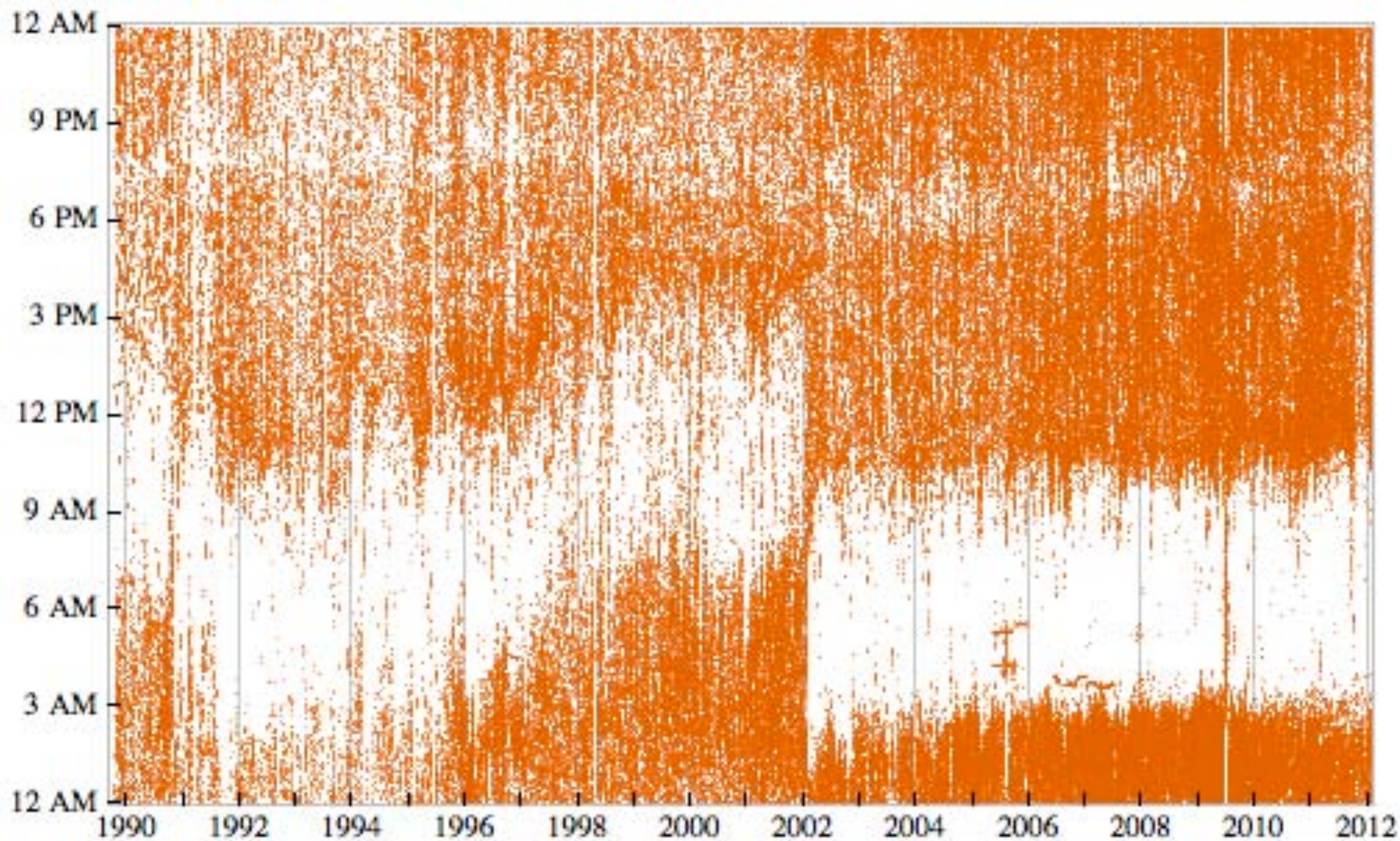
Modeling Individual-Level Data in the 21st Century

Padhraic Smyth

Department of Computer Science

University of California, Irvine

Presentation at SIAM Data Mining Conference, May 2013


12 AM
9 PM
6 PM
3 PM
12 PM
9 AM
6 AM
3 AM
12 AM
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012

Outline of Talk

- **What is individual level-data?**
  - Historical context and examples
- **How can we use individual-level data?**
  - Opportunities for machine learning and data mining
- **Research example:**
  - Modeling of personal archive data
- **Conclusions**

Individual Data: Demographics

- **1950's: availability of demographic data**
  - Age, zip-code, income, education, employment
- **Applications:**
  - Direct mail marketing
  - Consumer credit and loans
- **Example: Fair Isaac and FICO scores**

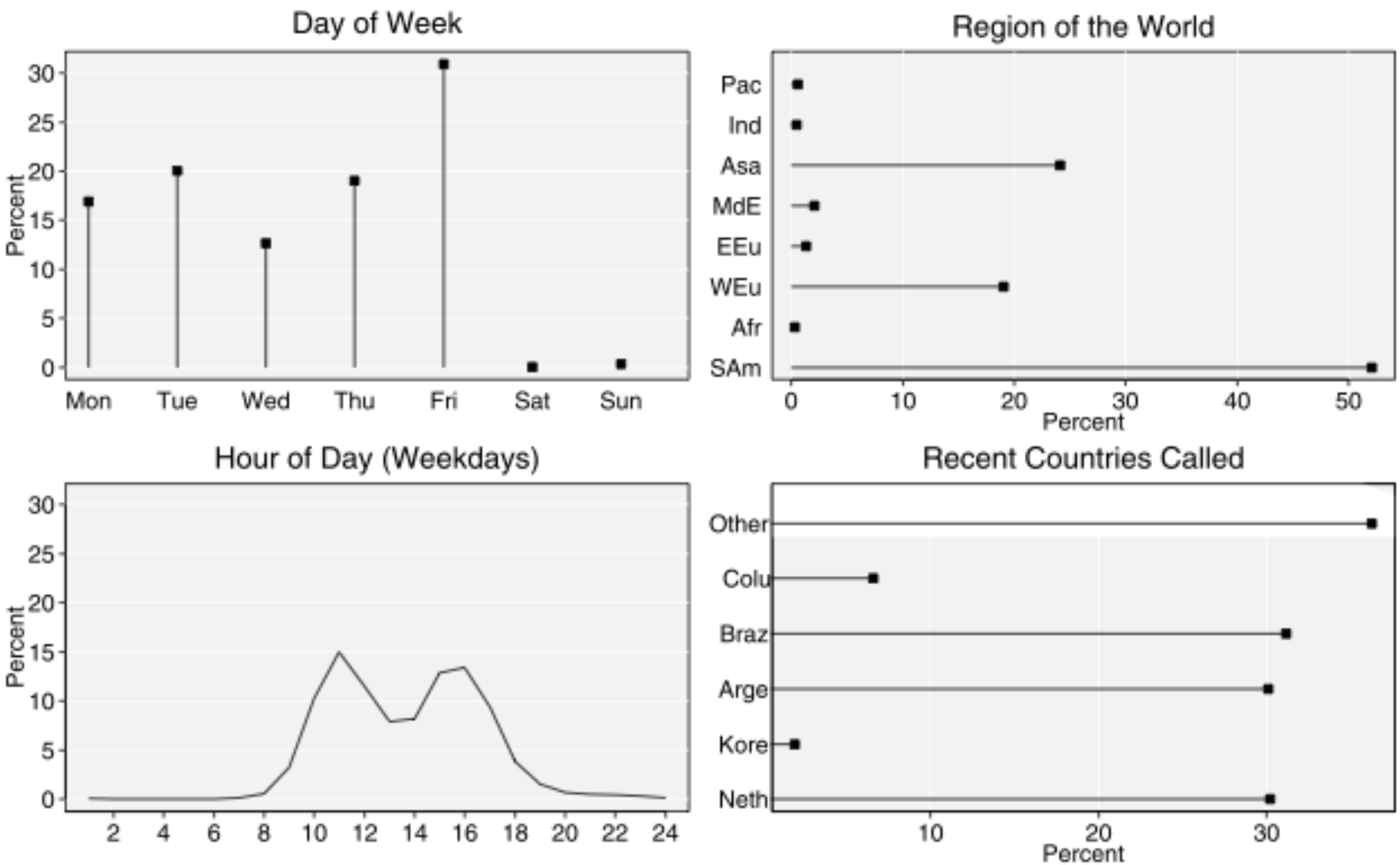
Individual Data: Transactions

- **1980's, 1990s'**
  - Billing and purchase transaction data

- **Applications**
  - Direct marketing/advertising
  - Fraud detection
  - .....

Fraud Detection at AT&T

From Becker, Volinsky, Wilks, Techometrics, 2010



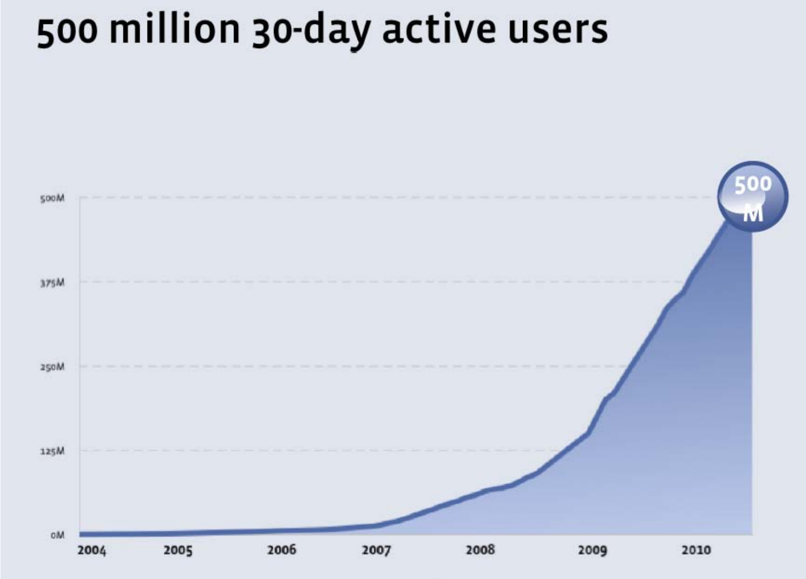
Individual Data: Internet

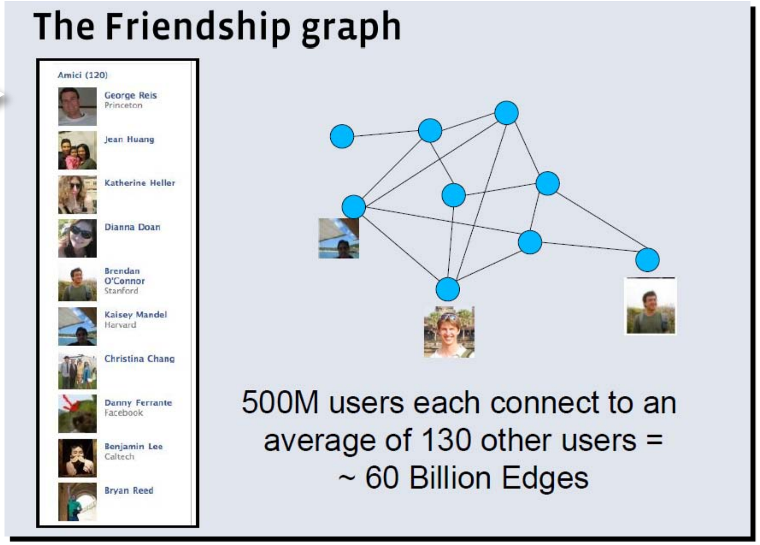
- **2000's**
  - Web pages visited, Web searches
  - Ads clicked on
  - Text (microblogs, emails)
  - Social networks (online, cell phones, etc)
  - Location (GPS, mobile phone)

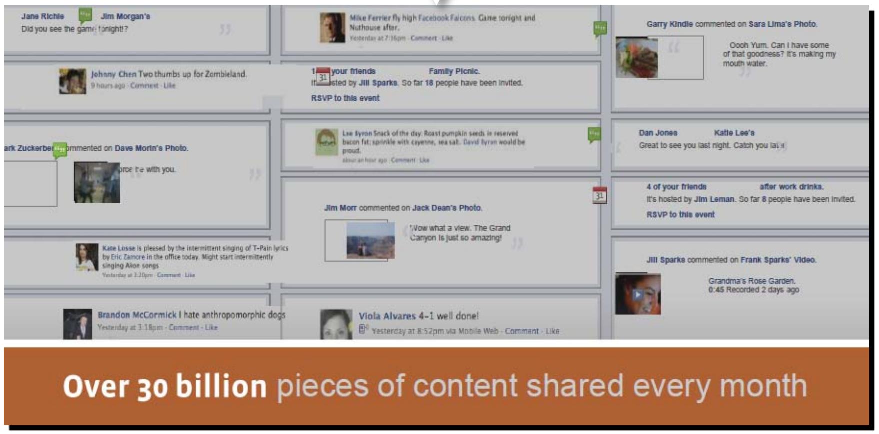
- **Applications**
  - Online advertising
  - Recommender systems
  - .....

facebook

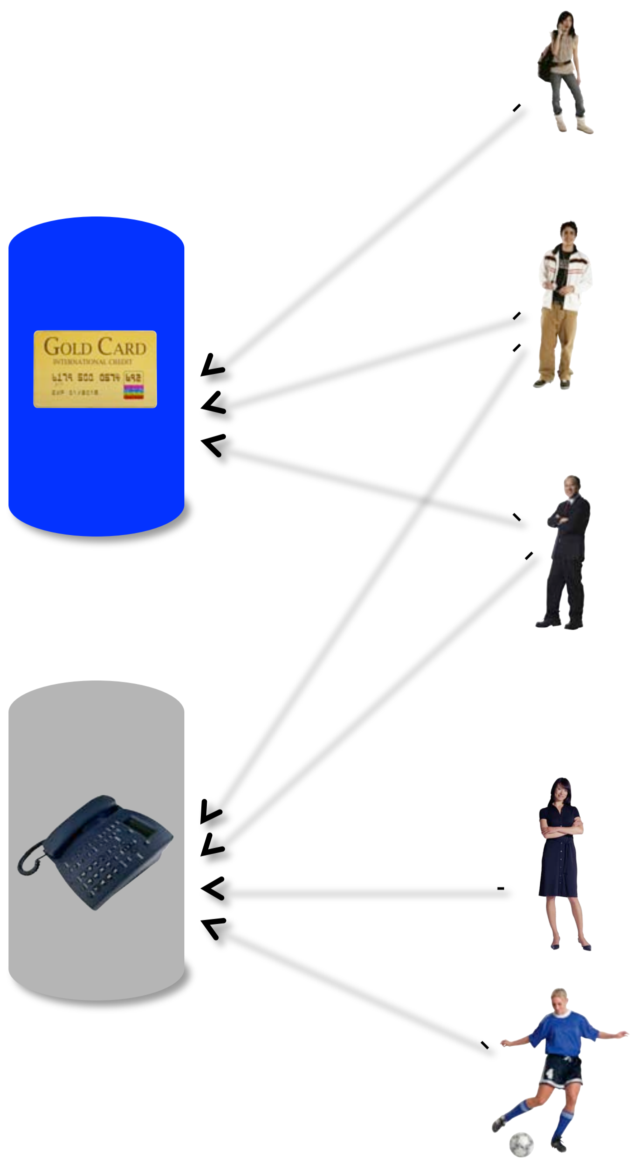
Graphics from Lars Backstrom, ESWC 2011



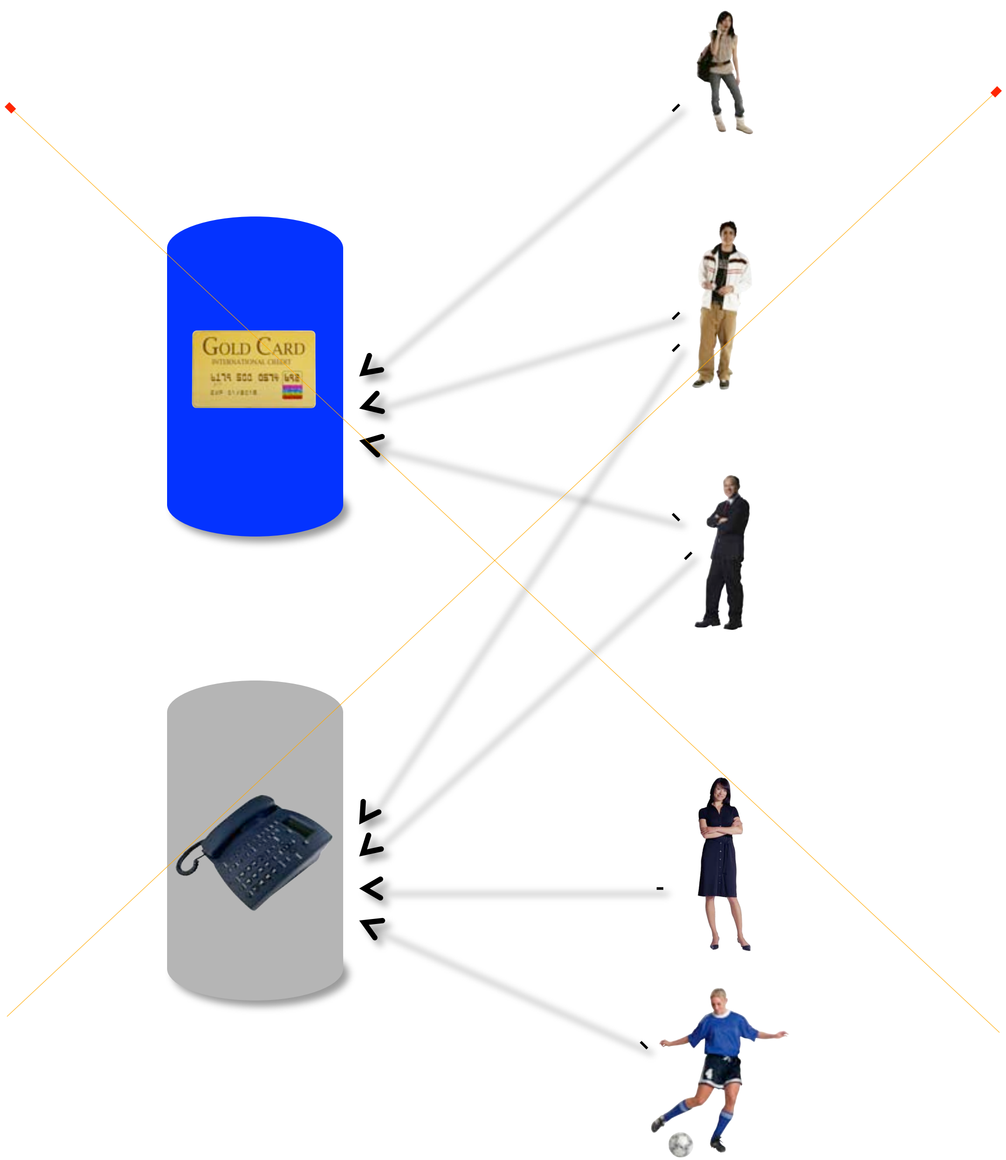




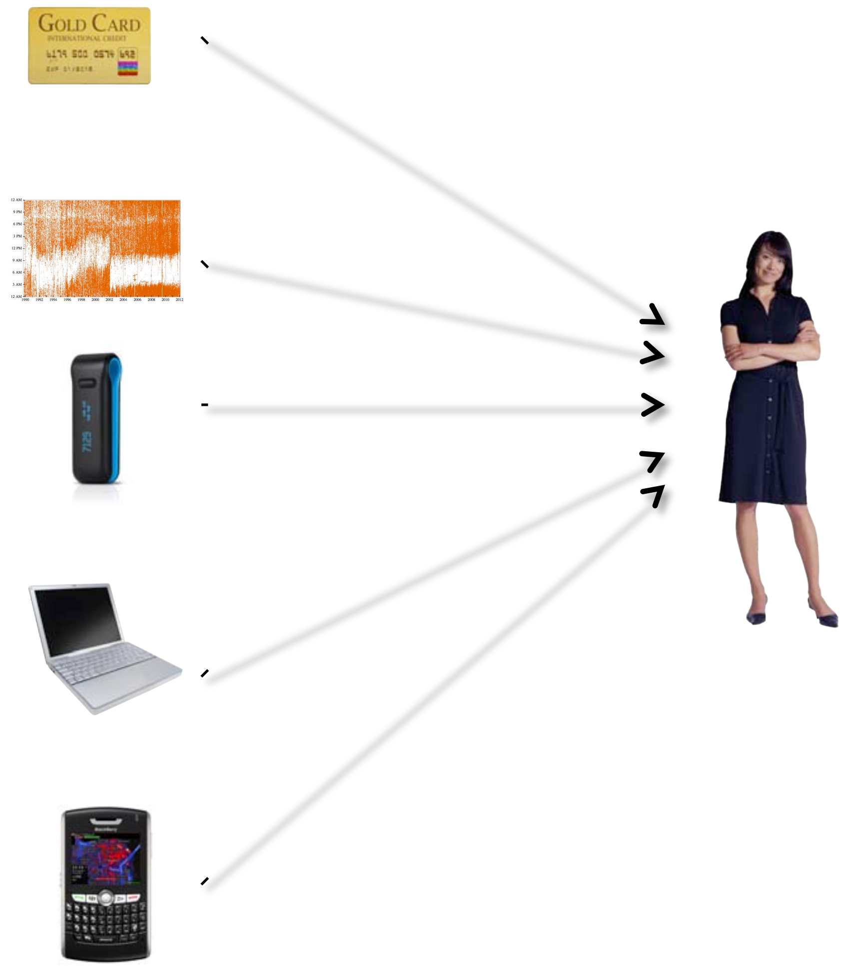
The Corporate View of Individual Data

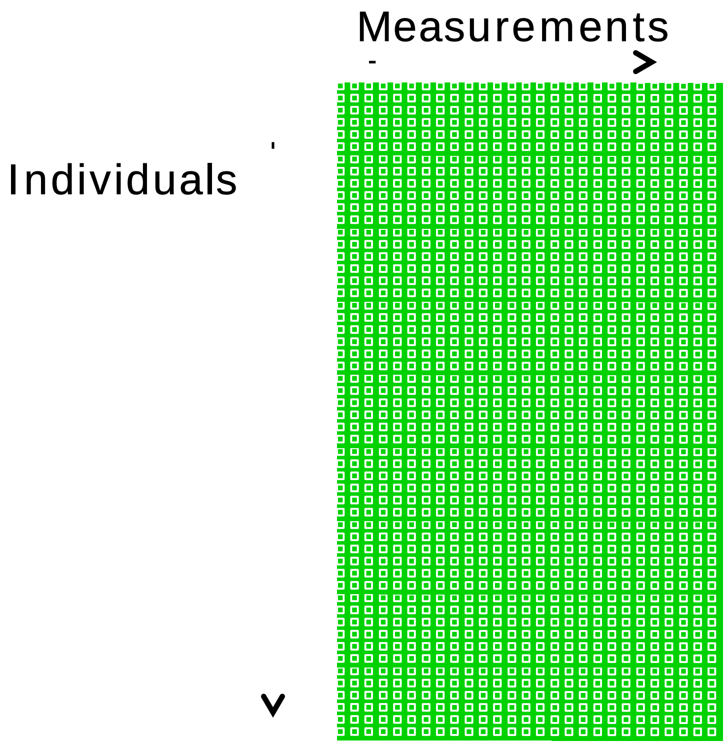


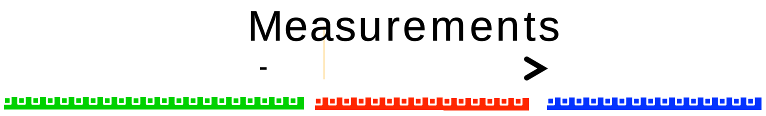
The Corporate View of Individual Data



The Individual's View of Individual Data




Measurements
Individuals

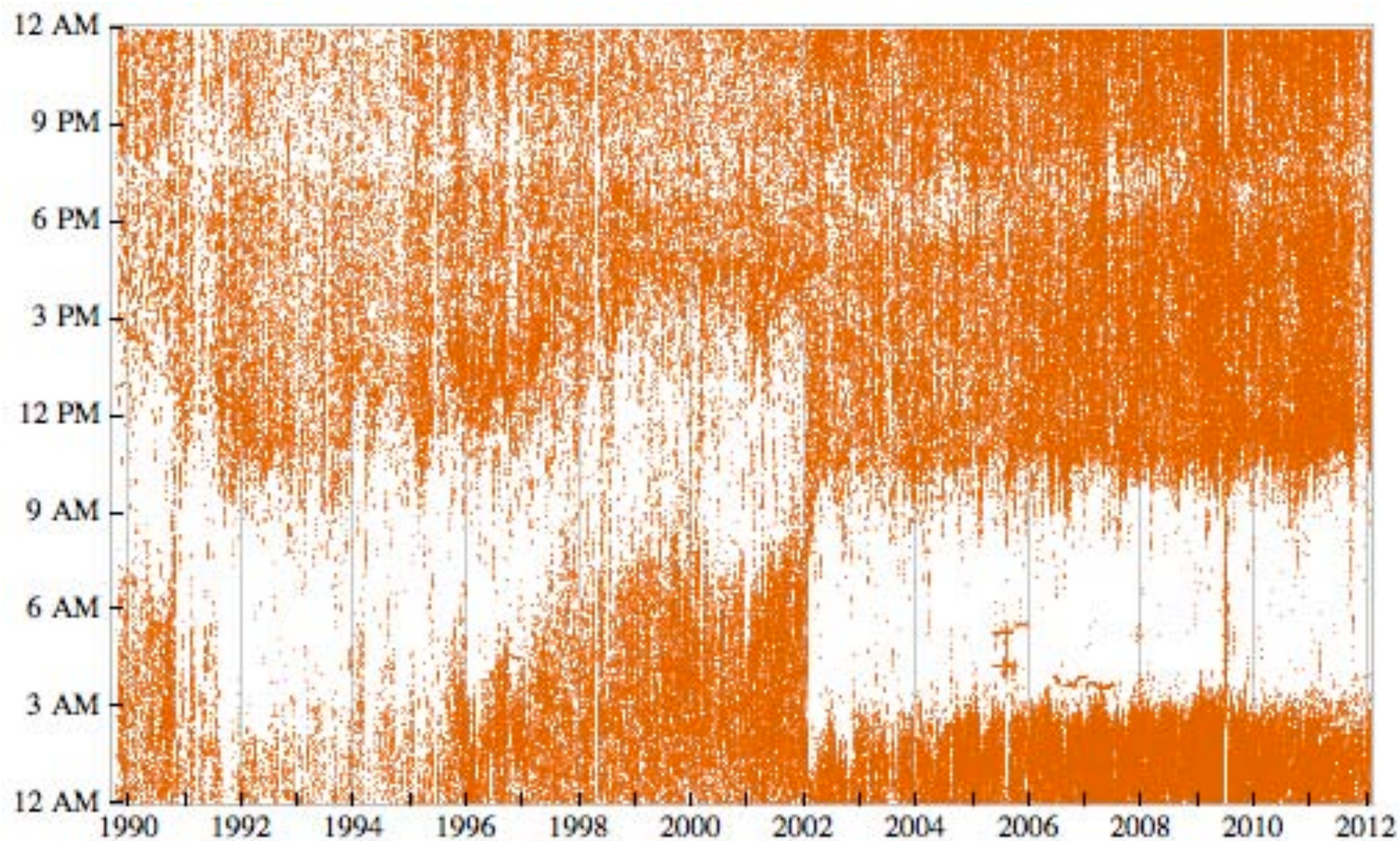

Measurements

Individual-Level Data

- **Digital Data**
  - Emails
  - Text messages
  - Phone calls
  - Location
  - Social media events

- **Physiological Data**
  - Activity
  - Exercise
  - Sleep
  - Blood pressure
  - Diet

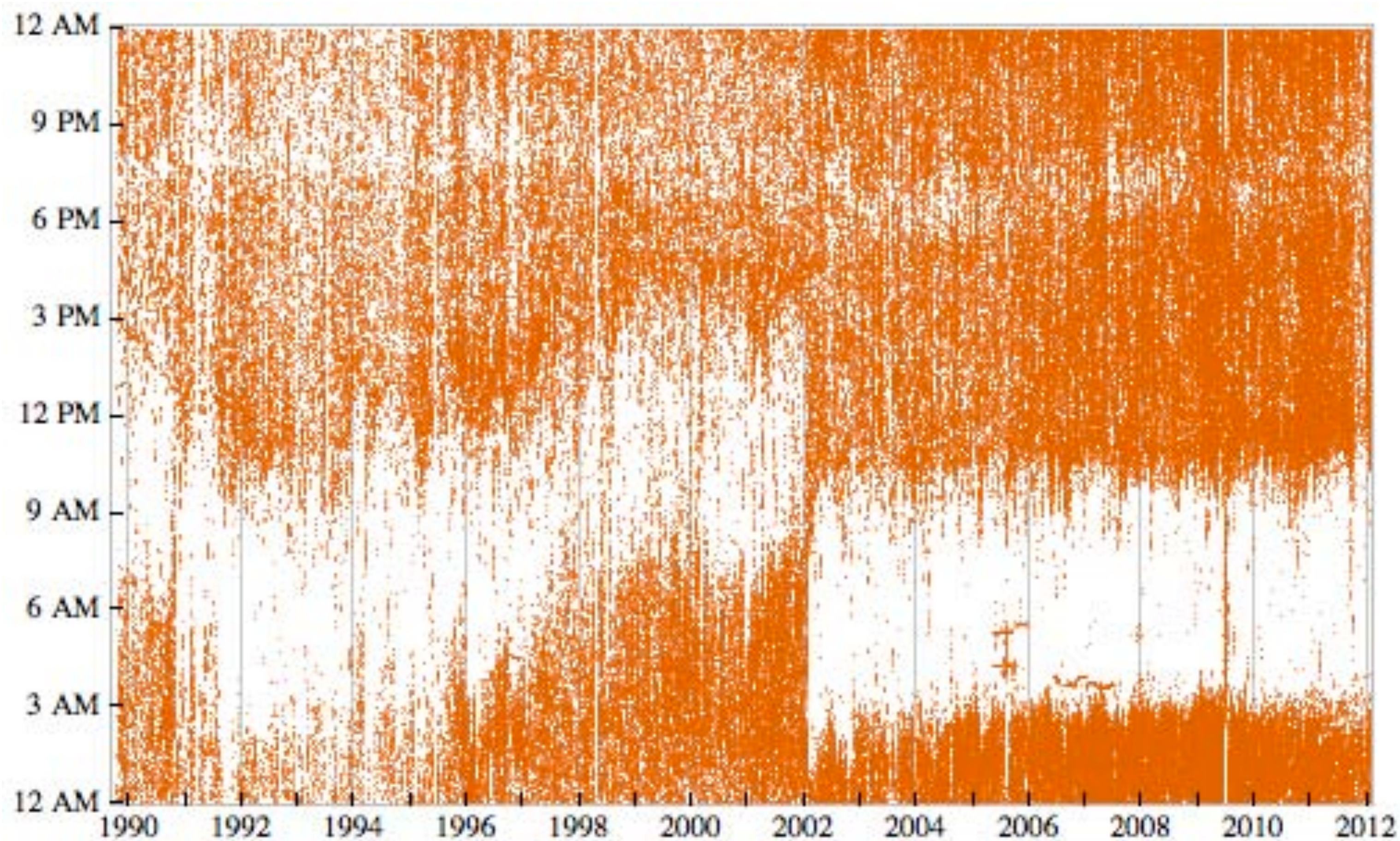
Email Data



Time plot of 1/3 million emails sent by Stephen Wolfram over 20 years

Figures from *The Personal Analytics of My Life*
blog.stephenwolfram.com, March 2012

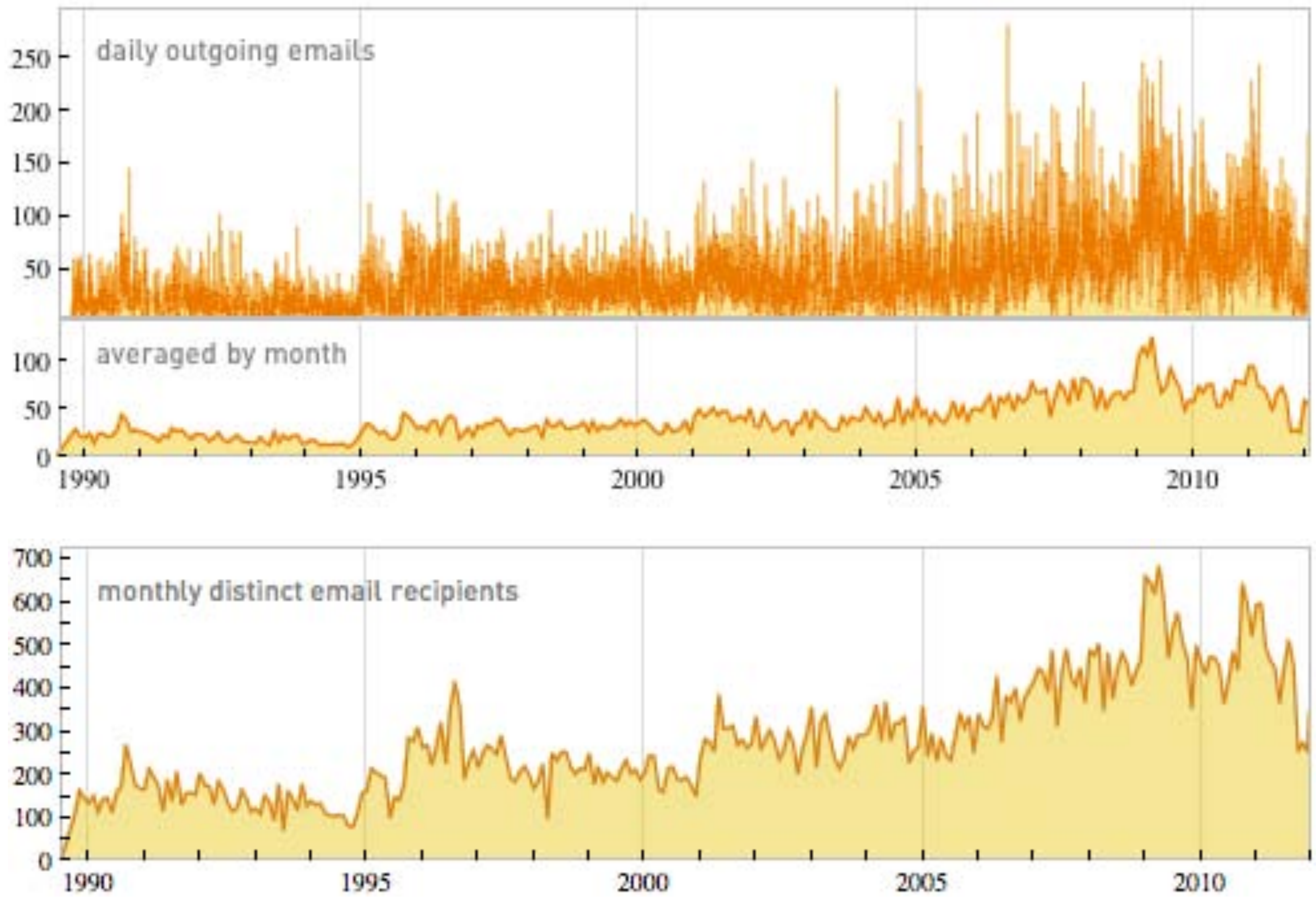
Email Data



Working on book **Return to “normal” life**

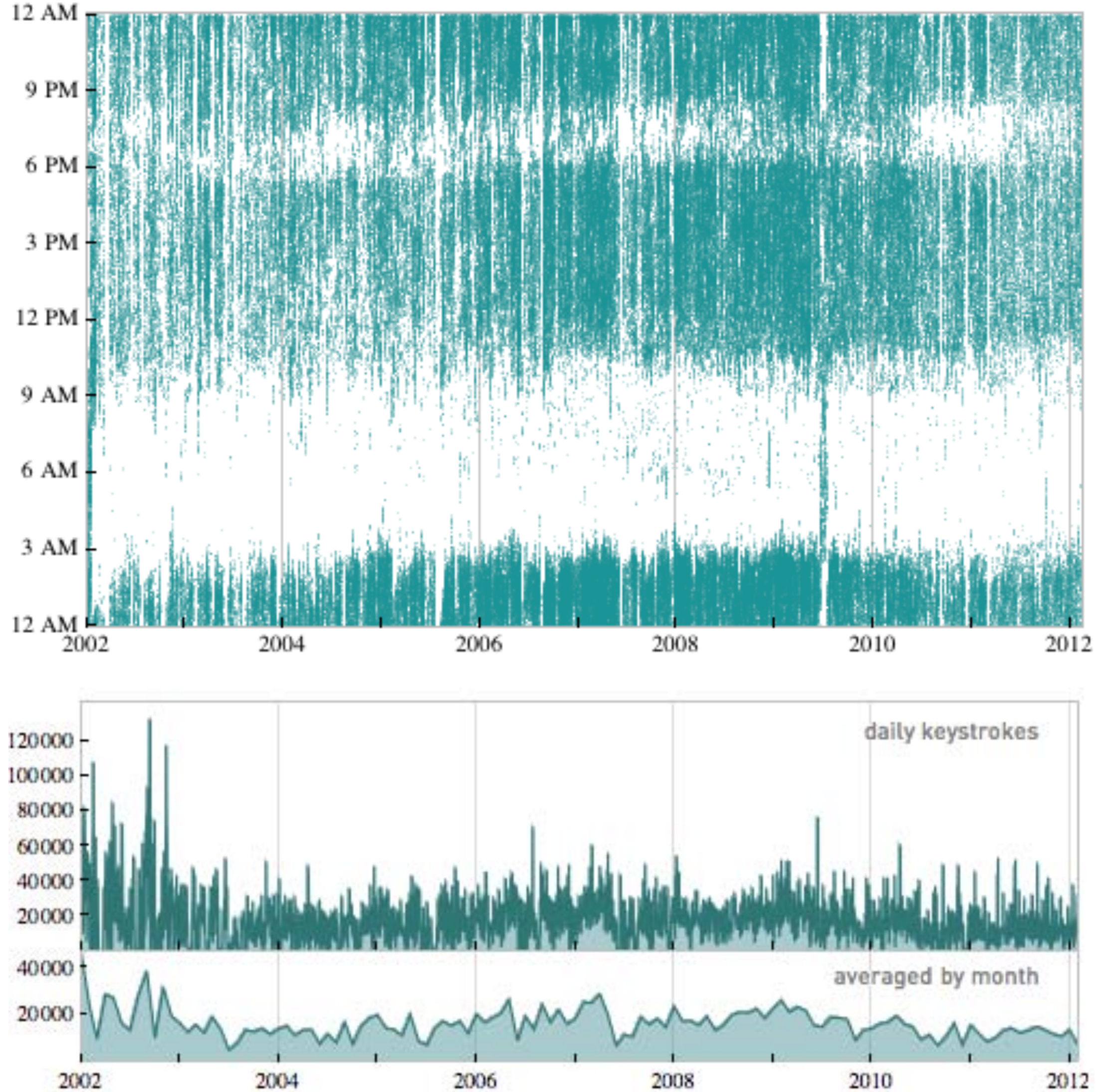
Figures from *The Personal Analytics of My Life*
blog.stephenwolfram.com, March 2012

Volume of Emails Sent



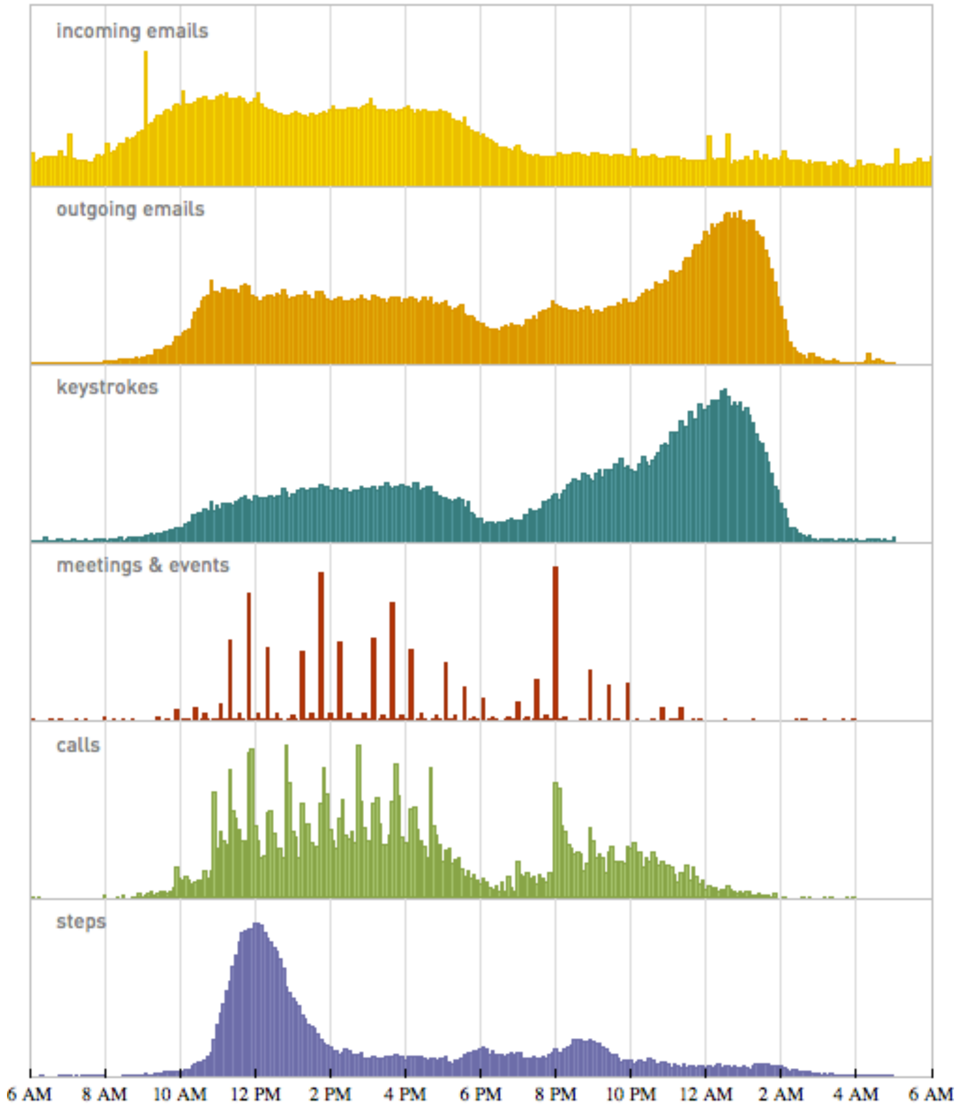
Figures from *The Personal Analytics of My Life*
blog.stephenwolfram.com, March 2012

Time plots of Keystrokes



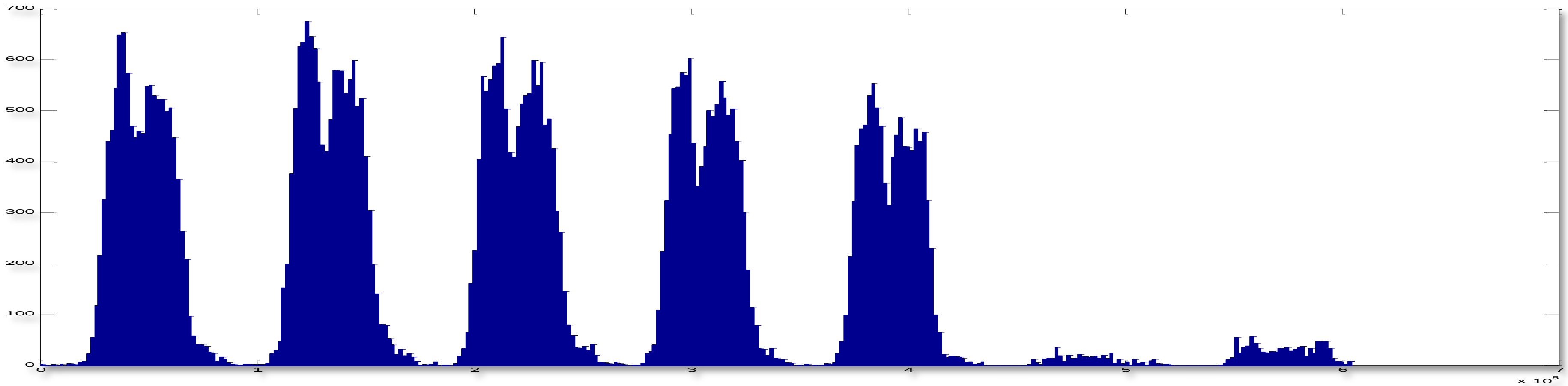
Figures from *The Personal Analytics of My Life*
blog.stephenwolfram.com, March 2012

Aggregated Daily Rhythms

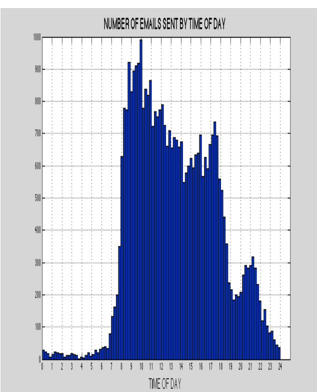


Figures from *The Personal Analytics of My Life*
blog.stephenwolfram.com, March 2012

Time of Day Variation in Enron Emails



Weekdays Weekend



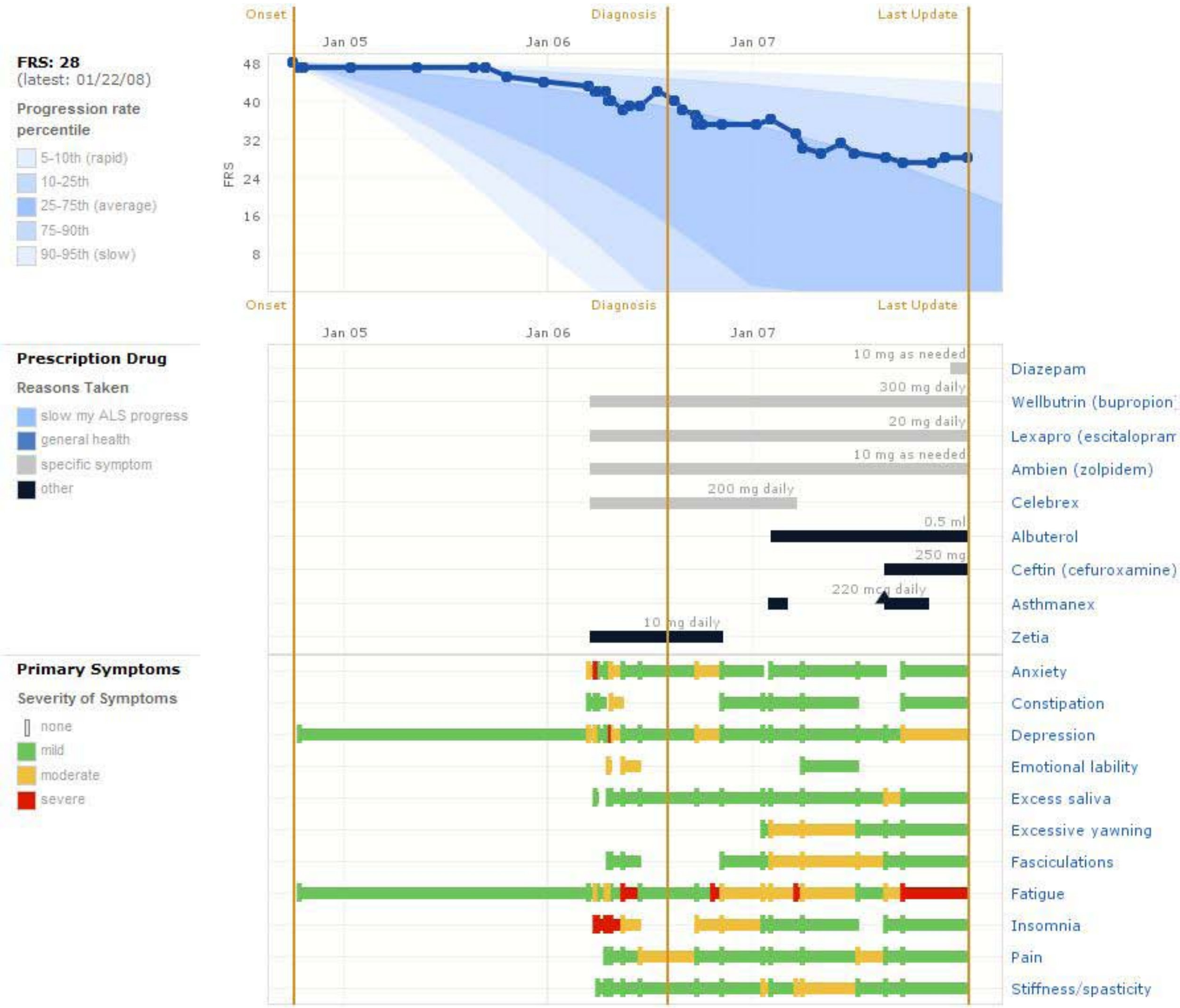
Time of Day Variation in Personal Email

What can we Measure?

- **Monitoring of the digital world**
  - Email, texts, Web clicks, searches, social media actions
  - Keystrokes, mouse movement, eye tracking
  - GPS location
  - And so on….

- **Monitoring of the physical world**
  - Heart-rate monitoring, skin conductance, etc
  - Acceleration/activity
  - Diet
  - Sleep patterns
  - Audio and speech
  - Video
  - And so on…

Medical Self-Reporting: PatientsLikeMe



Example of PatentsLikeMe chart
From Frost and Massagli, 2008

Quantified Self
self knowledge through numbers

MEETUPS · FOLLOW: RSS · ABOUT Search

European Conference Preview: Breakout Sessions

Posted on April 30, 2013 by Ernesto Ramirez

PREVIEW

Quantified Self
Europe Conference
May 11-12, 2013 Amsterdam, Netherlands

At its core, Quantified Self is a community-driven effort to extract personal meaning from personal data. Our conferences reflect that by providing opportunities to learn what others are doing in their Quantified Self practice. Through our Show & Tell presentations you get to see first-hand accounts of how data is being collected and put to use in order to understand and investigate personal phenomena, but that's not all our conference have to offer. In the spirit of collaborative learning we also schedule "Breakout Sessions" alongside our wonderful Show & Tell talks. These sessions, like all our conference programming, are developed and facilitated by our wonderful attendees. Here's a preview of just a few of the many fantastic Breakouts we have scheduled.

Title: The Self in Data
Breakout Leader: Sara Watson
Description: The self is the assumed center of all QS projects, but we don't often talk about what it means to make data about ourselves. What is my relationship to my data? Today, over time? What does my data tell me about myself? How do I derive meaning from my data?

About the Quantified Self

A place for people interested in self-tracking to gather, share knowledge and experiences, and discover resources. Learn more

Make a Sparktweet

QS Europe 2013 Conference Register Now!

Get Started Here...

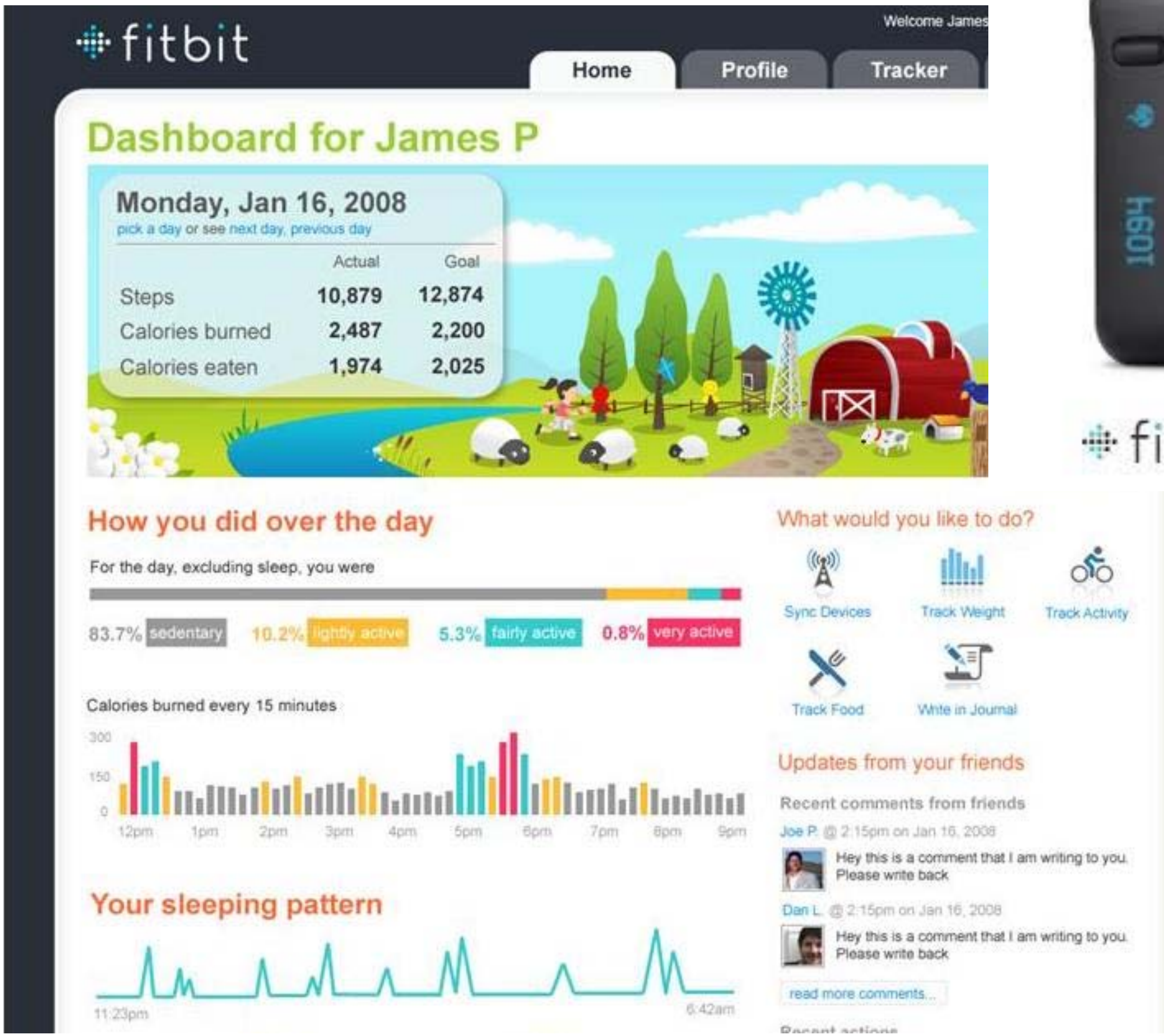
QS Show&Tell Videos

QS Forums

Global QS Event Calendar*

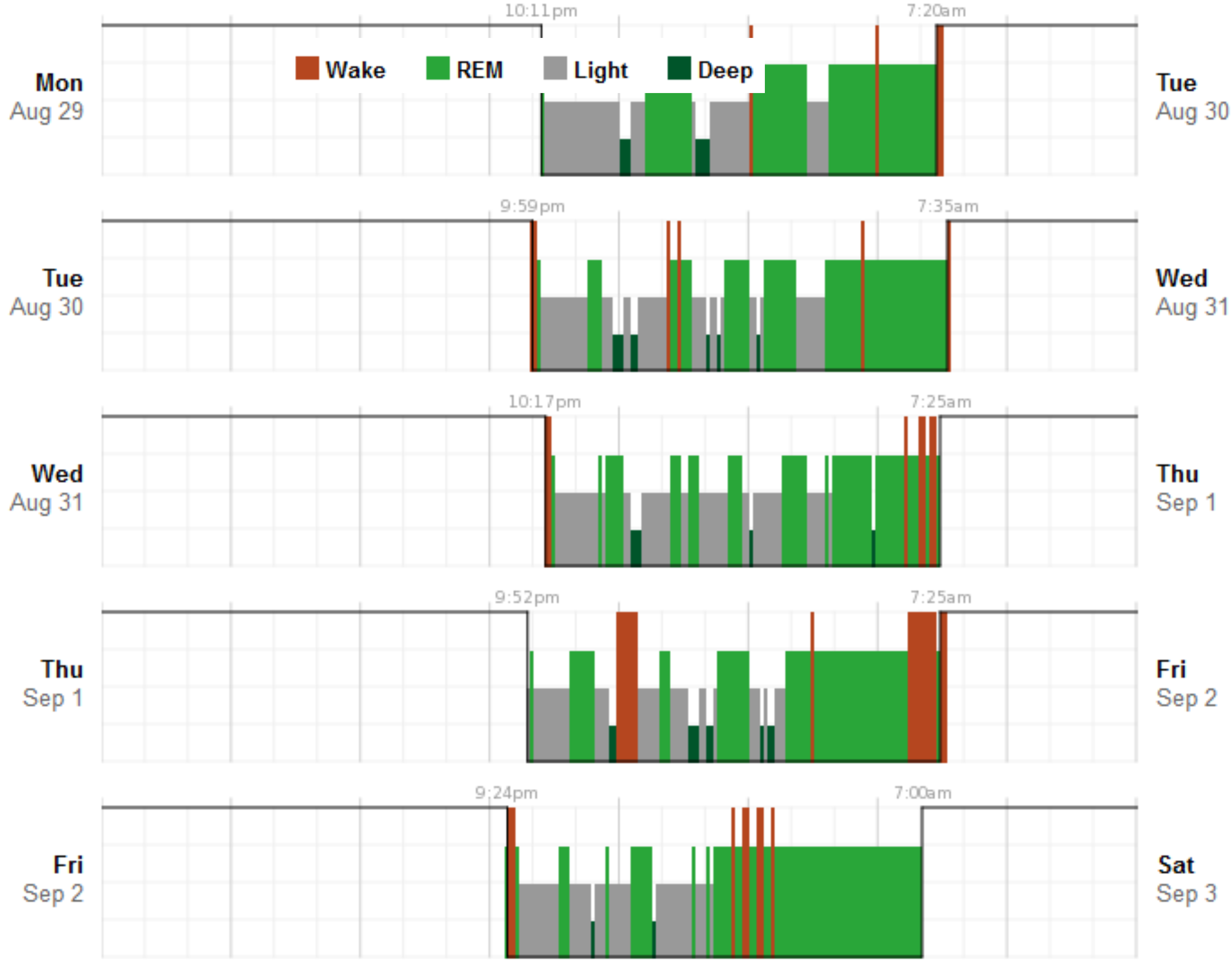
Thursday, May 2	
3:00pm	Boston QS meetup a
Monday, May 6	
10:30am	Hamburg QS meetup
3:00pm	Boston QS meetup a
Wednesday, May 8	

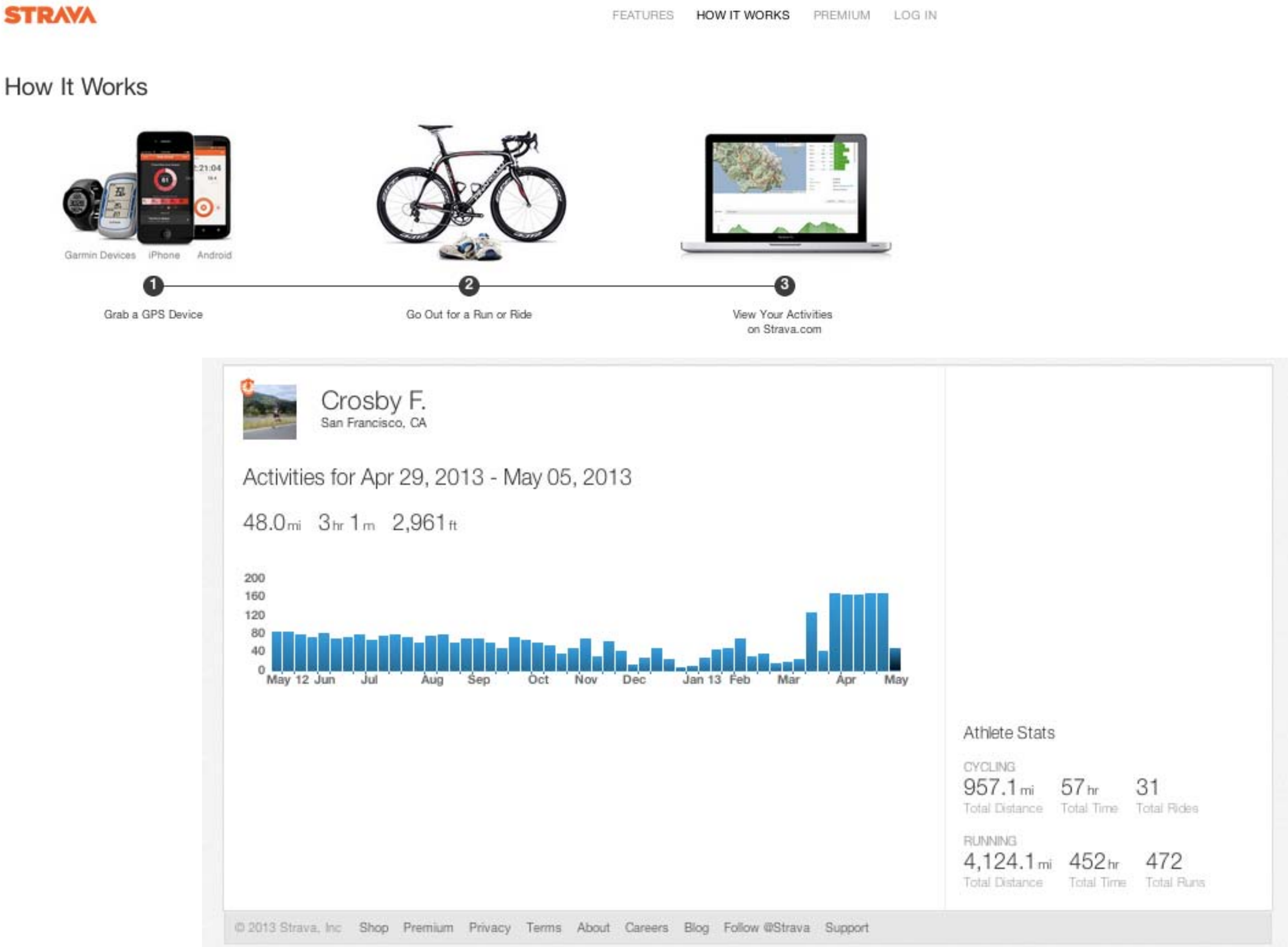
Exercise, Activity, Sleep Monitoring


fitbit
Welcome James
Home
Profile
Tracker
Dashboard for James P
Monday, Jan 16, 2008
pick a day or see next day, previous day
Actual Goal
Steps 10,879 12,874
Calories burned 2,487 2,200
Calories eaten 1,974 2,025
How you did over the day
For the day, excluding sleep, you were
83.7% sedentary
10.2% lightly active
5.3% fairly active
0.8% very active
Calories burned every 15 minutes
300
150
0
12pm 1pm 2pm 3pm 4pm 5pm 6pm 7pm 8pm 9pm
Your sleeping pattern
11:23pm
6:42am
What would you like to do?
Sync Devices
Track Weight
Track Activity
Track Food
Write in Journal
Updates from your friends
Recent comments from friends
Joe P. @ 2:15pm on Jan 16, 2008
Hey this is a comment that I am writing to you. Please write back
Dan L. @ 2:15pm on Jan 16, 2008
Hey this is a comment that I am writing to you. Please write back
read more comments...

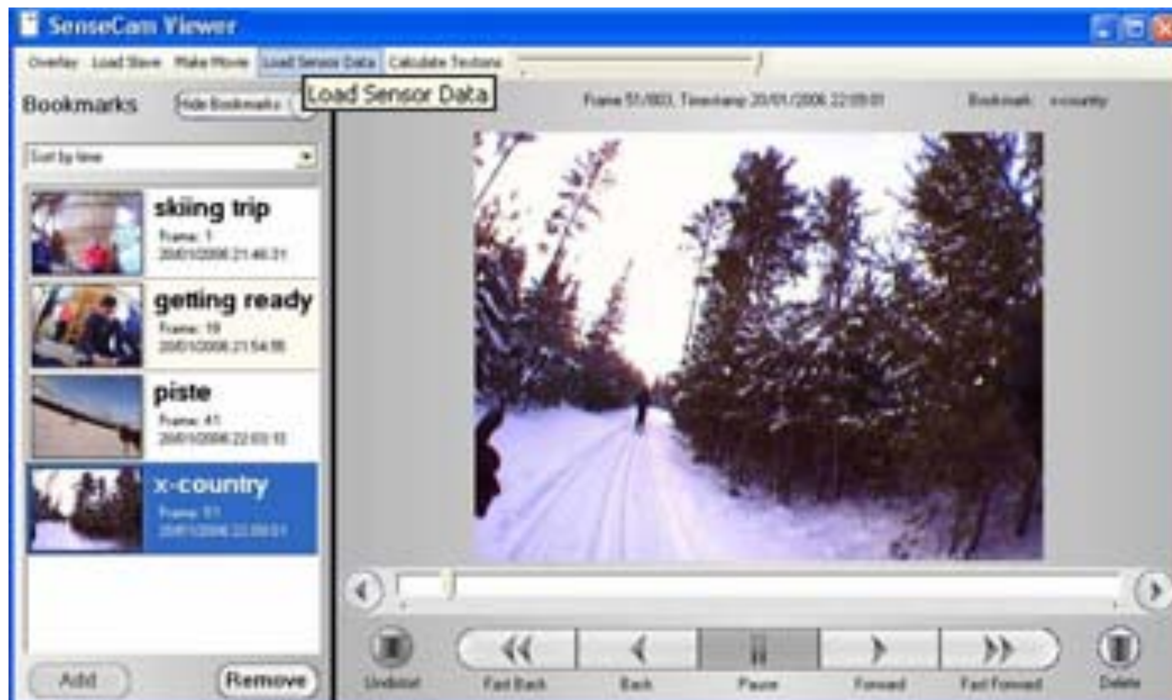

fitbit

Professor Larry Smarr, UCSD



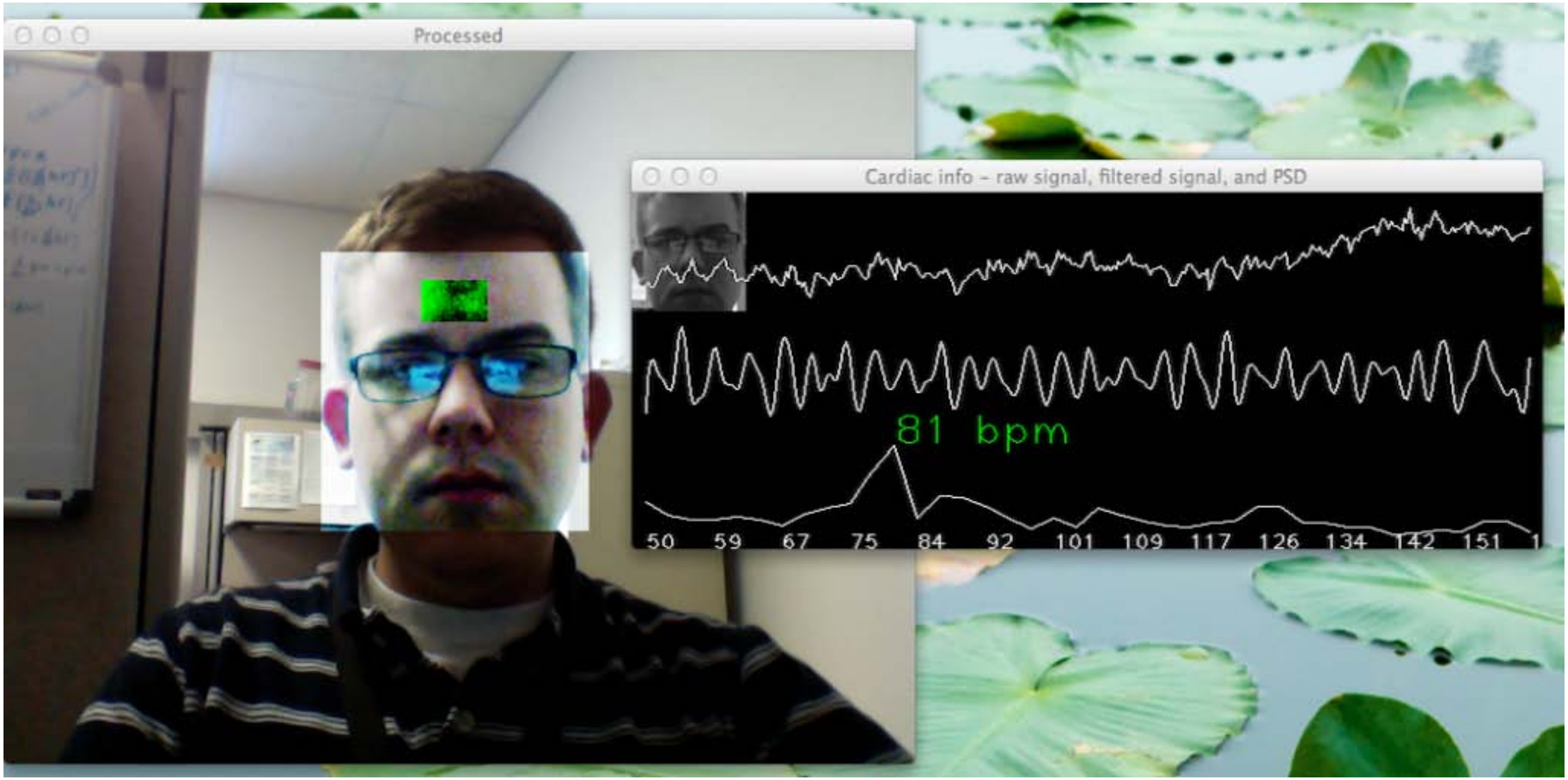

STRAVA
FEATURES
HOW IT WORKS
PREMIUM
LOG IN
How It Works
Garmin Devices
iPhone
Android
1
Grab a GPS Device
2
Go Out for a Run or Ride
3
View Your Activities on Strava.com
Crosby F.
San Francisco, CA
Activities for Apr 29, 2013 - May 05, 2013
48.0mi 3hr 1m 2,961ft
200
160
120
80
40
0
May 12 Jun Jul Aug Sep Oct Nov Dec Jan 13 Feb Mar Apr May
Athlete Stats
CYCLING
957.1mi Total Distance
57hr Total Time
31 Total Rides
RUNNING
4,124.1mi Total Distance
452hr Total Time
472 Total Runs
© 2013 Strava, Inc Shop Premium Privacy Terms About Careers Blog Follow @Strava Support

Microsoft SenseCam

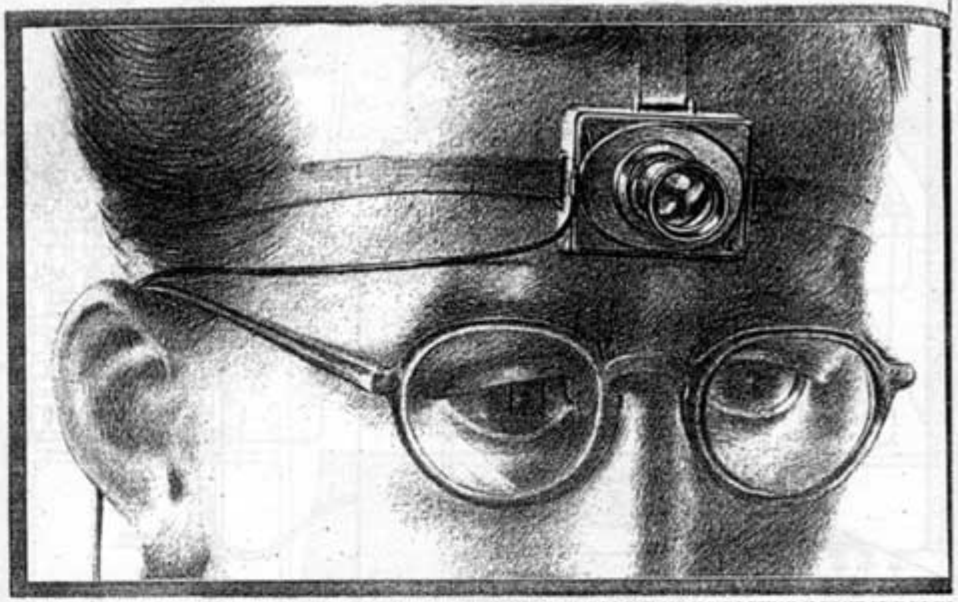


Measuring Blood Flow from Video Images

From Wu et al, MIT/Quanta, SIGGRAPH 2012



Where does data mining and machine learning fit?



A SCIENTIST OF THE FUTURE RECORDS EXPERIMENTS WITH A TINY CAMERA FITTED WITH UNIVERSAL-FOCUS LENS. THE SMALL SQUARE IN THE EYEGLASS AT THE LEFT SIGHTS THE OB

AS WE MAY THINK

A TOP U. S. SCIENTIST FORESEES A POSSIBLE FUTURE WORLI IN WHICH MAN-MADE MACHINES WILL START TO THINI

by VANNEVAR BUSH

DIRECTOR OF THE OFFICE OF SCIENTIFIC RESEARCH AND DEVELOPMENT

Condensed from the *Atlantic Monthly*, July 1945

This has not been a scientists' war; it has been a war in which all have had a part. The scientists, burying their old professional competition in the demand of a common cause, have shared greatly and learned much. It has been exhilarating to work in effective partnership. What are the scientists to do next?

For the biologists, and particularly for the medical scientists, there can be little indecision, for their war work has hardly required them to leave the old paths. Many indeed have been able to carry on their war research in their familiar peacetime laboratories. Their objectives remain much the same.

It is the physicists who have been thrown most violently off stride, who have left academic pursuits for the making of strange destructive gadgets, who have had to devise new methods for their unanticipated assignments. They have done their part on the devices that made it possible to turn back the enemy. They have worked in combined effort with the physicists of our allies. They have felt within themselves the stir of achievement. They have been part of a great team. Now one asks where they will find objectives worthy of their best.

• • •

There is a growing mountain of research. But there is increased evidence that we are being bogged down today as specialization extends. The investigator is staggered by the findings and conclusions of thousands of other workers—conclusions which he cannot find time to grasp, much less to remember, as they appear. Yet specialization becomes increasingly necessary for progress, and the effort to bridge between disciplines is correspondingly sup ficial.

Professionally our methods of transmitting and reviewing the result research are generations old and by now are totally inadequate for their p pose. If the aggregate time spent in writing scholarly works and in read them could be evaluated, the ratio between these amounts of time mi well be startling. Those who conscientiously attempt to keep abreast of c rent thought, even in restricted fields, by close and continuous reading mi well shy away from an examination calculated to show how much of the p vious month's efforts could be produced on call.

Mendel's concept of the laws of genetics was lost to the world for a g eration because his publication did not reach the few who were cap of grasping and extending it. This sort of catastrophe is undoubtedly be repeated all about us as truly significant attainments become lost in the m of the inconsequential.

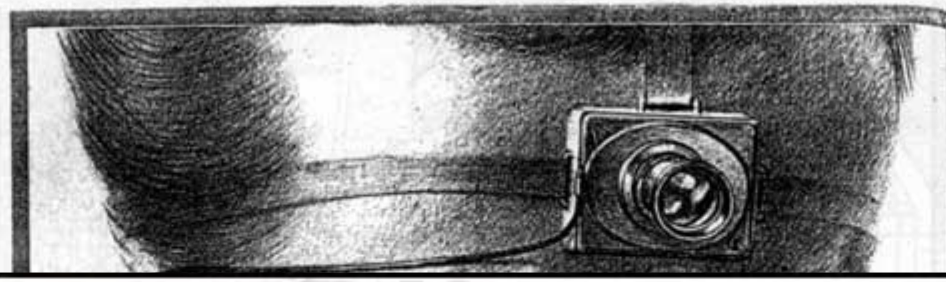
Publication has been extended far beyond our present ability to make t use of the record. The summation of human experience is being expande a prodigious rate, and the means we use for threading through the co quent maze to the momentarily important item is the same as was used the days of square-rigged ships.

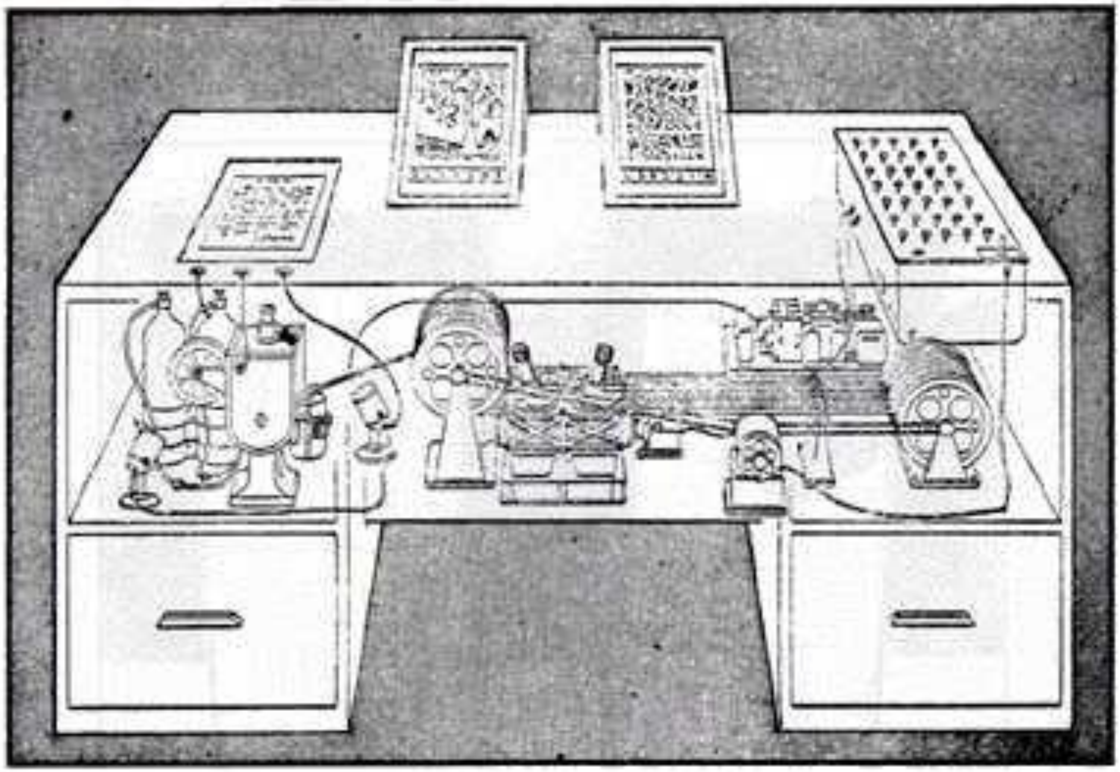
But there are signs of a change as new and powerful instrumentali come into use. Photocells capable of seeing things in a physical sense, vanced photography which can record what is seen or even what is t thermionic tubes capable of controlling potent forces under the guidanc

「LIFE」1945年9月10日号に特集された「AS WE MAY THINK」のページ

「LIFE」1945年9月10日号より引用

「AS WE MAY THINK」 From the Atlantic Monthly, July 1945





Memex in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference (*LIFE 19*(11), p. 123).

that we are being bogged down today as specialization extends. The investigator is staggered by the findings and conclusions of thousands of other workers—conclusions which he cannot find time to grasp, much less to remember, as they appear. Yet specialization becomes increasingly necessary for prog-

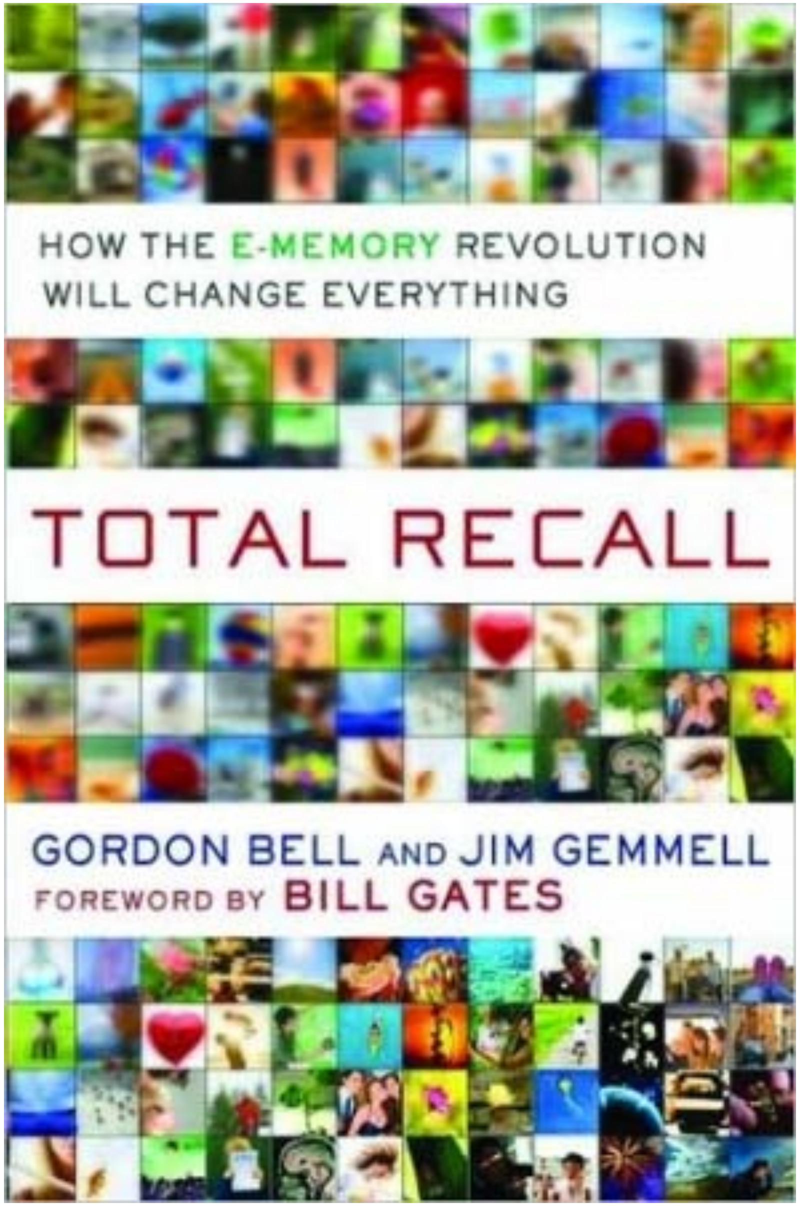
But there are signs of a change as new and powerful instrumentali come into use. Photocells capable of seeing things in a physical sense, vanced photography which can record what is seen or even what is thermionic tubes capable of controlling potent forces under the guidan

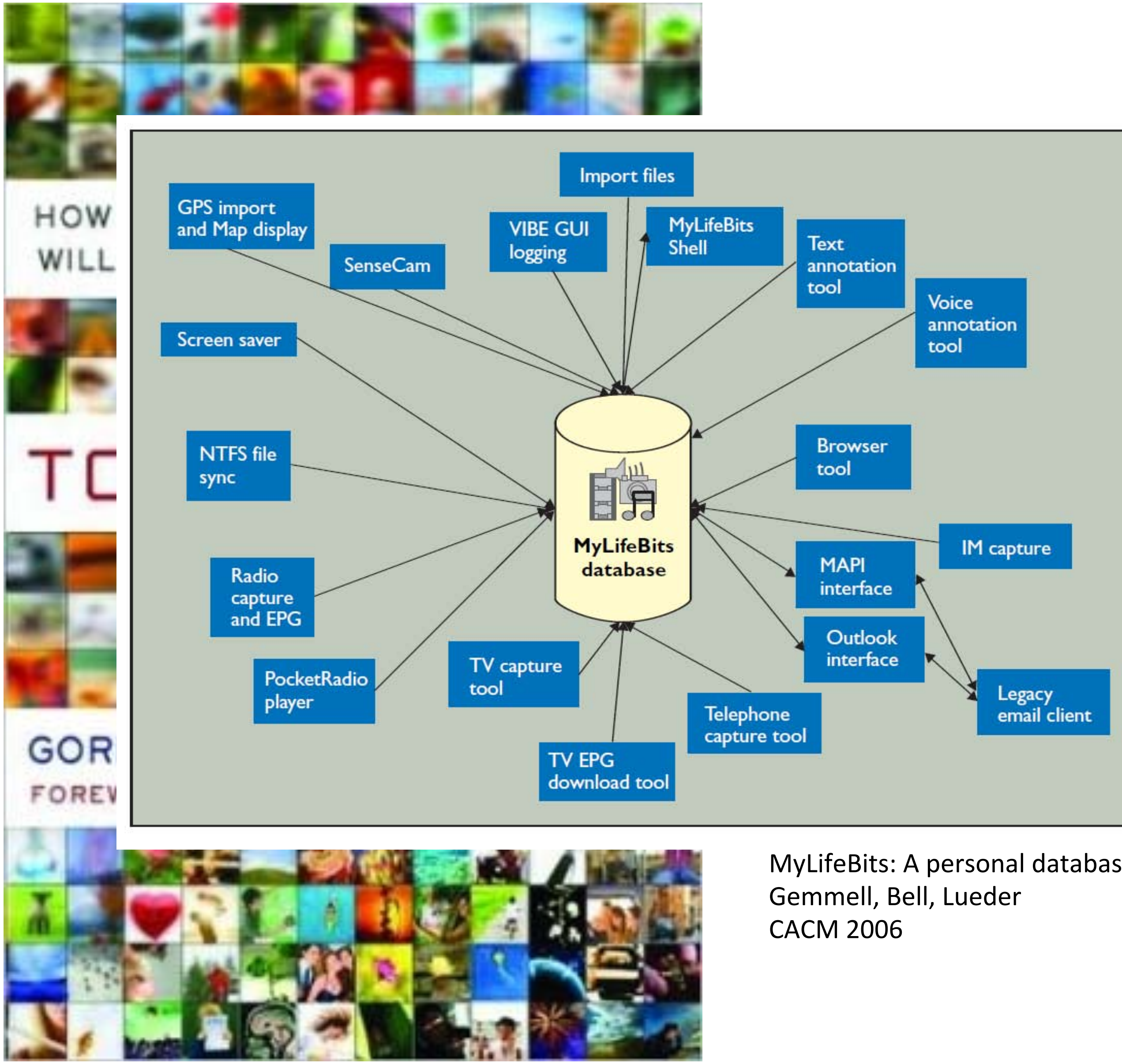
112

「LIFE」1945年9月10日号に特集された「AS WE MAY THINK」のページ

「LIFE」1945年9月10日号より引用

「AS WE MAY THINK」 From the Atlantic Monthly, July 1945


HOW THE E-MEMORY REVOLUTION
WILL CHANGE EVERYTHING
TOTAL RECALL
GORDON BELL AND JIM GEMMELL
FOREWORD BY BILL GATES



MyLifeBits: A personal database for everything
Gemmell, Bell, Lueder
CACM 2006

DOI:10.1145/1735223.1735243

Rather than try to capture everything, system design should focus on the psychological basis of human memory.

BY ABIGAIL SELLEN AND STEVE WHITTAKER

Beyond Total Capture: A Constructive Critique of Lifelogging

WHAT IF WE could digitally capture *everything* we do and see? What if we could save every bit of information we touch and record every event we experience? What would such a personal digital archive be like, and how might it affect the way we live? This vision of a complete "lifelog" is the holy grail for many technologists and researchers who consider us to be on the brink of an "e-memory" revolution.

In the past few years, capturing "Memories for Life" has become a U.K. Grand Challenge in Computing (http://www.nesc.ac.uk/esi/events/Grand_Challenges/proposals/), and many research programs today are dedicated to developing technologies to support the archiving of vast amounts of personal data. A 2009 book[3] by Gordon Bell and Jim Gemmell outlined an enthusiastic view of a future in which technology enables "total recall" of our lives through "total capture" of personally relevant information. Such information includes the paper and digital documents we work on or look at; email, paper mail, and instant messages sent and received; content of telephone conversations; Web sites visited; and charge-card transactions. Also included are data from other everyday activities (such as still images, video, ambient sound and location data). Finally, these personal archives might also be supplemented with environmental measures (such as light intensity and temperature variation) and even internal, biosensor data (such as heart rate and galvanic skin-response measures) reflecting our physical and emotional state.

Constructing such a diverse archive of personal information requires a range of technologies for its capture, management, and storage. Today's advances in wearable sensors, networking capabilities, and massive increases in digital-storage capacity mean this vision is feasible, fueling enthusiasm for the possibilities offered by the technology itself.

Further impetus comes from speculation about what a comprehensive lifelog might do and how it might change our lives. As outlined in 2006 by Czerwinski et al.,[9] lifelog-

» key insights

- Focusing on "total capture," current approaches to lifelogging have failed to explore what practical purpose such exhaustive personal digital records might actually serve.
- Evaluating new approaches, psychology has emerged as an underexploited resource in defining the nature of human memory and its key processes and weaknesses.
- Psychology as design framework could help define the types of memory such systems should support, along with their key interface properties and need to work in synergy with human memory, rather than as its replacement.

DOI:10.1145/1735223.1735243

Rather than try to capture everything, system design should focus on the psychological basis of human memory.

BY ABIGAIL SELLEN AND STEVE WHITTAKER

book[1] by Gordon Bell and Jim Gemmell outlined an enthusiastic view of a future in which technology enables "total recall" of our lives through "total capture" of personally relevant information. Such information includes the paper and digital documents we work on or look at; email, paper mail, and instant messages sent and received; content of telephone conversations;

- **Focusing on "total capture," current approaches to lifelogging have failed to explore what practical purpose such exhaustive personal digital records might actually serve.**

enthusiasm for the possibilities offered by the technology itself.

Further impetus comes from speculation about what a comprehensive lifelog might do and how it might change our lives. As outlined in 2006 by Czerwinski et al.,[9] lifelog-

WHAT IF WE could digitally capture *everything* we do and see? What if we could save every bit of information we touch and record every event we experience? What would such a personal digital archive be like, and how might it affect the way we live? This vision of a complete "lifelog" is the holy grail for many technologists and researchers who consider us to be on the brink of an "e-memory" revolution.

In the past few years, capturing "Memories for Life" has become a U.K. Grand Challenge in Computing (http://www.nesc.ac.uk/esi/events/Grand_Challenges/proposals/), and many research programs today are dedicated to developing technologies to support the archiving of vast amounts of personal data. A 2009

» key insights

- Focusing on "total capture," current approaches to lifelogging have failed to explore what practical purpose such exhaustive personal digital records might actually serve.
- Evaluating new approaches, psychology has emerged as an underexploited resource in defining the nature of human memory and its key processes and weaknesses.
- Psychology as design framework could help define the types of memory such systems should support, along with their key interface properties and need to work in synergy with human memory, rather than as its replacement.

Potential Applications?

- **Physical and Psychological Health Monitoring**
  - Behavioral modification, e.g., monitoring + feedback to reduce stress
  - Monitoring of existing conditions, e.g., depression
  - Early-warning via symptoms, e.g., Alzheimer's
- **Information Management Tools**
  - Search and retrieval of personal information
  - Ranking and prioritizing (e.g., email)
- **Sustainability**
  - Monitoring and feedback of energy consumption
- **Education**
  - Skills assessment, ntegrated with online learning

Opportunities for Data Mining and Machine Learning

- **Exploratory Data Analysis**
  - Visualization, Clustering, Summarization
- **Social Network Analysis**
  - Analyzing ego-networks over time
- **Time-Series Modeling**
  - Change detection, segmentation, trend analysis
- **Text Analysis**
  - sentiment classification, dialog analysis
- **Prediction**
  - Activity classification, ranking/prioritizing activities

2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops

Fitbit+: A behavior-based intervention system to reduce sedentary behavior

Laura R. Pina
Department of Computer Science & Engineering
University of California, San Diego
9500 Gilman Drive, # 0404
Email: lrpina@cs.ucsd.edu

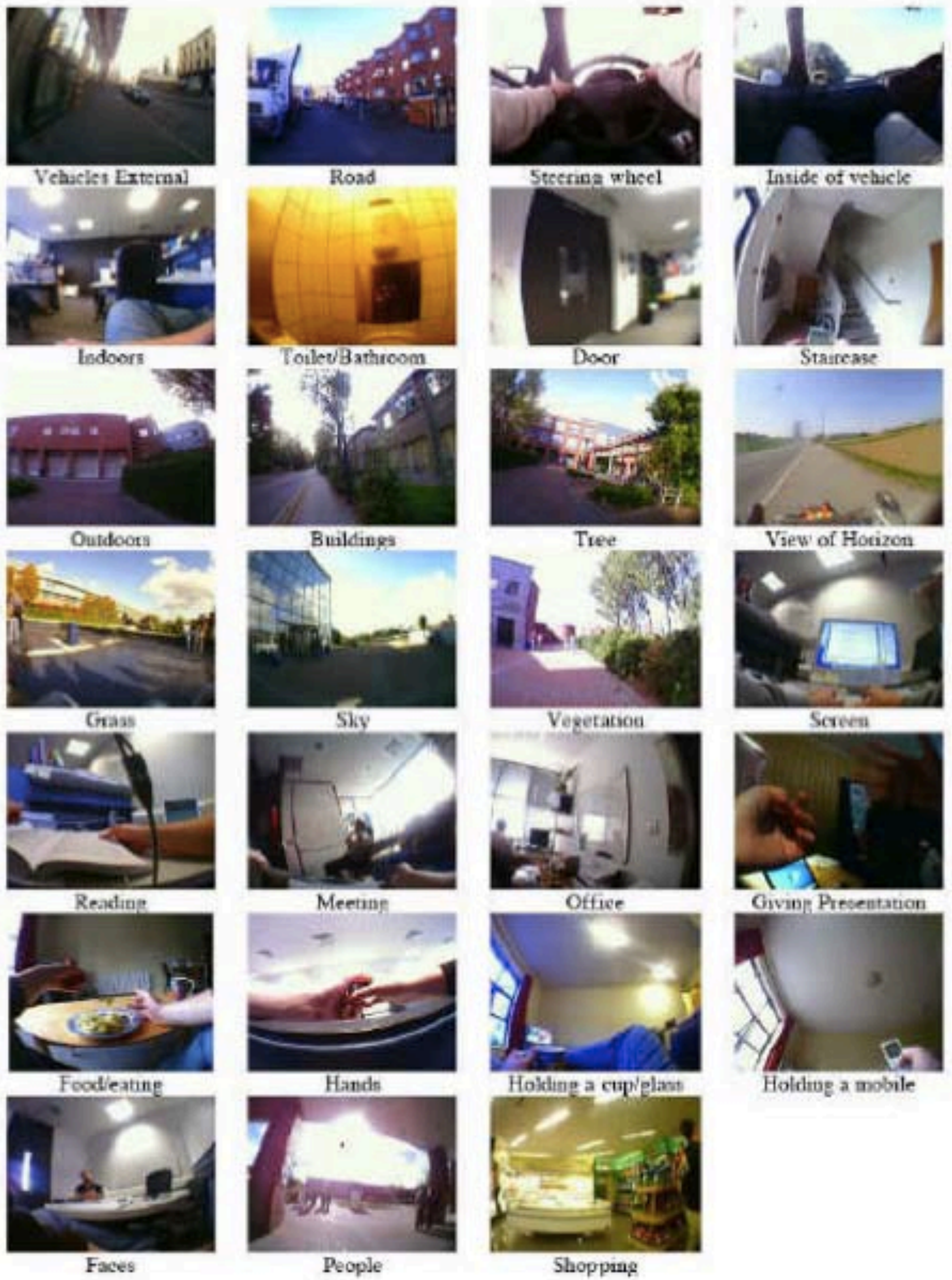
Ernesto Ramirez
Center for Wireless and Population Health Systems
University of California, San Diego
9500 Gilman Drive, # 0404
Email: erramirez@ucsd.edu

William G. Griswold
Department of Computer Science & Engineering
University of California, San Dieg
9500 Gilman Drive, # 0404
Email: wgg@cs.ucsd.edu

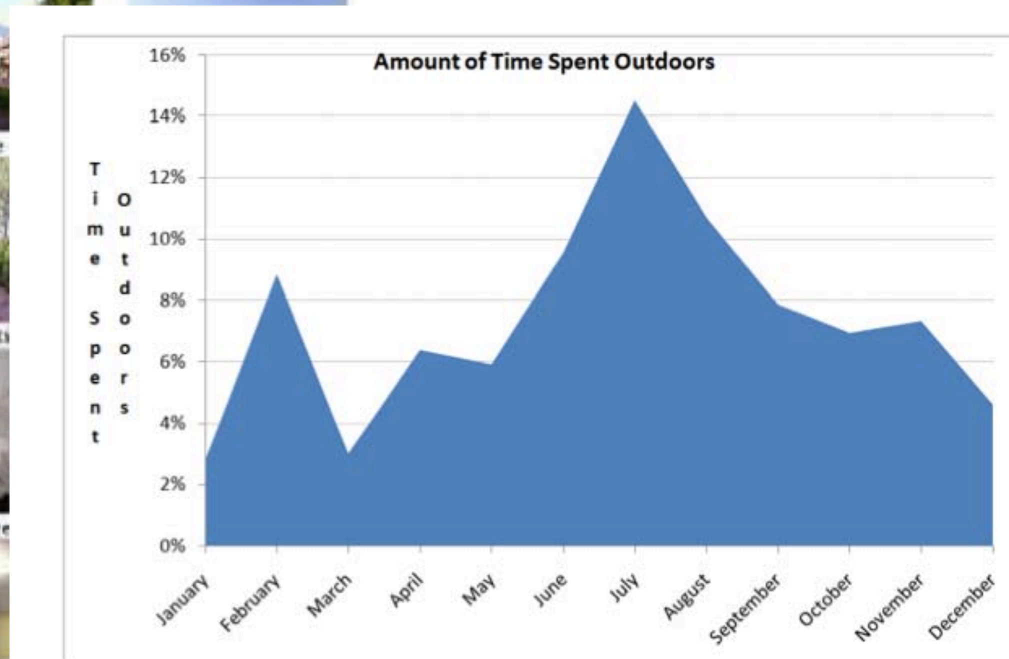
fitbit

***Abstract*—Self-tracking wearable devices are capable of tracking calorie consumption and inferring physical activity physical activity to support self-awareness and healthy behavior. These devices automatically capture human behavior (such as walking) but do not typically make the user aware detected unhealthy behaviors . Furthermore, these devices cannot intervene in the moment to make users aware they are engaging in unhealthy behavior (such as sitting for a long period of time) and persuade them to correct these unhealthy behaviors (e.g., by taking a break**

intensity of physical activity, duration of activity , distance travelled, and estimated caloric expenditure. This data is relayed to the user in two ways. The first, is through the device's screen which gives displays current state of activity, calories, and steps. The second through a website in the way of charts and graphs not only the current date but also summarizing past days.

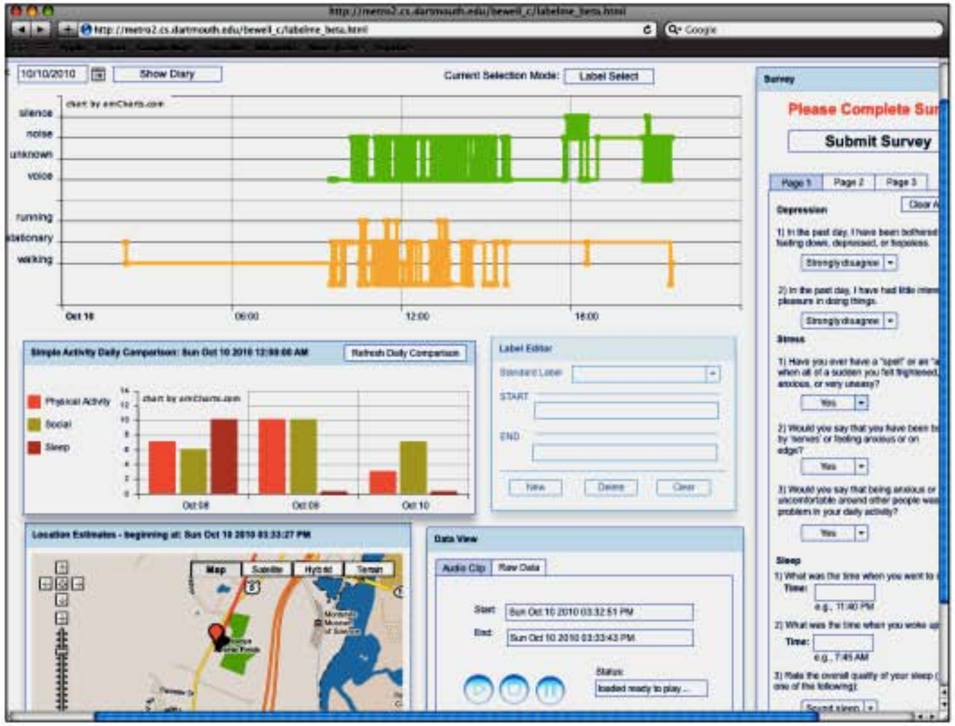


From Doherty et al,
Computers in Human Behavior, 2011

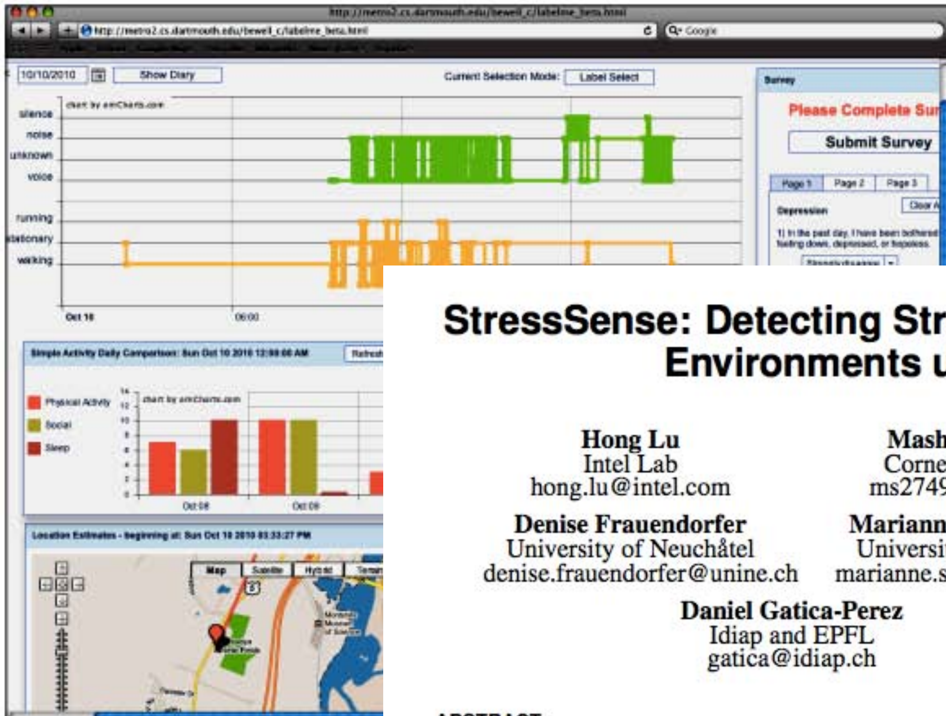


From Doherty et al,
Computers in Human Behavior, 2011

BeWell System, Andrew Campbell, Dartmouth



BeWell System, Andrew Campbell, Dartmouth



StressSense: Detecting Stress in Unconstrained Acoustic Environments using Smartphones

Hong Lu
Intel Lab
hong.lu@intel.com

Mashfiqui Rabbi
Cornell University
ms2749@cornell.edu

Gokul T. Chittaranjan
EPFL
gokul.thattaguppa@idiap.ch

Denise Frauendorfer
University of Neuchâtel
denise.frauendorfer@unine.ch

Marianne Schmid Mast
University of Neuchâtel
marianne.schmid@unine.ch

Andrew T. Campbell
Dartmouth College
campbell@cs.dartmouth.edu

Daniel Gatica-Perez
Idiap and EPFL
gatica@idiap.ch

Tanzeem Choudhury
Cornell University
tanzeem.choudhury@cornell.edu

ABSTRACT

Stress can have long term adverse effects on individuals' physical and mental well-being. Changes in the speech production process is one of many physiological changes that happen during stress. Microphones, embedded in mobile phones and carried ubiquitously by people, provide the opportunity to continuously and non-invasively monitor stress in real-life situations. We propose *StressSense* for unobtrusively recognizing stress from human voice using smartphones. We investigate methods for adapting a one-size-fits-all stress model to individual speakers and scenarios. We demonstrate that the StressSense classifier can robustly iden-

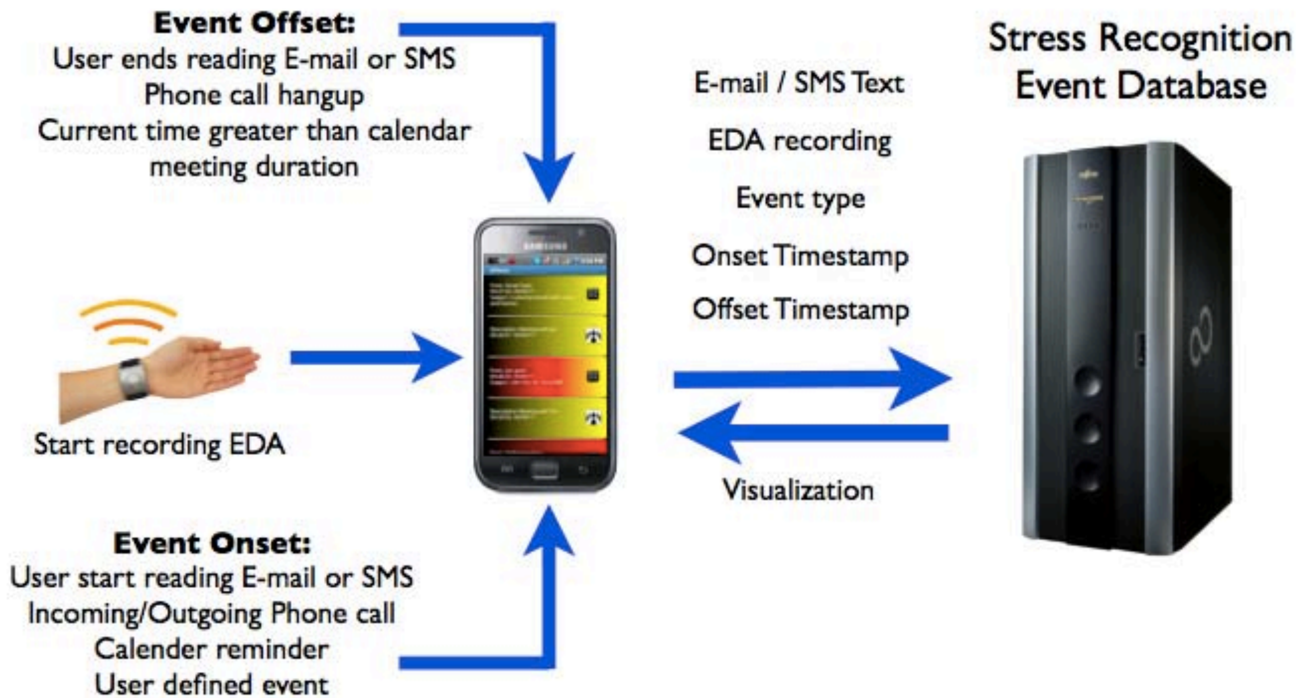
INTRODUCTION

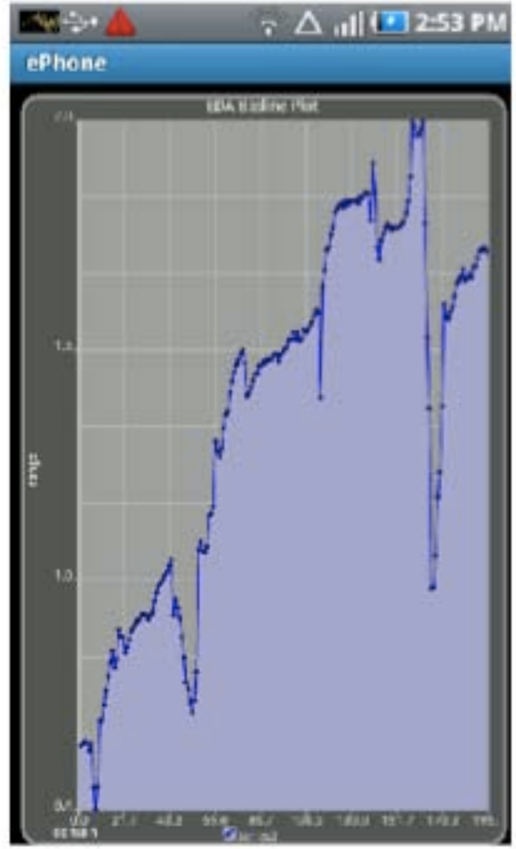
Stress is a universally experienced phenomenon in our modern lives. According to a 2007 study by the American Psychological Association, three quarters of Americans experience stress-related symptoms [1]. Studies have shown that stress can play a role in psychological or behavioral disorders, such as depression, and anxiety [2]. The amount of cumulative stress in daily life may have broad consequences on societal well-being, as stress-causing events have negative impact upon daily health and mood [2] and also contributes significantly to health care costs [3].

Inferring What is Stressful

Ayzenberg, Hernandez, Picard, CHI 2012

Figure 1. Affective Q^{TM} Sensor is a wrist-worn commercial sensor that monitors: 1) EDA, 2) skin temperature and 3) 3-axis accelerometer.





Dear X,

We have just been notified that the deadline for your grant request was not extended. Please submit the request according to the original schedule.

Thanks,
John Smith
Senior Administrator

From Ayzenberg, Hernandez, Picard, CHI 2012

A Wandering Mind Is an Unhappy Mind

Matthew A. Killingsworth* and Daniel T. Gilbert

Unlike other animals, human beings spend a lot of time thinking about what is not going on around them, contemplating events that happened in the past, might happen in the future, or will never happen at all. Indeed, "stimulus-independent thought" or "mind wandering" appears to be the brain's default mode of operation (*1–3*). Although this ability is a remarkable evolutionary achievement that allows people to learn, reason, and plan, it may have an emotional cost. Many philosophical and religious traditions teach that happiness is to be found by living in the moment, and practitioners are trained to resist mind wandering and "to be here now." These traditions suggest that a wandering mind is an unhappy mind. Are they right?

Laboratory experiments have revealed a great deal about the cognitive and neural bases of mind wandering (*3–7*), but little about its emotional consequences in everyday life. The most reliable method for investigating real-world emotion is experience sampling, which involves contacting people as they engage in their everyday activities and asking them to report their thoughts, feelings, and actions at that moment. Unfortunately, collecting real-time reports from large numbers of people as they go about their daily lives is so cumbersome and expensive that experience sampling has rarely been used to investigate the relationship between mind wandering and happiness and has always been limited to very small samples (*8*, *9*).

We solved this problem by developing a Web application for the iPhone (Apple Incorporated, Cupertino, California), which we used to create an unusually large database of real-time reports of thoughts, feelings, and actions of a broad range of people as they went about their daily activities. The application contacts participants through their iPhones at random moments during their waking hours, presents them with questions, and records their answers to a database at www.trackyourhappiness.org. The database currently contains nearly a quarter of a million samples from about 5000 people from 83 different countries who range in age from 18 to 88 and who collectively represent every one of 86 major occupational categories.

To find out how often people's minds wander, what topics they wander to, and how those wanderings affect their happiness, we analyzed samples from 2250 adults (58.8% male, 73.9% residing in the United States, mean age of 34 years) who were randomly assigned to answer a happiness question ("How are you feeling right now?") answered on a continuous sliding scale from very bad (0) to very good (100), an activity question ("What are you doing right now?") answered by endorsing one or more of 22 activities adapted from the day reconstruction method (*10*, *11*), and a mind-wandering question ("Are you thinking about something other than what you're currently doing?") answered with one of four options: no; yes, something pleasant; yes, something neutral; or yes, something unpleasant. Our analyses revealed three facts.

First, people's minds wandered frequently, regardless of what they were doing. Mind wandering occurred in 46.9% of the samples and in at least 30% of the samples taken during every activity except making love. The frequency of mind wandering in our real-world sample was considerably higher than is typically seen in laboratory experiments. Surprisingly, the nature of people's activities had only a modest impact on whether their minds wandered and had almost no impact on the pleasantness of the topics to which their minds wandered (*12*).

Second, multilevel regression revealed that people were less happy when their minds were wandering than when they were not [slope (b) = –8.79, $P < 0.001$], and this was true during all activities, including the least enjoyable. Although people's minds were more likely to wander to pleasant topics (42.5% of samples) than to unpleasant topics (26.5% of samples) or neutral topics (31% of samples), people were no happier when thinking about pleasant topics than about their current activity (b = –0.52, not significant) and were considerably unhappier when thinking about neutral topics (b = –7.2, $P < 0.001$) or unpleasant topics (b = –23.9, $P < 0.001$) than about their current activity (Fig. 1, bottom). Although negative moods are known to cause mind wandering (*13*), time-lag analyses strongly suggested that mind wandering in our sample was generally the cause, and not merely the consequence, of unhappiness (*12*).

Third, what people were thinking was a better predictor of their happiness than was what they were doing. The nature of people's activities explained 4.6% of the within-person variance in happiness and 3.2% of the between-person variance in happiness, but mind wandering explained 10.8% of within-person variance in happiness and 17.7% of between-person variance in happiness. The variance explained by mind wandering was largely independent of the variance explained by the nature of activities, suggesting that the two were independent influences on happiness.

In conclusion, a human mind is a wandering mind, and a wandering mind is an unhappy mind. The ability to think about what is not happening is a cognitive achievement that comes at an emotional cost.

Fig. 1. Mean happiness reported during each activity (**top**) and while mind wandering to unpleasant topics, neutral topics, pleasant topics or not mind wandering (**bottom**). Dashed line indicates mean of happiness across all samples. Bubble area indicates the frequency of occurrence. The largest bubble ("not mind wandering") corresponds to 53.1% of the samples, and the smallest bubble ("praying/worshipping/meditating") corresponds to 0.1% of the samples.

References and Notes

1. M. E. Raichle *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 676 (2001).
2. K. Christoff, A. M. Gordon, J. Smallwood, R. Smith, J. W. Schooler, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8719 (2009).
3. R. L. Buckner, J. R. Andrews-Hanna, D. L. Schacter, *Ann. N. Y. Acad. Sci.* **1124**, 1 (2008).
4. J. Smallwood, J. W. Schooler, *Psychol. Bull.* **132**, 946 (2006).
5. M. F. Mason *et al.*, *Science* **315**, 393 (2007).
6. J. Smallwood, E. Beach, J. W. Schooler, T. C. Handy, *J. Cogn. Neurosci.* **20**, 458 (2008).
7. R. L. Buckner, D. C. Carroll, *Trends Cogn. Sci.* **11**, 49 (2007).
8. J. C. McVay, M. J. Kane, T. R. Kwapil, *Psychon. Bull. Rev.* **16**, 857 (2009).
9. M. J. Kane *et al.*, *Psychol. Sci.* **18**, 614 (2007).
10. D. Kahneman, A. B. Krueger, D. A. Schkade, N. Schwarz, A. A. Stone, *Science* **306**, 1776 (2004).
11. A. B. Krueger, D. A. Schkade, *J. Public Econ.* **92**, 1833 (2008).
12. Materials and methods are available as supporting material on *Science* Online.
13. J. Smallwood, A. Fitzgerald, L. K. Miles, L. H. Phillips, *Emotion* **9**, 271 (2009).
14. We thank V. Pitiyanuvath for engineering www.trackyourhappiness.org and R. Hackman, A. Jenkins, W. Mendes, A. Oswald, and T. Wilson for helpful comments.

Supporting Online Material
www.sciencemag.org/cgi/content/full/330/6006/932/DC1
Materials and Methods
Table S1
References

18 May 2010; accepted 29 September 2010
10.1126/science.1192439

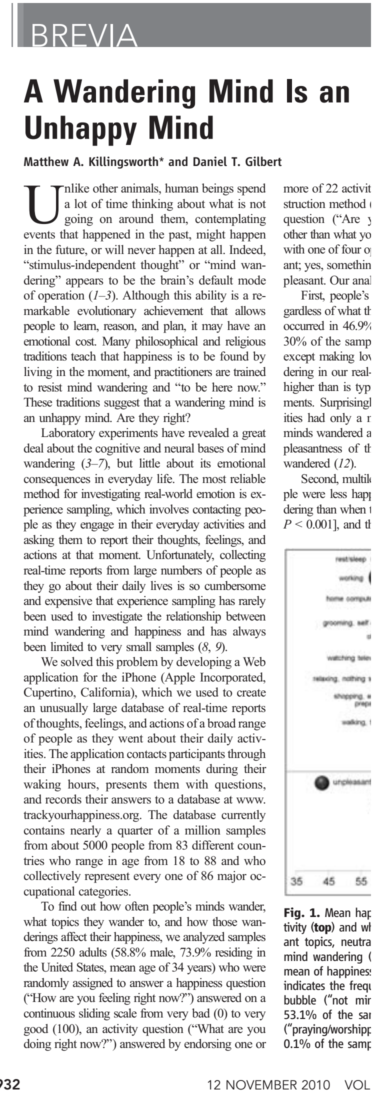
Harvard University, Cambridge, MA 02138, USA.

*To whom correspondence should be addressed. E-mail: mkilling@fas.harvard.edu

Killingworth and Gilbert, Science, 2010

5000 individuals

250,000 self-reports from a Web app



BREVIA

A Wandering Mind Is an Unhappy Mind

Matthew A. Killingsworth* and Daniel T. Gilbert

Unlike other animals, human beings spend a lot of time thinking about what is not going on around them, contemplating events that happened in the past, might happen in the future, or will never happen at all. Indeed, “stimulus-independent thought” or “mind wandering” appears to be the brain’s default mode of operation (1–3). Although this ability is a remarkable evolutionary achievement that allows people to learn, reason, and plan, it may have an emotional cost. Many philosophical and religious traditions teach that happiness is to be found by living in the moment, and practitioners are trained to resist mind wandering and “to be here now.” These traditions suggest that a wandering mind is an unhappy mind. Are they right?

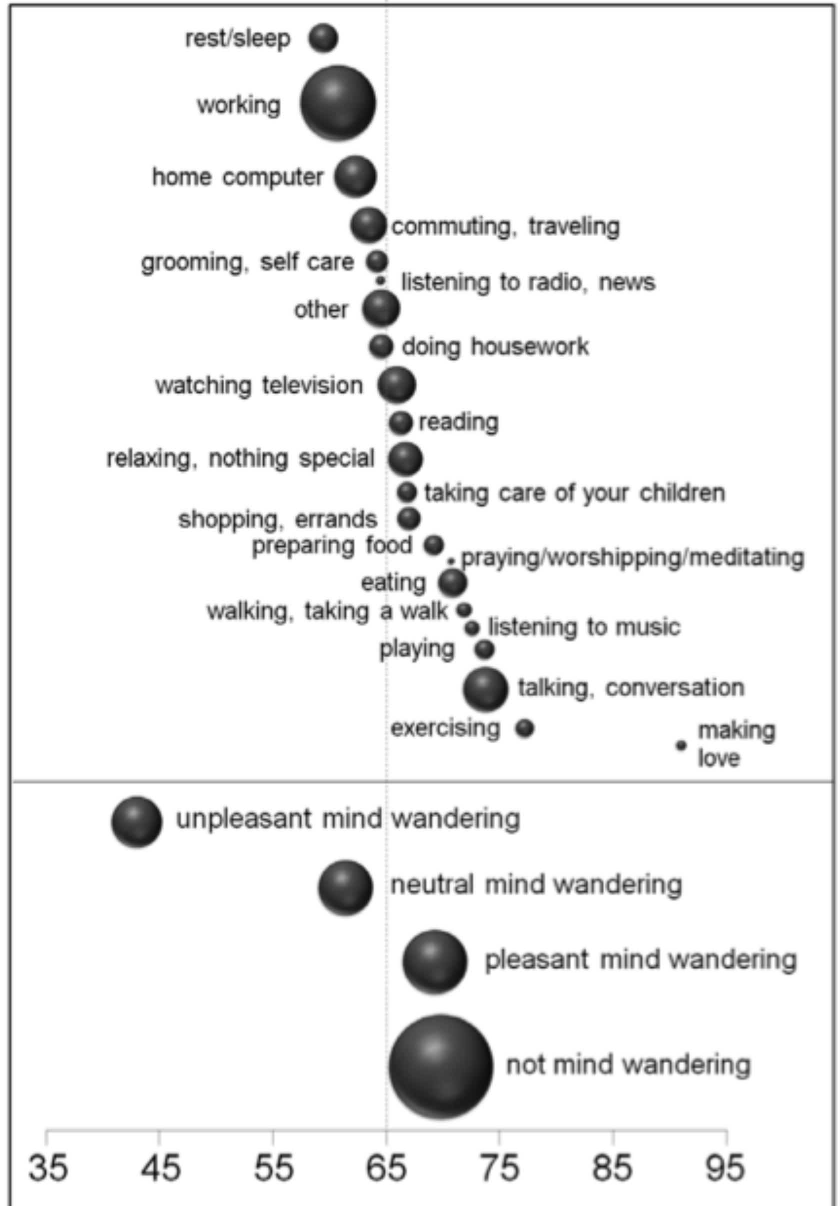
Laboratory experiments have revealed a great deal about the cognitive and neural bases of mind wandering (3–7), but little about its emotional consequences in everyday life. The most reliable method for investigating real-world emotion is experience sampling, which involves contacting people as they engage in their everyday activities and asking them to report their thoughts, feelings, and actions at that moment. Unfortunately, collecting real-time reports from large numbers of people as they go about their daily lives is so cumbersome and expensive that experience sampling has rarely been used to investigate the relationship between mind wandering and happiness and has always been limited to very small samples (8, 9).

We solved this problem by developing a Web application for the iPhone (Apple Incorporated, Cupertino, California), which we used to create an unusually large database of real-time reports of thoughts, feelings, and actions of a broad range of people as they went about their daily activities. The application contacts participants through their iPhones at random moments during their waking hours, presents them with questions, and records their answers to a database at www.trackyourhappiness.org. The database currently contains nearly a quarter of a million samples from about 5000 people from 83 different countries who range in age from 18 to 88 and who collectively represent every one of 86 major occupational categories.

To find out how often people’s minds wander, what topics they wander to, and how those wanderings affect their happiness, we analyzed samples from 2250 adults (58.8% male, 73.9% residing in the United States, mean age of 34 years) who were randomly assigned to answer a happiness question (“How are you feeling right now?”) answered on a continuous sliding scale from very bad (0) to very good (100), an activity question (“What are you doing right now?”) answered by endorsing one or

more of 22 activit
struction method (
question (“Are y
other than what yo
with one of four o
ant; yes, somethin
pleasant. Our anal
First, people’s
gardless of what th
occurred in 46.9%
30% of the samp
except making lov
dering in our real-
higher than is typ
ments. Surprisingl
ities had only a r
minds wandered a
pleasantness of th
wandered (12).
Second, multil
ple were less hap
dering than when t
$P < 0.001$], and th

Fig. 1. Mean hap
tivity (**top**) and wh
ant topics, neutra
mind wandering (
mean of happiness
indicates the frequ
bubble (“not min
53.1% of the sam
(“praying/worshipp
0.1% of the samp



, Science, 2010

m a Web app

What makes you happy?

Get started
(Sign in if you already have an account)

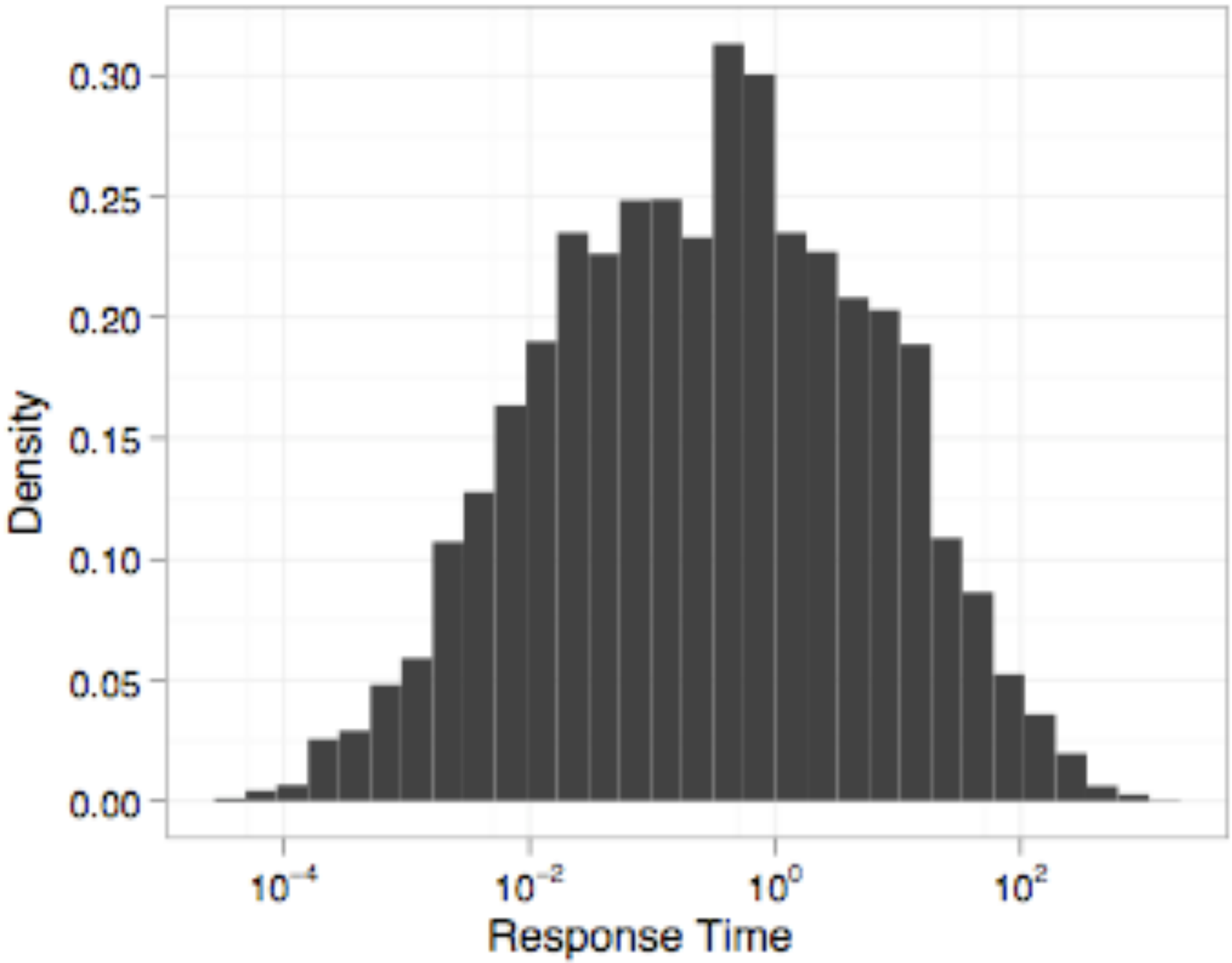
Track Your Happiness.org is a new scientific research project that investigates what makes life worth living.

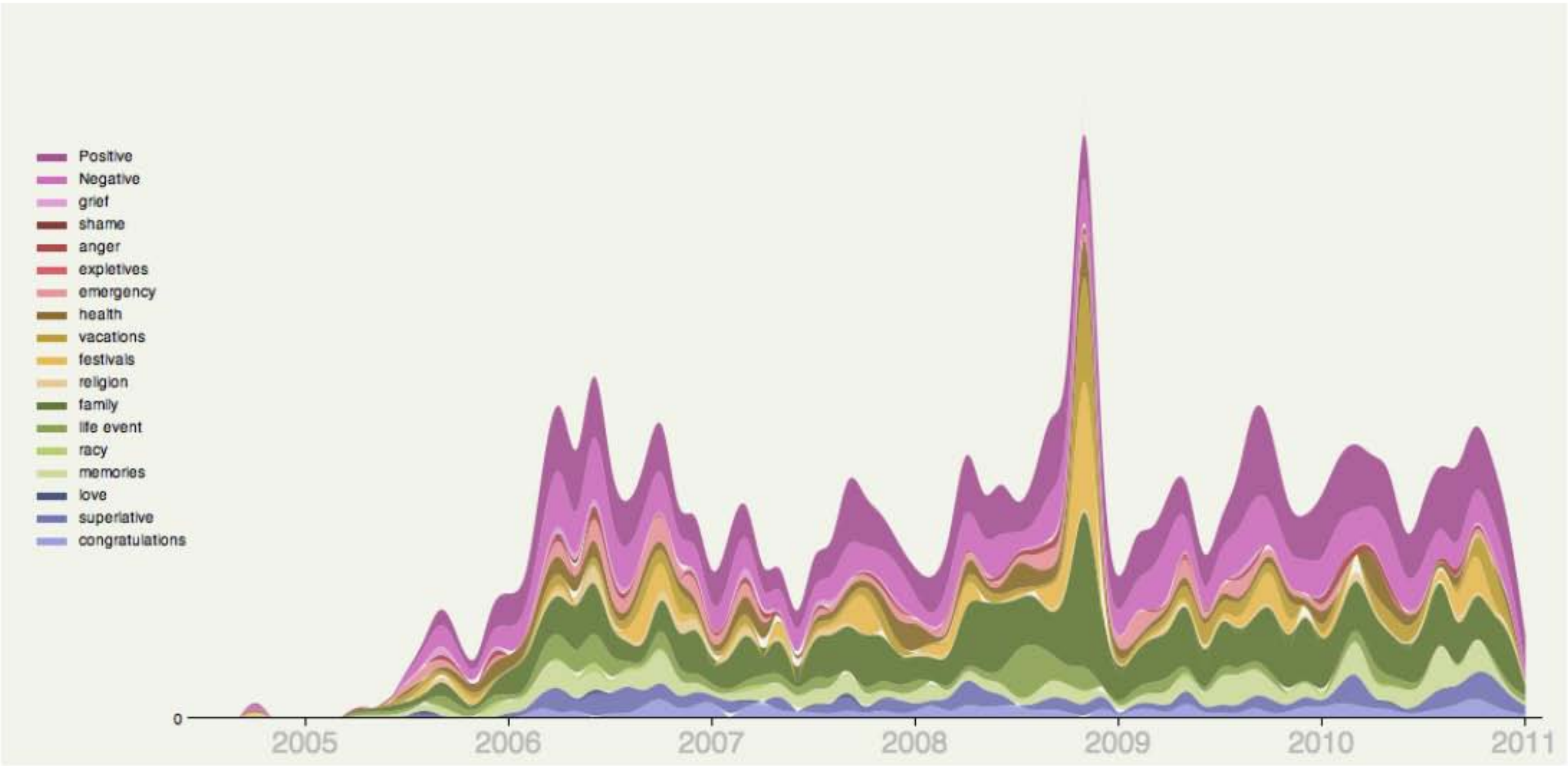
Using this site, you'll be able to track your happiness and find out what factors – for you personally – are associated with greater happiness. You'll also contribute to our scientific understanding of happiness.

How does it work?

1. **Answer a few questions**
   First we'll ask you some questions for statistical purposes. This will take about 10 minutes.
2. **Track your happiness**
   Using your iPhone, you'll be notified by email or text message and asked to report how you are feeling and what you are doing. You decide when and how often you want to be notified.
3. **Your Happiness Report**
   This report will show how your happiness varies depending on what you are doing, who you are with, where you are, what time of day it is, and a variety of other factors.

Email Response Time (log-scale)





From Hangal, Lam, Heer, UIST 2011
MUSE: Reviving memories using email archives

A Random Selection of Personal Photos

P. Sinha, WWW 2011


Xian, China - Daytrip Museums, 2009:09:20


London, UK - Daytrip Sight Seeing, 2009:05:06


Beijing, China - Daytrip Sight Seeing, 2009:09:16


Angkor Wat, Cambodia - Vacation/ Travel, 2009:08:23


London, UK - Daytrip Sight Seeing, 2009:05:06


Xian, China - Daytrip Sight Seeing, 2009:09:19


Beijing, China - Daytrip Sight Seeing, 2009:09:16


Xian, China - Daytrip Museums, 2009:09:20


Shanghai, China - Daytrip Sight Seeing, 2009:09:12


Nagpur, India - Ceremonies Anniversary, 2009:04:13


Beijing, China - Daytrip Sight Seeing, 2009:09:17


Xian, China - Daytrip Museums, 2009:09:20

System-Generated Photo Summary

P. Sinha, WWW 2011

Beijing, China - Daytrip Sight Seeing, 2009:09:14


Irvine, California - Party Birthday, 2009:02:07


New Delhi, India - Prof Trip, 2009:12:18


London, UK - Daytrip Sight Seeing, 2009:05:06


Irvine, California - Holiday Christmas, 2009:12:25


Nagpur, India - Ceremonies Anniversary, 2009:04:13


Angkor Wat, Cambodia - Vacation/ Travel, 2009:08:23


San Diego, California - Party Others, 2009:03:08


Irvine, California - Holiday Thanksgiving, 2009:11:26


Bath, UK - Daytrip Museums, 2009:05:09


Shanghai, China - Daytrip Sight Seeing, 2009:09:13


Irvine, California, California - Family/ Friends, 2009:02:14

JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1979, **12**, 173-184 NUMBER 2 (SUMMER 1979)

EFFECTS OF SELF-MONITORING AND FEEDBACK ON RESIDENTIAL ELECTRICITY CONSUMPTION

RICHARD A. WINETT, MICHAEL S. NEALE, AND H. CANNON GRIER

INSTITUTE FOR BEHAVIORAL RESEARCH, INC., SILVER SPRING, MARYLAND

Prior research has indicated that frequent feedback could reduce residential electricity consumption by 10% to 15%. However, because feedback was primarily given in written form, this procedure might not be practical. The present study evaluated a potentially more practical feedback procedure during peak-use periods with high electricity consuming households. The study was conducted during the winter in an upper-middle class neighborhood of almost identical, all-electric townhouses (N = 71) that averaged about 170 KWH per day per household for a monthly bill of over $200. Twelve households received daily written feedback. Sixteen households (self-monitoring) were taught to read their outdoor electricity meter and to record KWH used every day. A comparison group was composed of 14 households that had volunteered to participate and 29 others that had only given permission to have their meters read. During a 1-month period that the procedures were in effect, the feedback group reduced consumption by 13% and the self-monitoring group by about 7%. These reductions, relative to the comparison group, were maintained during an early spring 1-month follow-up period and, to a lesser extent, during a 6-week warm spring period. Self-monitoring participants were highly reliable and persistent meter readers. Reductions in electricity use were reported by households to be largely attributable to lowering of the heat thermostat, and large monetary and KWH savings were found. Techniques to make self-monitoring cost-effective important components of the self-monitoring procedure, methods to apply self-monitoring more broadly, and plans to combine behavioral procedures with physical technology are discussed.

DESCRIPTORS: behavioral community psychology, feedback, self-monitoring, energy consumption, energy conservation, households

Research Challenges

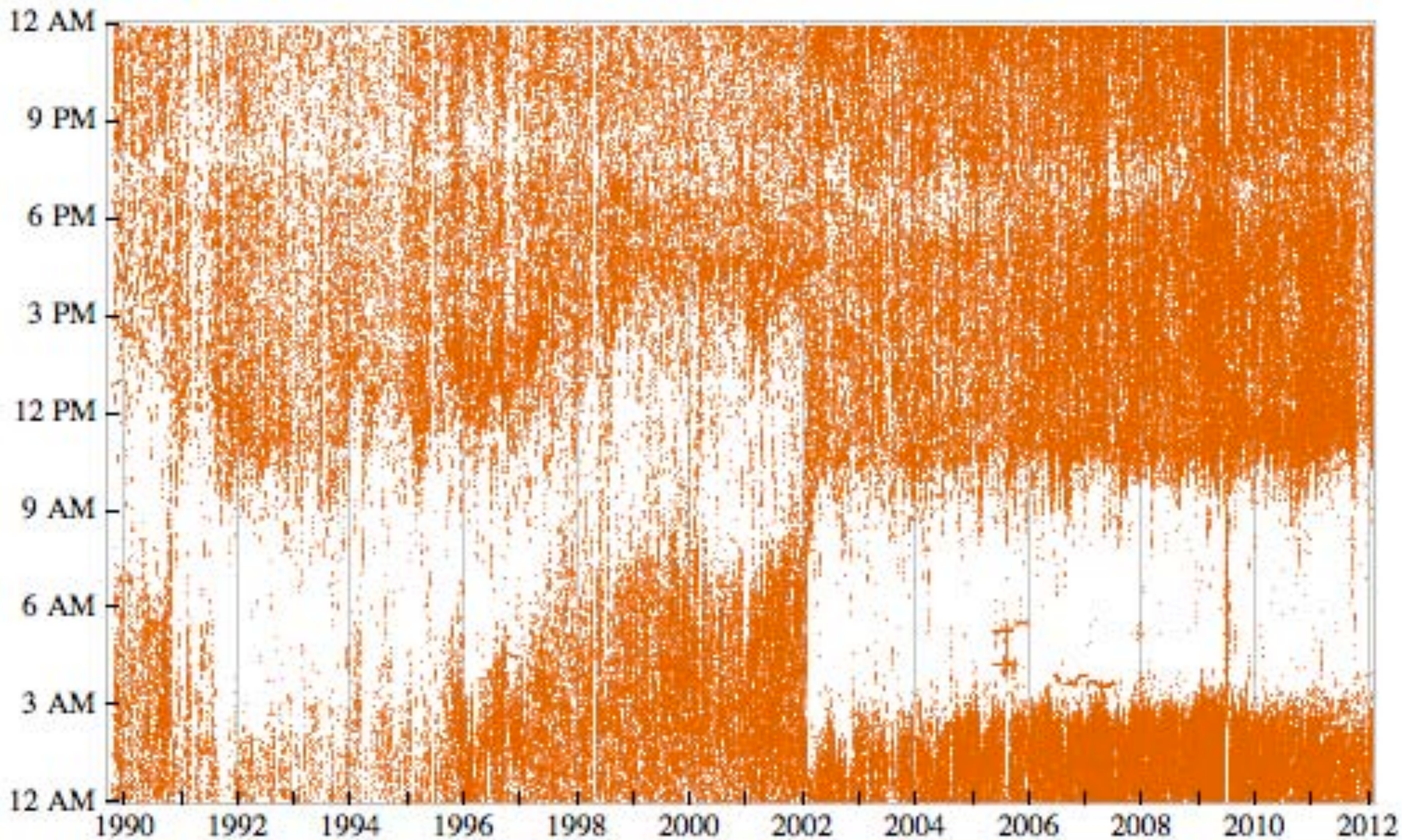
- Non-IID data
- Non-stationary, temporal variability
- Context (e.g., time of day, calendar effects)
- Multi-modal data
- Privacy issues
- And more…..

Example: Analyzing Personal Email Histories

For more details see

Navaroli, Dubois, Smyth, ACML 2012/ML Journal 2013

Email Data


12 AM
9 PM
6 PM
3 PM
12 PM
9 AM
6 AM
3 AM
12 AM
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012

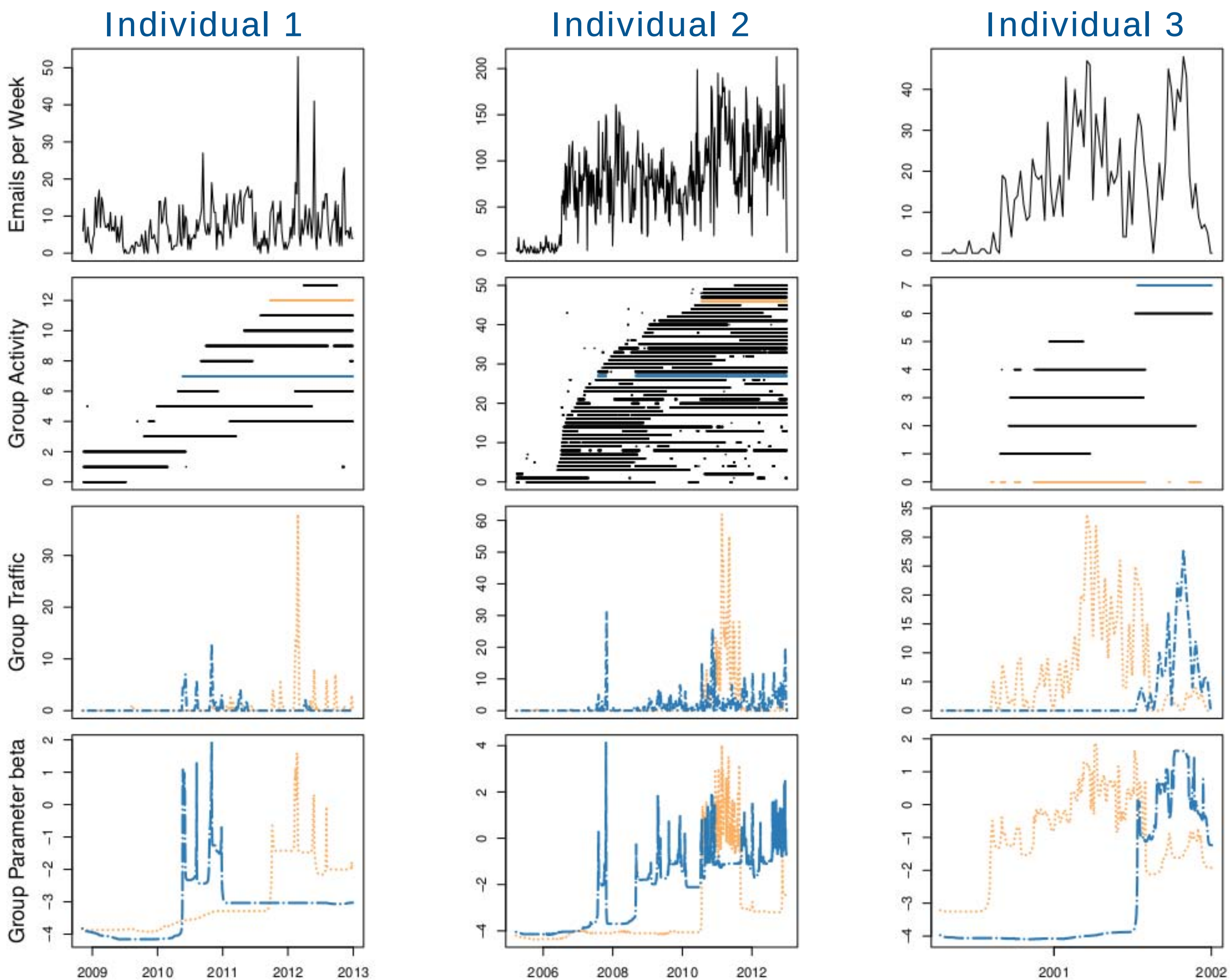
Email Recipient Data

Email ID	Day	Recipient IDs
1	t	{1,3,5}
2	t	{3}
3	t+1	{5, 9}
4	t+2	{1, 3, 4, 6, 8}
5	t+2	{2, 5}
...	...	....

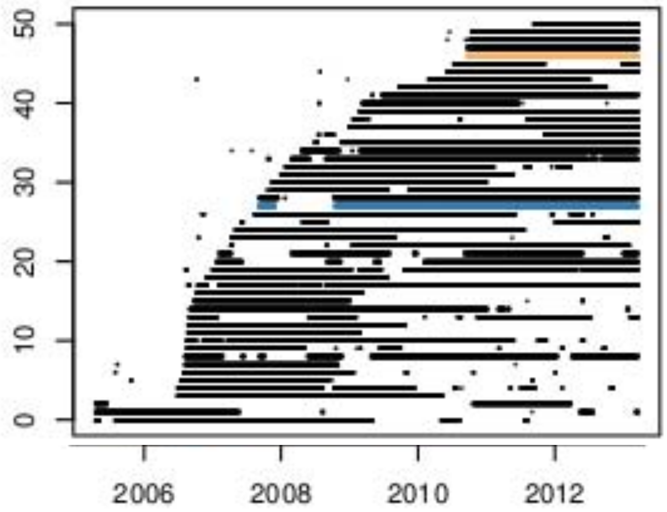
Learning Groups and Segments

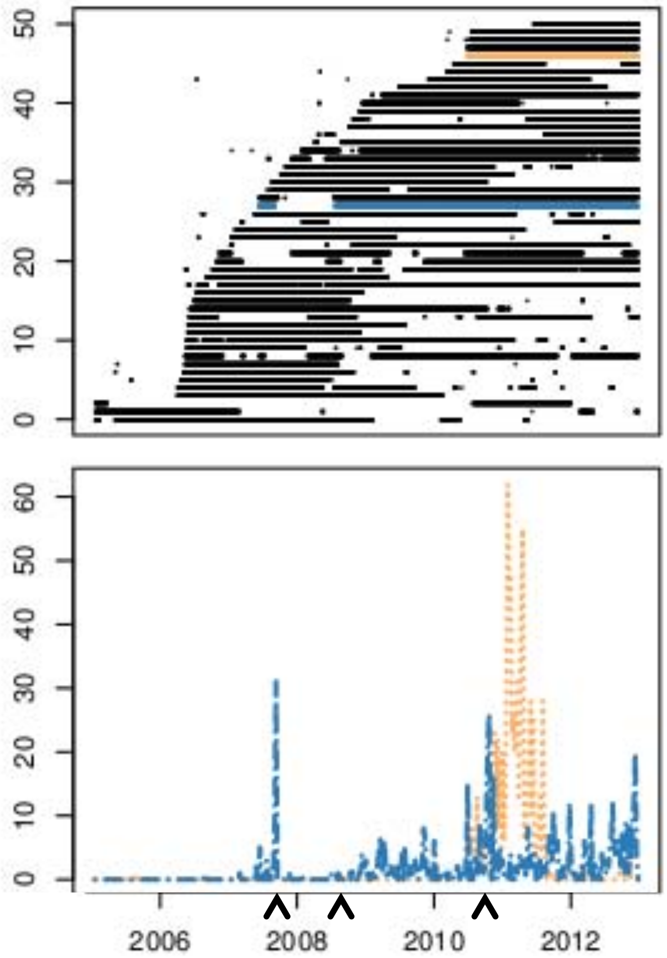
Navaroli, Dubois, Smyth, ACML 2012/ML Journal 2013

- **Each email is assumed to come from 1 of K latent groups**
  - Group k = set of conditionally independent Bernoullis over recipients
- **Group k has a Poisson rate λ_{kt} for day t**
  - P(email is sent to group k | day t) proportional to λ_{kt}
- **Group rates λ_{kt} are piecewise constant over time**
  - Unobserved number and location of segment boundaries, per group
- **Learning via Markov Chain Monte Carlo**
  - Algorithm learns groups, Poisson rates over time, and segment boundaries



Individual 1
Individual 2
Individual 3
Emails per Week
Group Activity
Group Traffic
Group Parameter beta
2009
2010
2011
2012
2013
2006
2008
2010
2012
2001
2002

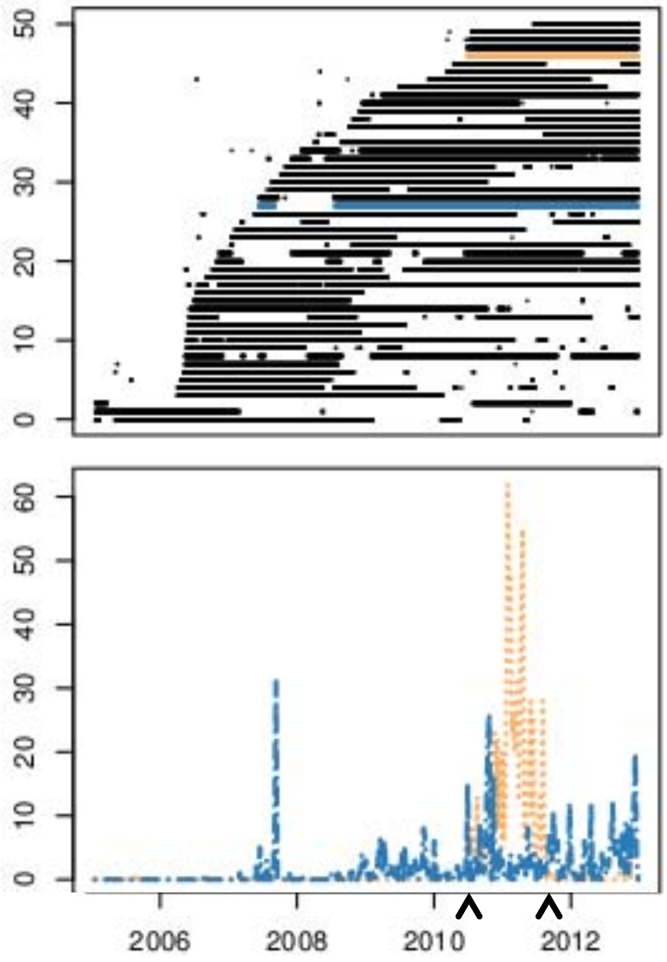

50
40
30
20
10
0
2006
2008
2010
2012



Proposal writing

Proposal awarded

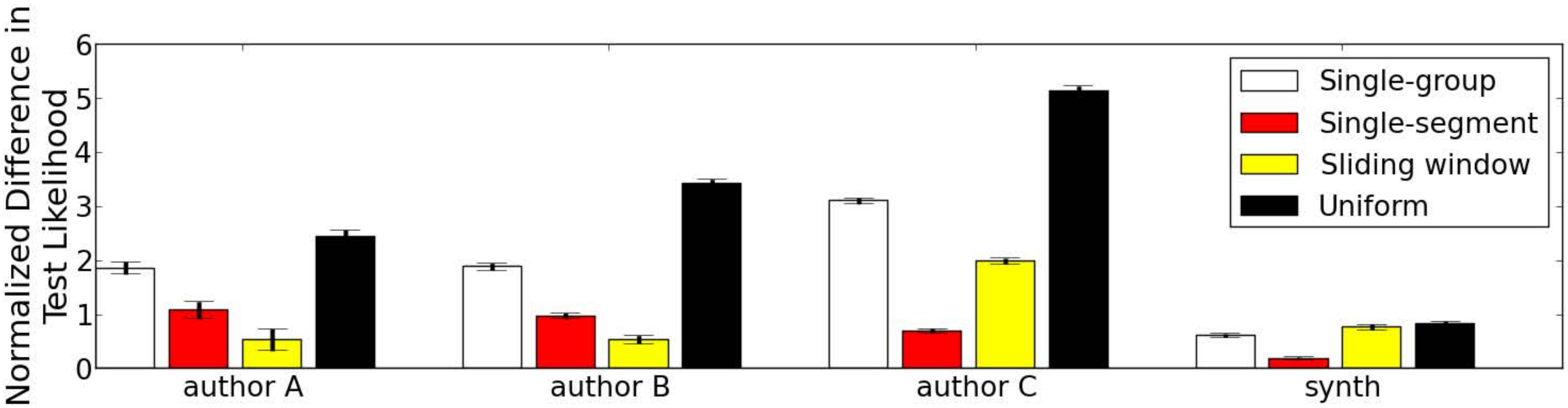
Project activities



KDD 2011 Planning Kickoff

KDD 2011 Conference

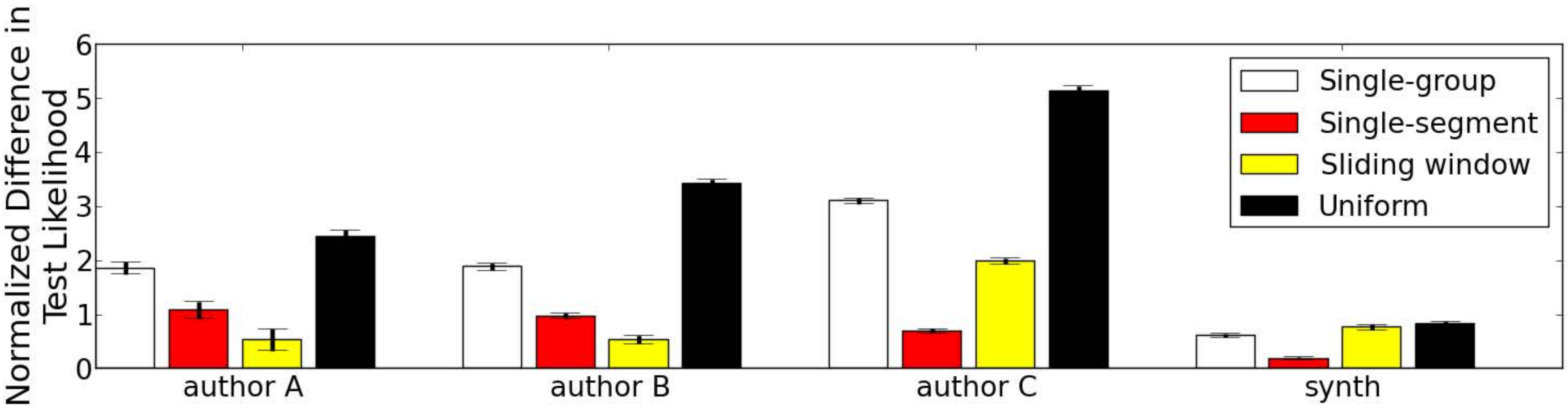
Predictive Performance

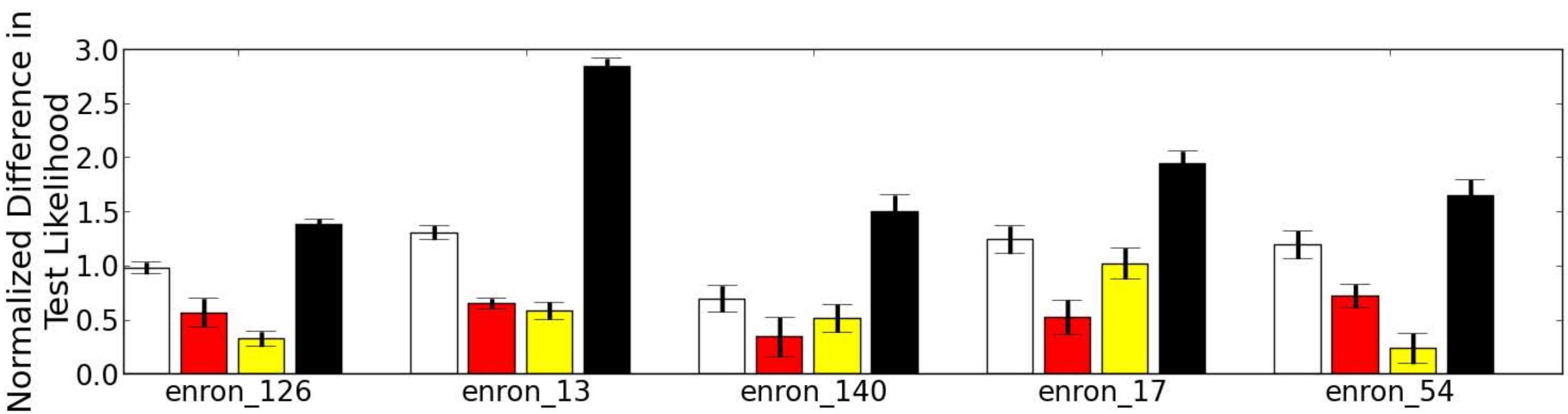


- **Y-axis: difference in log-likelihood relative to proposed model**
  - Smaller is better
  - Zero = proposed model

- **Baselines:**
  - Uniform (overly simple…but calibrates y-axis)
  - Single group
  - Single time segment
  - Sliding window

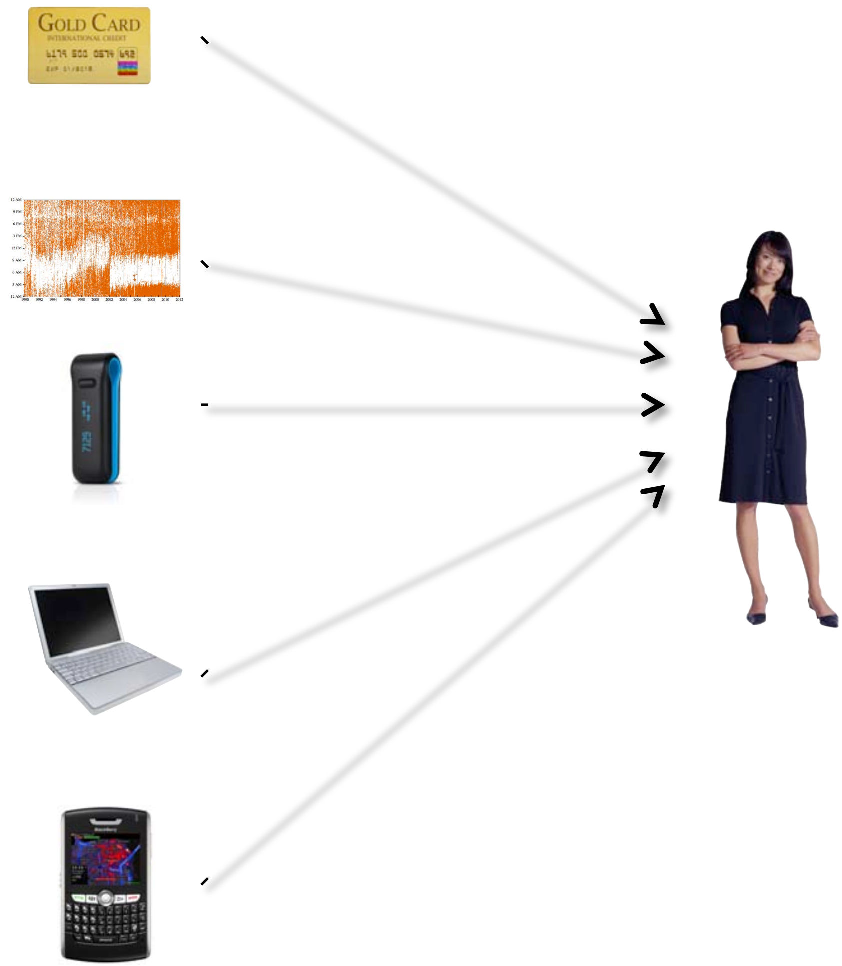
Predictive Performance





Concluding Comments

The Individual's View of Individual Data



Privacy and Data Sharing

- **Data sharing**
  - Surprising willingness of individuals to share data
  - Medical/health data is however more sensitive than Web clicks
  - Legal limits on data sharing between companies
- **Opt-in models seem likely**
  - Default is that only the individual gets to see and analyze their combined data
  - modeling/analysis is local, no sharing of data across individuals
  - Individuals may be willing to share their combined data on an opt-in basis
- **Not clear yet the balance between open-source/research and commercial involvement**

Conclusions and Predictions

- **The technology exists to measure and record every aspect of our daily lives**
- **Potentially tremendous benefits in physiological and behavioral health**
- **However, we do not know how to adequately analyze and make predictions with this type of data**
  - Ample opportunities for data mining and machine learning
- **This is a brave new world … its coming whether we like it or not**