fatigue is a stock of recent effort with decay. Next, we impose three basic properties—continuity,[1] monotonicity, and increasing marginal cost—and one structural principle—conditional stationarity. Proposition 1 shows that these properties uniquely characterize the so called fatigue disutility model. Later, we also add a productivity component whereby marginal productivity decreases with fatigue.

The fatigue disutility model is characterized by a cost function and/or a productivity function, a discount factor, the initial fatigue level, and a fatigue recovery rate. The model converges to discounted disutility as the fatigue recovery rate grows large, provided some qualifications. Our fatigue model is intimately related to the satiation model for consumption proposed by Baucells

[1] We adopt continuity with respect to convergence in distribution, which guarantees the local substitution property.

and Sarin (2007). Baucells and Sarin (2007) only analyze optimal consumption in discrete-time. Here, we deal with a continuous-time setup, which allows for a richer analysis that can be adapted to a variety of contexts (effort rates, on/off rate, or rates and impulses).[2]

The fatigue disutility model is meant to be prescriptive. The model, however, is able to provide a rational explanation to some key anomalies observed experimentally when individuals face temporal trade-offs (Frederick et al. 2002, Bleichrodt et al. 2016). The common finding is that individuals exhibit excessive discounting, magnitude effects (larger amounts are discounted less), decreasing impatience, and high response variability. To be clear, our model only speaks for the domain of effort, that is, the trade-off between doing a costly task now, or postponing the task to later, perhaps at a higher cost. In such domain, we show that all four patterns may be explained by the influence that fatigue and rest produces on the desire to postpone a task.

Under fatigue, different temporal profiles of effort with equal total output may induce different disutility costs. We find the optimal effort profile that minimizes disutility in three domains of interest: bounded effort rates, on/off effort rate, and any combination of effort rates and impulses. In all three domains, we show that the solution to minimizing disutility is structurally identical to the solution to maximizing productivity, or to maximizing productivity net of disutility.

Roughly speaking, the optimal pattern is *high-low-high*. Thus, we obtain three extensions of the discrete-time result that effort in the first and last periods ought to be higher (Baucells and Sarin 2007). In the on/off domain, we generalize Bechtold et al. (1984)'s result that work output can always be increased if more breaks are allowed. Using parametric examples, we illustrate the beneficial effects of breaks to reduce the worker's cost, increase the employer's profit, and increase overall productivity (while working less time).

The high-low-high pattern is consistent with the mid-workday break around most countries (McMenamin 2007), and the "end spurt" effect whereby workers increase their effort just prior to a break (Vernon 1921). More revealing is the high-low-high pacing observed in professional sports, such as rowing (Garland 2005) and mid-distance running (Abbiss and Laursen 2008).

Finally, to highlight the practical relevance of our work, we develop an energy-speed model for athletes in a time trial competition. We show that the athlete's time minimization problem can be reformulated as an equivalent disutility minimization problem. We assume optimal behavior and calibrate the model parameters to predict the speeds observed in professional swimming competitions. The predictions fit the high-low-high profile observed in the data very well.

[2] On the technical side, we also contribute by establishing necessary and sufficient conditions for optimality. We cannot use off-the-shelf conditions because our instant disutility is linear in the current effort. Instead, we transform our objective into an equivalent one that only depends on fatigue and not on effort; and then implement Bank and Riedel (2000)'s approach, which we also extend to the more realistic case where rates are upper bounded.

2. Disutility of Effort

Let $x(t)$, $t \in [0,T]$, describe the effort rate an individual exerts at each instant during some bounded time interval. We admit any profile of effort whose cumulative $X(t) = \int_0^t x(s)ds$ is absolutely continuous and bounded. We let X be short for an effort profile, and $\mathscr{X}$ the set of admissible profiles.

Associated with each effort profile is a corresponding evolution of the fatigue level. The fatigue level is a stock of past effort with decay. For simplicity, we assume the stationary specification

$$Y(t) = Y_0 e^{-\gamma t} + \int_0^t e^{-\gamma(t-s)} x(s)ds. \tag{1}$$

Here $Y_0 \geq 0$ is the *initial fatigue level* and $\gamma > 0$ is the *recovery rate.* A larger γ indicates that fatigue decays faster after effort is set to zero. Equivalently, fatigue satisfies $Y'(t) = x(t) - \gamma Y(t)$. Thus, the fatigue level remains constant if $x(t)$, which contributes to fatigue builds up, equals the natural decay, $\gamma Y(t)$.

We propose a particular disutility function to represent preferences over effort profiles. As a preliminary, let a *cost function* $g(Y) : \mathbb{R}_+ \to \mathbb{R}_+$ be a twice continuously differentiable, strictly increasing and strictly convex function. We normalize $g(0) = 0$. In a one-period setting, a natural way to account for fatigue is to stipulate that disutility be given by $g(Y_0 + k) - g(Y_0)$, where k is the total effort exerted, and Y_0 is the fatigue level from "the past" (Fudenberg and Levine 2006, p. 1467). Disutility is given by $g(k)$ if the individual is rested, but otherwise increases with Y_0. Our model is the continuous-time formulation of this idea.[3]

Definition 1. The **fatigue disutility model** for effort profile $X \in \mathscr{X}$ is given by

$$D(X) = \int_0^T e^{-\delta t} g'(Y(t))x(t)dt, \tag{2}$$

where δ is the discount rate, $g(Y)$ is a cost function, and $Y(t)$ is as in (1). The preferences from the point of view of time $\tau \geq 0$ are time consistent and represented by $\int_\tau^T e^{-\delta t} g'(Y(t))x(t)dt$.

2.1. Properties and Characterization

The fatigue disutility model is a special case of $D(X) = \int_0^T v(Y(t), x(t), t)dt$, a general specification of the notion that total utility is the time integral of instant utility (Kahneman et al. 1997), and that instant utility may depend on fatigue, the current effort rate, and time (Hindy et al. 1992).

We consider four desirable properties for this general model. The first is continuity with respect to convergence in distribution. This ensures local substitution (see §2.3) and will allow us to define a discrete-time version as a byproduct of the continuous-time version.

[3] Consider a smooth effort rate $x(t)$. During a small time interval, $[t, t+\Delta)$, the total effort is $k = \bar{x}(t)\Delta$. Thus, $e^{-\delta t}[g(Y(t)+k) - g(Y(t))] = e^{-\delta t}[g(Y(t)+k) - g(Y(t))](1/k)\bar{x}(t)\Delta$. As $\Delta \to dt$, we have that $\bar{x}(t) \to x(t)$, and $e^{-\delta t}[g(Y(t)+k) - g(Y(t))] \to e^{-\delta t} g'(Y(t))x(t)dt$.

P1. **Continuity**: *If a sequence of profiles* $X_n(t) \to X(t)$ *for all* t*, then* $D(X_n) \to D(X)$.

The next two properties are naturally appealing: instant disutility should increase with instant effort; and marginal instant disutility should increase with fatigue.

P2. **Monotonicity**: $v(Y,x,t)$ *is strictly increasing in* x.

P3. **Increasing marginal cost**: $\partial v(Y,x,t)/\partial x$ *is strictly increasing in* Y.

The final property captures the view that, although preferences may depend on state variables that change over time (e.g. fatigue, temperature, age), time *per se* does not alter preferences.

P4. **Conditional stationarity**. *If the fatigue level at times* t_1 *and* t_2 *is the same, then the preference from the point of view of* t_1 *and* t_2*, respectively, is also the same.*

Formally, P4 can be spelled out as follows. Given times $t_1 \neq t_2$, let $x(t)$ be any effort rate with support contained on $[0, T - \max\{t_1, t_2\}]$. Consider two profiles. The first one begins with some arbitrary effort rate between 0 and t_1, followed by $x(t)$ shifted by t_1, or $x_1(t) = x(t - t_1)$, $t \in [t_1, T]$. The second profile begins with some other arbitrary rate between 0 and t_2 followed by $x(t)$ shifted by t_2, or $x_2(t) = x(t - t_2)$, $t \in [t_2, T]$. P4 postulates that if $Y_1(t_1) = Y_2(t_2)$, then the preference represented by $\int_{t_1}^{T} v\left(Y_1(t), x_1(t), t\right) dt$ is identical to the preference represented by $\int_{t_2}^{T} v\left(Y_2(t), x_2(t), t\right) dt$. Thus, the current fatigue level encapsulates the effect of past effort on current preferences.

We verify that $D(X)$ satisfies P1 to P4. Indeed, P1 holds precisely because instant disutility is *linear* in the current effort rate (Hindy et al. 1992, Prop. 6). Incidentally, the time separable model $\int_0^T e^{-\delta t} g(x(t))dt$ violates P1, unless g is linear. P2 and P3 hold because g is strictly increasing and strictly convex. P4 holds due to exponential discounting and stationary fatigue. Conversely, the disutility fatigue model is uniquely pinned down by P1-P4.

PROPOSITION 1. *Assume* $D(X) = \int_0^T v(Y(t), x(t), t)dt$*, where* $Y(t)$ *is given by (1) and* $v: \mathbb{R}_+^3 \to \mathbb{R}_+$ *is twice continuously differentiable.*[4] *P1 to P4 hold if and only if there exist a cost function* g*, unique up to positive affine transformation, and a discount rate* δ *such that* $D(X)$ *is given by (2).*

Proofs of all propositions are relegated to Appendix A. P4 implies exponential discounting. In combination with P1-P3, P4 also implies that at times when the effort rate is zero, instant disutility is zero. If we relax P4, then we obtain $v(Y,x,t) = a(Y,t)x + b(Y,t)$, where $a(Y,t)$ generalizes our fatigue component, and $b(Y,t)$ corresponds to the durability component proposed by Hindy et al. (1992). Durability captures the lingering effect of fatigue such as disutility after the effort has stopped, or positive effects of habituation on effort. Exploring the model with durability is technically similar, but falls outside the scope of this paper.

[4] The state variable could be assumed to take the more general form $Y(t) = Y_0\theta(t,0) + \int_0^t \theta(t,s)x(s)ds$ where $\theta(t,s)$ is continuous on $s \leq t$ with $\theta(0,0) = 1$, and consider the properties of time invariance—if we shift a profile to the right by t, and $Y(t)$ is made the same as Y_0, then $Y(t+s)$ equals $Y(s)$ for all $s \geq 0$—and recursivity—if two effort profiles yield the same fatigue level at t and produces identical effort rate afterwards, then the fatigue levels at $t+s$ are also the same for all $s \geq 0$. Time invariance and recursivity would then imply that $Y(t)$ is as in (1).

2.2. Effort Impulses

We now introduce the notion of an effort impulse, understood as the limit of an effort rate of higher and higher intensity, exerted over a shorter and shorter time. For example, the sequence $x_n(t) = n \cdot 1_{t\in[t_1,t_1+1/n]}$, $n = 1, 2, ...$, keeps total effort constant, and tends towards an *effort impulse* of size 1 at time t_1, which we denote by $\mathbf{1}_{t=t_1}$.

An impulse creates a jump in the cumulative effort profile. To allow for jumps, we consider a larger set of cumulative profiles, $\overline{\mathscr{X}}$, comprising all non-decreasing and right-continuous functions. The fatigue disutility model extended to $\overline{\mathscr{X}}$ is defined as

$$Y(t) = Y_0 e^{-\gamma t} + \int_0^t e^{-\gamma(t-s)} dX(s), \text{ and} \tag{3}$$

$$D(X) = \int_0^T e^{-\delta t} g'(Y(t)) dX(t), \tag{4}$$

where the integrals involving $dX(t)$ are defined in the usual Lebesgue-Stieltjes sense.[5]

2.3. The Discrete-Time Version

We seek to evaluate a discrete profile of the form $x(t) = \sum_{i=1}^m k_i \cdot \mathbf{1}_{t=t_i}$, indicating that an amount of effort k_i is exerted at time t_i, with $0 \le t_1 < t_2 < ... < t_m$. Thus, cumulative effort is piecewise constant with jumps of size k_i at times t_i. Let $t_0 = 0$ and $k_0 = Y_0$. By continuity (see proof of Proposition 1), the fatigue level just before t_i and the disutility is given by

$$Y_i = \sum_{j=0}^{i-1} k_j e^{-\gamma(t_i - t_j)}, \ i = 1, ..., m, \text{ and} \tag{5}$$

$$D(X) = \sum_{i=1}^{m} e^{-\delta t_i} [g(Y_i + k_i) - g(Y_i)]. \tag{6}$$

Expression (5) indicates that at the moment of exerting k_i units of effort, the fatigue level jumps from Y_i to $Y_i + k_i$, and otherwise decays at a rate γ. In Expression (6), the term $g(Y_i + k_i) - g(Y_i)$ agrees with the aforementioned one-period case. Note that if we increase the time separation between t_i and t_{i-1}, then Y_i decreases and so does the cost.

A key property of (5)-(6) is local substitution. Suppose an individual plans to exert efforts k_1 and k_2 at times t_1 and t_2. If the individual brings t_2 closer to t_1, then it is trivial to verify that $D(X) \to e^{-\delta t_1}[g(Y_1 + k_1 + k_2) - g(Y_1)]$. As one would expect, k_1 and k_2 become perfect substitutes. In contrast, carrying this thought experiment in the time separable model yields the non-sensical limit $e^{-\delta t_1}[g(k_1) + g(k_2)]$.

[5] Convergence in distribution in $\overline{\mathscr{X}}$ occurs if $X_n(t) \to X(t)$ at all points of continuity of $X(t)$ and at $t = T$. Because $\mathscr{X}$ is dense in $\overline{\mathscr{X}}$, (3)-(4) is the unique continuous extension of (1)-(2) to $\overline{\mathscr{X}}$.

2.4. The Productivity Component

We have considered the negative effects of fatigue on the disutility cost. We now include the negative effect of fatigue on productivity. To do so, we complement the cost function with a *productivity function*, $h(Y):\mathbb{R}_+\to\mathbb{R}$, which is twice continuously differentiable, strictly increasing and concave, with $h(0)=0$. Associated with an effort rate $x(t)$ there is a discounted output rate $e^{-rt}h'(Y(t))x(t)$. Thus, the effectiveness of effort is dampened by a productivity factor which decreases with fatigue. Formally, the discounted output associated with an effort profile $X\in\overline{\mathscr{X}}$ is given by:

$$P(X)=\int_0^T e^{-rt}h'(Y(t))dX(t). \tag{7}$$

Note that this specification satisfies continuity (for $x\in\mathscr{X}$, we have that $dX(t)=x(t)dt$ and instant productivity is linear in the effort rate), monotonicity, decreasing marginal productivity with fatigue, and conditional stationarity. For discrete-time profiles, the discounted productivity becomes $P(X)=\sum_{i=1}^m e^{-rt_i}[h(Y_i+k_i)-h(Y_i)]$, where Y_i is as in (5).

In some settings (§5.1 and §7), the goal of the individual will be to maximize the total output, $P(X)$. In other applications, the individual receives a wage of w per unit of outcome, paid at the moment the output is produced. Then, r is the discount rate for money and $wP(X)$ the discounted income. A self-employed individual will seek to maximize discounted income minus discounted cost, $wP(X)-D(X)$ (see §5.3 and §6.2).

The two negative effects of fatigue, increasing marginal cost and decreasing marginal productivity, are substitutes of sorts. Assume $h'(0)$ is finite and normalized to 1.

- If the goal is to maximize $P(X)$, then we can define a convex proxy cost function $\tilde{g}(Y)=Y-h(Y)$ and rewrite $P(X)=\int_0^T e^{-rt}dX(t)-\int_0^T e^{-rt}\tilde{g}'(Y(t))dX(t)$.
- If the goal is to maximize $wP(X)-D(X)$ and $r=\delta$, then we can define $\tilde{g}(Y)=wY-wh(Y)+g(Y)$ and rewrite $wP(X)-D(X)=w\int_0^T e^{-rt}dX(t)-\int_0^T e^{-rt}\tilde{g}'(Y(t))dX(t)$. A particularly convenient assumption is $h(Y)=Y-\varphi g(Y)$, $\varphi\geq 0$, yielding $\tilde{g}(Y)=(1+\varphi w)g(Y)$.

In either case, the productivity in the modified program is linear. Consequently, the structure of the optimal effort profile will be determined by the minimization of the disutility component. Thus, our initial focus will be on the disutility component.

2.5. The Power Family

We address the qualitative properties and optimal solutions for general cost and productivity functions. To illustrate the solutions with closed-form expressions, we employ the important family of *power* cost functions, $g(Y)=Y^\phi/\phi$, $\phi>1$. The productivity counterpart is the *linear minus power* family, $h(Y)=Y-\varphi Y^\phi/\phi$, $\phi>1$, $\varphi>0$.

Among the power family, the *quadratic* cost function has a special property, namely that instant utility is bilinear in fatigue and effort. In turn, this implies time symmtery; which then translates into symmetric optimal solutions (see Examples 1, 2, 5).

Remark 1. For $X \in \overline{\mathscr{X}}$, let $\overleftarrow{x}(t) = x(T-t)$ be the profile of X running backwards in time.[6]
$D(X) = D(\overleftarrow{X})$ for all $X \in \mathscr{X}$ if and only if $Y_0 = 0$, $\delta = 0$ and $g(Y) = Y^2/2$.
$P(X) = P(\overleftarrow{X})$ for all $X \in \mathscr{X}$ if and only if $Y_0 = 0$, $r = 0$ and $h(Y) = Y - \varphi Y^2/2$.

A calibration of the power cost function for our fatigue model was performed, somehow unintentionally, by Augenblick et al. (2015). The authors elicit preferences over work amounts using a separation of one week between working days. Importantly, they set Mondays as the working days, ensuring the same rest during the weekend, and impose a minimum amount of work at the beginning of each working day. Luckily, they estimate a time separable model with a Stone-Geary background parameter that acts as our initial fatigue level. While they do not subtract $g(Y_i)$, such term would be constant in their case because they employ the same minimum work across days. They assume a power cost function and their data yields an estimate of $\phi = 1.6$.

3. Implied Time Preference

A core question in time preference is the propensity to postpone a costly task to a later time. To keep things simple, consider the prospect of exerting an impulse of effort k at a particular time t, and resting otherwise. The disutility associated with this simple prospect is

$$D(k,t) = e^{-\delta t}\left[g\left(Y_0 e^{-\gamma t} + k\right) - g\left(Y_0 e^{-\gamma t}\right)\right].$$

We employ $\beta(k,t) = -\frac{1}{D(k,t)}\frac{\partial D(k,t)}{\partial t}$ to measure the relative desire to postpone effort by a little amount of time. Simple manipulations yield

$$\beta(k,t) = \delta + \gamma Y_0 e^{-\gamma t}\frac{g'\left(Y_0 e^{-\gamma t} + k\right) - g'\left(Y_0 e^{-\gamma t}\right)}{g\left(Y_0 e^{-\gamma t} + k\right) - g\left(Y_0 e^{-\gamma t}\right)}.$$

As expected, if the initial fatigue is zero, then $\beta(k,t)$ is equal to the discount rate. If $Y_0 > 0$, however, then the passage of time reduces fatigue and makes future effort less costly. The rate of time preference $\beta(k,t)$ takes this into account, and predicts four patterns of behavior.

Proposition 2. *Assume $Y_0 > 0$, g power, and $\delta \geq 0$. Time preferences exhibits*

i) excessive discounting: $\beta(k,t) > \delta$.

ii) magnitude effects: $\beta(k,t)$ is strictly decreasing in k for all t.

iii) decreasing impatience: $\beta(k,t)$ decreases towards δ as time increases.

iv) fatigue dependence: $\beta(k,t)$ is strictly increasing in Y_0.

Moreover, i) holds for all cost functions, ii) holds whenever $g''(Y)/g'(Y)$ is strictly decreasing, and iii-iv) hold if, in addition, $g''(Y)Y/g'(Y)$ and $g'''(Y)/g''(Y)$ are increasing.

[6] We verify the if part using a change of variables. As for the only if, that $Y_0 = \delta = 0$ follows from considering $x = k \cdot \mathbf{1}_t$ and imposing $D(X) = D(\overleftarrow{X})$ for all $k \geq 0$. That g is quadratic, follows from considering $x = k \cdot \mathbf{1}_0 + k' \cdot \mathbf{1}_t$ and imposing $D(X) = D(\overleftarrow{X})$ for all $k' \geq 0$ and $t \geq 0$.

Observed discount rates are usually much higher than interest rates or mortality risks (Loewenstein and Thaler 1989). High discount rates are usually seen as a sign of neglect for the future selves. When it comes to delaying a costly task under fatigue, however, a high desire to postpone a task can be perfectly rational. Indeed, if my current fatigue level is high, then delay provides the opportunity to rest, and complete the task later at a much reduced cost.

That larger outcomes are discounted less is a robust finding (Thaler 1981, Frederick et al. 2002). For small tasks the effect of the current fatigue on overall disutility is relatively large and induces a strong preference for resting. For large tasks, in contrast, the initial fatigue level will be a smaller component of the total cost, and the preference for resting (while positive) is relatively weaker.

Patterns of decreasing impatience are commonly observed in experiments involving delayed gratification (Frederick et al. 2002), as well as the delay of unpleasant tasks (Augenblick et al. 2015). As shown in Halevy (2015), if preferences are stationary, then decreasing impatience produces time inconsistencies and preference reversals. Here, however, preferences are not stationary because they depend on the fatigue level. Indeed, the ability to manage fatigue in the future—but not now—can easily explain a rational preference for supplying 2 units of effort next week rather than 1 unit today, while at the same time preferring to supply 1 unit four weeks from now, rather than 2 units five weeks from now.

Finally, Frederick et al. (2002) observed a very high variability in elicited discount rates. If fatigue levels vary over time, then it is not surprising that discount rates vary as well. Our model emphasizes that to obtain more stable discount rates the experimenter needs to ensure participants are fully rested, or otherwise find a way to control for fatigue.

The insights of Proposition 2 hold for more general effort profiles. Regarding the magnitude effect, we show in §6.1 that for g power, the optimal temporal distribution of effort shifts to sooner as we increase the total effort requirement. To generalize the other three effects, consider any effort profile $X \in \overline{\mathscr{X}}$ with support contained on $[t, t+\tau]$, $t, \tau \geq 0$; and denote by X^s the postponement of X to $[s, s+\tau]$, $s > t$. The implied discount rate, $\beta(X,t) = -\frac{1}{D(X^s)} \frac{\partial}{\partial s}\big|_{s=t} D(X^s)$, equals

$$\beta(X,t) = \delta + \gamma Y_0 e^{-\gamma t} \frac{\int_t^{t+\tau} e^{-\delta s} g''(Y(s)) dX(s)}{\int_t^{t+\tau} e^{-\delta s} g'(Y(s)) dX(s)}.$$

The monotonic dependence of $\beta(X,t)$ with respect to t and Y_0 cannot always be guaranteed. The following, however, is apparent from examining $\beta(X,t)$.

Remark 2. Let $Y_0 > 0$ and $\delta \geq 0$. For all cost functions, we have $i)$ excessive discounting, $\beta(X,t) > \delta$; $iii)$ that decreasing impatience eventually holds, $\lim_{t\to\infty} \beta(X,t) \to \delta$; and $iv)$ that fatigue dependence holds, in the sense that $\lim_{Y_0\to\infty} \beta(X,t) \to \infty$ and $\lim_{Y_0\to 0} \beta(X,t) \to \delta$.

4. The Optimal Effort Rate

Consider a task that requires a total of $K > 0$ units of effort. Effort can be supplied in any form or manner during $[0, T]$ as long as $\int_0^T x(t)dt = K$. We seek the temporal profile that minimizes $D(X)$. In this section, we seek a solution in effort rates (no impulses allowed). Unless noted otherwise, and to concentrate on the effect of fatigue, we set the discount rate to zero.

The optimal profile may (and in fact does) involve short periods with a high effort rate. To guarantee the existence of a solution with rates only, we must specify a maximal effort rate, c, the individual can afford. Let $\mathscr{X}_c = \{x \in \mathscr{X} \mid 0 \le x(t) \le c\}$, and set $K < cT$ to ensure feasibility. The goal is to solve

$$\min_{X \in \mathscr{X}_c} D(X) \quad s.t.\ X(T) = K. \tag{8}$$

The optimal profile, $x^*(t)$, is not constant. Instead, it is piecewise constant over three time intervals. In the third interval, the effort rate is c. In the second interval, the effort rate is below c. In the first interval, the effort rate is either c or 0, depending on whether the initial fatigue level is low or high. The intervals do not need to be of equal duration.

To describe the fatigue level during the second interval, let

$$\Delta(Y) = (1/\gamma)\ln(c/(c - \gamma Y))$$

be the amount of time it takes to bring the fatigue level from 0 to Y using the maximal rate, c. Thus, $\Delta(Y) - \Delta(Y_0)$ is the amount of time it takes to bring the fatigue level from Y_0 to some $Y > Y_0$. Next, and to determine the terminal fatigue, define the function $Y_f(Y)$ implicitly by[7]

$$g'(Y_f) - g'(Y) = (1 - \gamma Y/c)g''(Y)Y. \tag{9}$$

Proposition 3. *Assume $\delta = 0$ and $g''(Y)Y$ is strictly increasing. The effort profile that solves (8) is unique, and depends on how the initial fatigue level compares to thresholds $0 < \overline{Y} \le \overline{\overline{Y}}$.*

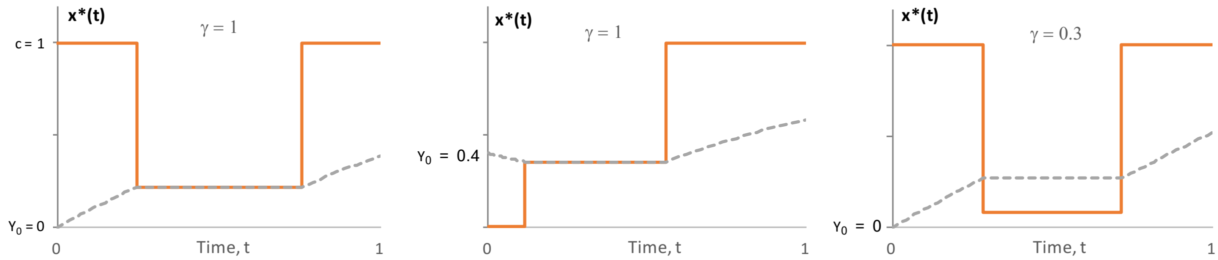
- *If $Y_0 \le \overline{Y}$, then $x^*(t) = c \cdot 1_{t \in [0, t_1^*]} + \gamma Y^* \cdot 1_{t \in (t_1^*, t_2^*)} + c \cdot 1_{t \in [t_2^*, T]}$.*
- *If $\overline{Y} < Y_0 < \overline{\overline{Y}}$, then $x^*(t) = \gamma Y^* \cdot 1_{t \in (t_1^*, t_2^*)} + c \cdot 1_{t \in [t_2^*, T]}$.*
- *If $Y_0 \ge \overline{\overline{Y}}$, then $x^*(t) = c \cdot 1_{t \in [T - K/c, T]}$.*

The first case is pinned down by $t_1^* = \Delta(Y^*) - \Delta(Y_0)$, $T - t_2^* = \Delta(Y_f(Y^*)) - \Delta(Y^*)$, and $X(T) = K$; the second by $Y^* = Y_0 e^{-\gamma t_1^*}$, $T - t_2^* = \Delta(Y_f(Y^*)) - \Delta(Y^*)$, and $X(T) = K$; and the third by $X(T) = K$. The proof contains the equations determining $\overline{Y}$, $\overline{\overline{Y}}$, Y^*, t_1^*, and t_2^*.

The *initialization period* brings the fatigue level to a steady level Y^* as fast as possible by using a maximal rate if $Y_0 \le \overline{Y}$ (Figure 1, left), or resting whenever $Y_0 > \overline{Y}$ (Figure 1, middle). If the

[7] For g power, we have that $Y_f(Y) = Y\,[1 + (\phi - 1)(1 - \gamma Y/c)]^{1/(\phi - 1)}$.

Figure 1 Optimal profiles for different values of Y_0 and γ [g quadratic, $c = T = 1$, and $K = 0.6$].



initial fatigue level is very high, then sustentation also disappears and all effort is provided during termination. Intuitively, one likes to set a high effort rate to profit from the lower marginal cost when one is rested; and desires a period of rest when one is tired.

The *sustentation period*, $\gamma Y^* \cdot 1_{t\in\left(t_1^*,t_2^*\right)}$, maintains the fatigue level at Y^*. During sustentation the individual balances two forces: smoothing out fatigue levels and moderating fatigue in anticipation of future effort. The *termination period* is a concentration of effort near the end, $c \cdot 1_{t\in\left[t_2^*,T\right]}$. Intuitively, we do not mind the final sprint because increasing the fatigue level has no subsequent effect once we reach T. Technically, termination imposes an extra cost to elevating fatigue while approaching t_2^*, and maintains the desire for moderation constant during $[t_1^*, t_2^*]$.[8]

The recovery rate γ greatly influences the shape of the optimal solution. As γ increases, the duration of both the initialization and termination periods shrinks to zero, and $x^*(t)$ converges to a constant rate K/T. A constant effort rate is the solution advocated by the undiscounted time separable model. Later, in footnote 12, we confirm the intuition that the time separable model can be thought as the limit of $D(X)$ when γ goes to infinity. Conversely, if γ is small relative to T and fatigue has no time to decay, then the sustentation phase involves close to zero effort because the fatigue level stays automatically constant. Instead, it is optimal to work only at the beginning and at the end, and take it easy in the middle (see Figure 1, right).

[8] Heuristically, for the proposed pattern to be optimal the effect of any feasible perturbation should be locally flat. Let $D(X_\varepsilon)$ denote the disutility generated by $X_\varepsilon(t) = X(t) + \varepsilon \int_0^t \zeta(s)ds$, where $\zeta(t) : [t_1^*, t_2^*] \to \mathbb{R}$ is a perturbation of the sustentation period satisfying $\int_{t_1^*}^{t_2^*} \zeta(t)dt = 0$. We have that

$$\left.\frac{dD(X_\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0} = \int_0^T \mathcal{M}\left(t; Y(t)\right)\zeta(t)dt, \quad \text{where} \quad \mathcal{M}(t;Y) = g'(Y(t)) + \int_t^T g''(Y(s))e^{-\gamma(s-t)}x(s)ds.$$

The term $g'(Y(t))$ is the direct effect of the perturbation on instant disutility at time t, whereas the integral term is the indirect effect of changing $Y(t)$ on subsequent effort. Setting X at the proposed profile yields

$$\mathcal{M}(t;Y) = g'(Y) + g''(Y)Y + e^{-\gamma(t_2^*-t)}\left[(g'(Y_f) - g'(Y))/(1-\gamma Y/c) - g''(Y)Y\right].$$

For the first-order condition $dD(X_\varepsilon)/d\varepsilon|_{\varepsilon=0} = 0$ to hold, $\mathcal{M}(t;Y)$ must be independent of t. This sets the term in brackets to zero and yields (9).

The quadratic cost function with $Y_0=0$ yields an explicit solution with symmetric durations of the initialization and termination periods involving the Lambert-W function.[9]

Example 1. For g quadratic and $Y_0=0$, the solution to (8) is given by

$$t^* = T/2 - W(\gamma\,(T-K/c)\,e^{\gamma T/2}/2)/\gamma, \text{ and}$$
$$x^*(t) = c\cdot 1_{t\in[0,t^*]} + c(1-e^{-\gamma t^*})\cdot 1_{t\in(t^*,T-t^*)} + c\cdot 1_{t\in[T-t^*,T]}. \tag{10}$$

For $\phi>2$, the termination period is shorter than the initialization period, and vanishes as ϕ becomes large. Intuitively, if the cost function is very steep, then the termination period need not be long in order to make the desire for moderation constant. For $1<\phi<2$, the termination period is longer than the initialization period, but the initialization period does not vanish as $\phi\to 1$.

From the inspection of the necessary and sufficient conditions that minimize $D(X)$ used in Propositions 3, we verify that the same high-low-high structure obtains when the objective is to maximize $P(X)$, provided $h''(Y)Y$ is strictly decreasing; and when the objective is to maximize $wP(X)-D(X)$, provided $h''(Y)Y$ is decreasing and $g''(Y)Y$ is increasing, with at least one condition strict. This holds even if $r\geq 0$, $\delta\geq 0$, or $h'(0)$ is not finite.

5. On/Off Effort Rate

We now examine a setup where the effort rate is either $c>0$ (On) or 0 (Off) to capture the realistic situations in which a steady effort rate is required while working (e.g., operating a machine or vehicle, or attending customers). It also applies to mental task such as grading exams or reviewing papers that require constant concentration. The Off times, of course, correspond to taking breaks. This setup naturally leads to the problem of finding the optimal location of breaks during a given time window $[0,T]$.

We denote by $\mathscr{S}_n$ $(n=0,1,...)$ the set of On/Off schedules in $\mathscr{X}$ with at most n intermediate breaks and $n+1$ disjoint working intervals. Specifically, $X\in\mathscr{S}_n$ can be characterized by $(\ell_0,s_1,\ell_1,...,s_n,\ell_n,s_{n+1},\ell_{n+1})$, where ℓ_0 denotes a possible initial period of rest, $s_1\geq 0$ denotes the duration of the first work interval, followed by a break of duration $\ell_1\geq 0$, followed by a working interval of duration $s_2\geq 0$, followed by a break of duration $\ell_2\geq 0$, etc. Of course, $\ell_0+\sum_{j=0}^{n+1}(s_j+\ell_j)=T$. Note that for all $0\leq n<n'$, we have that $\mathscr{S}_n\subset\mathscr{S}_{n'}\subset\mathscr{X}_c$.

As in Bechtold et al. (1984), throughout this section we set $Y_0=0$ and $\delta=0$. Our setup is more general than theirs, and we do not impose that a nonzero working interval must precede the first break. Given n and $K<cT$, we seek to solve

$$D_n^*(K) = \min_{X\in\mathscr{S}_n} \; D(X) \quad s.t.\; X(T)=K. \tag{11}$$

For $n=0$, we trivially obtain $\ell_0^*=0$, $s_1^*=K/c$, and $\ell_1^*=T-K/c$.

[9] The positive Lambert-W function is defined by $x=W(x)e^{W(x)}$, $x\geq 0$. $W(x)$ can be calculated using the recursion $W_n=W_{n-1}-(W_{n-1}-x/e^{W_{n-1}})/(1+W_{n-1})$ and $W_0=\ln(1+x)$.

Proposition 4. *Assume $Y_0 = \delta = 0$, $n \geq 1$, and $g''(Y)Y$ strictly increasing. The solution to (11) exhibits $\ell_0^* = \ell_{n+1}^* = 0$, $s_1^*, ..., s_{n+1}^* > 0$, and $\ell_1^*, ..., \ell_n^* > 0$. Moreover, $D_n^*(K)$ is strictly decreasing in n and tends to $D^*(K)$, the solution to (8). If $g(Y)$ further satisfies two well-posedness conditions (see proof), then the solution is unique and satisfies $s_2^* = ... = s_n^*$, and $\ell_1^* = ... = \ell_n^* = (T - K/c)/n$.*

The optimal schedule can be characterized by four durations: initialization, termination, intermediate breaks, and intermediate working times. The proof contains the equations determining these. Of course, the role of $s_1^* > 0$ and $s_{n+1}^* > 0$ is to recreate the initialization and termination periods of the $\mathscr{X}_c$ setup. That disutility can be decreased by allowing more and more breaks is because frequent short breaks smooth out the fatigue level and recreate the sustentation period. Thus, breaks allow the individual to approach the flexible high-low-high profile.

Note that it is optimal to begin work at time 0, and end at time T; so that the breaks in the middle can be spaced out in order to reduce fatigue. A practical recommendation would be to do some work in the morning before showering (or even breakfast), as otherwise showering becomes a wasted opportunity to reduce fatigue that has not accumulated yet.

The quadratic cost function is well-posed, and yields a unique and symmetric solution.

Example 2. For g quadratic, the solution of (11) is as follows. For $n = 1$, $s_1^* = s_2^* = K/(2c)$. For $n \geq 2$, $s_1^* = s_{n+1}^* = \Delta(Y^*)$ and $s_2^* = ... = s_n^* = \Delta(Y^*) - \Delta(Y^* e^{-\gamma \ell_1^*})$, where Y^* uniquely solves

$$\left(\frac{c - \gamma Y}{c}\right)^2 = e^{-\gamma K/c} \left[(1 - e^{-\gamma \ell_1^*}) \frac{c}{c - \gamma Y} + e^{-\gamma \ell_1^*}\right]^{n-1}.$$

In either case, $D_n^*(K) = c[K + Y^*(n e^{-\gamma \ell_1^*} - n - e^{-\gamma s_1^*}) - c(1 - e^{-\gamma s_1^*})/\gamma]/\gamma$.

5.1. Work Less and Be More Productive

To illustrate how Proposition 4 carries over to the problem of maximizing productivity, consider a task such as grading exams, which requires full concentration and creates fatigue.

Example 3. Assume the maximum rate is $c = 6$ exams per hour. The fatigue recovery rate is $\gamma = 1.5$, so that fatigue drops by 78% ($= 1 - e^{-\gamma}$) after one hour of rest. The productivity function is $h(Y) = Y - \varphi Y^2/2$, so that the pace at which grading advances decreases linearly with fatigue, as given by $h'(Y(t))c = (1 - \varphi Y(t))c$. We set $\varphi < \gamma/c$ to ensure that productivity stays positive. The goal is to maximize the exams graded in $T = 10$ hours, stoically ignoring the disutility cost.

To solve such problems, we decompose $\max_{X \in \mathscr{S}_n} P(X)$ into two programs:

$$P_n^* = \max_K \ P_n^*(K), \text{ where}$$

$$P_n^*(K) = \max_{X \in \mathscr{S}_n} \ P(X) \ s.t. \ c \sum_{i=1}^{n+1} s_i = K.$$

By our convenient choice for $h(Y)$, the second program becomes $P_n^*(K) = K - \varphi D_n^*(K)$, where $D_n^*(K)$ is as in Example 2. We numerically solve the first program to find K_n^*.

Table 1 Optimal schedules for a given number of breaks [$c=6$, $\gamma=1.5$, $\varphi=0.2$ and $T=10$].

Number of breaks	Duration of work and rest periods: Initialization	Int. break	Int. work	Termination	Total working time	Total productivity
0	10h	-	-	-	10h	15.2
1	4h 32m	56m	-	4h 32m	9h 04m	16.5
3	2h 03m	46m	1h 48m	2h 03m	7h 42m	18.6
5	1h 25m	34m	1h 05m	1h 25m	7h 10m	19.6
10	59m	19m	33m	59m	6h 50m	20.3
flexible	39m	8h 42m at rate $0.62c$		39m	$K^*/c=$ 6h 44m	20.6

Table 1 exhibits the results for $n=0,1,3,5$, and 10 intermediate breaks. With no breaks, it is not optimal to quit before T because marginal productivity is always positive. The total output after with working for 10 hours is $P_0^*=15.2$ exams. With one break, the total working time drops to 9h 4min, while productivity increases to $P_1^*=16.5$ exams. With three breaks, where the total working time drops to 7h 42min hours and productivity increases to $P_3^*=18.6$ exams. Here, the optimal plan is to work for 2h 3min, take a 46min break, work for 1h 48min minutes, another 46min break, another 1h 48min minutes of work, another 46min break, and end with 2h 3min work. As the number of breaks increases, total working time decreases and productivity increases towards the flexible solution in $\mathscr{X}_c$ (see last row).

5.2. Breaks Increase Wellbeing

Consider a self-employed individual facing a wage $w>0$ per unit of output. The individual is not stoic, and wishes to balance income and effort cost. His goal is to maximize $wP(X)-D(X)$. In order to use Proposition 4, we decompose this goal into

$$\begin{aligned} U_n^* &= \max_K \; U_n^*(K), \text{ where} \\ U_n^*(K) &= \max_{X\in\mathscr{S}_n} wP(X)-D(X) \; s.t. \; c\sum_{i=1}^{n+1} s_i = K. \end{aligned} \tag{12}$$

Our focus is on K_n^*, the endogenous total effort, as a function of w. We compare three situations: work during one continuous interval and optimize over $\mathscr{S}_0$; take one break and optimize over $\mathscr{S}_1$; and the hypothetical first best when optimizing over $\mathscr{X}_c$ (flexible).

Example 4. Assume h linear and g quadratic. The solution to (12) for $n=0,1$ is as follows.

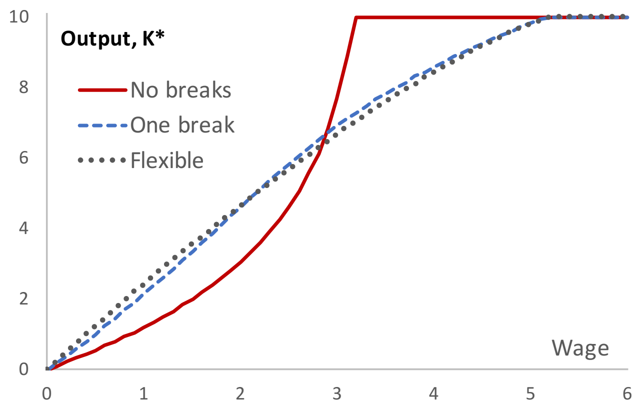
- $\mathscr{S}_0$: With no breaks, if $w\geq c(1-e^{-\gamma T})/\gamma$, then $K_0^*=cT$. Otherwise, $K_0^*=(c/\gamma)\ln(c/(c-\gamma w))$.

The individual stops working exactly when the fatigue level reaches the wage.

- $\mathscr{S}_1$: With one break, by our convenient choice for $h(Y)$, the second program becomes $U_1^*(K)=wK-D_1^*(K)$, where $D_1^*(K)$ follows from Example 2. K_1^* then follows from the first-order condition.
- $\mathscr{X}_c$: In the flexible case, we numerically calculate K_{flex}^* and U_{flex}^* for different value of w.

Figure 2 plots total output as a function of wage. In all three situations, total output is increasing with wage. Of course, U_{flex}^*, U_1^*, and U_0^* are also increasing with w. Because of the set inclusion

Figure 2 **Optimal total output as a function of wage [$c = 1$, $\gamma = 0.3$ and $T = 10$].**



$\mathscr{X}_c \supset \mathscr{S}_1 \supset \mathscr{S}_0$, we have that $U^*_{\text{flex}} > U^*_1 > U^*_0$ for all w. One would guess that the possibility of resting will induce the worker to work more. When the wage is low, we indeed observe that $K^*_0 < K^*_1$. For high wages, however, the opposite is true. To understand why, note that with one break the individual operates at lower fatigue levels. Increasing the wage pushes towards working more by shortening the break, but the individual resists the push because shortening the break elevates fatigue and makes work most costly. With no breaks, the individual operates at high fatigue levels. The fatigue level plateaus for w relatively large, and the marginal cost of extending the working time is not that high.

Importantly, for all wage levels, K^*_1 and K^*_{flex} are very close, and so are U^*_1 and U^*_{flex}. Hence, just one break brings the individual very close to the first best. This is to be expected whenever the fatigue recovery rate is small relative to the time window, in which case the first best solution over $\mathscr{X}_c$ has a low middle (see Figure 1, right), and is close to the optimal solution over $\mathscr{S}_1$.

5.3. Breaks Increase Profits

Consider now the business perspective. A company can hire a worker who can verifiably supply any effort profile in $\mathscr{S}_0$, $\mathscr{S}_1$, or $\mathscr{X}_c$, as long as he is compensated with at least $D(X)$. The price of the product is w per unit of output. The company's revenue is $w\int_0^T x(t)dt$, and the joint surplus is $w\int_0^T x(t)dt - D(X)$. The latter is equal to the company's profit is the worker is paid $D(X)$. The problem of maximizing the joint surplus is formally identical to (12), and so is the solution. In Figure 2 we then have the optimal amount of contracted work as a function of price; and $U^*_{\text{flex}} > U^*_1 > U^*_0$ would be the joint surplus as a function of price.

Thus, our model provides practical recommendations on how to efficiently manage fatigue using breaks in order to maximize productivity, increases worker's wellbeing, and employer's profits.

6. Optimal Profile with Rates and Impulses

In some applications, it may be convenient to use $\overline{\mathscr{X}}$, the domain that includes any effort rate and effort impulses. As usual, we begin with considering

$$\min_{X \in \overline{\mathscr{X}}} D(X) \quad s.t. \quad X(T) = K. \tag{13}$$

The optimal solution is similar to that over $\mathscr{X}_c$, but with the initialization and termination periods replaced by effort impulses. Indeed, as c grows large, the duration of the initialization and termination periods shrink, and the total effort provided converges towards an impulse.

PROPOSITION 5. *Assume $\delta \geq 0$ and $g''(Y)Y$ is strictly increasing. The solution to (13) is unique, and depends on how Y_0 compares to thresholds $0 < \overline{Y} \leq \overline{\overline{Y}}$.*

- *If $Y_0 \leq \overline{Y}$, then $x^*(t) = k_1 \cdot \mathbf{1}_{t=0} + k(t) \cdot 1_{t \in (0,T)} + k_2 \cdot \mathbf{1}_{t=T}$.*
- *If $\overline{Y} < Y_0 < \overline{\overline{Y}}$, then $x^*(t) = k(t) \cdot 1_{t \in \left(t_1^*, T\right)} + k_2 \cdot \mathbf{1}_{t=T}$.*
- *If $Y_0 \geq \overline{\overline{Y}}$, then $x^*(t) = K \cdot \mathbf{1}_{t=T}$.*

The proof contains the equations determining $\overline{Y}$, $\overline{\overline{Y}}$, t_1^*, k_1, $k(t)$, k_2. The sustentation rate $k(t)$ increases over time if $\delta > 0$, and stays constant if $\delta = 0$. If the domains $\mathscr{X}_c$ and $\mathscr{S}_n$ are only tractable for g quadratic and no discounting, $\overline{\mathscr{X}}$ yields closed form solutions for g power and discounting.

EXAMPLE 5. Assume g is power and let $\rho = \delta/(\phi - 1)$, $\eta = (\phi + \delta/\gamma)^{1/(\phi-1)}$, $L(T) = \gamma\left(e^{\rho T} - 1\right)/\rho + \eta e^{\rho T}$. We have that $\overline{Y} = K/(L(T) - 1)$ and $\overline{\overline{Y}} = Ke^{\gamma T}/(\eta - 1)$. If $\overline{Y} < Y_0 < \overline{\overline{Y}}$, then $t_1^* \in (0, T)$ is the solution to $Y_0 = Ke^{\gamma t}/(L(T-t) - 1)$; otherwise set $t_1^* = 0$. Then,

$$\begin{aligned} k_1 &= (K + Y_0)/L(T) - Y_0, \\ k(t) &= (K + Y_0 e^{-\gamma t_1^*})(\gamma + \rho) e^{\rho(t - t_1^*)}/L(T - t_1^*), \text{ and} \\ k_2 &= (K + Y_0 e^{-\gamma t_1^*})(\eta - 1) e^{\rho(T - t_1^*)}/L(T - t_1^*). \end{aligned}$$

6.1. A Slow Sparrow Has to Do an Early Start

Rewriting Example 5 in terms of the total effort required, K, creates a more elaborated version of the magnitude effect discussed in §3. To make the discussion more vivid, imagine a student who faces an exam at the end of the term, and let K be the total amount of material to be absorbed. Assume $Y_0 > 0$ and let $\overline{K} = Y_0(\eta - 1)e^{-\gamma T}$ and $\overline{\overline{K}} = Y_0(L(T) - 1)$. We momentarily set $\delta = 0$.

- If the course is easy, $K \leq \overline{K}$, then $Y_0 \geq \overline{\overline{Y}}$ and postponing study to the last minute is optimal even with no discounting. The reason, again, is to let fatigue decay.
- If $\overline{K} < K < \overline{\overline{K}}$, then $\overline{Y} < Y_0 < \overline{\overline{Y}}$ and the best is to slack off for a while, and start studying at some point in the middle of the term, setting a constant pace, and provide a final push.
- If the course is difficult, $K \geq \overline{\overline{K}}$, then $Y_0 \leq \overline{Y}$ and the student better begin right away with an initial impulse, followed by a constant pace, complemented with a final push.

If $\delta > 0$, then $\overline{K}$ and $\overline{\overline{K}}$ increase, intensifying the tendency towards postponement. For moderate to difficult courses, the pace of study—once initiated—will progressively increase as given by $k(t)$.

Note that if $Y_0 = 0$, then k_1, $k(t)$ and k_2 all increase in proportion to K. Thus, the fraction of work done up to time t is independent of K. If $Y_0 > 0$, however, the work distribution shifts to earlier, as if the individual were more patient. The reason has nothing to do with δ, but with the effect of fatigue. The result is elementary to verify.

Remark 3. Consider the solution to (13) for $Y_0 > 0$, $\delta \geq 0$, and g power. For all $t < T$, the fraction of work done up to time t, given by $X^*(t)/K$, is strictly increasing in K.

6.2. The Tension between Labor Supply and Wage

We now examine the maximization of $wP(X) - D(X)$ over $\overline{\mathscr{X}}$. Thus, the self-employed individual can choose how much to work, and regulate the pace of work in any way and manner, and use impulses if necessary. We initially assume h is linear, but allow for r and δ to be different.

Example 6. Assume h linear and g power. If $r < \delta + \gamma(\phi - 1)$, then let $\rho = (\delta - r)/(\phi - 1)$, $\eta = [(\delta + \gamma\phi)/(r + \gamma)]^{1/(\phi-1)}$, and $L = (r + \gamma)\left[1 - e^{-(r-\rho)T}\right]/(r - \rho) + \eta e^{-(r-\rho)T}$; otherwise set $\eta = L = 1$. The solution to $\max_{X \in \overline{\mathscr{X}}} wP(X) - D(X)$ produces $D^*(K) = \frac{1}{\phi}\left[\eta^{\phi-1}(K + Y_0)^{\phi}/L^{\phi-1} - Y_0^{\phi}\right]$ and[10]

$$K^* = w^{1/(\phi-1)}L/\eta - Y_0, \quad \text{provided } w \geq \eta^{\phi-1}Y_0^{\phi-1}.$$

Observe that K^* and Y_0 are perfect substitutes: if Y_0 were to increase by one unit, the individual would optimally decrease the initial impulse by one unit. The elasticity of $K^* + Y_0$ with respect to wage is constant and equal to $1/(\phi - 1)$. Hence, K^* is concave in w for $\phi > 2$, linear for $\phi = 2$, and convex for $1 < \phi < 2$.

This picture changes when productivity decays with fatigue. To retain tractability, let $h(Y) = Y - \varphi g(Y)$, $\varphi \geq 0$, g power, and set $r = \delta$. Then, we can rewrite the objective as $[w/(1 + \varphi w)]\int_0^T e^{-rt}dX(t) - \int_0^T e^{-rt}g'(Y(t))dX(t)$. Thus, the solution is identical, provided we replace w by a productivity-adjusted wage of $w/(1 + \varphi w)$. The optimal labor supply is

$$K^* = \left[\frac{w}{1 + \varphi w}\right]^{1/(\phi-1)}\frac{L}{\eta} - Y_0, \quad \text{provided } w \geq \frac{\eta^{\phi-1}Y_0^{\phi-1}}{1 - \varphi\eta^{\phi-1}Y_0^{\phi-1}}.$$

Thus, the elasticity of $K^* + Y_0$ with respect to wage is now decreasing, and equal to $1/[(\phi - 1)(1 + \varphi w)]$. As before, labor-wage elasticity is independent of γ. Labor supply K^* is less responsive to w, and is guaranteed to be eventually concave.

Remark 4. Let $\varphi > 0$. For $\phi \geq 2$, K^* is strictly concave in w. For $1 < \phi < 2$, K^* is an S-shaped function of w, and the inflection point is $\tilde{w} = (2 - \phi)/[2\varphi(\phi - 1)] > 0$.

[10] If $r < \delta + \gamma(\phi - 1)$, then the optimal profile is $x^*(t) = k_1 \cdot \mathbf{1}_{t=0} + k(t) \cdot 1_{t \in (0,T)} + k_2 \cdot \mathbf{1}_{t=T}$, where $k_1 = (K + Y_0)/L - Y_0$, $k(t) = (K + Y_0)(\gamma + \rho)e^{\rho t}/L$ and $k_2 = (K + Y_0)(\eta - 1)e^{\rho T}/L$; otherwise all effort is provided at time 0.

7. The Time Trial Problem: Theory and Calibration

In a time trial, a solo athlete seeks to minimize the time it takes to cover a fixed distance, d. The athlete needs to economize the energy resources to produce external power. In particular, the athlete experiences fatigue, namely, past effort reduces the rate at which energy burned in the muscles is transformed into external power. We formalize the time trial problem as follows.

7.1. The Energy-Speed Model

The control variable of the athlete is $x(t)$, the internal burning rate of energy by the muscles. The athlete faces two constraints. First, $x(t)$ cannot exceed an upper bound c. Second, the athlete's total energy expenditure at time t must not exceed $k_0 + k_1 t$, where $k_0 > 0$ is an initial buffer of energy (e.g., ATP and glycogen in anaerobic processes), and $0 \leq k_1 < c$ is a replenishment rate (e.g., oxygen to support aerobic processes). The athlete can maintain the maximum effort rate c as long as $ct \leq k_0 + k_1 t$, i.e., only for a maximum time $\overline{T} = k_0/(c - k_1)$.

Let $v(t)$ denote the athlete's speed. The *time trial problem* is

$$\min_{X \in \mathscr{X}_c} T, \ s.t. \int_0^t x(s)ds \leq k_0 + k_1 t \ \text{ and } \int_0^T v(t)dt = d.$$

For this problem to be well-defined, we need to specify how $v(t)$ varies with $x(t)$. Following (7), we assume the internal burning rate $x(t)$ translate into an external power rate rate of $h'(Y(t))x(t)$, where h is a productivity function, and $Y(t) = \int_0^t e^{-\gamma(t-s)} x(s)ds$ is the fatigue level at time t. Note that $Y_0 = 0$ because the athlete begins afresh. For concreteness, let $h(Y) = \varrho\left(1 - e^{-Y/\varrho}\right)$, implying that efficiency $h'(Y) = e^{-Y/\varrho} \in [0,1]$ decays exponentially with fatigue. The condition $\varrho > c/\gamma$ ensures that $h''(Y)Y$ is strictly decreasing.

Power produces speed by counteracting the drag against the environmental fluid such as air or water, or the friction with the ground. The relationship between speed and power is generally non-linear. Because the athlete's speed during the race takes a narrow set of values we estimate a local linear approximation, $v(t) = a + bh'(Y(t))x(t)$.

If the athlete sets the maximum rate $x(t) = c$, then fatigue gradually increases according to $Y(t) = \int_0^t e^{-\gamma(t-s)} c ds = (c/\gamma)(1 - e^{-\gamma t})$, speed gradually decreases according to $v(t) = a + be^{-Y(t)/\varrho}c$, and the total distance covered by the time the energy buffer is exhausted is

$$\overline{d} = a\overline{T} + bc \int_0^{\overline{T}} h'\left((c/\gamma)\left(1 - e^{-\gamma t}\right)\right) dt.$$

If the race is short, $d \leq \overline{d}$, then the myopic solution is optimal. Henceforth, we focus on non-short races, $d > \overline{d}$, where the athlete ought to be strategic and economize energy.

We transform the time trial problem into a disutility minimization problem. First, we use the linearity of $v(t)$ to replace the constraint $\int_0^T v(t)dt = d$ for $\int_0^T h'(Y(t))x(t)dt = (d - aT)/b$. Next, we make the educated guess (which we later verify) that the energy constraint is binding only at $t=T$. The guess is reasonable unless d is very large. These two steps imply that the time trial problem is equivalent to first solve, for any $T>\overline{T}$, $\max_{X\in\mathscr{X}}\int_0^T h'(Y(t))x(t)dt$ subject to $0\le x(t)\le c$ and $\int_0^T x(t)dt=k_0+k_1T$, and then find the value of T for which the objective takes the value $(d-aT)/b$. Letting $\tilde{g}(Y)=Y-h(Y)$ we can equivalently solve, for any T,

$$\min_{X\in\mathscr{X}_c}\int_0^T \tilde{g}'(Y(t))x(t)dt,\ s.t.\ \int_0^T x(t)dt=k_0+k_1T, \tag{14}$$

then find the T for which the objective takes the value $(k_0+k_1T)-(d-aT)/b$; and verify that $\int_0^t x(s)ds\le k_0+k_1t$ holds for $t\le T$. After having successfully transformed the time minimization into a disutility minimization problem, we can apply Proposition 3 to conclude that *the solution to the time trial problem is a high-low-high effort profile.*

Thus, the athlete must begin releasing the maximum rate of energy. Passed this initialization period, which has a precise duration, the athlete must lower the internal burning rate so as to maintain the fatigue level and the external power rate constant, perhaps for a long time. If all goes according to plan, the energy reserve should be sufficient to set the internal burning rate at the maximum near completion. If the athlete sets a high pace for too long at the beginning, then his budget of energy will deplete too soon, causing him to inefficiently slow down later or shorten the termination period. The timing of when to initiate the final attack is also crucial.

7.2. Calibration with Swimming Data

As an empirical exercise, we seek data from a professional sport well suited for the calibration of our energy-speed model. An ideal one is swimming, where the physical and strategic interaction between athletes is minimal; and the external power rate translates into speed via the drag equation. Because each swimmer has his own lane, it is reasonable to think that the swimmer is a solo athlete facing a time trial. We admit that other factors may play a role (physical or psychological interactions between swimmers, the effect of the cheering of the public, etc.), but hope that their effect is small. Given that professional athletes devote a great amount of time and effort to optimize their performance under the guidance of experts, and the rewards of fame and riches are quite significant, we expect the observed profiles of effort to be close to optimal.

Baucells and Vangelis (2012) gathered data for swimming competitions from the website *www.omegatiming.com*, which includes data from the Olympics, FINA world cups and European. Swimming races are broken in laps of 50m, and partial times for each lap are made available. The only style with races exceeding 200m is freestyle swimming, for which the available races are 200m, 400m and 1500m for men; and 200m, 400m and 800m for women. Focusing attention on finals only, the data set consists of 60 races, 10 of each kind. In races of 200m (4 laps), the averages speeds in laps 2, 3, and 4 are not significantly different (lap 1 is faster, probably due to the jump start). This suggests that 200m may fall below the threshold $\overline{d}$. Using paired t-test for every two consecutive laps, Baucells and Vangelis (2012) confirm the high-low-high pattern in races of 400m or more for both men and women (see Figure 3).

Consistent with sustentation, the intermediate speeds are very steady. Indeed, for 1500m, the average speed from lap 3 to 28 stays within the narrow range of 1.65m/sec and 1.66m/sec; and for women 800m, the average speed from lap 3 to 15 stays within the narrow range of 1.56m/sec and 1.57m/sec. Consistent with termination, the speed increases significantly in the last lap in all races. Unfortunately, lap 1 needs to be disregarded due to the jump start, but evidence for initialization comes from lap 2, which is faster than lap 3 in 70% of the races. Consistent with prediction, the duration of the initialization and termination periods seems shorter for longer races.[11]

For given parameters $(\gamma, \varrho, k_0, k_1, c)$, we can solve (14) and calculate the optimal effort profile and the predicted instantaneous speeds. Conversely, we use the observed mid-lap speeds and completion times averaged over 10 races to calibrate the parameters for the representative swimmer. Appendix B describes the details of the calibration.

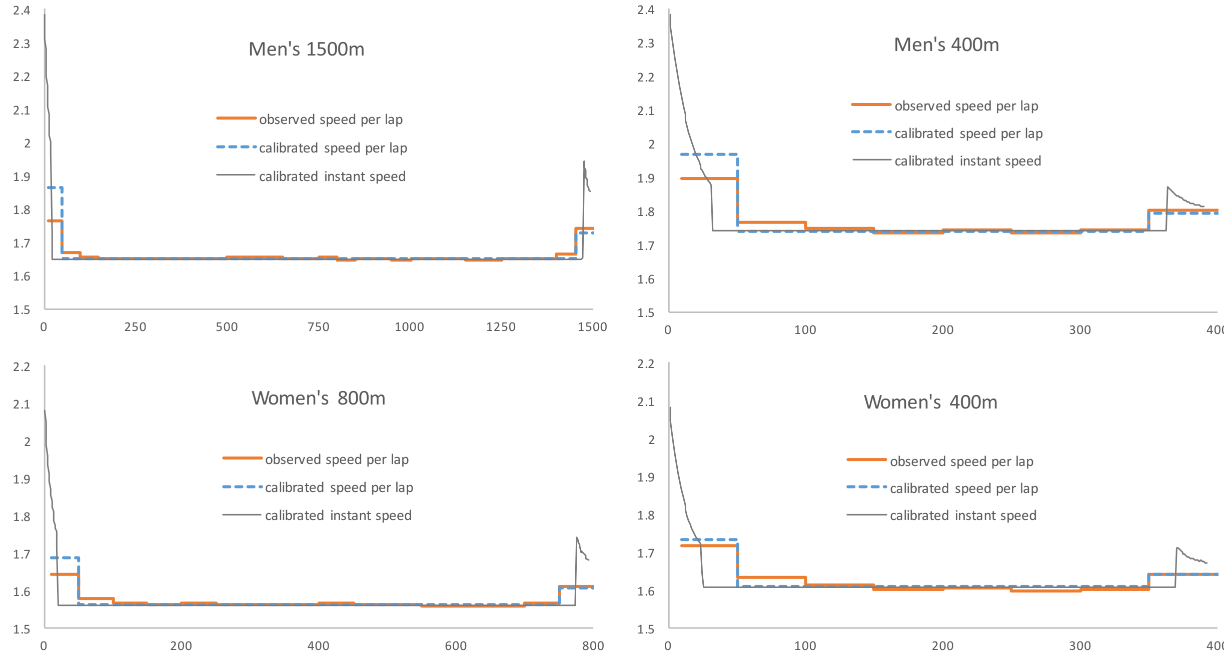
We obtain the following set of parameters. For men, $(\gamma, \varrho, k_0, k_1, c) = (0.133, 2249, 7996, 111, 168)$, yielding $\overline{d}$=259m. For women, $(\gamma, \varrho, k_0, k_1, c) = (0.161, 1208, 3521, 65.2, 89.3)$, yielding $\overline{d}$=246m. Using these parameters, the optimal time trial strategies are:

- Men 1500m: swim at full rate for 9.5 seconds at the beginning and at the end, covering a distance of 20.3m and 18.0m, respectively; and set the sustentation speed at 1.65m/s.
- Men 400m: swim at full rate for 15.1 seconds at the beginning and the end, covering a distance of 31.0m and 27.7m, respectively; and set the sustentation speed at 1.74m/s.
- Women 800m: swim at full rate for 10.1 seconds at the beginning and at the end, covering a distance of 18.8m and 17.2m, respectively; and set the sustentation speed at 1.56m/s.
- Women 400m: swim at full rate for 13.3 seconds at the beginning and the end, covering a distance of 23.4m and 22.3m, respectively; and set the sustentation speed at 1.61m/s.

In Figure 3 we compare the observed speeds per lap with the predicted speeds (instantaneous and per lap) using the calibrated parameters.

[11] In 400m for men, lap 2 is significantly faster than laps 4-5-6-7 in 7 out of 10 races; and for women in 9 out of 10 races. Lap 3 is also faster in 6 races for men and 8 races for women. Comparing laps 4-5, 5-6, and 6-7 over 10 races (30 pairs), the hypothesis of constant speed cannot be rejected in 24 out of 30 cases (for men) and 23 out of 30 (for women). Lap 8 is significantly faster than lap 7 in all 10 races for both men and women. For 1500m men, and all the 25 consecutive pairs 3-4, 4-5, ..., 26-27, 27-28, the differences in speed are not significantly different in 228 of the 250 pairs. In contrast, the difference between the 2nd and the 3rd lap is significant in 7 of 10 races with the 2nd lap being always faster than the 3rd. Lap 30 is significantly faster than lap 29 in all races, and lap 29 is faster than lap 28 in 4 of 10 races. For 800m women, and all the 12 consecutive pairs of laps 3-4, 4-5, ..., 13-14, 14-15, the differences in speeds are not significantly different in 117 of the 120 pairs. Lap 2 is significantly faster than lap 3 in 7 out of 10 races. Lap 16 is significantly faster than lap 15 in all 10 races.

Figure 3 Observed versus Calibrated Speeds [m/s]. Source: www.omegatiming.com



8. Concluding Remarks

The managerial literature on work and productivity has employed ad-hoc continuous-time models to capture the notion that fatigue increases cost and reduces productivity (Bechtold et al. 1984). The economics literature uses discrete-time model of fatigue, which cannot be readily applied to resolve practical sequencing or scheduling problems; or uses the time separable model. In this paper, we begin with the preference framework of Hindy et al. (1992), impose four appealing properties, and derive the fatigue disutility model, and its extension to productivity.

The model addresses a serious shortcoming of the time separable model, namely, the disregard for the effect of past effort. The time separable model would predict, in the On/Off case, that only the total working time matters. If the total is fixed, then breaks play no role in reducing cost or increasing productivity. Similarly, the time separable model cannot easily explain the abrupt transitions from high to steady at initiation, and from steady to high near termination, observed in athletic competitions. The proposed fatigue model is a relatively minor modification of the time separable model that remedies this shortcoming. We can think of the time separable model as the limiting case of the fatigue model when the recovery rate of fatigue is large.[12]

[12] In discrete-time, as the recovery rate increases, the individual enters each period fully rested, and past effort does not influence current disutility. To see this, set $Y_0 = 0$ and write $Y_i = (Y_{i-1} + k_{i-1})e^{-\gamma(t_i - t_{i-1})}$ and assume $t_i - t_{i-1} \geq \Delta$, for some $\Delta > 0$. Then, $Y_i \to 0$, $i = 1, \ldots, m$; and $D(X) \to \sum_{i=1}^{m} e^{-\delta t_i} g(k_i)$. In the case of effort rates, and assuming the effort rate is piecewise continuous, as the recovery rate increases, the fatigue level becomes small and converges to $x(t)/\gamma$ at all continuity points. For g power, we use $\gamma^{\phi-1} g'(x/\gamma)x = \phi g(x)$ to conclude that $\gamma^{\phi-1} D(X) \to \phi \int_0^T e^{-\delta t} g(x(t))dt$.

Table 2 Programs we have solved explicitly.

Baseline Program	Domain	Assumptions	Situations
$\min_{X(T)=k} D(X)$	$\mathscr{X}_c$	g quadratic, $Y_0 = 0$, $\delta = 0$	Example 1, §4
	$\mathscr{S}_n$	g quadratic, $Y_0 = 0$, $\delta = 0$	Example 2, §5
	$\overline{\mathscr{X}}$	g power, $Y_0 \geq 0$, $\delta \geq 0$	Example 5, §6
Variant Program			
$\max P(X)$	$\mathscr{S}_n$	g quadratic, h linear, $Y_0 = 0$, $\delta = r = 0$	Example 3, §5.1
$\max wP(X) - D(X)$	$\mathscr{S}_n$	g quadratic, h linear, $Y_0 = 0$, $\delta = r = 0$	Example 4, §5.3
	$\overline{\mathscr{X}}$	g power, h linear, $Y_0 \geq 0$, $\delta \geq 0$, any r g power, $h = Y - \varphi g$, $Y_0 \geq 0$, $\delta = r \geq 0$	Example 6, §6.2
$\min_{X(t) \leq k_0 + k_1 t} T$	$\mathscr{X}_c$	$h(Y) = \varrho\left(1 - e^{-Y/\varrho}\right)$, $Y_0 = 0$, $\delta = r = 0$ quadratic approximation for $Y - h(Y)$	Time trial, §7

We have considered three domains: bounded effort rates, $\mathscr{X}_c$; On/Off work with n breaks, $\mathscr{S}_n$; and general rates and impulses, $\overline{\mathscr{X}}$. We have also considered three objectives, namely, (i) minimize disutility $D(X)$ subject to a total effort requirement; (ii) maximize productivity $P(X)$; and (iii) maximize income minus effort cost, $wP(X) - D(X)$.

Our first central message is to provide and compare the optimal solution to the disutility minimization problem over the three domains. For $\mathscr{X}_c$ the effort rate is high-low-high, with initialization, sustentation, and termination periods (see Proposition 3). For $\mathscr{S}_n$ we have that having n breaks of equal duration during the workday, but away from the beginning and the end, is optimal (see Proposition 4). These two patterns are consistent with observation made in sports, including rowing, cycling, and running, as well as behavior of industry workers. Finally, for the idealized $\overline{\mathscr{X}}$ domain we have the impulse-rate-impulse solution (see Proposition 5).

Our second central messages is that, for each domain, the aforementioned structure applies to all three objectives, provided (i) $g''(Y)Y$ is strictly increasing, (ii) $h''(Y)Y$ is strictly decreasing, or (iii) $h''(Y)Y$ is decreasing and $g''(Y)Y$ is increasing, with at least one condition strict.

Table 2 summarizes the different programs we have solved in closed form or numerically. It is apparent from the table that closed-form solutions for $\mathscr{X}_c$ and $\mathscr{S}_n$ are only available for g quadratic and no discounting; but available for $\overline{\mathscr{X}}$ for g power and discounting. Thus, allowing for effort impulses is important for model tractability and insight.

To conclude, the fatigue disutility model produces sensible solutions to work scheduling problems, guidance for the design of services that require a supply of effort from the customer (e.g., educational programs), prescription for when to take breaks, and strategies for athletes under time trial. While the model is meant to be prescriptive, it also provides sensible rationalizations of observed preferences for delaying costly tasks. In short, the proposed fatigue-productivity model produces new insights, predictions, and prescriptions.

Acknowledgments

The authors benefited from comments at Informs Hawaii 2016. Manel Baucells acknowledges financial support from the Spanish Ministerio de Economía y Competitividad [Grant PSI2013-41909-P]. Lin Zhao acknowledges financial support from the National Center for Mathematics and Interdisciplinary Sciences, CAS, and National Science Foundation of China [Grant 71301161 and 71532013].

Appendix A. Proofs

Proof of Proposition 1. We trivially verify that $v(Y,x,t)=e^{-\delta t}g'(Y)x$ with $g'(Y)>0$ and $g''(Y)>0$ satisfies P1-P4. As for the converse, Hindy et al. (1992, Prop. 6) show that P1 implies the existence of two continuous functions $a(Y,t)$ and $b(Y,t)$ such that $D(X)=\int_0^T a(Y(t),t)dt+\int_0^T b(Y(t),t)x(t)dt$. Define $\bar{a}(Y,t)=\int_{Y_0}^{Y} a(y,t)dy$, $\bar{b}(Y,t)=\int_t^T b\left(Ye^{-\gamma(s-t)},s\right)ds$, and $g(Y,t)=\bar{a}(Y,t)+\bar{b}(Y,t)$. We first claim that for $X\in\mathscr{X}$,

$$D(X)=\bar{b}(Y_0,0)+\int_0^T g_Y(Y(t),t)x(t)dt.$$

This equation holds if $\int_0^T b(Y(t),t)dt=\bar{b}(Y_0,0)+\int_0^T \bar{b}_Y(Y(t),t)x(t)dt$. We calculate $\bar{b}_Y(Y,t)$ and $\bar{b}_t(Y,t)$, and verify that $b(Y,t)=\gamma Y\bar{b}_Y(Y,t)-\bar{b}_t(Y,t)$. Integrating over time, inserting $\gamma Y(t)dt=x(t)dt-dY(t)$, replacing $\int_0^T\left[\bar{b}_Y(Y(t),t)dY(t)+\bar{b}_t(Y(t),t)dt\right]=\bar{b}(Y(T),T)-\bar{b}(Y_0,0)$ and using $\bar{b}(Y(T),T)=0$, establishes the claim.

Next, we claim for any discrete-time profile $x(t)=\sum_{i=1}^m k_i\cdot\mathbf{1}_{t=t_i}$,

$$D(X)=\bar{b}(Y_0,0)+\sum_{i=1}^m[g(Y_i+k_i,t_i)-g(Y_i,t_i)].$$

Indeed, let $x_n(t)=\sum_{i=1}^m nk_i\cdot 1_{t\in[t_i,t_i+1/n]}$ and $Y_n(t)$ be the fatigue generated by $X_n(t)$. Then, $Y_n(t)$ is continuous on $[0,T]$ and converge in distribution to $Y(t)=Y_0e^{-\gamma t}+\sum_{i=1}^m k_ie^{-\gamma(t-t_i)}\cdot 1_{t\in[t_i,T]}$ as n increases. Since $x_n(t)=0$ outside $[t_i,t_i+1/n]$, we have

$$D(X_n)=\bar{b}(Y_0,0)+\sum_{i=1}^m\int_{t_i}^{t_i+1/n} g_Y(Y_n(t),t)x_n(t)dt.$$

Using $x_n(t)dt=dY_n(t)+\gamma Y_n(t)dt$ and $dg=g_YdY+g_tdt$ produces $\int_{t_i}^{t_i+1/n} g_Y(Y_n(t),t)x_n(t)dt=g\left(Y_n\left(t_i+1/n\right),t_i+1/n\right)-g(Y_n(t_i),t_i)+\int_{t_i}^{t_i+1/n}[\gamma g_Y(Y_n(t),t)Y_n(t)-g_t(Y_n(t),t)]dt$. The claim then follows from P1. As Y_0 is fixed, we normalize $D(\mathbf{0})=0$ by setting $\bar{b}(Y_0,0)=0$.

By continuity, P4 also applies to discrete profiles. Therefore, $\sum_{i=1}^m[g(Y_i+k_i,t_i)-g(Y_i,t_i)]$ is an order-preserving transformation of $\sum_{i=1}^m[g(Y_i+k_i,t_i+\Delta)-g(Y_i,t_i+\Delta)]$, provided Y_0 in the second case is set to $Y_0e^{\gamma\Delta}$ to ensure that the fatigue levels are the same. Replacing $Y_i+k_i=Y_{i+1}e^{\gamma(t_{i+1}-t_i)}$

shows that $D(X)$ is additive over $(Y_1,...,Y_m)$. By Fishburn (1970, p. 54), two additive preferences are identical if and only if one of them can be expressed as a positive linear transformation of the other. Therefore, for $f(Y,k,t)=g(Y+k,t)-g(Y,t)$, there exist two functions $A(\Delta)$ and $B(t,\Delta)$ such that $f(Y,k,t+\Delta)=A(\Delta)f(Y,k,t)+B(t,\Delta)$. Setting $k=0$ implies $B(t,\Delta)=0$ and $f(Y,k,t+\Delta)=A(\Delta)f(Y,k,t)$. By Aczél (1966, p. 141), the unique solution is $f(Y,k,t)=f(Y,k)e^{-\delta t}$ for some function $f(Y,k)$ and a constant δ. Let $g(Y)=f(0,Y)$ to conclude $g(Y,t)=e^{-\delta t}g(Y)$.

Finally, P2 and P3 imply $g'(Y)>0$ and $g''(Y)>0$. □

Proof of Proposition 2. (i) is obvious. (ii) Let $y=Y_0e^{-\gamma t}$. Then,

$$\frac{\partial\beta(k,t)}{\partial k}=\frac{\gamma y g'(y+k)}{g(y+k)-g(y)}\left[\frac{g''(y+k)}{g'(y+k)}-\frac{g''(y+\theta k)}{g'(y+\theta k)}\right]<0,$$

where $\theta\in(0,1)$ follows from the mean value theorem. (iii) and (iv) is equivalent to prove that $H(y)=\ln\left(\frac{y(g'(y+k)-g'(y))}{g(y+k)-g(y)}\right)$ is strictly increasing in y. Note that

$$H'(y)=\frac{g''(y+k)-g''(y)}{g'(y+k)-g'(y)}+\frac{1}{y}-\frac{g'(y+k)-g'(y)}{g(y+k)-g(y)}.$$

Assuming $g''(y)y/g'(y)$ increasing, $\left[\ln\left(\frac{g''(y)y}{g'(y)}\right)\right]'=\frac{1}{y}+\frac{g'''(y)}{g''(y)}-\frac{g''(y)}{g'(y)}\geq 0$, and

$$H'(y)\geq\frac{g'''(y+\theta_1k)}{g''(y+\theta_1k)}-\frac{g'''(y)}{g''(y)}+\frac{g''(y)}{g'(y)}-\frac{g''(y+\theta_2k)}{g'(y+\theta_2k)},\ \theta_1,\theta_2\in(0,1).$$

Hence, if $g''(y)/g'(y)$ is strictly decreasing in y and $g'''(y)/g''(y)$ is increasing in y, then $H'(y)>0$. The cost function $g(y)=y^\phi/\phi$ satisfies the conditions on $g''(y)y/g'(y)$, $g''(y)/g'(y)$ and $g'''(y)/g''(y)$ when $\phi\leq 2$. For $\phi>2$, $H(y)=\ln\phi+\ln\left[\frac{(1+k/y)^{\phi-1}-1}{(1+k/y)^{\phi}-1}\right]$ is also increasing in y. □

Proof of Proposition 3. We include time value and generalize (8) to

$$\min_{X\in\mathscr{X}}D(X)=\int_0^T e^{-\delta t}g'(Y(t))dX(t),\ s.t.\ 0\leq x(t)\leq c\text{ and }\int_0^T e^{-rt}x(t)dt=K.$$

Our goal is to replace $dX(t)$ for dt. Use $dX(t)=dY(t)+\gamma Y(t)$ and integration by parts to obtain

$$D(X)=e^{-\delta T}g(Y(T))-g(Y_0)+\int_0^T e^{-\delta t}\hat{g}(Y(t))dt,$$

where $\hat{g}(Y)=\delta g(Y)+\gamma g'(Y)Y$. The requirement that $\delta\geq 0$ and $g''(Y)Y$ is strictly increasing ensures that $\hat{g}(Y)$ is strictly convex. As both $g(Y)$ and $\hat{g}(Y)$ are strictly convex in Y, and Y is linear in X, $D(X)$ is strictly convex in X. Because the feasible domain is compact and the objective function is strictly convex, the solution exists and is unique (Bank and Riedel 2000, Theorem 3.3). Following Bank and Riedel (2000, Theorem 4.2), let

$$\mathcal{M}(t;Y)=g'(Y(T))e^{-\delta T-\gamma(T-t)}+\int_t^T\hat{g}'(Y(s))e^{-\delta s-\gamma(s-t)}ds,$$

which can be rewritten as $\mathcal{M}(t;Y)=g'(Y(t))+\int_t^T g''(Y(s))e^{-\gamma(s-t)}x(s)ds$. Then, x^* is optimal if and only if for some $M\geq 0$:

(i) $\int_0^T e^{-rt}x^*(t)d(t) = K$;

(ii) $\mathcal{M}(t;Y^*) \geq Me^{-rt}$ if $x^*(t) = 0$;

(iii) $\mathcal{M}(t;Y^*) = Me^{-rt}$ if $0 < x^*(t) < c$;

(iv) $\mathcal{M}(t;Y^*) \leq Me^{-rt}$ if $x^*(t) = c$.

We verify that the proposed solution satisfies (i)-(iv). Next, we characterize the solution parameters and boundaries. Consider the high-low-high solution structure proposed for $Y_0 \leq \overline{Y}$ based on $0 < t_1^* < t_2^* < T$. Applying (iii) using $r = \delta = 0$ yields $\mathcal{M}(t;Y^*) = M$ and $Y^* = \hat{g}'^{-1}(\gamma M)$. The equation $Y(t_1) = Y_0 e^{-\gamma t_1} + (c/\gamma)(1 - e^{-\gamma t_1}) = Y^*$ determines t_1^*, and is equivalent to $t_1^* = \Delta(Y^*) - \Delta(Y_0)$. The equation $\mathcal{M}(t_2;Y^*) = M$ with $Y(t) = Y^* e^{-\gamma(t-t_2)} + (c/\gamma)(1 - e^{-\gamma(t-t_2)})$, $t \in [t_2, T]$, determines t_2^*, and it is equivalent to $t_2^* = T - [\Delta(Y_f(Y^*)) - \Delta(Y^*)]$. Inserting $x^*(t)$ into (i) determines M. As a result, we obtain that $Y^* > 0$ is the solution to

$$[\Delta(Y) - \Delta(Y_0)]c + [T - \Delta(Y_f(Y)) + \Delta(Y_0)]\gamma Y + [\Delta(Y_f(Y)) - \Delta(Y)]c = K.$$

Setting $Y_0 = Y$ yields the equation that determines $\overline{Y}$.

For $Y_0 > \overline{Y}$, consider the proposed break-low-high profile. Clearly, $Y^* = Y_0 e^{-\gamma t_1^*}$. Let $t_2(t_1) = T - [\Delta(Y_f(Y_0 e^{-\gamma t_1})) - \Delta(Y_0 e^{-\gamma t_1})]$. By (i), we conclude that t_1^* solves

$$[t_2(t_1) - t_1]\gamma Y_0 e^{-\gamma t_1} + [T - t_2(t_1)]c = K.$$

Finally, setting Y_0 equal to the solution to $\Delta(Y_f(Y e^{-\gamma(T-K/c)})) - \Delta(Y e^{-\gamma(T-K/c)}) = K/c$ sets $t_2(t_1) = t_1 = T - K/c$ and determines $\overline{\overline{Y}}$. □

Proof of Proposition 4. It is obvious that $\ell_0^* = 0$ and $\ell_{n+1}^* = 0$. Otherwise, moving the first working interval to sooner or the last working interval to later decreases disutility. Next, we argue that $(\ell_1^*, \ell_2^*, ..., \ell_n^*) \gg 0$. Assume we use less than n breaks. The sequence will contain a work-rest-work episode $(s, \ell, s') \gg 0$ which starts with fatigue level Y. Consider the perturbation $(\hat{s}, \hat{\ell}, \hat{s}', \hat{\ell}', \hat{s}'') = (s - \varepsilon, \ell - \varepsilon - x, 2\varepsilon, x + \varepsilon, s' - \varepsilon)$. Choose x so that the fatigue level at time $s + \ell + \varepsilon$ is identical in both schedules. The marginal disutility created by the perturbation at $\varepsilon = 0$ equals $2g'\left(\frac{2\overline{Y}e^{-\gamma\ell}}{1+e^{-\gamma\ell}}\right) - g'\left(\overline{Y}\right) - g'\left(\overline{Y}e^{-\gamma\ell}\right)$, where $\overline{Y} = Ye^{-\gamma s} + \frac{c}{\gamma}(1 - e^{-\gamma s})$. Because $g''(Y)Y$ is strictly increasing and $g''(Y) > 0$, we verify that for some $\varepsilon > 0$, the perturbation is a strict improvement, contradicting that less than n breaks is optimal. This argument also establishes that $D_n^*(K)$ is strictly decreasing in n.

To obtain the optimal schedule, let Y_{s_j} and Y_{ℓ_j} denote the fatigue level at the end of s_j and ℓ_j, respectively. Let $\hat{g}(y) = -c\int_{\frac{c}{\gamma}(1-y)}^{\frac{c}{\gamma}} \frac{g'(c/\gamma) - g'(x)}{c - \gamma x} dx$. We use

$$c\int_0^{s_j} g'\left(Y_{\ell_{j-1}}e^{-\gamma s} + \frac{c}{\gamma}\left(1 - e^{-\gamma s}\right)\right) ds = cg'\left(\frac{c}{\gamma}\right)s_j - \left[\hat{g}\left(\frac{\gamma}{c}\left(\frac{c}{\gamma} - Y_{s_j}\right)\right) - \hat{g}\left(\frac{\gamma}{c}\left(\frac{c}{\gamma} - Y_{\ell_{j-1}}\right)\right)\right],$$

to translate (11) into

$$\min_{Y_{s_j},Y_{\ell_j}} D_n^*(K) = g'\left(\frac{c}{\gamma}\right)K + \hat{g}(1) - \sum_{j=1}^{n+1}\hat{g}\left(\frac{\gamma}{c}\left(\frac{c}{\gamma}-Y_{s_j}\right)\right) + \sum_{j=1}^{n}\hat{g}\left(\frac{\gamma}{c}\left(\frac{c}{\gamma}-Y_{\ell_j}\right)\right)$$

$$s.t. \prod_{j=1}^{n} Y_{\ell_j} = e^{-\gamma(T-K/c)}\prod_{j=1}^{n} Y_{s_j},\ \prod_{j=1}^{n+1}\left(\frac{c}{\gamma}-Y_{s_j}\right) = \frac{c}{\gamma}e^{-\gamma K/c}\prod_{j=1}^{n}\left(\frac{c}{\gamma}-Y_{\ell_j}\right).$$

To ensure well-posedness, we introduce two conditions:

(i) $\forall a,b>0,\ d\in\mathbb{R},\ \frac{\hat{g}'(y)-\hat{g}'\left(a\left(1-\frac{d}{a-y}\right)\right)\frac{ad}{(a-y)^2}}{1-\frac{ad}{(a-y)^2}} - \hat{g}'\left(by\left(1-\frac{d}{a-y}\right)\right)b$ never switches sign;

(ii) $\forall a,b\in(0,1),\ \min_{y\in(0,1)} -n\hat{g}(y)+n\hat{g}\left(1-a(1-y)\right)-\hat{g}\left(b\left(\frac{1-a}{y}+a\right)^n\right)$ has a unique solution.

Condition (i) ensures that any deviation from $Y_{s_1}=...=Y_{s_n}$ and $Y_{\ell_1}=...=Y_{\ell_n}$ cannot be optimal. Based on this symmetry, problem (11) further simplifies to

$$\min_{Y_{s_1}\in\left[0,\frac{c}{\gamma}\left(1-e^{-\gamma K/c}\right)\right]} \ -n\hat{g}\left(\frac{\gamma}{c}\left(\frac{c}{\gamma}-Y_{s_1}\right)\right)+n\hat{g}\left(\frac{\gamma}{c}\left(\frac{c}{\gamma}-e^{-\gamma(T-K/c)/n}Y_{s_1}\right)\right)$$
$$-\hat{g}\left(e^{-\gamma K/c}\left(\frac{c/\gamma-e^{-\gamma(T-K/c)/n}Y_{s_1}}{c/\gamma-Y_{s_1}}\right)^n\right).$$

Condition (ii) ensures that $Y_{s_1}^*$ is unique. After obtaining $Y_{s_1}^*$, we have $Y_{\ell_1}^*=e^{-\gamma(T-K/c)/n}Y_{s_1}^*$, $Y_{s_{n+1}}^* = c/\gamma - e^{-\gamma K/c}\left(\frac{c/\gamma-Y_{\ell_1}^*}{c/\gamma-Y_{s_1}^*}\right)^n$. Accordingly, $s_1^*=\Delta(Y_{s_1}^*)$, $s_2^*=...=s_n^*=\Delta(Y_{s_1}^*)-\Delta(Y_{\ell_1}^*)$, and $s_{n+1}^* = \Delta(Y_{s_{n+1}}^*)-\Delta(Y_{\ell_1}^*)$.

Finally, to see that $D_n^*(K)$ tends to $D^*(K)$, let $x^*(t)=c\cdot 1_{t\in\left[0,t_1^*\right]}+\gamma Y^*\cdot 1_{t\in\left(t_1^*,t_2^*\right)}+c\cdot 1_{t\in\left[t_2^*,T\right]}$ be the profile achieving $D^*(K)$, and $x_n(t)=c$ on $[0,t_1^*]\cup(T_{j1},T_{j2})|_{j=1}^n\cup[t_2^*,T]$ and $x_n(t)=0$ otherwise, where $T_{j1}=t_1^*+(j-1)(t_2^*-t_1^*)/n$, $T_{j2}=t_1^*+(j-1+\gamma Y^*/c)(t_2^*-t_1^*)/n$. By continuity, $X_n\to X^*$ as $n\to\infty$ and $D(X_n)\to D(X^*)=D^*(K)$. The result then follows from $D(X_n)\ge D_n^*(K)>D^*(K)$. □

Proof of Proposition 5. We include time value and generalize (13) to

$$\min_{X\in\overline{\mathscr{X}}} D(X),\ s.t.\ \int_0^T e^{-rt}dX(t)=K.$$

As in Proposition 3, the feasible domain is compact and the objective function is strictly convex, and hence the solution exists and is unique. Following Bank and Riedel (2000, Theorem 4.2), $X^*\in\overline{\mathscr{X}}$ is optimal if and only if for some $M\ge 0$:

(i) $\int_0^T e^{-rt}dX^*(t)=K$;

(ii) $\mathcal{M}\left(t;Y^*\right)\ge Me^{-rt}$, for all $t\in[0,T]$;

(iii) $\mathcal{M}\left(t;Y^*\right)=Me^{-rt}$ for all $t\in\text{support}(dX^*)$.

We confirm that the proposed solution satisfies (i) to (iii). Next, we characterize the solution parameters and boundaries. To simplify exposition, we illustrate the case of $r=0$ (for $r>0$ the same logic applies). If $r=0$ and the solution has full support, then (iii) becomes $g'(Y(T))e^{-(\gamma+\delta)T}+$

$\int_t^T \hat{g}'(Y(\tau))e^{-(\gamma+\delta)\tau}d\tau = Me^{-\gamma t}$, $t \in [0,T]$. Therefore, $\hat{g}'(Y(t)) = \gamma M e^{\delta t}$ for $t \in [0,T)$ and $g'(Y(T)) = Me^{\delta T}$. Setting $t=0$, $t \in (0,T)$ and $t \to T$, we have that

$$\begin{aligned} k_1 &= \hat{g}'^{-1}(\gamma M) - Y_0, \\ x(t) &= \gamma Y(t) + Y'(t), \text{ and} \\ k_2 &= g'^{-1}\left(Me^{\delta T}\right) - \hat{g}'^{-1}\left(\gamma M e^{\delta T}\right). \end{aligned}$$

Imposing $\lim_{t\to 0+} Y^*(t) = Y_0$ yields $\overline{Y}$ as the solution to $G(Y,T) = K + Y$, where

$$G(Y,T) = g'^{-1}\left(\hat{g}'(Y)e^{\delta T}/\gamma\right) + \gamma \int_0^T \hat{g}'^{-1}\left(\hat{g}'(Y)e^{\delta t}\right) dt.$$

For $Y_0 \le \overline{Y}$, we obtain $Y^* > 0$, $k_1 \ge 0$, $k(t)$, and $k_2 > 0$ as the solution to

$$\begin{aligned} &G(Y,T) = K + Y_0,\ k_1 = Y^* - Y_0,\ k(t) = \gamma Y^*(t) + Y^{*\prime}(t), \\ &Y^*(t) = \hat{g}'^{-1}\left(\hat{g}'(Y^*)e^{\delta t}\right), \text{ and } k_2 = g'^{-1}\left(\hat{g}'(Y^*)e^{\delta T}/\gamma\right) - Y^*(T^-). \end{aligned}$$

For $Y_0 > \overline{Y}$, the threshold $\overline{\overline{Y}}$ solves $G(Ye^{-\gamma T}, 0) = K + Ye^{-\gamma T}$. Finally, for $\overline{Y} < Y_0 < \overline{\overline{Y}}$, we have that t_1^* solves $G(Y_0 e^{-\gamma t}, T-t) = K + Y_0 e^{-\gamma t}$, yielding $k(t) = \gamma Y^*(t) + Y^{*\prime}(t)$, where

$$\begin{aligned} Y^*(t) &= \hat{g}'^{-1}\left(\hat{g}'(Y_0 e^{-\gamma t_1^*})e^{\delta(t-t_1^*)}\right), \text{ and} \\ k_2 &= g'^{-1}\left(\hat{g}'(Y_0 e^{-\gamma t_1^*})e^{\delta(T-t_1^*)}/\gamma\right) - Y^*(T^-). \end{aligned}$$

□

Appendix B. Calibration

We derive the optimal profile after applying two approximation, namely, $h'(Y) = e^{-Y/\varrho} \approx 1 - Y/\varrho$ and $v(t) \approx a + bE(t)$, where $E(t) = h'(Y(t))x(t)$. Then, problem (14) reduces to minimizing disutility with a quadratic cost function, whose solution is given by (10). Thus, $x^*(t) = c \cdot 1_{t\in[0,t^*]} + c(1 - e^{-\gamma t^*}) \cdot 1_{t\in(t^*, T-t^*)} + c \cdot 1_{t\in[T-t^*,T]}$, where

$$t^* = \frac{T^*}{2} - \frac{1}{\gamma} W\left(\frac{\gamma}{2}\left(T^* - \frac{k_0}{c} - \frac{k_1 T^*}{c}\right) e^{\gamma T^*/2}\right).$$

The athlete's fatigue level increases during $[0,t^*]$, stays constant on $(t^*, T-t^*)$, and further increases during $[T^* - t^*, T]$. Accordingly, the observable speed gradually decreases during $[0,t^*]$, settles at a lower but constant value on $(t^*, T-t^*)$, increases suddenly at $t = T^* - t^*$ to gradually decrease over the final meters. The speed at t^* matches the speed at $T^* - t^*$.

To calculate the speeds predicted by the optimal profile, however, we use the more precise equations $v(t) = \sqrt{2E(t)/B}$, with $E(t) = e^{-Y(t)/\varrho}x(t)$ and B equal to a drag coefficient times the density of the water (1000kg/m^3) times the cross sectional area of the athlete. We employ Toussaint et al. (1988, p. 435-437) estimates (cross sectional area 0.091m^2 for men and 0.075m^2 for women; and a drag coefficient of 0.64 for men and 0.54 for women), yielding $B = 58.2\text{kg/m}$ for men and $B = 40.5\text{kg/m}$ for women. Thus, the predicted speeds are

$$v^*(t) = \begin{cases} e^{-c\left(1-e^{-\gamma t}\right)/(2\gamma\varrho)}\sqrt{2c/B}, & 0 \le t \le t^*, \\ e^{-c\left(1-e^{-\gamma t^*}\right)/(2\gamma\varrho)}\sqrt{2c\left(1-e^{-\gamma t^*}\right)/B}, & t^* < t < T^* - t^*, \\ e^{-c\left[1-e^{-\gamma(t-T^*+2t^*)}\right]/(2\gamma\varrho)}\sqrt{2c/B}, & T^* - t^* \le t \le T^*, \end{cases}$$

where T^* solves $\int_0^{T^*} v^*(t)dt = d$.

We calibrate the five unknown parameters $(\gamma, \varrho, k_0, k_1, c)$ separately for men and women in three steps. First, note that $v^*(0) = \sqrt{2c/B}$ is independent of ϱ, γ, k_0 and k_1. Next, we guess estimate the initial speed v_0^o and calculate c using $\sqrt{2c/B} = v_0^o$. Noting that $v^*(t)$ depends on $(t^*, \gamma, \varrho, c)$, and that c is already available, the next goal is to estimate the four parameters $(t_1^*, t_2^*, \gamma, \varrho)$, where the subscript 1 refers to 1500m for men and 800m for women, and subscript 2 refers to 400m.

To calibrate the parameters we use the observed mid-lap speeds in m/s and completion times in seconds, averaged over the 10 races of each type, namely

- $(v^o_{\text{mid}1}, v^o_{\text{mid}2}, T_1^o, T_2^o) = (1.65, 1.74, 899, 221)$ for men, and
- $(v^o_{\text{mid}1}, v^o_{\text{mid}2}, T_1^o, T_2^o) = (1.56, 1.61, 504, 241)$ for women.

To remove the effect of the jump start we assume for men that the first dive takes 5sec and lasts for 10m; and for women that the first dive takes 4sec and lasts for 8m. Thus, the effective length of the races is 1490m and 390m for men, and 792m and 392m for women.

The four parameters and four observations are linked by the equations:

$$v^*(t_1^*) = v^o_{\text{mid}1}, \;\; v^*(t_2^*) = v^o_{\text{mid}2}, \;\; \int_0^{T_1^o} v^*(t)dt = d_1 \;\; \text{and} \;\; \int_0^{T_2^o} v^*(t)dt = d_2.$$

We solve this system and get $(t_1^*, t_2^*, \gamma, \varrho)$. Next, we obtain (k_0, k_1) by solving the system

$$t_i^* = T_i^o/2 - W\left(\gamma\left(T_i^o - k_0/c - k_1 T_i^o/c\right) e^{\gamma T_i^o/2}/2\right)/\gamma, \;\; i = 1, 2.$$

Finally, because v_0^o is not directly available in our data, we vary v_0^o from 1.80m/s to 2.50m/s and repeat the above procedure until we minimize the squared differences between the predicted speed and the observed speed in each lap j and each race i, measured by $\sum_j \sum_i |\hat{v}_{ji} - v^o_{ji}|^2$. The estimated values are $v_0^o = 2.38$m/s for men and $v_0^o = 2.08$m/s for women, yielding a calibration residual very small, of the order of 10^{-3}.

We verify that ϱ is well above the fatigue upper bound c/γ for all races, justifying $h'(Y) \approx 1 - Y/\varrho$. We also check that the energy budget is not binding for $t < T$. Finally, for 1500m men, the difference between $v(t) = a + bE(t)$ and $\sqrt{2E(t)/B}$ is at most 1.7% in the predicted range of speeds; and for the rest of races the difference is 1.2% or less.

References

Abbiss CR, Laursen PB (2008) Describing and understanding pacing strategies during athletic competition. *Sports Medicine* 38(3):239–252.

Aczél J (1966) *Lectures on functional equations and their applications* (New York: Academic Press).

Augenblick N, Niederle M, Sprenger C (2015) Working over time: Dynamic inconsistency in real effort tasks. *The Quarterly Journal of Economics* 130(3):1067–1115.

Bank P, Riedel F (2000) Non-time additive utility optimization –the case of certainty. *Journal of Mathematical Economics* 33(3):271–290.

Baucells M, Sarin RK (2007) Satiation in discounted utility. *Operations Research* 55(1):170–181.

Baucells M, Vangelis T (2012) Effort allocation in competitive swimming. Working paper.

Bechtold SE, Janaro RE, Sumners DWL (1984) Maximization of labor productivity through optimal rest-break schedules. *Management Science* 30(12):1442–1458.

Bechtold SE, Sumners DL (1988) Noteoptimal work-rest scheduling with exponential work-rate decay. *Management Science* 34(4):547–552.

Bleichrodt H, Gao Y, Rohde KI (2016) A measurement of decreasing impatience for health and money. *Journal of Risk and Uncertainty* 52(3):213–231.

Brachet T, David G, Drechsler AM (2012) The effect of shift structure on performance. *American Economic Journal: Applied Economics* 4(2):219–246.

Dragone D (2009) I am getting tired: Effort and fatigue in intertemporal decision-making. *Journal of Economic Psychology* 30(4):552–562.

Fehr E, Goette L (2007) Do workers work more if wages are high? Evidence from a randomized field experiment. *American Economic Review* 97(1):298–317.

Fishburn PC (1970) *Utility theory for decision making* (Operations Research Society of America).

Frederick S, Loewenstein G, O'donoghue T (2002) Time discounting and time preference: A critical review. *Journal of Economic Literature* 40(2):351–401.

Fudenberg D, Levine DK (2006) A dual self model of impulse control. *American Economic Review* 96(5):1449–1476.

Garland S (2005) An analysis of the pacing strategy adopted by elite competitors in 2000 m rowing. *British Journal of Sports Medicine* 39(1):39–42.

Hagger MS, Wood C, Stiff C, Chatzisarantis NL (2010) Ego depletion and the strength model of self-control: a meta-analysis. *Psychological Bulletin* 136(4):495–525.

Halevy Y (2015) Time consistency: Stationarity and time invariance. *Econometrica* 83(1):335–352.

Hawkins D, Hull M (1993) Muscle force as affected by fatigue: Mathematical model and experimental verification. *Journal of Biomechanics* 26(9):1117–1128.

Hindy A, Huang CF, Kreps D (1992) On intertemporal preferences in continuous time: The case of certainty. *Journal of Mathematical Economics* 21(5):401–440.

Hunter EM, Wu C (2016) Give me a better break: Choosing workday break activities to maximize resource recovery. *Journal of Applied Psychology* 101(2):302–311.

Kahneman D, Wakker PP, Sarin RK (1997) Back to bentham? Explorations of experienced utility. *Quarterly Journal of Economics* 112(2):375–405.

Kanai A (2009) "karoshi (work to death)" in Japan. *Journal of Business Ethics* 84(2):209–216.

Kc DS, Terwiesch C (2009) Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science* 55(9):1486–1498.

Loewenstein G, Thaler RH (1989) Anomalies: Intertemporal choice. *The Journal of Economic Perspectives* 3(4):181–193.

McMenamin TM (2007) A time to work: recent trends in shift work and flexible schedules. *Monthly Lab. Rev.* 130(12):3–15.

Meijman TF, Mulder G, Drenth P, Thierry H (1998) Psychological aspects of workload. *Handbook of Work and Organizational Psychology. Volume* 2.

Muraven M, Tice DM, Baumeister RF (1998) Self-control as a limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology* 74(3):774–789.

Ozdenoren E, Salant SW, Silverman D (2012) Willpower and the optimal control of visceral urges. *Journal of the European Economic Association* 10(2):342–368.

Samuelson P (1937) A note on measurement of utility. *Review of Economic Studies* 4:155–161.

Thaler RH (1981) Some empirical evidence on time inconsistency. *Review of Economic Studies* 23:165–180.

Toussaint H, De Groot G, Savelberg H, Vervoorn K, Hollander A, van Ingen Schenau G (1988) Active drag related to velocity in male and female swimmers. *Journal of Biomechanics* 21(5):435–438.

Veasey S, Rosen R, Barzansky B, Rosen I, Owens J (2002) Sleep loss and fatigue in residency training. *JAMA: The Journal of the American Medical Association* 288(9):1116–1124.

Vernon HM (1921) *Industrial fatigue and efficiency* (G. Routledge).