

Self-paced rest decisions: stable but ineffective for recovery

Nastasia Klevak, Joshua W. Buckholtz, and Russell A. Poldrack

Department of Psychology, Stanford University

{nklevak,buckholtz,poldrack}@stanford.edu

Abstract

Rest is critical for fatigue recovery following periods of cognitive effort. Yet, the inter and intra-individual dynamics of rest decisions—and how they relate to performance—remain poorly characterized. In the present study, we analyze two experimental datasets in which participants alternate between epochs of two cognitively effortful tasks. After each epoch, subjects were given a costly and self-paced rest opportunity (performing a low-effort task). Using predictive modeling and latent profile analysis, we find that subjects' decisions to rest are stable and auto-correlated, potentially reflecting an individualized value policy rather than tracking in-task performance dynamics. By integrating a time-varying hidden markov model fit on within-subject performance metrics, we also find that distinct subject profiles have significantly different state occupancy patterns, and that, across all profiles, resting for longer does not improve subsequent performance. Overall, costly rest decisions appear to be a stable individual preference that is neither driven by nor beneficial to task performance.

Introduction

Cognitive effort expenditure is crucial for adaptive behavior, yet costly (Boksem & Tops, 2008; Inzlicht et al., 2018; Kool et al., 2010; Shenhav et al., 2017), with cognitive fatigue a subjective "read-out" of this cost (Bijleveld, 2023; Matthews et al., 2023). Fatigue is associated with worse performance (Boksem & Tops, 2008), and subjectively increases following performance errors (Matthews et al., 2023; Xu et al., 2023). Recent work has used periodic rest—pre-determined or self-paced—as an intervention to reduce cognitive fatigue or improve performance in tasks, with mixed results: some find benefits (Patel et al., 2024) while others find no, or even detrimental, effects (Lim & Kwok, 2016; Perquin et al., 2020). Moreover, in the self-paced rest case, what drives rest decisions remains unclear. Ineffective self-paced rest decisions may reflect an inability to judge rest needs (Perquin et al., 2020), indicate preexisting subjective fatigue differences (Kanfer, 2011), or suggest that rest decisions are auto-correlated habits unaligned with task reward (Gershman, 2020). Here, we use computational modeling to explore the relationship between rest and recovery, and clarify whether participants' self-paced rest decisions reflect influences from performance dynamics versus stable preferences.

Methods

Data and experiment

The 55-minute experiment was conducted online (Prolific), and consisted of two high-effort tasks (spatial recall: 4 flashing squares, 4x4 grid; digit span: 4 flashing digits), and a lower-effort "rest task" (shape discrimination, no executive load). After practice, subjects played 10 total blocks of alternating tasks in a counter-balanced order. Each block comprised of three epochs of the corresponding high-effort task. After each (10 trial) epoch, subjects were presented with a cue (indicating the next epoch's task) and then proceeded to a self-paced rest (SPR) period consisting of up to 20 rest task trials (1.5s/trial). Participants were informed that rest was costly¹ and that they could "end rest" and return to the next high-effort epoch at any time. In total, participants completed 30 high-effort epochs and 30 SPR periods. For our analyses, we combined two experimental samples ("original": n=84 post-exclusion, "replication": n=103 post-exclusion)². On average, replication subjects performed worse on the high-effort tasks and chose to stay in rests for longer.

Modeling rest decisions

To identify features most predictive of subjects' rest-length decisions, we first trained Gradient Boosted Decision Trees (GBDT) and a Long Short Term Memory network (LSTM). Importantly, GBDTs capture non-linear feature relationships (from epoch-level summary features) while LSTMs use trial-level temporal dynamics (from subjects' 300-trial sequence)³. To isolate rest history's contribution and test auto-correlation, we trained a "no-rest" version of each model (only included performance and context features for each epoch/trial) and a "rest" version (adding the prior period's rest length and cumulative rest trials taken so far). This comparison,

¹Rest trials cost 1 point/trial (from a 600-point bank); leftover points converted into a bonus. No performance-based bonus.

²Participants with accuracies <40% or >95% on all three tasks were excluded: n=6 from original, n=24 from replication

³Training: 80% of subjects / dataset, test: 20%. GBDT: $lr = 0.05$, estimator number and tree depth were tuned via (subject-grouped) 15-fold cross-validation with early stopping. LSTM: (2 layers, 64 units/layer, dropout = 0.3, weight decay = 10^{-3} , batch size = 8), trained with Adam for 100 epochs with early stopping (patience = 15), and 20% of training subjects were held out as a validation set.



along with SHAP analyses, quantified how much of rest prediction depends on rest history versus task performance fluctuations.

Performance states across latent profiles

We then used latent profile analysis (LPA) in conjunction with a time-varying hidden markov model (TV-HMM)⁴ to investigate inter and intra-individual differences in rest decisions ($n=157$)⁵. The winning four-profile LPA model was estimated (Table 1a) using each subject’s summary rest (mean, SD, and slope) and performance (means and slopes of accuracy, RT, and RT SD) features. To characterize within-subject performance dynamics, a winning four-state TV-HMM was fit on within-subject z-scored emissions (epoch accuracy, RT, and RT SD) (Table 1b; covariates: epoch number). Finally, using mixed effects models, we analyzed the effect of state on subsequent rest length, and the effect of rest length on subsequent state improvement. Importantly, all models accounted for epoch number, game type, cue type, dataset, and had a random intercept for subject.

(A) LPA Profiles

Profile (n subjects)	DS acc	SR acc	RT m (SD)	Rest m (SD)
Good-LowRest (60)	0.922	0.800	2639 (376)	3.2 (2.2)
Good-HighRest (34)	0.931	0.836	2738 (351)	10.1 (6.1)
Okay-MaxRest (21)	0.768	0.585	3145 (390)	19.8 (0.9)
Bad-ModRest (42)	0.593	0.386	3402 (492)	10.0 (5.0)

(B) TV-HMM States

State	Acc (z)	RT (z)	RT SD (z)	Rest (raw)
S1 (Disengaged)	-0.792	0.843	0.830	10.7
S2 (Typical)	-0.307	0.329	0.181	8.9
S3 (Good)	0.518	-0.726	-0.444	9.0
S4 (Peak)	0.604	-0.086	-0.067	6.3

Table 1: (A) LPA profiles: DS/SR = digit span/spatial recall; RT (ms) and rest (# trials) entries are mean (mean SD) across epochs. Slopes omitted for simplicity. Profile names are formatted as performance-rest-amount. (B) TV-HMM within-subject z-scored emissions, trained with the LaMa package (Mews et al., 2025). Rest (mean # trials in associated epochs) was not part of the emission distribution.

Results and Discussion

Predictive accuracy was high for the rest-included models, and poor for the no-rest models ($GBDT_{rest}$: $R^2 = 0.74$, $GBDT_{no-rest}$: $R^2 = 0.17$ | $LSTM_{rest}$: $R^2 = 0.67$, $LSTM_{no-rest}$: $R^2 = 0.09$), indicating that rest-decisions are auto-correlated. SHAP analyses of rest-included models confirmed that predictions relied on prior rest and epoch number, not within-subject performance dynamics, and $LSTM_{rest}$ feature analysis furthered this finding: trial sequences added no predictive value beyond epoch-wide summaries, and hidden units corre-

lated with global performance traits instead of within-subject epoch-to-epoch performance.

The stable between-subject differences identified by LPA reinforce this finding: we found that Good-LowRest ($n=60$) and Good-HighRest ($n=34$) profiles have near-identical game performance but differ in their rest decisions (6.9 rest trials per SPR on average). The lowest-performing profile (Bad-ModRest) also rested less than the moderately performing profile (Okay-MaxRest). These distinct profiles, which also show significantly different TV-HMM state occupancy over time (Figure 1), corroborate that rest decisions reflect stable individual differences rather than compensatory patterns. Additionally, using the TV-HMM states, we found no ev-

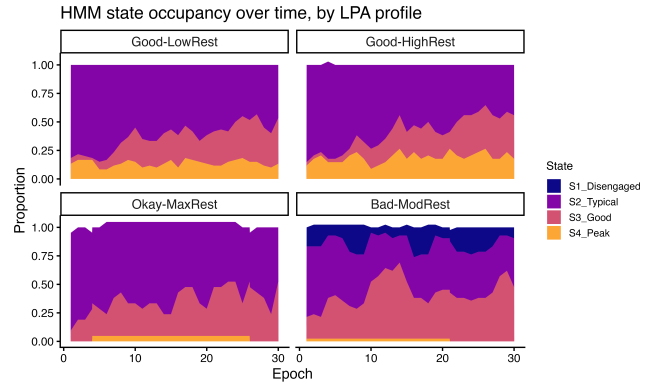


Figure 1: **Proportion of TV-HMM states occupied by each LPA profile group over time.** These profile occupancy patterns differ significantly ($\chi^2(9) = 972.91$, $p < .001$).

idence that longer rests lead to performance improvements, with (glmer, $z = -0.984$, $p = 0.325$, $BF_{01} = 8.86$) or without⁶ ($z = -0.031$, $p = 0.302$, $BF_{01} = 793$) adding profile as an interaction. We also found little evidence of a general compensatory relationship between lower performance and more rest, although there were profile differences. The effect of state on rest (using lmer) was significant ($F(3, 4631.6) = 4.25$, $p = 0.005$), but a Bayes Factor analysis found inconclusive evidence ($BF_{01}=1.26$) and pairwise comparisons showed that the significance was from a single contrast: $S2_{typical} - S3_{good}$ ($b = 0.60$, $p = 0.003$). This relationship differs by profile ($F(7, 4607.0) = 4.17$, $p < .001$): while three of the profiles rested consistently regardless of state, the Bad-ModRest group rested longer in $S2_{typical}$ than $S1_{disengaged}$ or $S3_{good}$ ($S1-S2:b = -2.05$, $p = .001$; $S2-S3:b = 1.97$, $p < 0.0001$).

Overall, decisions about how long to rest appear to primarily reflect stable, auto-correlated individual differences, and longer rest does not enhance recovery even when accounting for between-subject differences.

⁴LPA profile number selected with BIC (2-6 profiles tested); TV-HMM state number selected by AIC/BIC (2-4 states tested).

⁵We excluded additional subjects who had >80% zero-accuracy high-effort task epochs: they led to degenerate TV-HMM states.

⁶Due to convergence issues we removed states that never improved, and the Okay-MaxRest profile (no rest variance).

Acknowledgments

All authors contributed to this work. Claude Code was used to assist in drafting code for some of the initial analyses.

References

- Bijleveld, E. (2023). The ebb and flow of cognitive fatigue. *Trends in Cognitive Sciences*, 27(12), 1109–1110.
- Boksem, M. A., & Tops, M. (2008). Mental fatigue: Costs and benefits. *Brain research reviews*, 59(1), 125–139.
- Gershman, S. J. (2020). Origin of perseveration in the trade-off between reward and complexity. *Cognition*, 204, 104394.
- Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in cognitive sciences*, 22(4), 337–349.
- Kanfer, R. (2011). Determinants and consequences of subjective cognitive fatigue., 189–207.
- Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of experimental psychology: general*, 139(4), 665.
- Lim, J., & Kwok, K. (2016). The effects of varying break length on attention and time on task. *Human factors*, 58(3), 472–481.
- Matthews, J., Pisauro, M. A., Jurgelis, M., Müller, T., Vassena, E., Chong, T. T.-J., & Apps, M. A. (2023). Computational mechanisms underlying the dynamics of physical and cognitive fatigue. *Cognition*, 240, 105603.
- Mews, S., Koslik, J.-O., & Langrock, R. (2025). How to build your latent markov model: The role of time and space. *Statistical Modelling*, 25(6), 481–507.
- Patel, T. N., Yang, R., Steyvers, M., & Benjamin, A. S. (2024). The metacognition of vigilance: Using self-scheduled breaks to improve sustained attention. *Journal of Experimental Psychology: Applied*, 30(4), 554.
- Perquin, M. N., Yang, J., Teufel, C., Sumner, P., Hedge, C., & Bompas, A. (2020). Inability to improve performance with control shows limited access to inner states. *Journal of Experimental Psychology: General*, 149(2), 249.
- Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual review of neuroscience*, 40(1), 99–124.
- Xu, A., Frömer, R., Wolff, W., & Shenhav, A. (2023). Do you ever get tired of being wrong? the unique impact of feedback on subjective experiences of effort-based decision-making.