Breakout Session 4: Track B

Implementation of Cloud Based Computing in a Modern Systems/Behavioral-Neuroscience Laboratory

Dr. Matt Howe
Assistant Professor, Virginia Tech

Implementation of cloud based computing in a modern systems/behavioral-neuroscience laboratory

THE HOWE LAB
at Virginia Tech

COLLEGE OF SCIENCE
SCHOOL OF NEUROSCIENCE
VIRGINIA TECH.

What do we study?

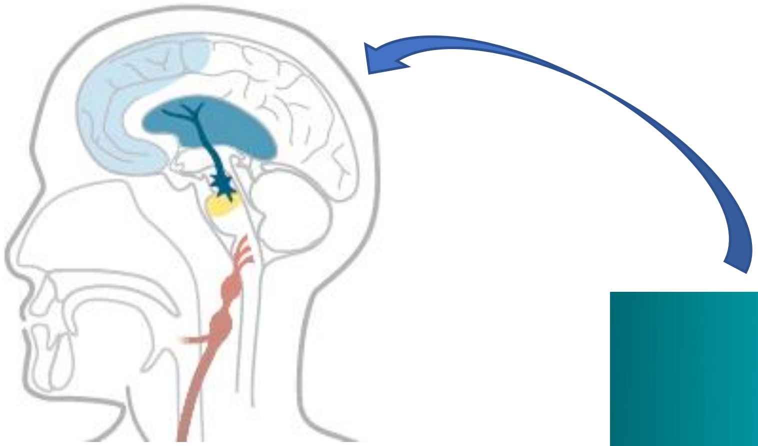
<u>Neurochemistry of primary reward</u>

- What are the brain mechanisms that underlie the pursuit of natural rewards (e.g. food).

- How do the properties of those rewards change their value and attractiveness (e.g. macronutrient or caloric content?)

Gut-brain circuits controlling food reward

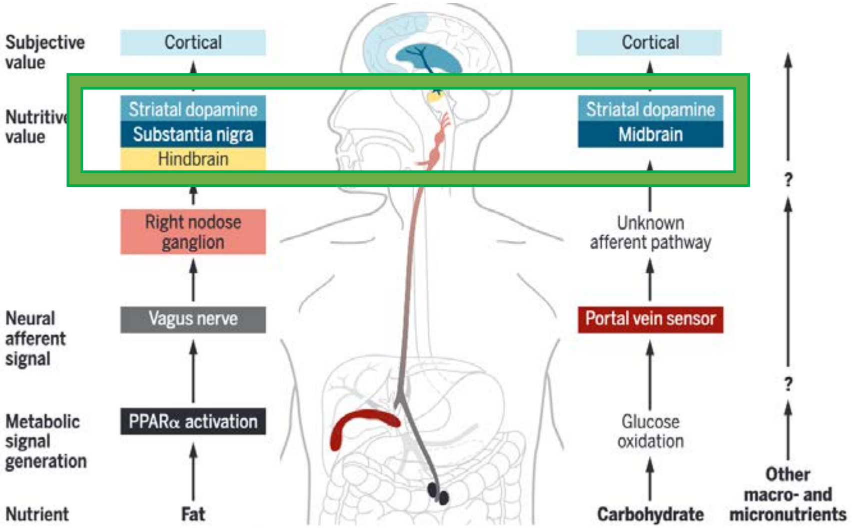
Post-ingestive: signals of the nutrient content of food generated in the gut



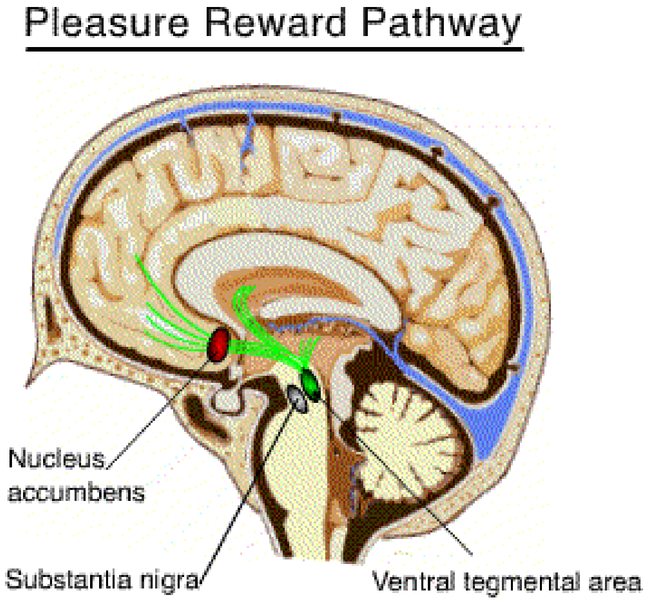
COLLEGE OF SCIENCE
SCHOOL OF NEUROSCIENCE
VIRGINIA TECH.

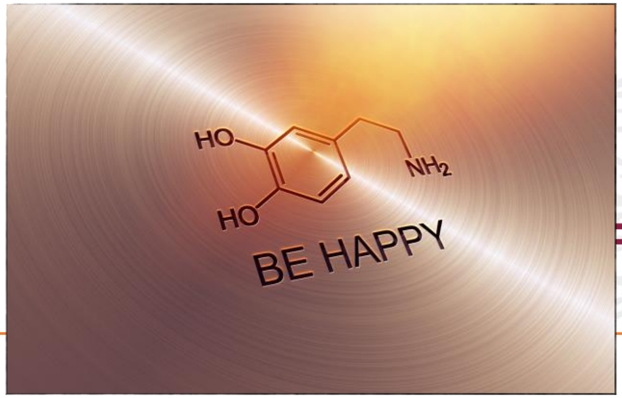
Post-ingestive modulation of dopamine

Nutrients delivered to the gut can evoke DA release in the striatum



Small and DiFeliceantonio, 2019

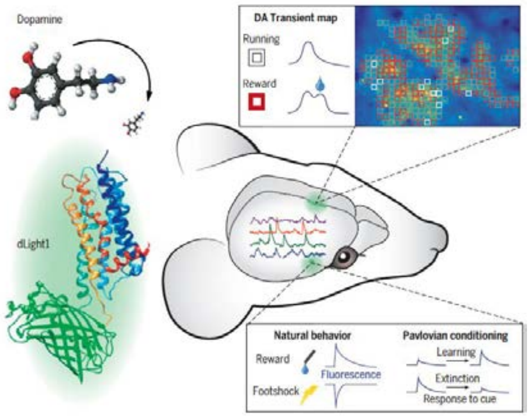




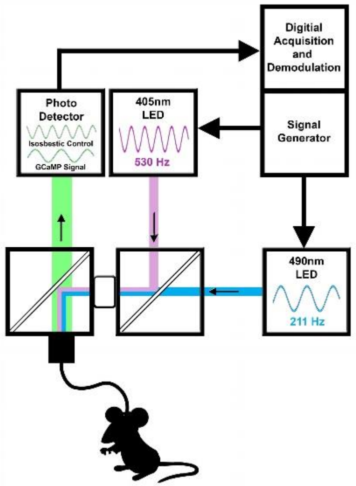
Post-ingestive signals are relayed to DA systems compute reward value, expectation, motivation, etc.

Optical tools for measuring neurotransmitter release

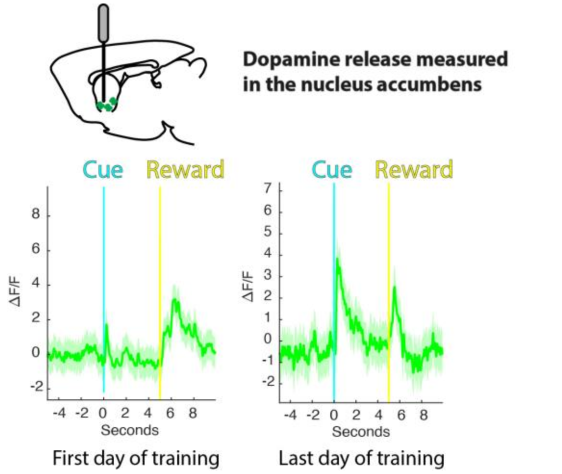
<u>Dlight: a new tool to measure fast DA release *in vivo*</u>

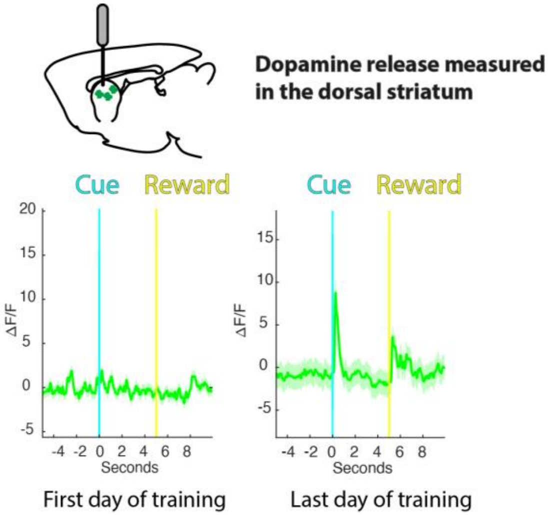


<u>Fiber photometry</u>



<u>Multi-region dopamine dynamics</u>

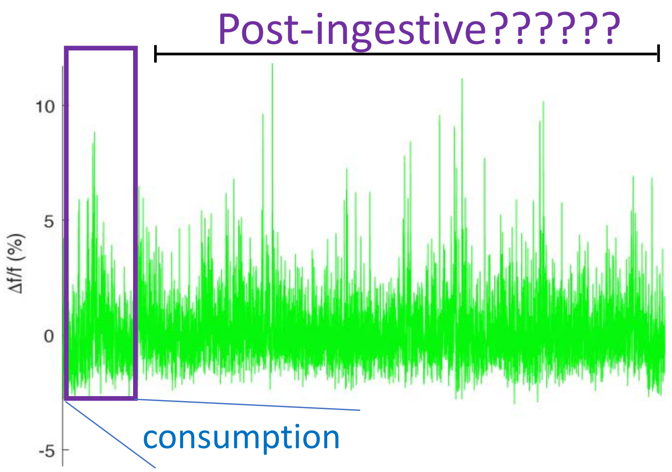




Fast sampling rates offer new insights to second by second DA encoding of reward

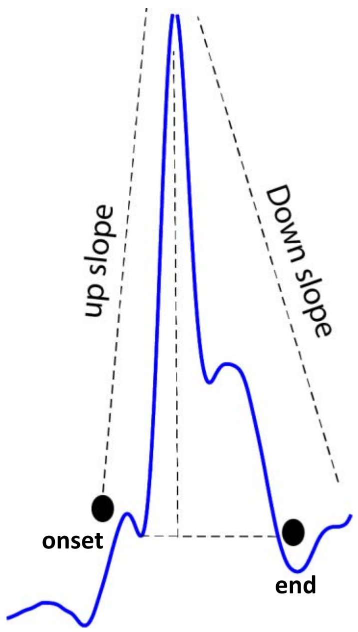
LEGE OF SCIENCE
HOOL OF NEUROSCIENCE
GINIA TECH.

The big data challenge: isolating the post-ingestive signal



Typical recording session: 45 minutes, 1kHz sampling rate. Multiple repeats within mouse (~5-10 recording sessions)

"spikes" = discrete release events

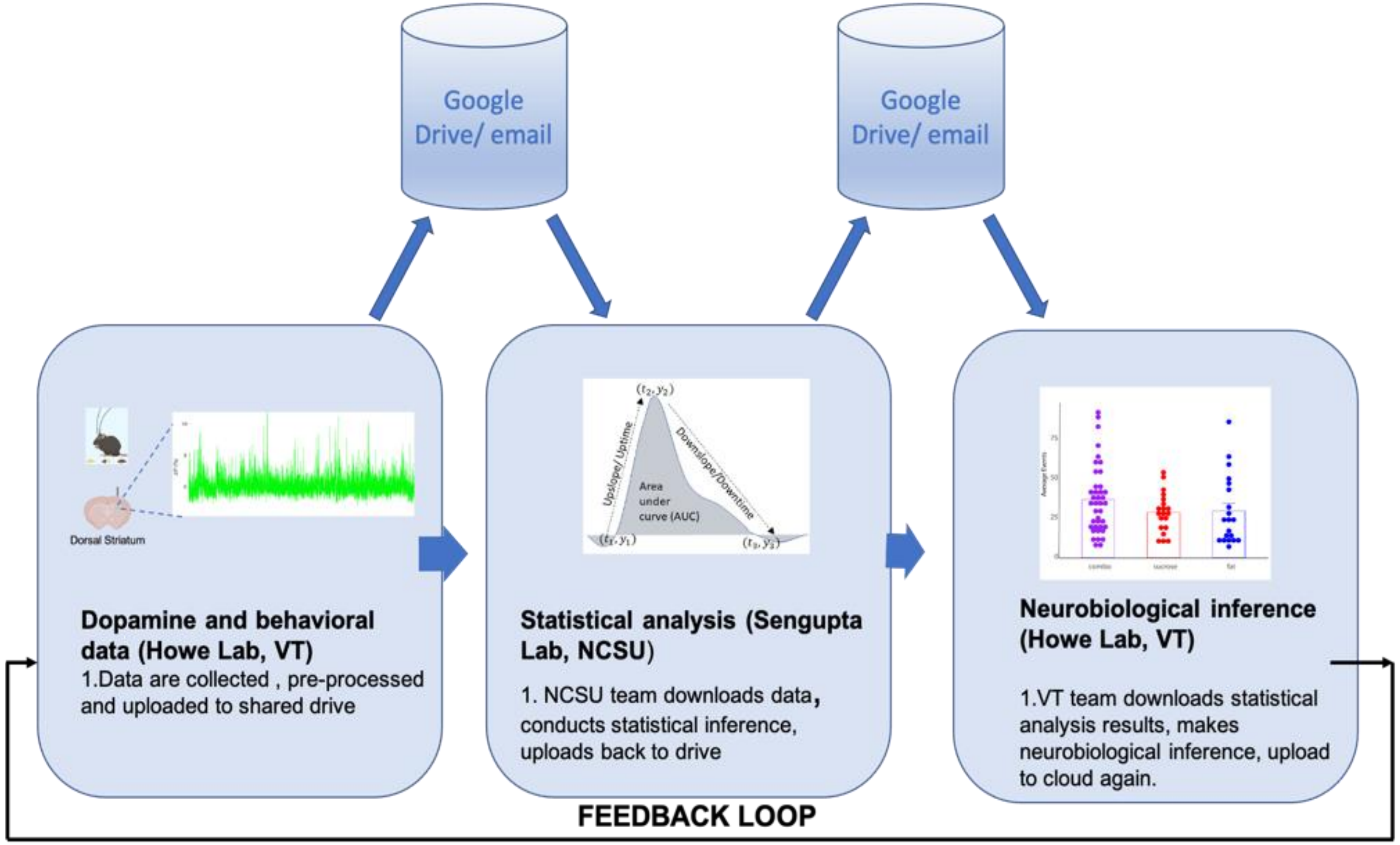


<u>Measures of interest</u>

1. Upslope
2. Downslope
3. Rise time
4. Decay time
5. Peak
6. AUC

Do these measures diverge at particular times within the recording session that may be a signature of the "post-ingestive" effect? Is this effect behaviorally meaningful?

Typical workflow



Major limitations and bottlenecks

1. ***Data integrity***: All done on personal computers running MATLAB. Team has no insight into the actions of other team members.

2. ***Computational costs:*** Analysis Involves converting the time-series data into temporal windows, computing statistical features (e.g., cross-correlation functions) of the functional data across these windows, computing test statistics, and carrying out hypothesis testing using these test statistics. The latter relies on bootstrap resampling, which requires repeated computations of the test statistic on random resamples of the data. Combined with the large size of the data, process leads to **impractically high runtimes and memory costs.**

3. ***Time costs***: In addition to the above, each data transfer step takes multiple hours.

4. ***ERROR!!!:*** Disjointed, personal machine based approach introduces more risk of human error. Individual cleaning and transformational steps cannot be audited by the team.

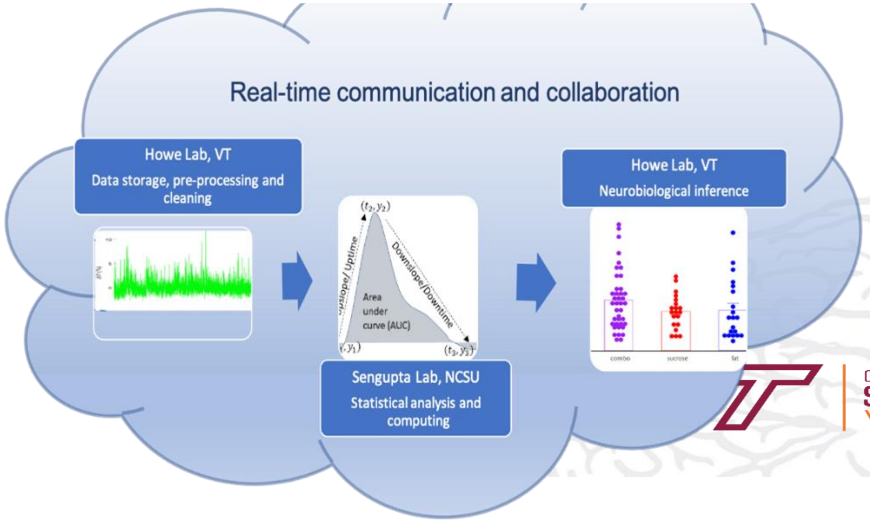
Perceived benefit of a cloud-based approach

In addition to overcoming the aforementioned issues.

1. ***Scalability:*** scale up and down storage and computing software as needed, as opposed to investing in on-premises infrastructure
2. ***Computational efficiency:*** potentially faster data analysis
3. ***Reduced costs:*** reduced costs for analysis tools (on-demand); reduced person hours per step, more effort back into primary research.
4. ***Increased collaboration****:* data can be shared in real time, collaborative computing and data auditing

5. ***Security***: enhanced security measures over what is available on current personal machines; sets the stage for future projects in human subjects.

6. ***Dissemination***: Projects make it out to the scientific community more rapidly.



COLLEGE OF SCIENCE
SCHOOL OF NEUROSCIENCE
VIRGINIA TECH.

Proposed plan

AIM 1: Compare data transfer and computational costs under the two workflows

- Reproduce our existing statistical analysis in the cloud environment, compare overall running time, total network data transfer, cumulative CPU and memory usage. Prediction: cloud will improve runtime, reduce need for data transfer, optimize compute expense

AIM 2: Determine personnel costs for each workflow and approach

- Start with raw datasets, quantify hours spent in learning how to use the cloud infrastructure vs. existing on-premises system for data management, storage, and analysis

AIM 3: Probe scalability costs between approaches

- Conduct a standard analysis on a typical sample of mice (6-10 mice); use this to project for dataset growth needed to account for known sources of variance in brain data that are typically ignored given computational costs (biological sex, time of day, mouse strain), compare estimated time and cost for cloud vs. on-premises approaches.

Implementation: VIRGINIA TECH CENTER FOR BIOSTATISTICS & HEALTH DATA SCIENCE

…EGE OF SCIENCE …HOOL OF NEUROSCIENCE …GINIA TECH.