Harnessing the Web for Population-Scale Physiological Sensing: A Case Study of Sleep and Performance

Tim Althoff*, Eric Horvitz, Ryen W. White, Jamie Zeitzer

* Research performed at Microsoft Research
Human Cognitive Performance

- Ability to perform mental actions and processes including attention, memory, reasoning, decision making, planning, etc.

- High cognitive performance important for:
  - Productivity [Colten & Altevogt, 2006]
  - Learning outcomes [Kelley et al., 2015]
  - Accident risk [Dinges, 1995]

- Laboratory setting: Performance varies throughout day [Van Dongen & Dinges, 2000] and is decreased after sleep loss [Dinges, 1995]
Existing Research

- Laboratory setting:
  - Induce sleep deprivation
  - Regular, intrusive, artificial performance tasks
  - Missing real-world influences incl. motivation, mood, illness, behavioral compensation (e.g., caffeine), and complex sleep patterns

Tim Althoff, Stanford University
Existing Research

- Laboratory setting:
  - Induce sleep deprivation
  - Regular, intrusive, artificial performance tasks
  - Missing real-world influences incl. motivation, mood, illness, behavioral compensation (e.g., caffeine), and complex sleep patterns

Lack of scalable methods to characterize real-world cognitive performance & sleep

[Roenneberg, 2013]
Open Research Questions

1. How does cognitive performance vary in the real world?

2. How do real-world sleep patterns impact performance?
Challenges

- Real-world more complex than laboratory
  → Need much larger dataset

- But existing methods don’t scale!

- Need cognitive performance measurements
  - Annoying: Regular, intrusive, artificial performance tests
  - Should use performance on real tasks

- Need sleep measurements
  - Can’t observe in lab or control as before
  - Can’t trust subjective reports

- How can research progress outside the laboratory?
Our Key Insight

Use existing interactions with technology as a sensor into real-world cognitive performance.
Harnessing Search Engine Interactions

- Search engines are used repeatedly every day, awake or sleepy, by billions of people
- Reframe everyday interactions with web search engine as series of performance tasks

Tim Althoff, Stanford University
Harnessing Search Engine Interactions

- Search engines are used repeatedly every day, awake or sleepy, by billions of people

- Reframe everyday interactions with web search engine as series of performance tasks
  - Query typing speed, or click on search result

\[
\begin{align*}
\Delta t(\text{“c”}) &= 237\text{ms} \\
\Delta t(\text{“e”}) &= 219\text{ms}
\end{align*}
\]
Harnessing Search Engine Interactions

- Search engines are used repeatedly every day, awake or sleepy, by billions of people
- Reframe everyday interactions with web search engine as series of performance tasks
  - Query typing speed, or click on search result

\[ \Delta t("c") = 237 \text{ms} \]
\[ \Delta t("e") = 219 \text{ms} \]
\[ \Delta t("f") < \Delta t("*") \]
Measuring Sleep

- Use wearable device
  - Many search engine users already own device
  - Objective measures of time in bed by clicking “Start” & “I’m awake” (plus accelerometer-based algorithm)
Real-World Sleep & Performance at Scale

- Our insights enable study of real-world performance & sleep at scale
- 400x larger study than ever before
Dataset

- **Cohort:** 32k users over 18 months
  - US representative age, BMI, sleep; mostly male (93%)
  - (Opt-in to link Bing searches & Band data)

- **Performance:** 75M interaction tasks
  - Keystroke time (and click time)
  - Bing search engine

- **Sleep:** 3M nights of sleep
  - Microsoft Band

Tim Althoff, Stanford University
How does cognitive performance vary in the real-world?
Diurnal Performance Variation

- Performance far from constant (31% variation)
- Slowest during typical sleep times (circadian rhythm)

Error bars (all figures): 95% confidence interval
Robustness of Results

Findings are robust. Not explained by…

- Effects of individual users / Population differences
  - Observe true within-person variation

- Type of query
  - Control for click entropy to capture query intent (navigational vs. informational)
  - Similar results for specific queries like “facebook”

- Learning effects
  - Few queries repeat; show no signs of learning effects

- Weekend vs. weekday effects

- Network latency dynamics
How can we model real-world performance variation?
Modeling Challenges

- Three biological processes drive performance variation
Modeling Challenges

- Three biological processes drive performance variation
  1. Circadian rhythm (C): time-dependent, behavior-independent, near 24h oscillations
Modeling Challenges

- Three biological processes drive performance variation
  1. Circadian rhythm (C): time-dependent, behavior-independent, near 24h oscillations
  2. Homeostatic sleep drive (H): the longer awake, the more tired you become
Modeling Challenges

- Three biological processes drive performance variation
  1. Circadian rhythm (C): time-dependent, behavior-independent, near 24h oscillations
  2. Homeostatic sleep drive (H): the longer awake, the more tired you become
  3. Sleep inertia (I): performance impairment experienced immediately after waking up
Modeling Challenges

- Three biological processes drive performance variation
  1. Circadian rhythm (C): time-dependent, behavior-independent, near 24h oscillations
  2. Homeostatic sleep drive (H): the longer awake, the more tired you become
  3. Sleep inertia (I): performance impairment experienced immediately after waking up

- Hard to disentangle effects
  - Many factors, highly correlated
  - Lab: Forced desynchrony protocol
  - Our method: Variation across millions of real-world interactions (web search)
Statistical Model

- Generalized Additive Model
  - Intercept
  - Keystroke (control for key pressed: “A”, “a”, “@”, …)
  - Time of day (circadian rhythm)
  - Time since wakeup (homeostatic sleep drive & sleep inertia)

\[ y_i = \alpha + f^k(x_i^k) + f^t(x_i^t) + f^w(x_i^w) + \epsilon_i \]

Keystroke time

Residual

Tim Althoff, Stanford University
Statistical Model

- Generalized Additive Model
  - Intercept
  - Keystroke (control for key pressed: “A”, “a”, “@”, …)
  - Time of day (circadian rhythm)
  - Time since wakeup (homeostatic sleep drive & sleep inertia)

\[ y_i = \alpha + f_k(x^k_i) + f_t(x^t_i) + f_w(x^w_i) + \epsilon_i \]

- Parameter learning
  - Fine-grained discretization functions (non-parametric)
  - Least squares optimization
Model Estimates: Time of Day

- Model identifies underlying circadian rhythm
- Consistent with lab-based studies

\[ f^t(x^t_i) \]

Habitual sleep time

Habitual sleep time
Model: Time After Wakeup

- Model identifies underlying homeostatic sleep drive and sleep inertia consistent with lab-based studies (validation).
- New insights: It was impossible to measure real-world cognitive performance at scale. Now we can!
How do real-world sleep patterns impact performance?
Sleep Loss over *Multiple* Nights

1. Can we observe an additive effect of multiple nights with little sleep?

2. How long does it take to recover from sleep loss? (Given real-world sleep patterns)

- Measure performance over 7 days after zero (SS), one (SI), or two (II) insufficient nights of sleep (less than 6h)
Recovering from Sleep Loss

Performance decreased further after two insufficient nights of sleep.

- **Baseline**: Slowest performance after SS
- **Performance decreased further after two insufficient nights of sleep**

![Graph showing performance decrease over days after sleep pattern](image-url)
Recovering from Sleep Loss

Performance decreased further after two insufficient nights of sleep

Baseline: Slowest performance after SS

Performance after two insufficient nights recovers only after six days!
Performance by Sleep Timing

We control for sleep duration (7-8h)
Our Contributions

- **New method:** Use existing technology interactions to study sleep and cognitive performance
  - Large-scale, real-world (outside of laboratory)
  - Continuous, non-intrusive measurements of realistic tasks

- **New insights:** Real-world performance is not constant but exhibits variation based on time of day and complex sleep patterns. We are the first to quantify these effects.
Our Contributions

- **New method**: Use existing technology interactions to study sleep and cognitive performance
  - Large-scale, real-world (outside of laboratory)
  - Continuous, non-intrusive measurements of realistic tasks

- **New insights**: Real-world performance is not constant but exhibits variation based on time of day and complex sleep patterns. We are the first to quantify these effects.

- **Population-scale Physiological Sensing**
  - Physiology: Branch of biology dealing with the functions and activities of living organisms and their parts
  - Learn about biological functions through user activity logs
Acknowledgments

- Research performed during internship at Microsoft Research

Joining faculty job market end of 2017. Please let me know about opportunities at your institution.

Ask me anything!

@timalthoff
althoff@cs.stanford.edu
www.timalthoff.com