

**DRAPER**

**Embedding Human-System  
Integration Metrics in Agile  
Software Evaluation  
Environments:**

**The Value of Opportunistic  
Behavioral Data Collection**

Dr. Joshua Poore

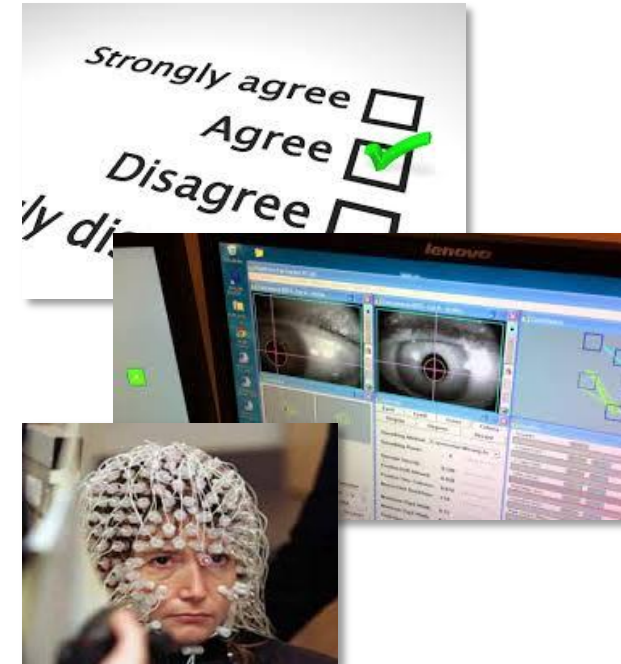
# Human System Integration Metrics Development

## ***Our methods for gauging human system integration can affect transitional value of technology***

- The laboratory's very purpose is to remove measurement from the wild.
- Experiments force specific ways of interacting with stimuli.
- Questionnaires, Surveys, and Interviews include both subjective and retrospective biases.
- Physiological methods provide convoluted information that is difficult to directly apply to relevant use cases.

## ***Any tool that people use to perform a task can be instrumented to collect data on how they perform that task within their respective environments***

- *We're turning software applications, mobile phones, games and simulators into "sensors" to understand how people behave not just observe what they do.*
- From models of how people behave we develop intuitive psychometrics.



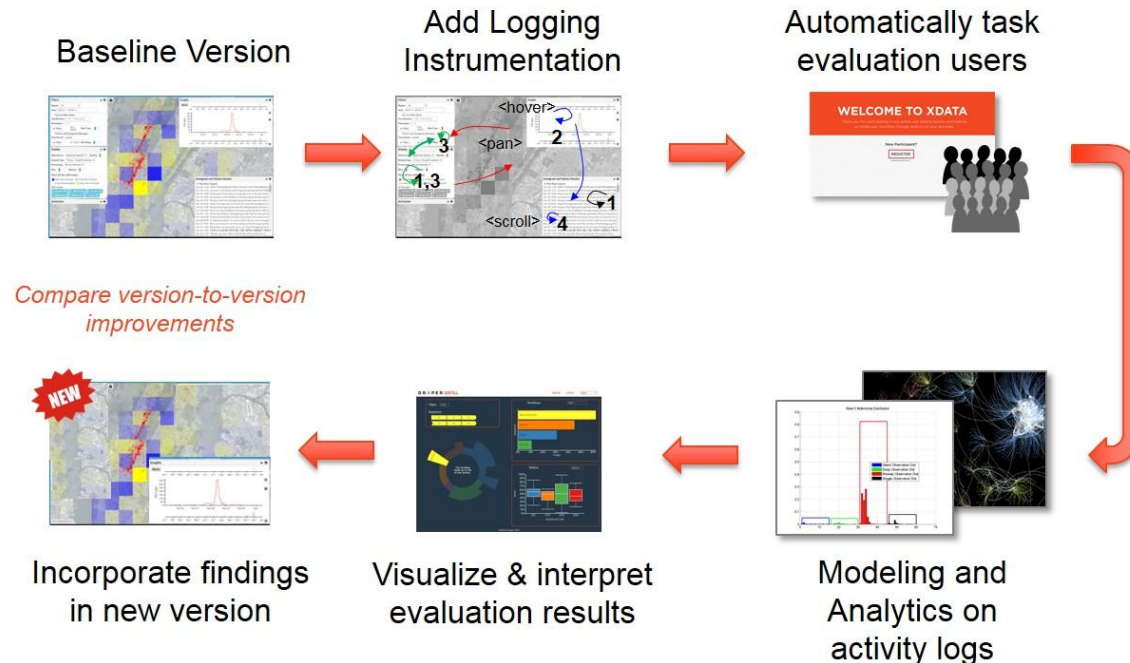
# Software as a Sensor™ Assessment

## Process

1. **Instrument Software**
2. **Let end-users use Software**
3. **Model** log data
4. **Visualize** models and derivative metrics
5. **Track** changes across versions

## Value

1. **Non-invasive, streaming data source**
2. **Collect Data in actual use cases**
3. **Generalized Models of “how”**
4. **Assessment tied to technology**
5. **Agile process**

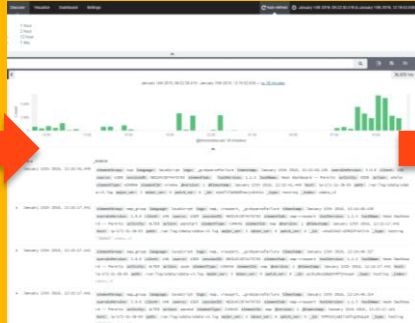


# Metrics Approach

## Data Collection

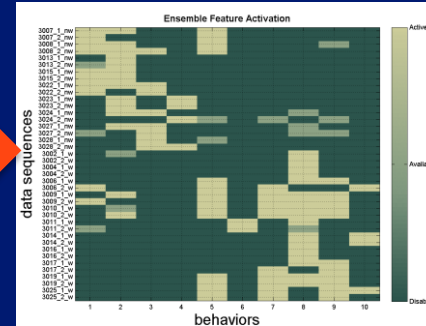
```
$('.btn').mouseover(function  
( ) { var msg = { activity:  
'inspect', action:  
'mouseover', elementId:  
this.getAttribute('id'),  
elementType: 'button',  
elementGroup: 'button_group',  
source: 'user', tags:  
['submit', 'query'] };  
ale.log(msg); } )
```

Instrumentation

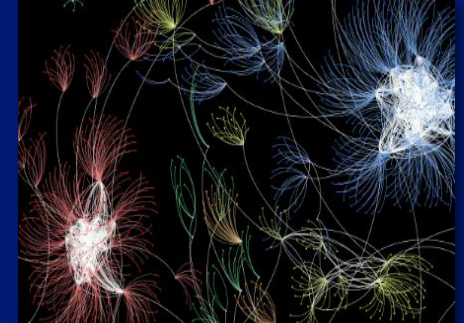


Activity Sequences

## Modeling



Beta Process HMM



Sub-sequences

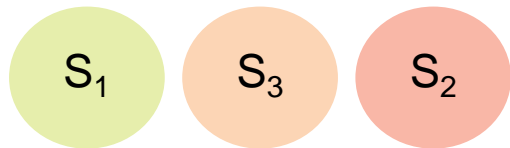
Proxies for Cognitive Load based  
*how software is used.*

- Integrated Use
- Strategy Shifts
- Fixedness in Use
- Specialized Workflows
- Workflow Variety

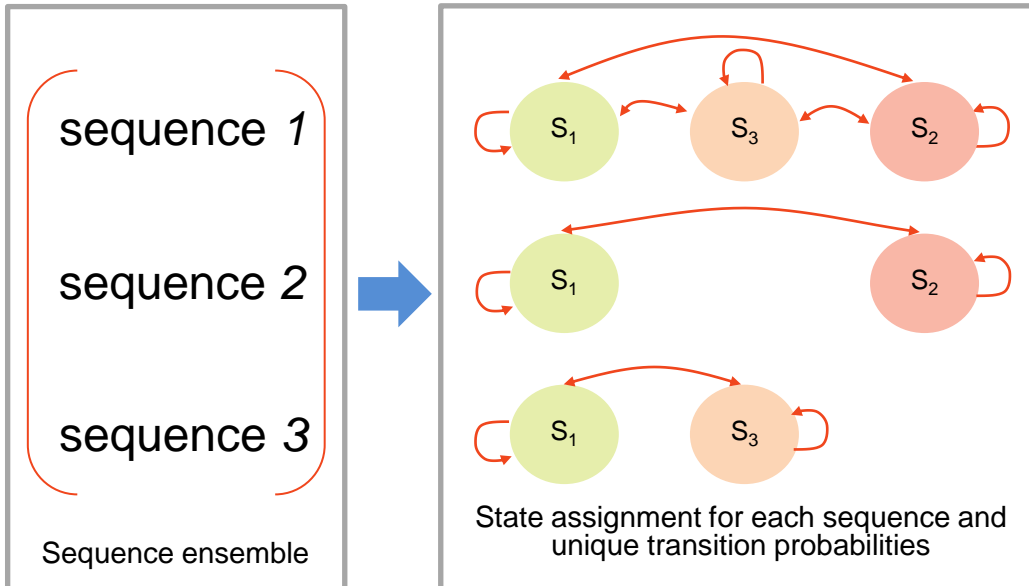
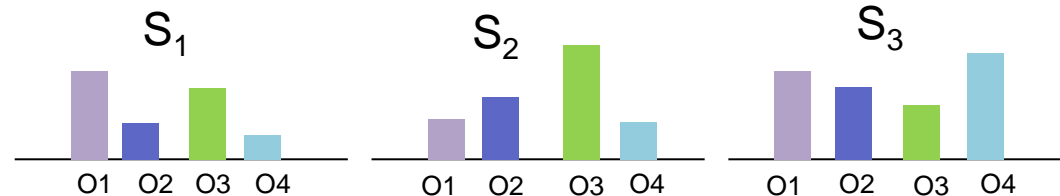
# Modeling HCI: BP-HMM

- The Beta-Process HMM (BP-HMM):
  - An unsupervised-learning method that identifies a *global* library of states (behaviors) from an ensemble of related time-series, and assigns a subset to each individual sequence.

*Global library of states*

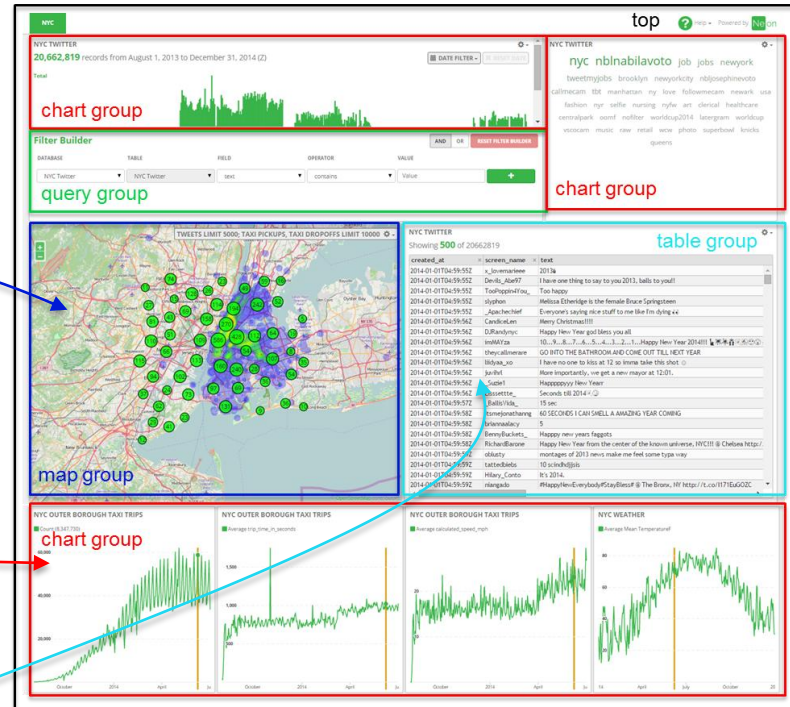
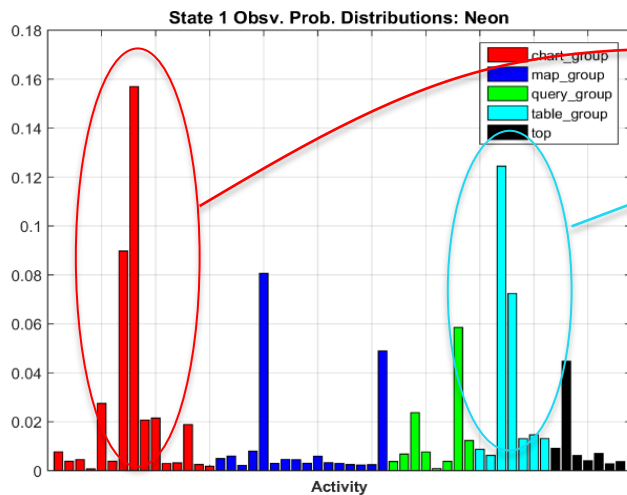
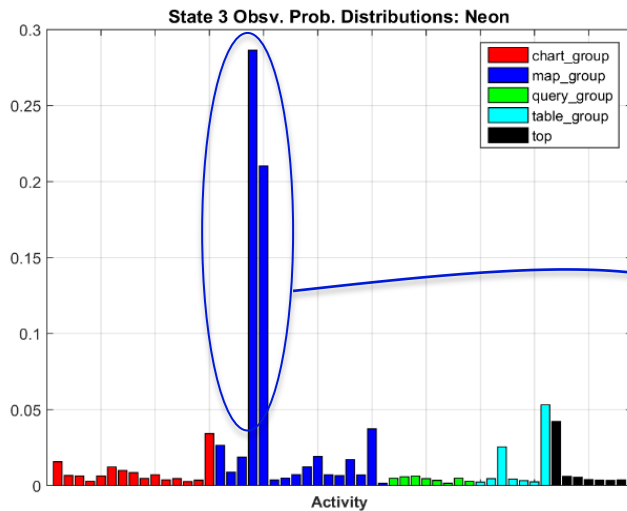


*Observation Probability Distributions*



- Example 3-state model of an ensemble of sequences encoding 4-categories of activities.
- The canonical 'modes' of behavior are encoded in the parameters of the hidden-state categorical distributions.

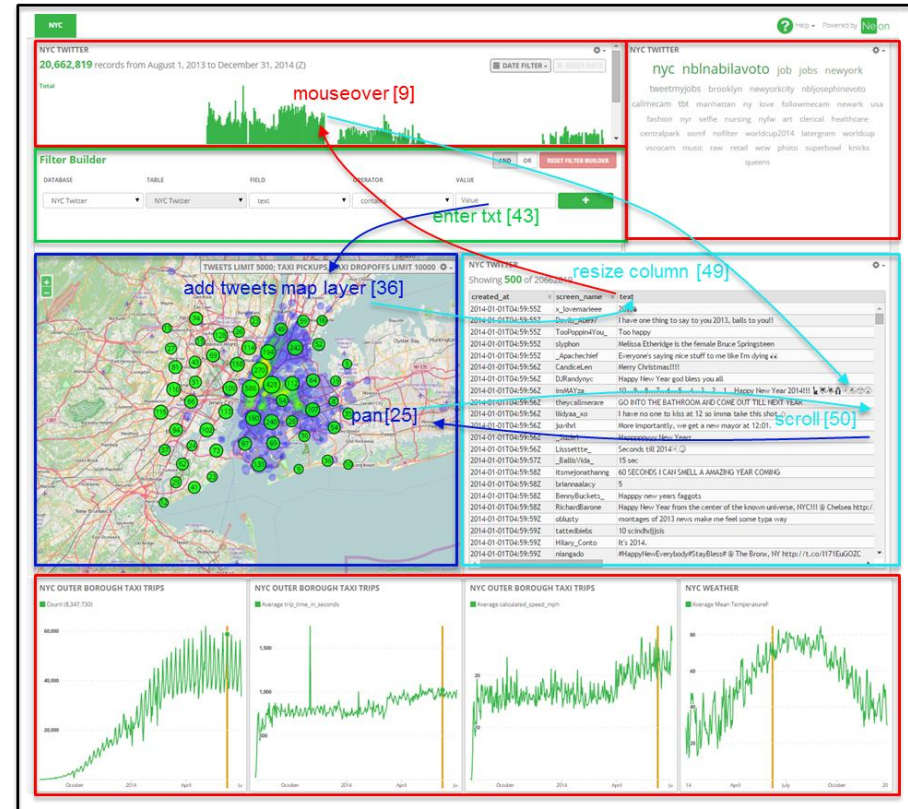
# Beta-Process HMM Metrics



- Activity distributions for each state capture the way different components of the application were used together in time.
- Metrics derived from the BP-HMM model parameters allow for comparison across applications.

# Modeling Workflow: Subsequences

- Sub-sequences are atomic-level workflows describing the patterns of software activities users integrate to perform tasks.
- For each user session, we extract overlapping sequences of lengths 3-6, creating a library of short activity kernels.



Sequence of activity logs from a Neon session

59, 25, 50, 25, 59, 25, 50, 43, 36, 50, 25, 25, 25, 24, 24, 25, 25, 24, 9, 49, 9, 14, 49, 8, 49, 25, 8, 25, 43, 36, 49, 9, 44, 49, 5, 49, 39, 49, 9, 59, 8, 50, 43, 36, 49, 8, 59, 25, 43, 36, 49, 9, 50, 9, 43, 36, 49, 39, 49, 9, 38, 49, 8, 49, 9, 2, 4, 1, 62, 60, 1, 63, 1, 14, 9, 49, 9

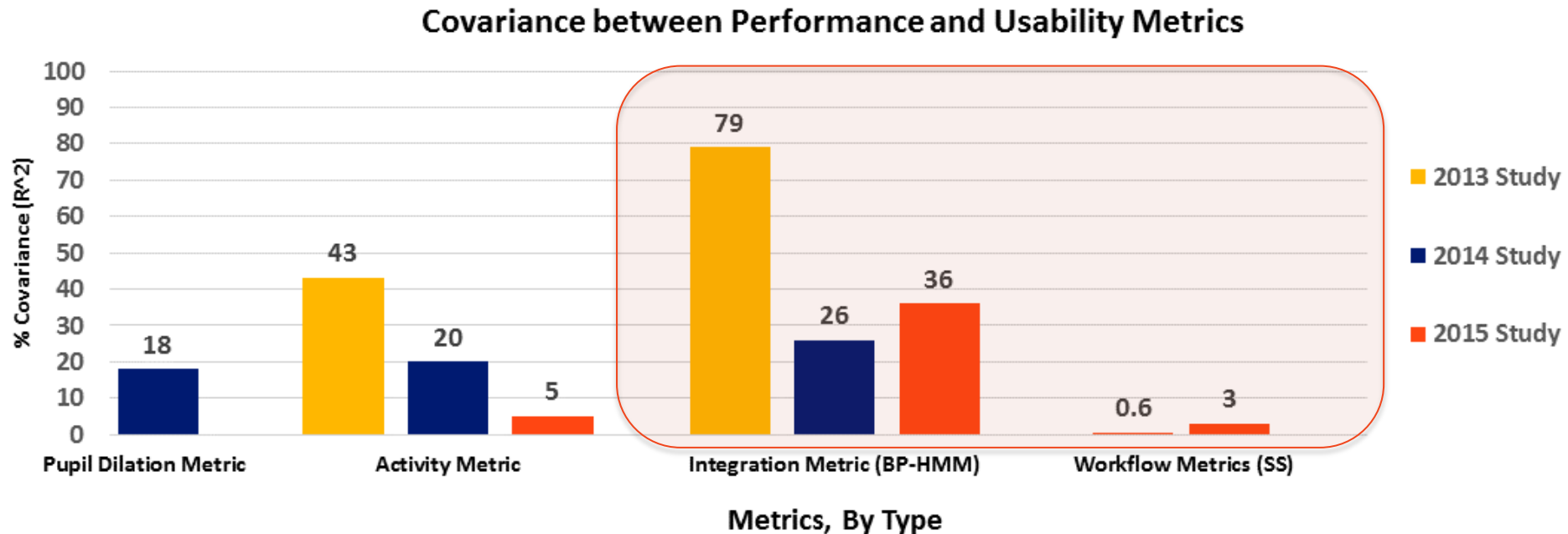
Extracted Sub-Sequences (length 3)

[59, 25, 50], [25, 50, 25], [50, 25, 59], [25, 59, 25], [59, 25, 50], [25, 50, 43], [50, 43, 36], [43, 36, 50], [36, 50, 25], [50, 25, 25], [25, 25, 25], [25, 25, 24], [25, 24, 24], [24, 24, 25], [24, 25, 25], [25, 25, 24], [25, 24, 9], [24, 9, 49], [9, 49, 9], [49, 9, 14], [9, 14, 49], [14, 49, 8], [49, 8, 49], [8, 49, 25], [49, 25, 8], [25, 8, 25], [8, 25, 43], [25, 43, 36], [43, 36, 49], [36, 49, 9], [49, 9, 44], [9, 44, 49], [44, 49, 5], etc....

- Screenshot of Neon with sample length-3 activity sub-sequences overlaid

# Year 3 Key Results - Performance

Using software as a sensor proves to be better than the current state-of-the-art in predicting **performance**.





# Data Collection at Scale

- Large scale data collection on a variety of applications provides data to move from metrics to assessment.
  - Assessments and labels applied to metrics based on large population distributions
    - Normalized Percentile Ranks
    - Good OR Bad

## Operation XDATA: A Public Event

- Register at <https://xdataonline.com>
- Participate in research!
- Download DARPA open source analytic tools and use with your own data!

