Experience Sampling

Simon Dennis

Complex Human Data Hub
University of Melbourne
Unforgettable Technologies Pty Ltd

Schedule for Today

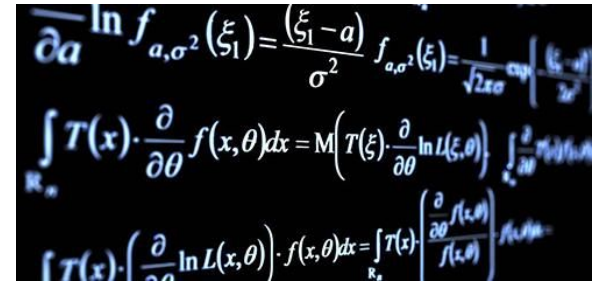
9:30 am	Experience Sampling: Utopian and Dystopian Views
10:30 am	Break
10:45 am	Data Collection Tutorials
11:30 am	Privacy Preserving Data Analysis
12 noon	Lunch
1 pm	Thinking about Privacy
1:30 pm	Experience Sampling Data Analysis
2:15 pm	Break
2:30 pm	Empirical Dynamic Modelling (Michael Zyphur)
3:30 pm	Research Meetings
4:30 pm	Close

Utopian and Dystopian Views

Simon Dennis

Complex Human Data Hub
University of Melbourne
Unforgettable Technologies Pty Ltd

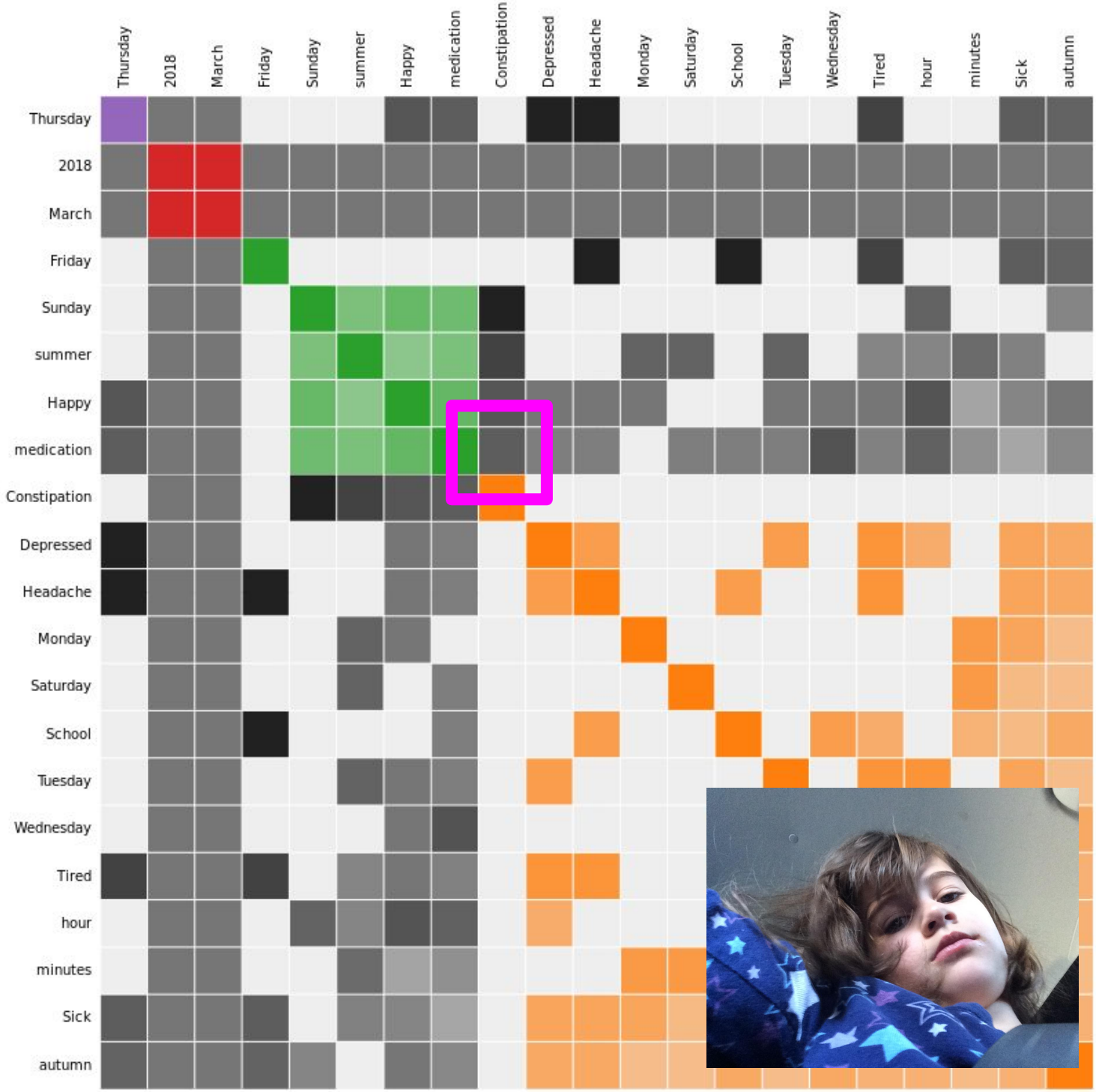

ONE BEAR. TWO SIDES.
YOU CHOOSE!
sky kids
STAR WARS


$\frac{\partial}{\partial a}\ln f_{a,\sigma^2}(\xi_1)=\frac{(\xi_1-a)}{\sigma^2}$
$f_{a,\sigma^2}(\xi_1)=$
$\int_{R_n} T(x)\cdot\frac{\partial}{\partial\theta}f(x,\theta)dx=\mathrm{M}\left(T(\xi)\cdot\frac{\partial}{\partial\theta}\ln L(\xi,\theta)\right)$
$\int T(x)\cdot\left(\frac{\partial}{\partial\theta}\ln L(x,\theta)\right)\cdot f(x,\theta)dx=\int_{R_n} T(x)\cdot$


764.4 mi
2.96

Is Sofie crying too much?

Is melatonin making Phoebe constipated?


Thursday
2018
March
Friday
Sunday
summer
Happy
medication
Constipation
Depressed
Headache
Monday
Saturday
School
Tuesday
Wednesday
Tired
hour
minutes
Sick
autumn

Alibi Retrieval and Memory for WHERE

Ronald Cotton was exonerated in 1995, after spending over 10 years in prison for crimes he did not commit.

His convictions were based largely on an eyewitness misidentification made by one of the victims, Jennifer Thompson-Cannino. However, he also gave a false alibi.

Click marker (📍) where you were on Tuesday 2018-02-13 at 20:00

What aspects of people's data (time, accelerometry, GPS, audio) predict the errors they make.

Mood Correlates in Bipolar Spectrum Disorder

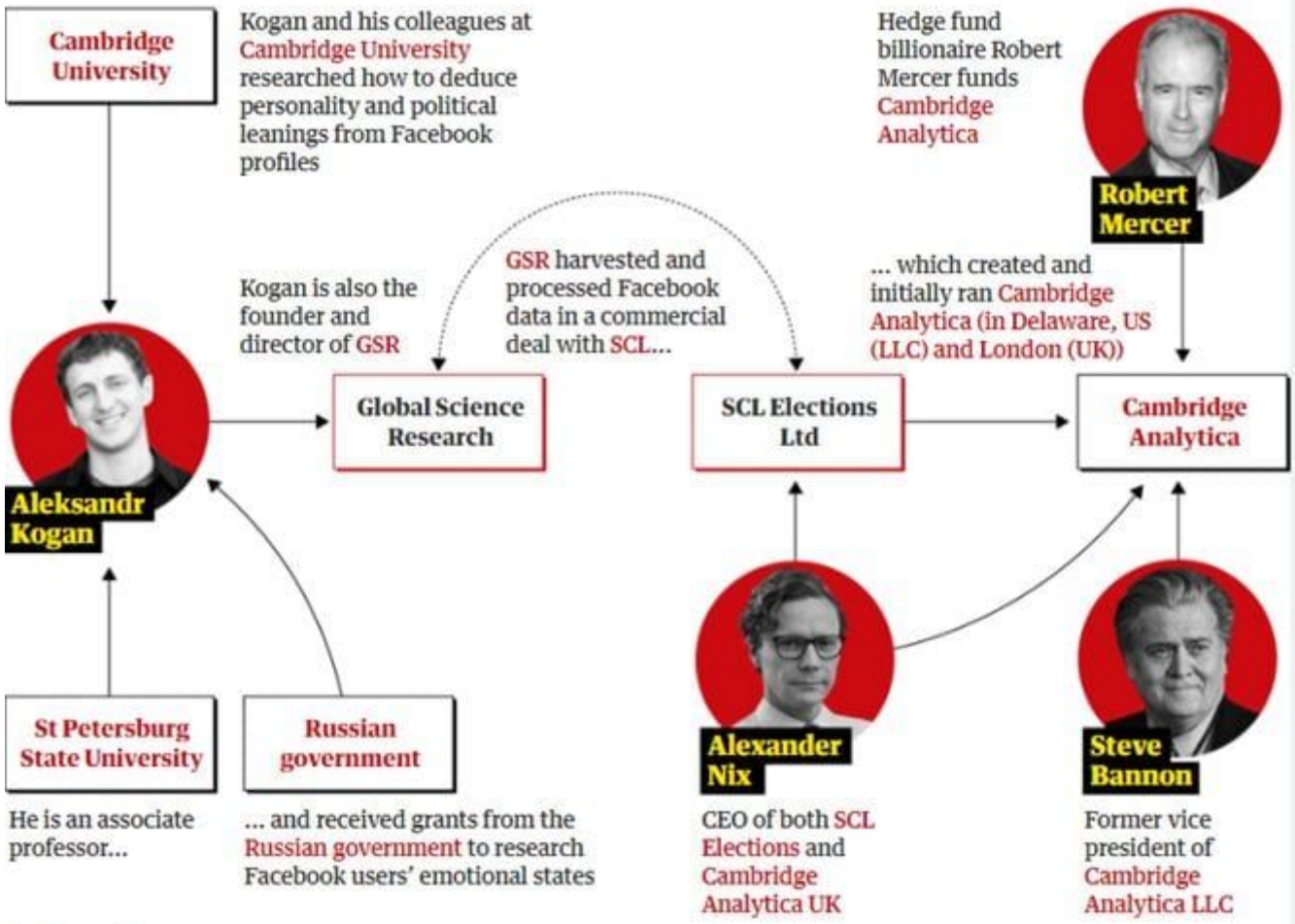
Tracking people with bipolar spectrum disorder over a year to see what aspects of their data (sleep, accelerometry, GPS, audio, phone calls and SMSs) predict mood elevated and depressed states.


ONE BEAR. TWO SIDES.
YOU CHOOSE!
sky kids
STAR WARS

87 million Facebook profiles taken
Used to influence the US presidential election??

Alexandr Kogan is a social psychologist who studies personality
Research associate at Cambridge University
“This is your digital life” app tested the big five
About 270,000 people used the app

Cambridge Analytica: how the key players are linked



Guardian graphic

There was no hack

Sandy Parakilas: The way it works is if you're using an app and I'm your friend, the app can say, "Hey, Lesley, we want to get your data for use in this app, and we also want to get your friends' data." If you say, "I will allow that," then the app gets my data, too.

Lesley Stahl: What you're saying is I give permission for the friend? The friend doesn't give permission?

Sandy Parakilas: Right. It doesn't feel right when you say it out loud?

Lesley Stahl: No, it doesn't feel right.

Sandy Parakilas: Right.

https://www.cbsnews.com/news/aleksandr-kogan-the-link-between-cambridge-analytica-and-facebook/


NBC NEWS
LONDON
NEG
NEWS BITE GLOBAL
HYPOCRITE
POWER LUNCH
FIRING BACK AT FACEBOOK
ALEKSANDR KOGAN SPEAKS OUT
CNBC

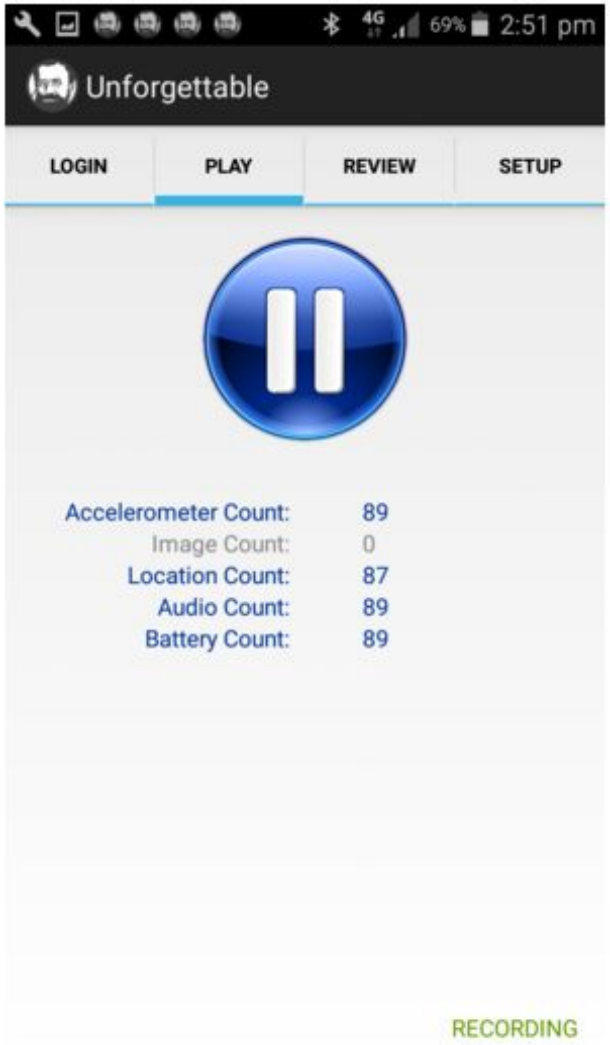
Collection Mechanisms

a) a prominent all stop button to allow users to cease all recording

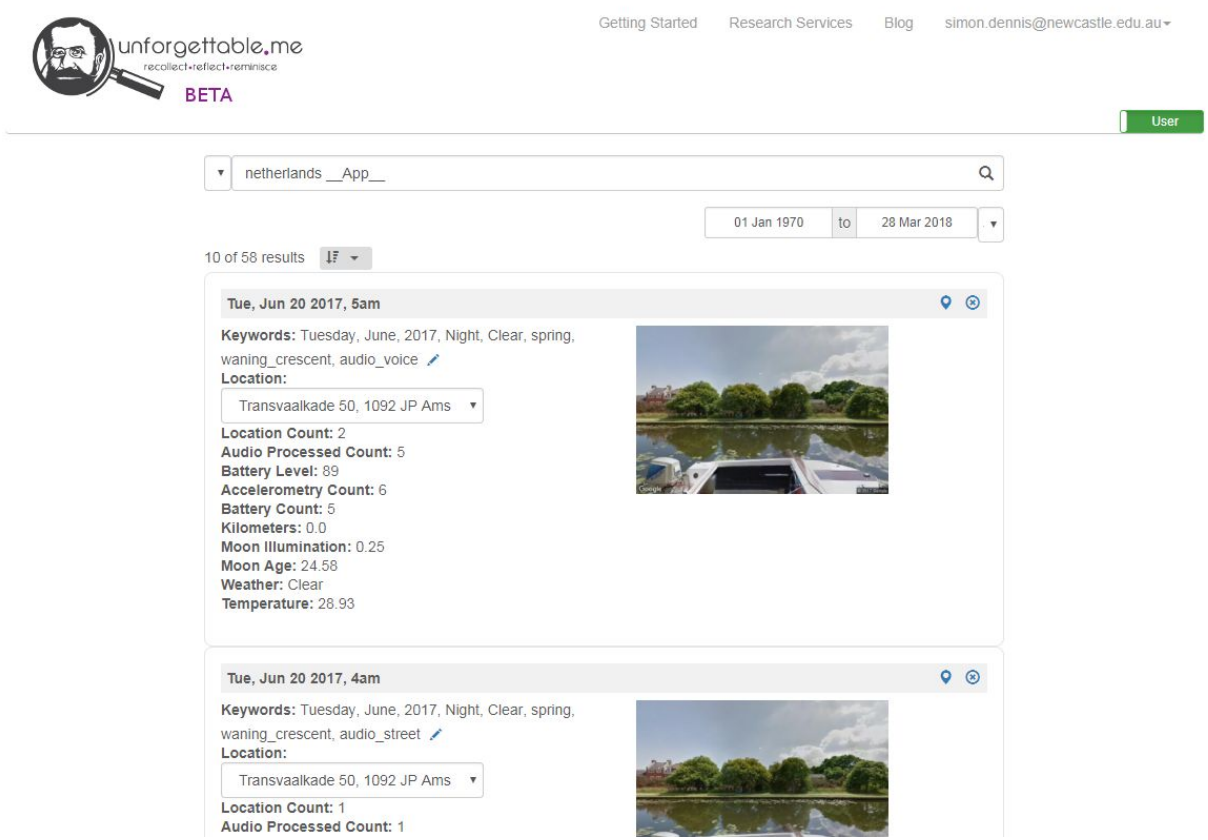
b) Users should be able to turn on and off individual data streams

c) Delay data upload so the user can delete it

d) Use a private format of the data



Search and Visualization Mechanisms



Most data collection exercises it is difficult to get access to data to delete it.

Need to make the interface intrinsically motivating.

Data Analysis

Researchers should not have access to raw data (c.f. open data).
Aggregate data can still be used to reveal information about people.

e.g. from mean of the ages of 20 people and the mean of the ages of 19 of those people can work out age.

Provide randomized results that have been tested to determine if the distribution changes if any person is removed. (c.f. Private language)

Whose data should it be?

Today data is owned by corporations, governments, and research institutions.
But it is generated by users, citizens and participants.
Perhaps they should own it.

They could then participate in a data marketplace:

- where they consent each time their data is used knowing what it will be used for
- where they are compensated each time

Australian Government
Australian Digital Health Agency

My Health Record

Benefits of user ownership

- people would know what their data is being used for
- participants would be incentivised to curate their data
- a data marketplace would produced a more nuanced understanding of privacy

Benefits of user ownership

- data would be submitted directly to the repository by the participant and would be available to the scientific community - not locked in a researcher’s draw
- participant ownership of data may lead to increased engagement in the scientific process
- people in lower socioeconomic classes often have the most valuable data - marketplace allows the accumulation of a data asset

Data Collection Tutorials

Simon Dennis

Complex Human Data Hub
University of Melbourne
Unforgettable Technologies Pty Ltd

unforgettable.me
recollect•reflect•reminisce

Getting Started

https://www.unforgettable.me/howto/tutorials/getting-started/

The Unforgettable App

https://www.unforgettable.me/howto/tutorials/app/

If This Then That
IFTTT

https://www.unforgettable.me/howto/tutorials/ifttt/

Privacy Preserving Data Analysis

Simon Dennis

Complex Human Data Hub
University of Melbourne
Unforgettable Technologies Pty Ltd

Although Kogan is a social psychologist, in every other way he presented as a perfectly legitimate researcher

Current data analysis languages don't preserve privacy

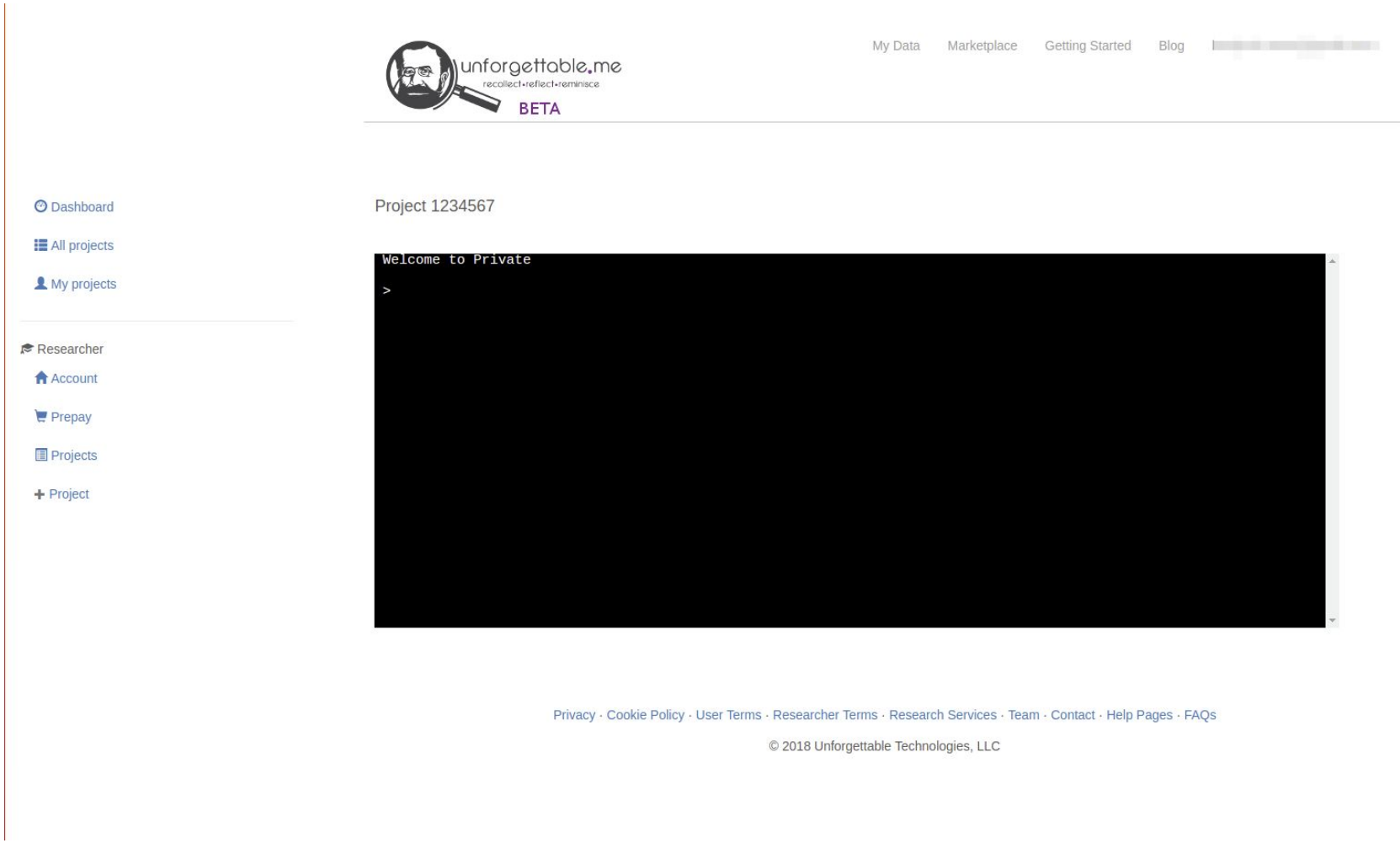
```
>>> print [email for email in emails if "Simon Dennis" in email][0]

Dear Lady Gaga,

Can I have your children?

Please?

Yours eternally,

Professor Simon Dennis
Director, Complex Human Data Hub,
Melbourne School of Psychological Sciences
University of Melbourne
>>>
```

Hard to retrofit existing languages to deal with privacy

Private

- Tightly couples data and analysis
- Tests to determine if any result should be released to coder
- Is both a probabilistic and deterministic language based on python and bugs/jags/stan
- Is declarative (like bugs/jags)
- Allows for incremental definition (like python?)
- Automatically parallelizes
- Built using pyMC3
- Is alpha code

Embedded in platform



No installation or set up with Private - just login.

Deterministic Definitions

Unlike jags/stan where data mangling mostly happens in R or python before the probabilistic code is called, Private embeds everything within the language:

```
> l = Bernoulli(0.3, 100)
> sv
l = Bernoulli(0.3, 100)  [1.000000 ... 0.000000]
> l
[1 0 0 0 0 0 0 0 1 0 0 1 0 1 0 0 1 0 0 0 0 1 0 0 0 0 1 1 0 0 0 0 1 1 0
 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 1 1 0 1 0 0 1 1 1
 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 1 1 1 1 0 1 0 0]
>
```

Probabilistic Definitions

Variables can have deterministic and probabilistic definitions

```
> l ~ Bernoulli(r)
> sv
l = Bernoulli(0.3, 100)    [0.000000 ... 0.000000]
l ~ Bernoulli(r)           [0.000000 ... 0.000000]
>
```

Sampling happens automagically

If we supply the prior the sampler has all of the definitions it needs and so begins sampling.

```
> r ~ Uniform(0,1)
> sv
l = Bernoulli(0.3, 100)   [0.000000 ... 0.000000]
l ~ Bernoulli(r)          [0.000000 ... 0.000000]
r ~ Uniform(0, 1)         Computing
```

The shell doesn't wait for the computation to complete as that may take a long time.

While we are waiting

We can set up some variables that we might be interested in:

```
> meanr = mean(r)
> stdr = std(r)
```

Sampler is finished

Samples have been calculated and meanr and stdr have been updated

```
> sv
l = Bernoulli(0.3, 100)   [1.000000 ... 0.000000]
meanr = mean(r)           0.35546615262834885
stdr = std(r)             0.046057887766870025
l ~ Bernoulli(r)          [1.000000 ... 0.000000]
r ~ Uniform(0, 1)         [0.340152 ... 0.381661]
>
```

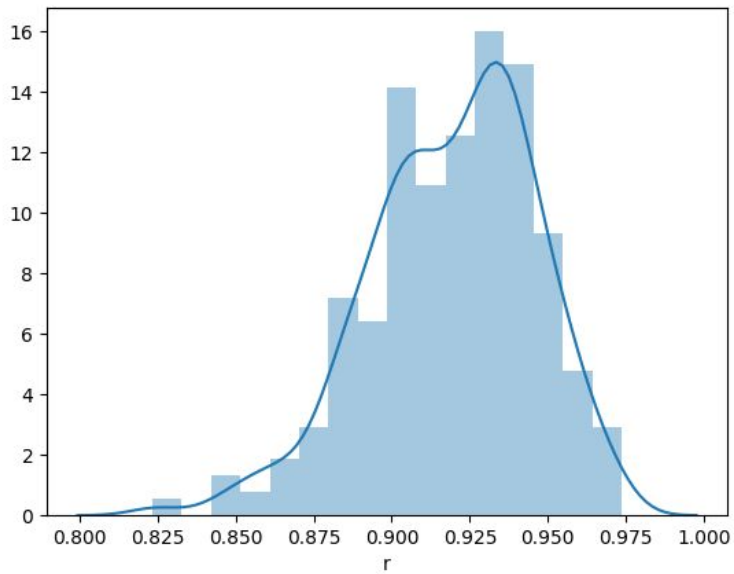
Declarative so variables are updated automagically

```
> l = Bernoulli(0.9, 100)

> sv
l = Bernoulli(0.9, 100)   [0.000000 ... 1.000000]
meanr = mean(r)           0.9193716800844921
stdr = std(r)             0.02647437986790655
l ~ Bernoulli(r)          [0.000000 ... 1.000000]
r ~ Uniform(0, 1)         [0.965137 ... 0.899771]
>
```

In R or python one often has to rerun code to update variables
That's error prone

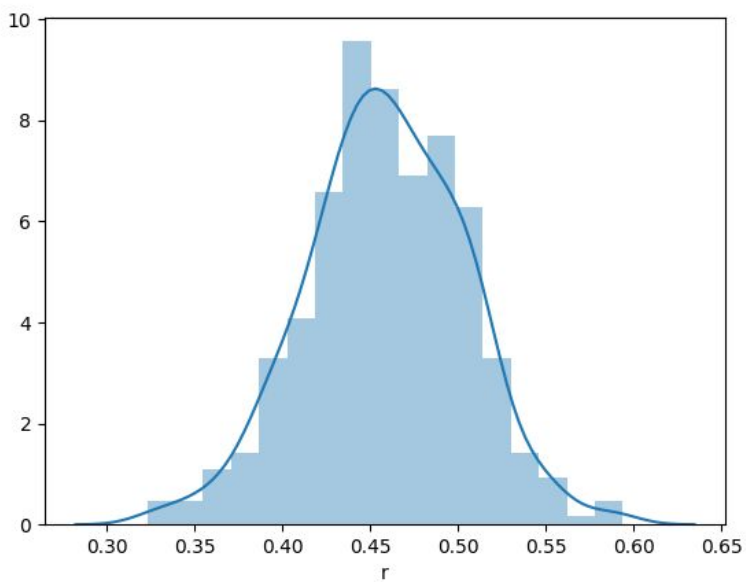
Plotting

```
rplot = distplot(r)
> sv
l = Bernoulli(0.9, 100)    [0.000000 ... 1.000000]
meanr = mean(r)            0.9193716800844921
stdr = std(r)              0.02647437986790655
rplot = distplot(r)        <_io.Byt...c9edb8f0>
l ~ Bernoulli(r)           [0.000000 ... 1.000000]
r ~ Uniform(0, 1)          [0.965137 ... 0.899771]
```



Plots are declarative too

```
l = Bernoulli(0.5, 100)
> sv
l = Bernoulli(0.5, 100)    [0.000000 ... 1.000000]
meanr = mean(r)            0.9193716800844921
stdr = std(r)              0.02647437986790655
rplot = distplot(r)        <_io.Byt...a16c9470>
l ~ Bernoulli(r)           [0.000000 ... 1.000000]
r ~ Uniform(0, 1)          [0.965137 ... 0.899771]
```



Controlling the sampler

```
> NumberOfSamples
200
> NumberOfChains
2
> NumberOfTuningSamples
200
>
```

Parallelism

Variables that can be calculated in parallel will be across the cluster

```
> n = 100000000
> t = Gumbel(10,2,n)
> t2 = Gumbel(20,4,n)
> sv
n = 100000000              100000000
t = Gumbel(10, 2, n)       Computing
t2 = Gumbel(20, 4, n)      Computing
>
```

Plan to distribute computation of list results too, but this is not implemented yet.

Builtins

Bernoulli
Beta
BetaBinomial
Binomial
Categorical
Cauchy
ChiSquared
Constant
DemoEvents
DiscreteUniform
DiscreteWeibull
Event
Events
ExGaussian
Exponential
Gamma
Geometric
Gumbel
HalfCauchy
HalfNormal
HalfStudentT
InverseGamma
Kumaraswamy
Laplace
LogNormal
Logistic
LogitNormal
NegativeBinomial
Normal
NumberOfChains
NumberOfSamples
NumberOfTuningSamples
Pareto
Poisson
SkewNormal
StudentT
Triangular
Uniform
Wald
Weibull
ZeroInflatedBinomial
ZeroInflatedNegativeBinomial
ZeroInflatedPoisson
abs
all
any
bin
bool
chr
cmp
complex
dict
distplot
divmod
enumerate
exp
filter
float
format
frozenset
getattr
hasattr
hex
int
isinstance
issubclass
iter
len
list
long
map
max
mean
min
object
oct
ord
pow
property
range
reduce
repr
reversed
round
set
slice
sorted
std
str
sum
tuple
type
unichr
unicode
vars
xrange
zip

Probabilistic Builtins

Bernoulli
Beta
Binomial
Categorical
Cauchy
ChiSquared
Constant
DiscreteUniform
DiscreteWeibull
Exgaussian
Exponential
Gamma
Geometric
Gumbel
HalfCauchy
HalfNormal
HalfStudentT
InverseGamma
Kumaraswamy
Laplace
Logistic
LogitNormal
Lognormal
NegativeBinomial
Normal
OrderedLogistic
Pareto
Poisson
SkewNormal
StudentT
Triangular
Uniform
VonMises
Wald
Weibull
ZeroInflatedBinomial
ZeroInflatedNegativeBinomial
ZeroInflatedPoisson

And its called Private because ...

> Events
Private
>

DemoEvents


```
{
  "AccelerometryCount": 1,
  "AudioProcessedCount": 9,
  "Temperature": 2.880355518947072,
  "MoonIllumination":
0.0707992850345782 2,
  "LocationCount": 9,
  "EndDateTime": "2012-05-21",
  "UserId":
"10dac7d9-b3ee-4a99-860a-24895806a032"
,
  "keywords": [
    "January",
    "Saturday",
    "1976",
    "summer",
    "audio_home",
    "audio_street",
    "rain",
    "waxing_gibbous"
  ],
  "Weather": "rain",
  "StartDateTime": "2012-05-21",
  "Kilometers": 0.379523036935310 5,
  "BatteryCount": 9,
  "MoonAge": 23.47574200905594 2,
  "latitude": -39.12135149340159,
  "longitude": 146.3565740873568 8,
  "BatteryLevel": 13,
  "type": "App",
  "address": "Level 6\n 94 Ritter
Roadside\nPort Stephanie, QLD, 2434"
}
```

Calculating the probability of rain
Public version

```
> rained = [e.Weather == "rain" for e in DemoEvents if e.hasField("Weather")]
> sv
rained = [e.Weather == "rain" for e in DemoEvents if e.hasField("Weather")]      [False, False, False, True, False, False, ...]
> rained ~ Bernoulli(r)
> r ~ Uniform(0,1)
> sv
rained = [e.Weather == "rain" for e in DemoEvents if e.hasField("Weather")]      [False, False, False, True, False, False, ...]
rained ~ Bernoulli(r)                                                           [False, False, False, True, False, False, ...]
r ~ Uniform(0, 1)                                                               Computing
> sv
rained = [e.Weather == "rain" for e in DemoEvents if e.hasField("Weather")]      [False, False, False, True, False, False, ...]
rained ~ Bernoulli(r)                                                           [False, False, False, True, False, False, ...]
r ~ Uniform(0, 1)                                                               [0.350420 ... 0.278729]
> meanr = mean(r)
> meanr
0.33446577061115873
```

Calculating the probability of rain Private version

```
> rained = [e.Weather == "rain" for e in Events if e.hasField("Weather")]
> sv
rained = [e.Weather == "rain" for e in Events if e.hasField("Weather")]        Private
> rained ~ Bernoulli(r)
> r ~ Uniform(0,1)
> sv
rained = [e.Weather == "rain" for e in Events if e.hasField("Weather")]        Private
rained ~ Bernoulli(r)                                                          Private
r ~ Uniform(0, 1)                                                              Computing
> sv
rained = [e.Weather == "rain" for e in Events if e.hasField("Weather")]        Private
rained ~ Bernoulli(r)                                                          Private
r ~ Uniform(0, 1)                                                              Privacy Unknown
> sv
rained = [e.Weather == "rain" for e in Events if e.hasField("Weather")]        Private
rained ~ Bernoulli(r)                                                          Private
r ~ Uniform(0, 1)                                                              [0.350420 ... 0.278729]
> meanr = mean(r)
> meanr
0.33446577061115873
```


0 5 10 15 20 25 30
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7

Experience Sampling: Data Analysis

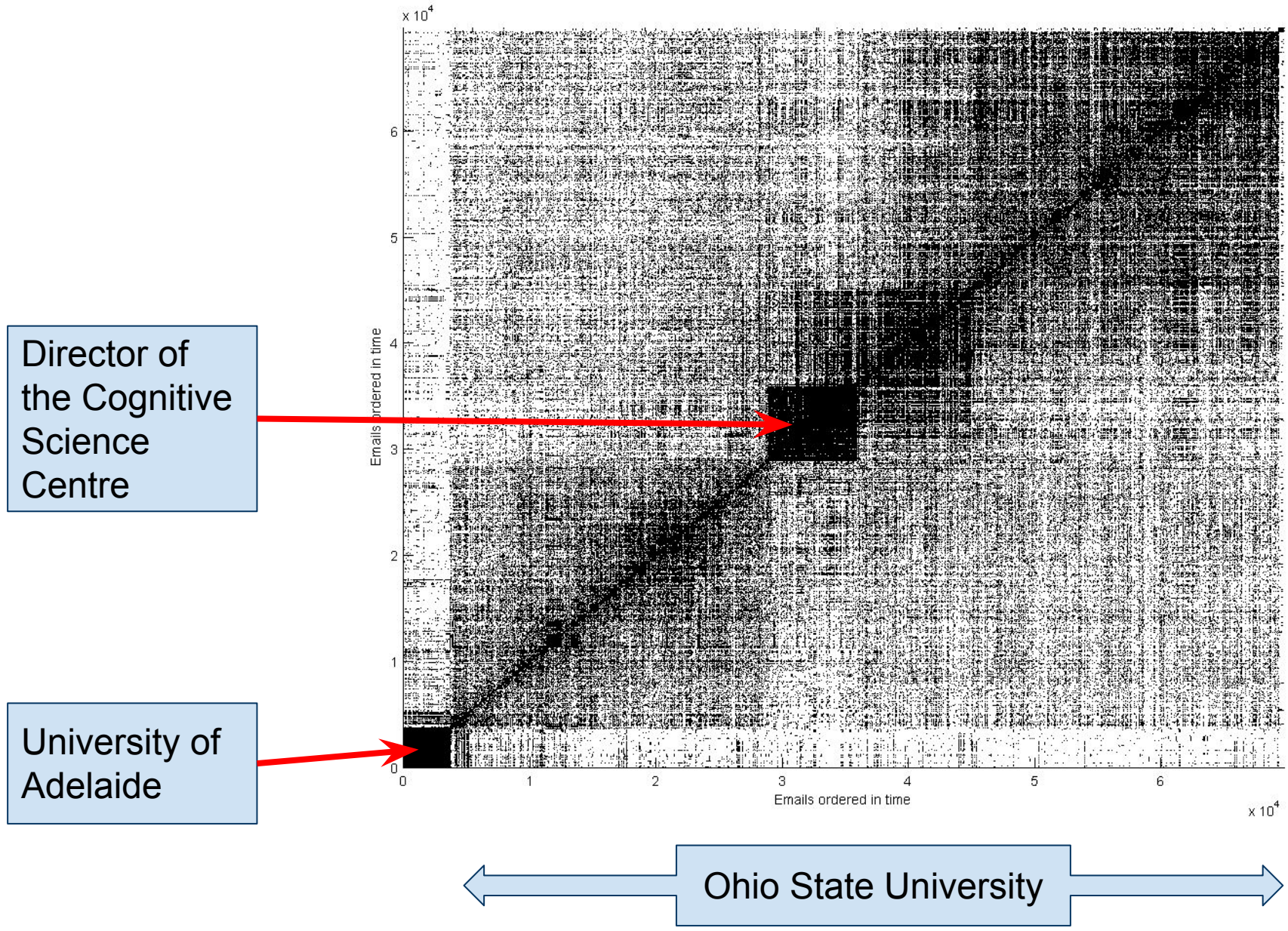
Simon Dennis

Complex Human Data Hub
University of Melbourne
Unforgettable Technologies Pty Ltd

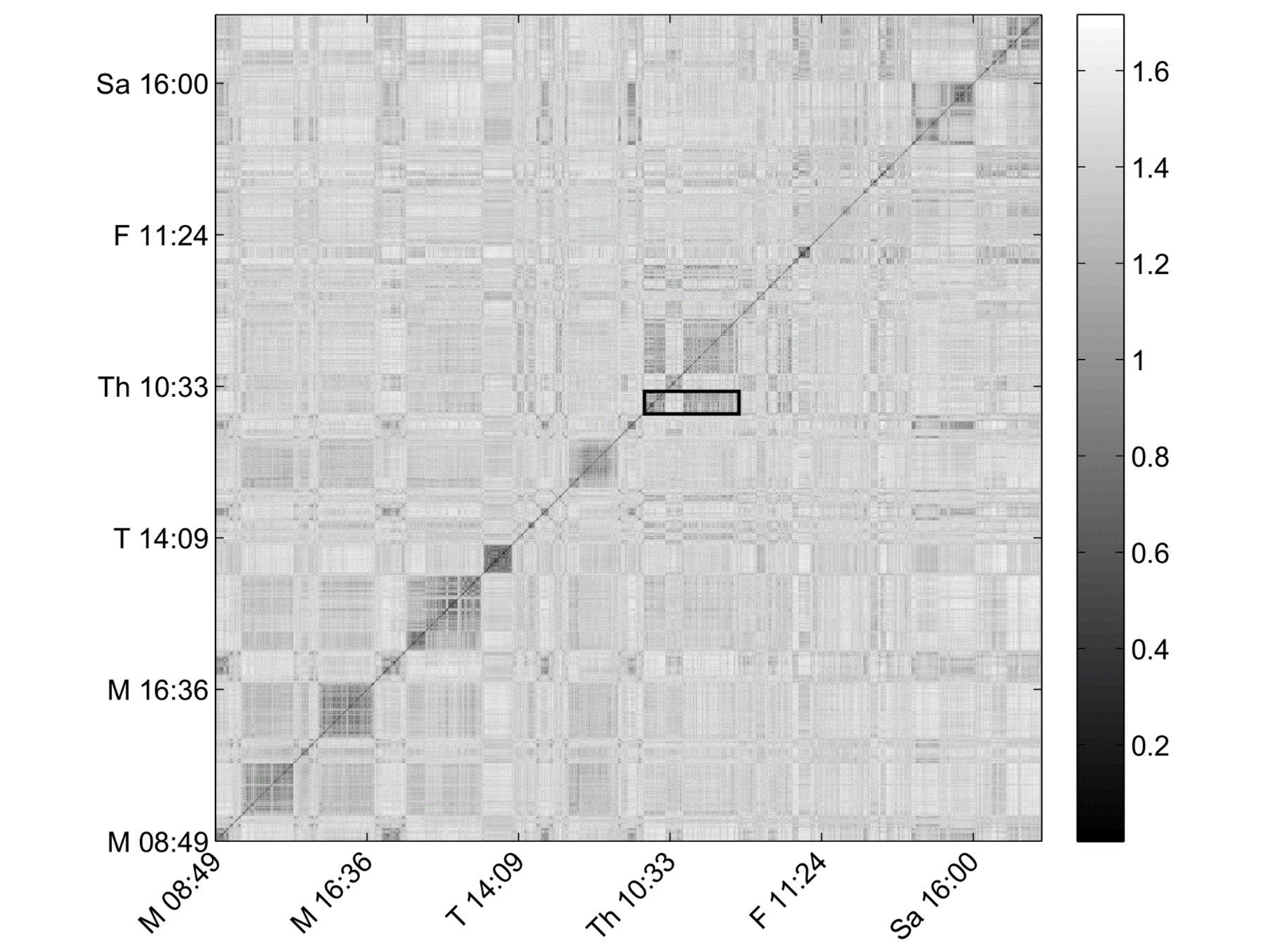
Recurrence Plots and the Correlation Dimension

Sreekumar, V., Dennis, S., Doxas, I., Zhuang, Y., & Belkin, M. (2014). The geometry and dynamics of lifelogs: discovering the organizational principles of human experience. PloS one, 9(5), e97166.

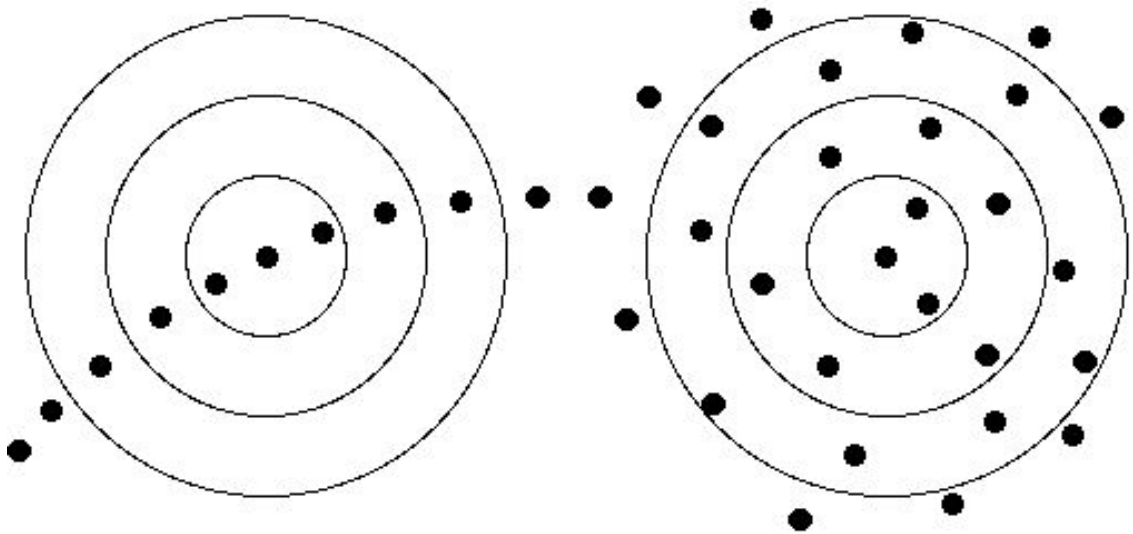
Sreekumar, V., Dennis, S., & Doxas, I. (2017). The episodic nature of experience: a dynamical systems analysis. Cognitive Science, 41(5), 1377-1393.



Recurrence plot of 30823 SD emails from 2007 to 2012


Sa 16:00
F 11:24
Th 10:33
T 14:09
M 16:36
M 08:49
M 08:49
M 16:36
T 14:09
Th 10:33
F 11:24
Sa 16:00
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2

Correlation Dimension

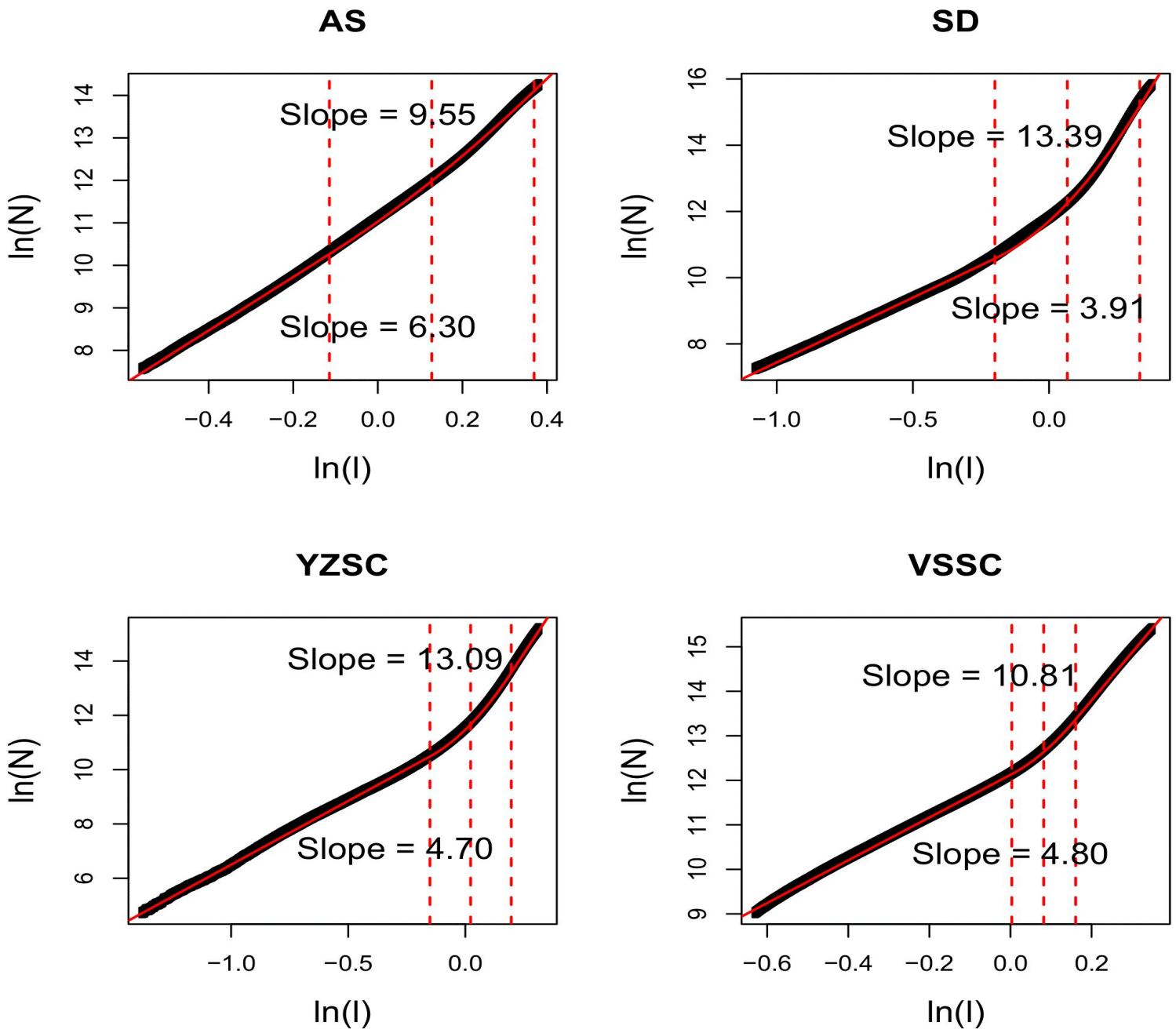


$$C(r) = \frac{1}{N^2} \sum_{i,j}^{N} \theta(r - \Delta r)$$

$$\text{where } \theta(x) = \begin{cases} 1 \text{ for } x > 0 \\ 0 \text{ for } x < 0 \end{cases}$$

(Heaviside function)

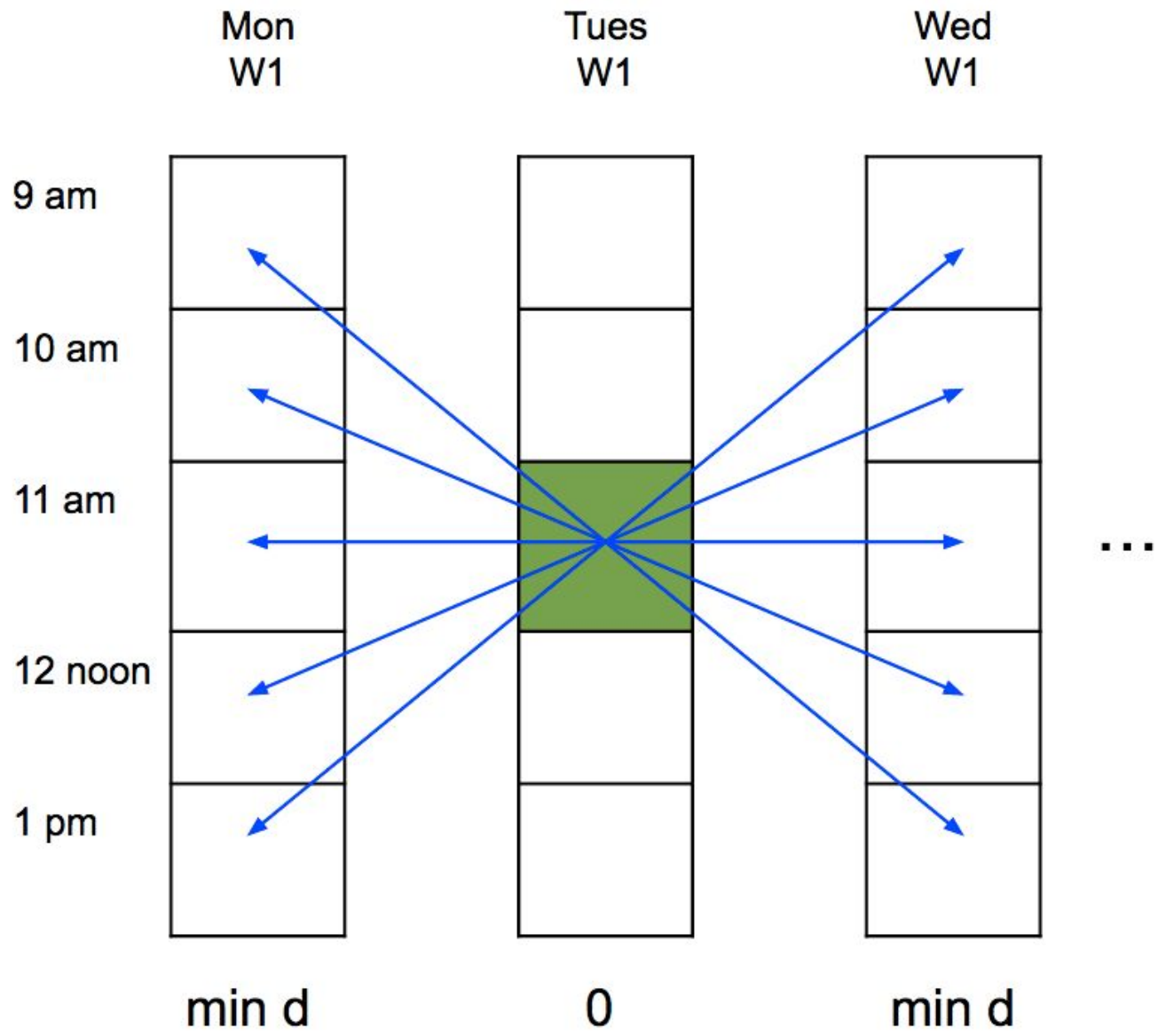
Ref: P. Grassberger and I. Procaccia, Phys. Rev. **Lett. 50**, 346 (1983)

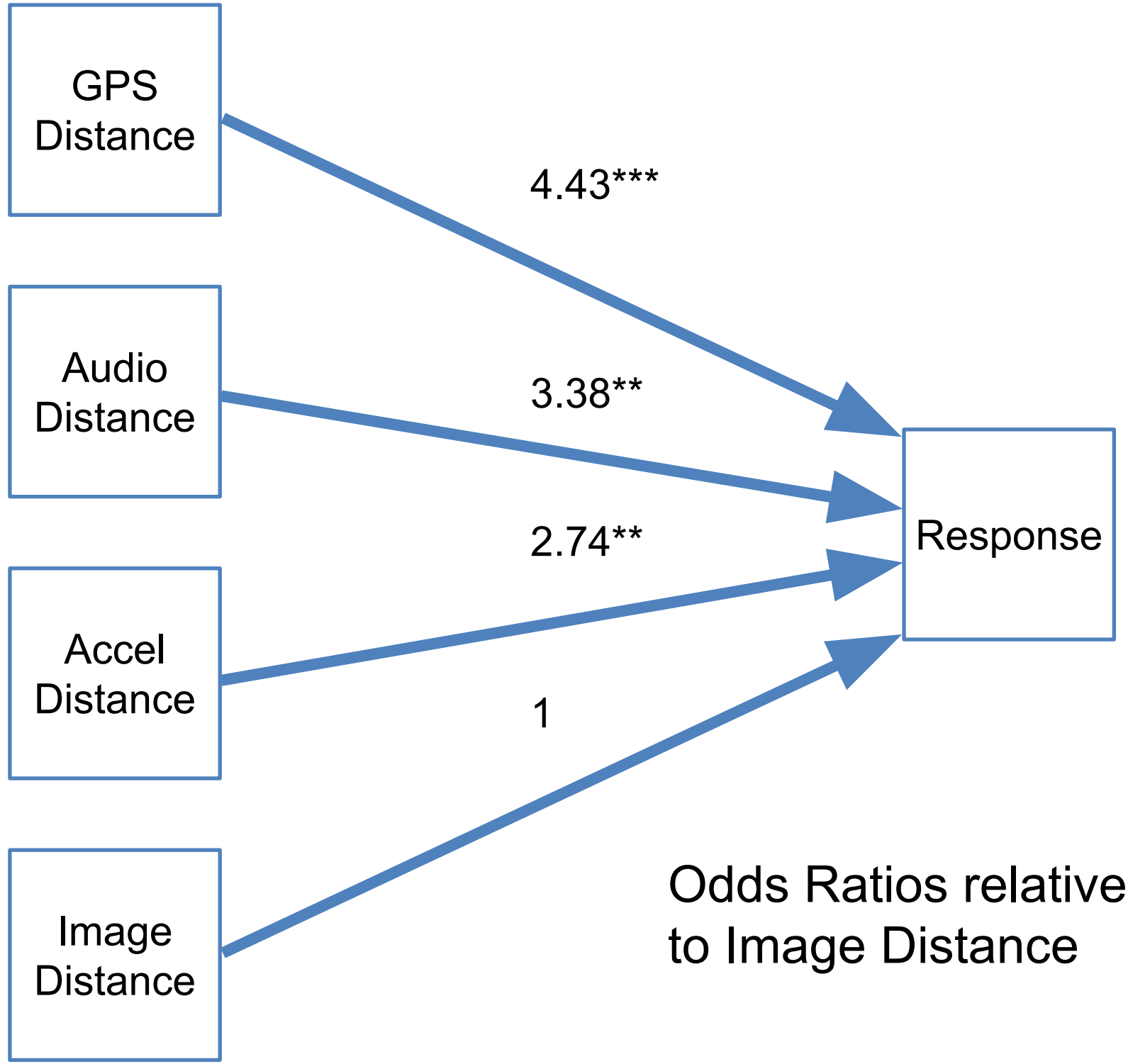


Although context representation can be hierarchical in structure, there is a basic level which might be analogous to the basic level of object representation.

Bayesian Models

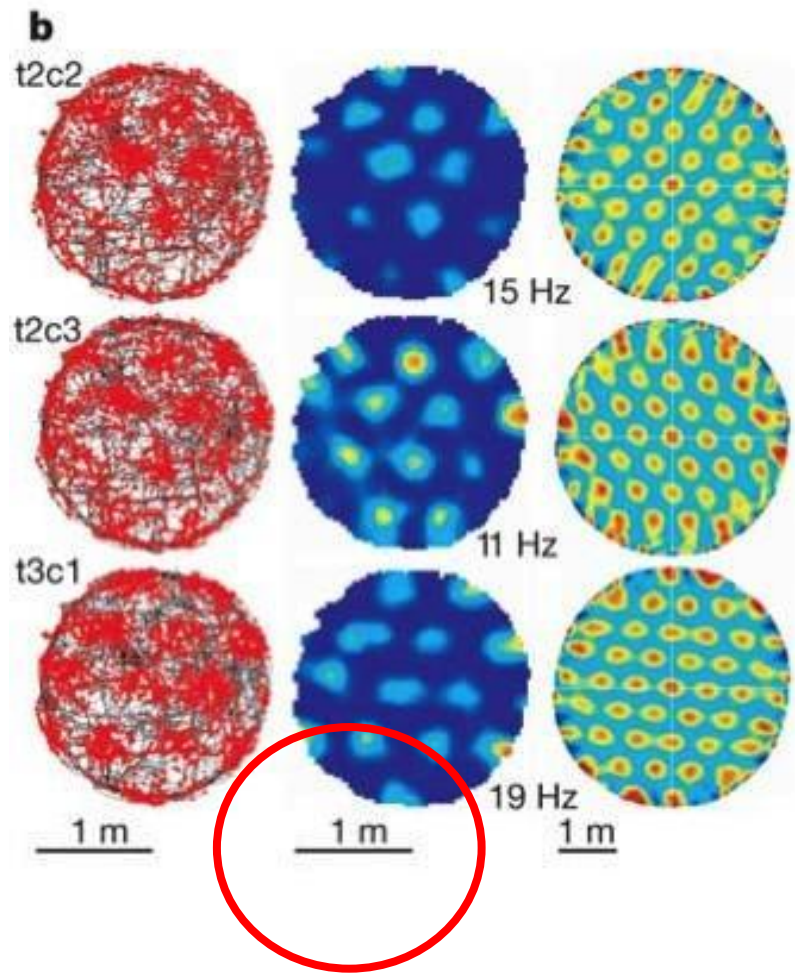
Dennis, S., Yim, H., Sreekumar, V., Evans, N. J., Garrett, P., & Sederberg, P. (2017). A hierarchical Bayesian model of “memory for when” based on experience sampling data. Cogn Sci.

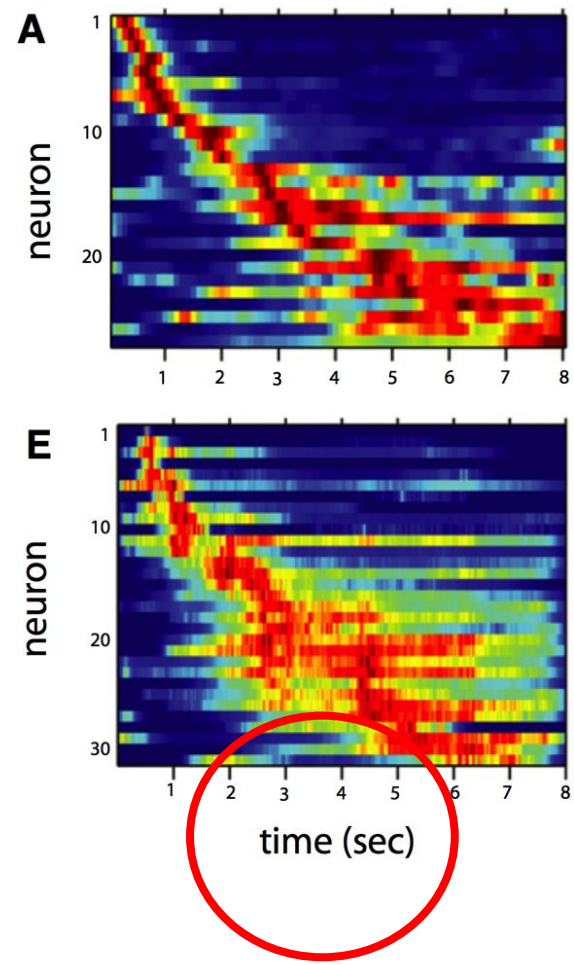

Mon
W1
Tues
W1
Wed
W1
9 am
10 am
11 am
12 noon
1 pm
...
min d
0
min d

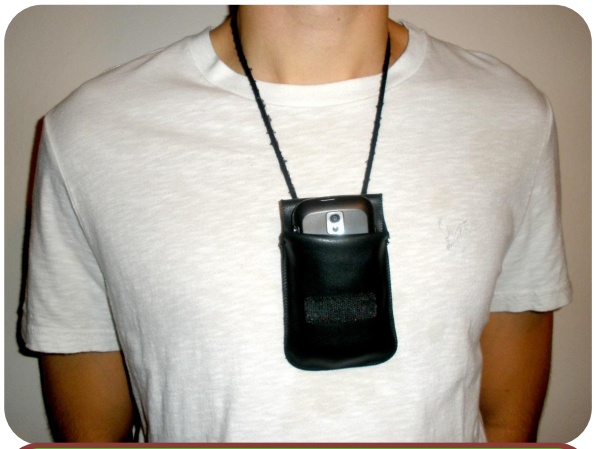

GPS Distance
Audio Distance
Accel Distance
Image Distance
4.43***
3.38**
2.74**
1
Response
Odds Ratios relative to Image Distance

Representational Similarity Analysis

Nielson, D. M., Smith, T. A., Sreekumar, V., Dennis, S., & Sederberg, P. B. (2015). Human hippocampus represents space and time during retrieval of real-world memories. Proceedings of the National Academy of Sciences, 112(35), 11078-11083.

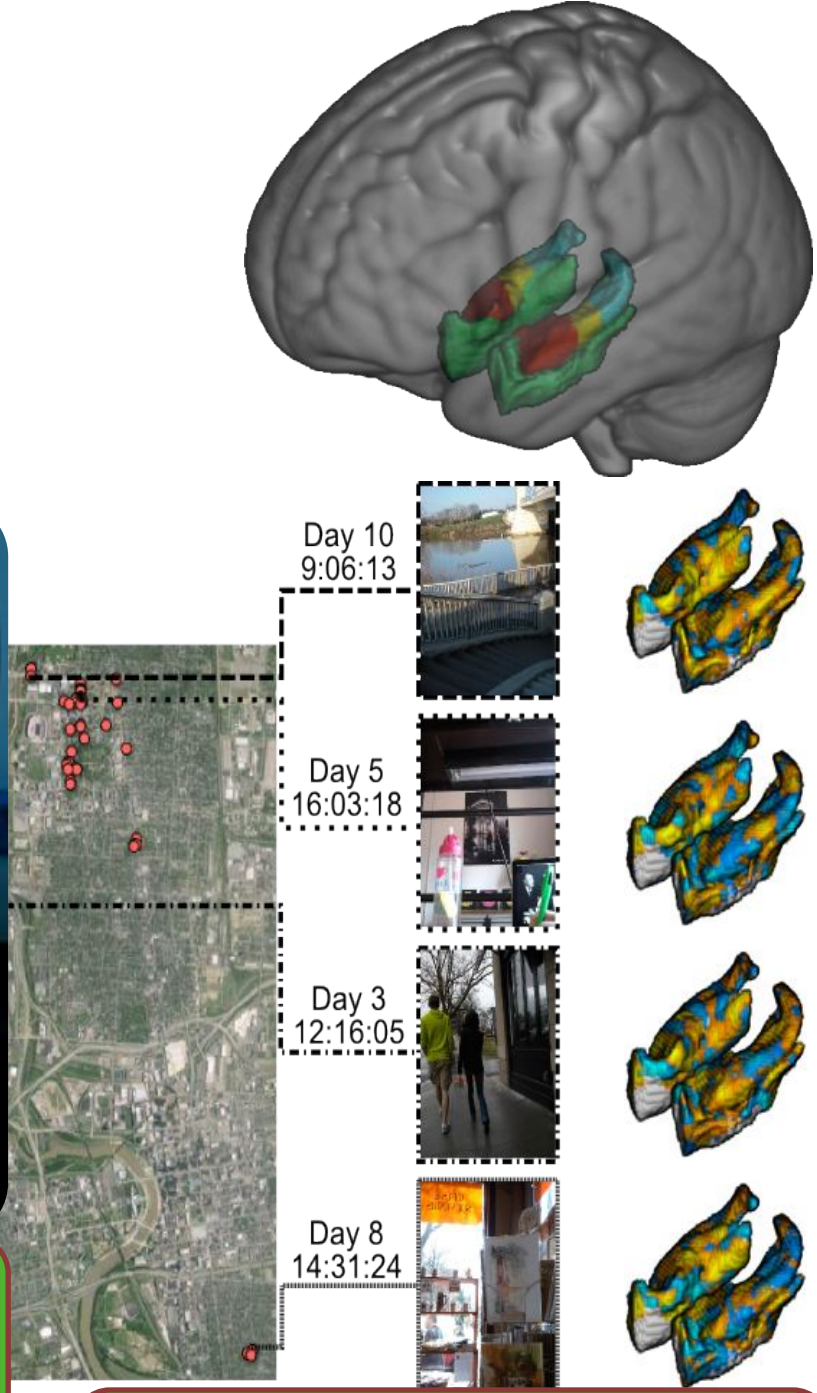

b
t2c2
15 Hz
t2c3
11 Hz
t3c1
19 Hz
1 m
1 m
1 m


A
neuron
1
10
20
1
2
3
4
5
6
7
8
E
neuron
1
10
20
30
1
2
3
4
5
6
7
8
time (sec)

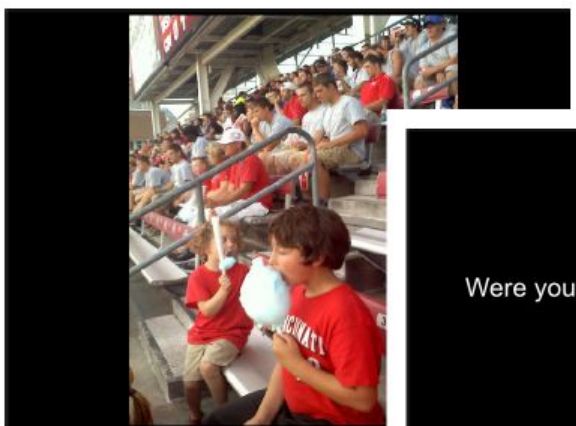


Experience Sampling
(2-4 weeks)

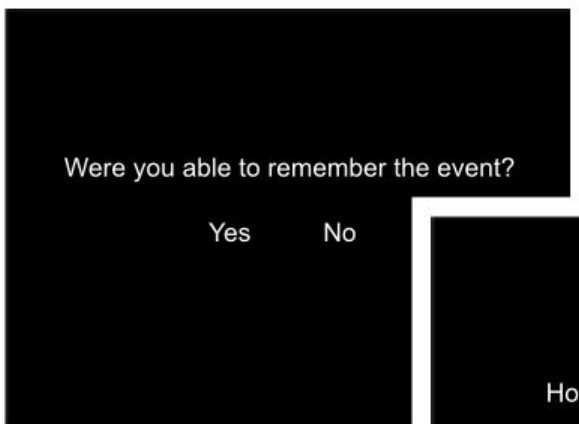
Reminiscence Task
in fMRI



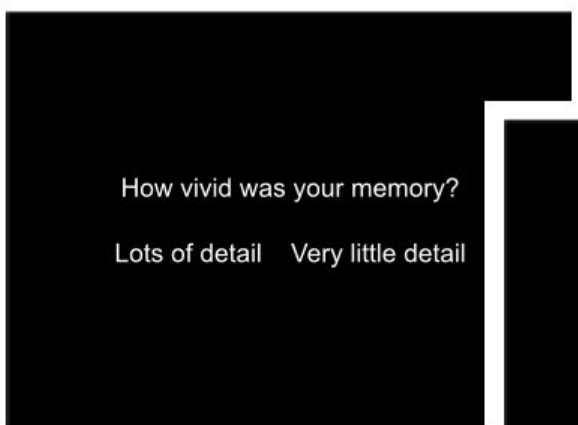
Compare Patterns of Brain Activity to Structure of the World



8 s



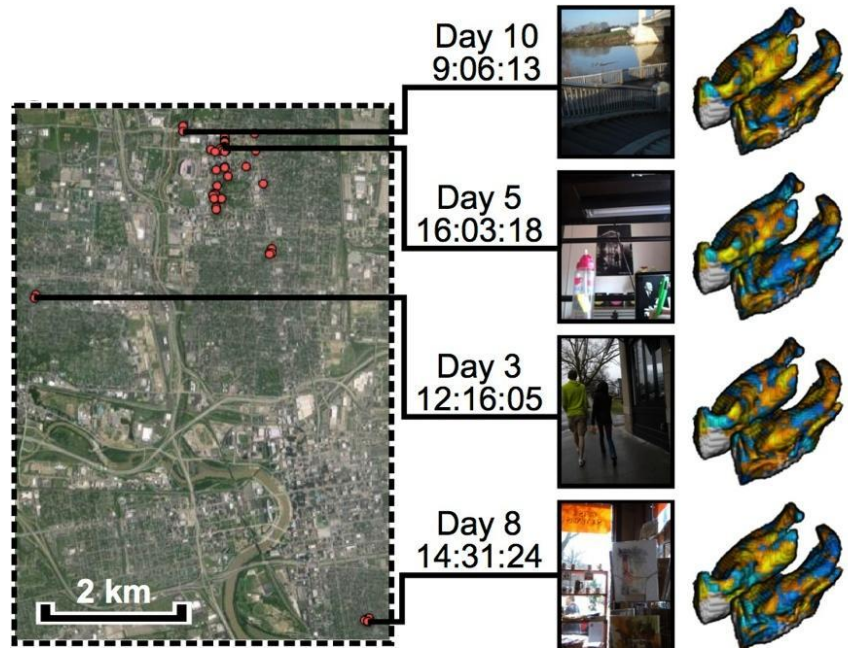
3s

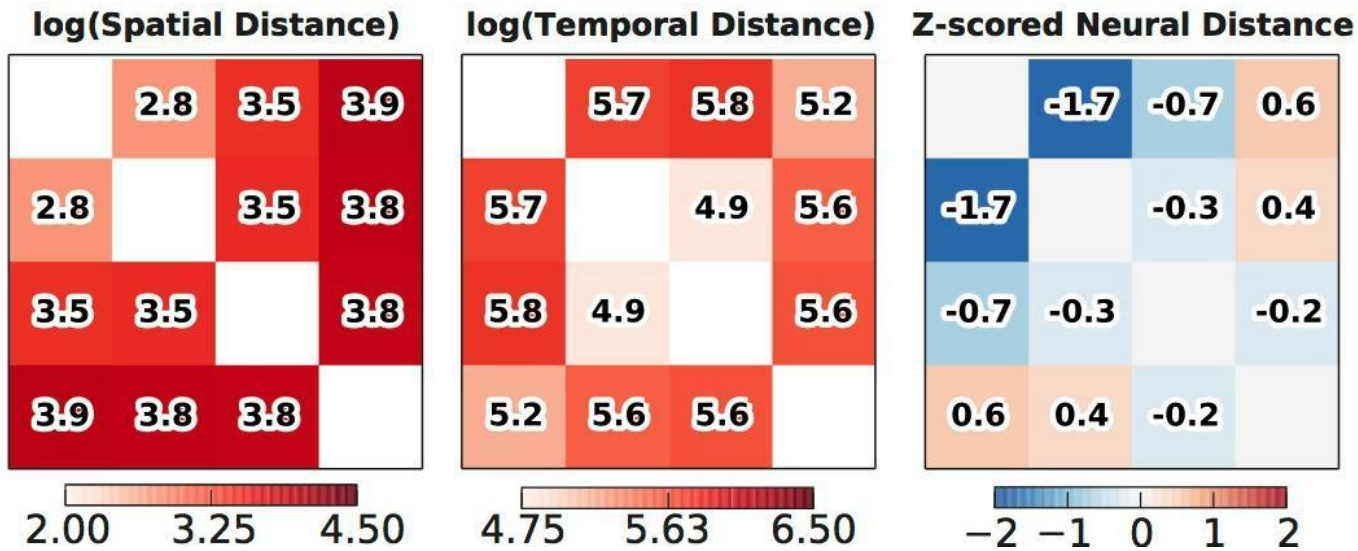


3s

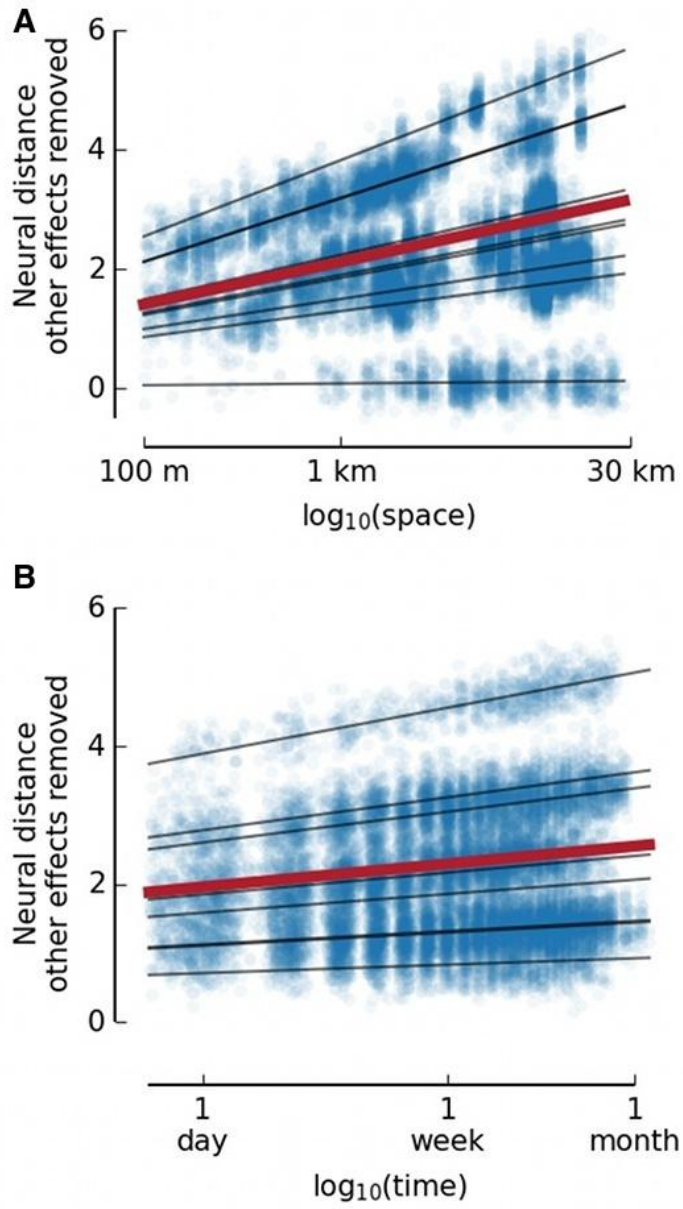
4-10 s ITI

Representational Similarity Analysis (RSA)

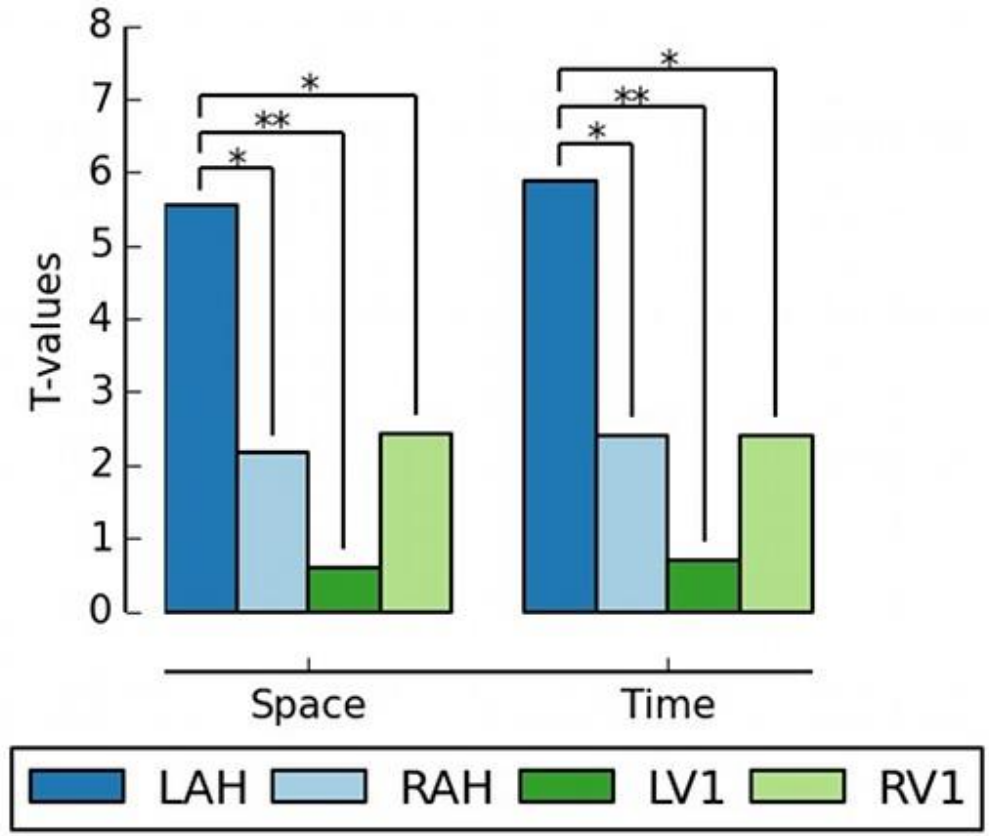




Significant in Left Ant. Hippocampus



More than in other control regions



Convergent Cross Mapping

Sugihara, G., May, R., Ye, H., Hsieh, C. H., Deyle, E., Fogarty, M., & Munch, S. (2012). Detecting causality in complex ecosystems. *Science*, 1227079.

Tanya Preminger

Thanks to

Grant number FA9550-09-1-0614

Ben Stone
Jihun Hamm
Mikhail Belkin

Vishnu Sreekumar

Per Sederberg
Troy Smith
Dylan Neilson

Hyungwook Yim
Paul Garrett
Nathan Evans

Students & Volunteers

Corrine McGann
Gabrielle Matthews
Kay Chan
Samuel Collison
Mhairi Gray
Anita Gray
Megan Baker
Viviana Pinzon Morales
Rachael Vince
Alison Rasmussen
Ruby Walker