Information Visualization

eXplainable Artificial Intelligence

Dr. David Koop

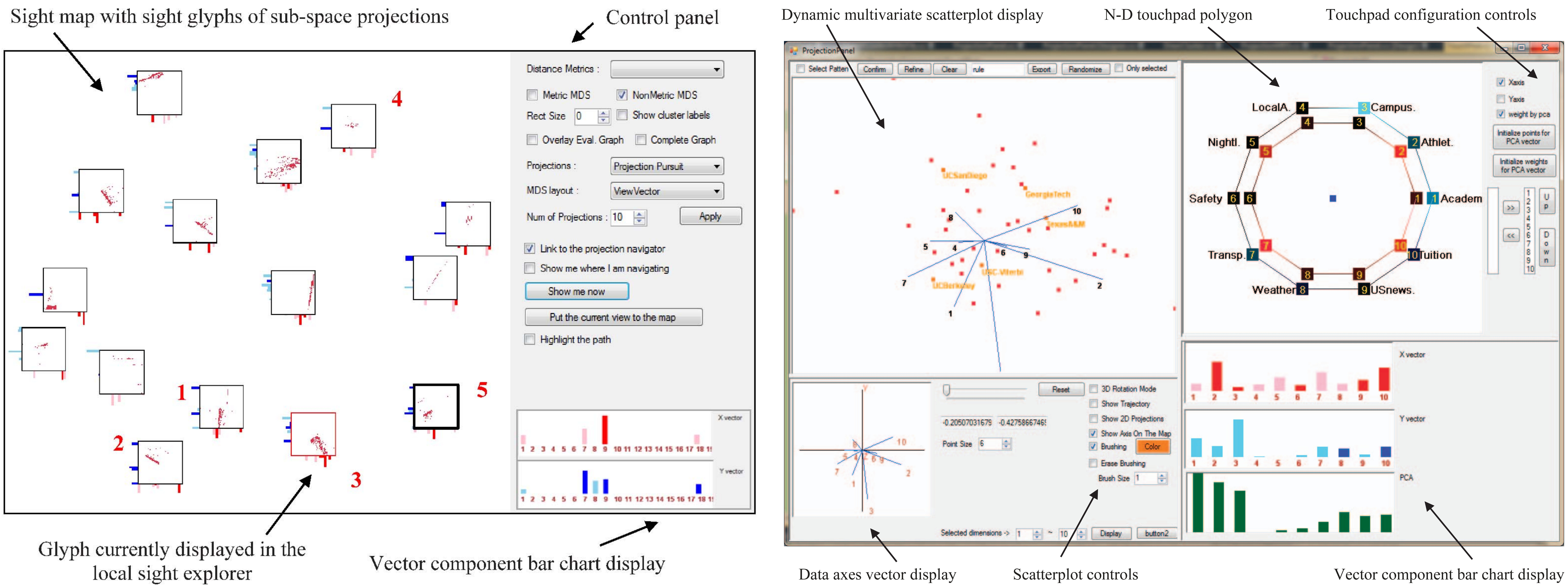
Northern Illinois University

High-Dimensional Data Exploration

- What are the tasks?
  - Discovering data configurations according to personal preference
  - Understanding the tradeoffs involved in such configurations
  - Partition the data (or views) to help with exploration
- Goals of TripAdvisorND & Subspace Voyager:
  - Facilitate examination of key projection and key clustering
  - Let users explore and tweak

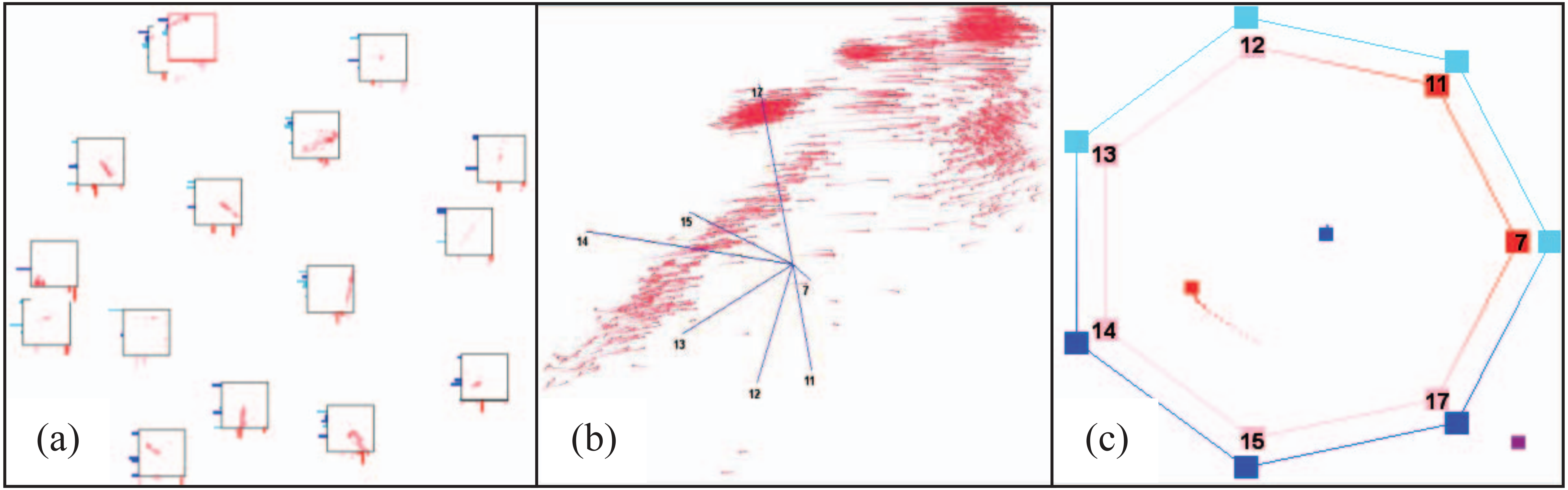
[J. E. Nam & K. Mueller, 2013]

TripAdvisor-ND: Global Sight Map & Local Sight Explorer



[J. E. Nam & K. Mueller, 2013]

Coordination of Views



[J. E. Nam & K. Mueller, 2013]

N-D Touchpad Polygon

- 2 polygons, one for each axis (inner = x, outer = y)
- Controls the orientation of the two PPA vectors
- Shading of vertices indicates weight
- Move the vertices around to change the weights

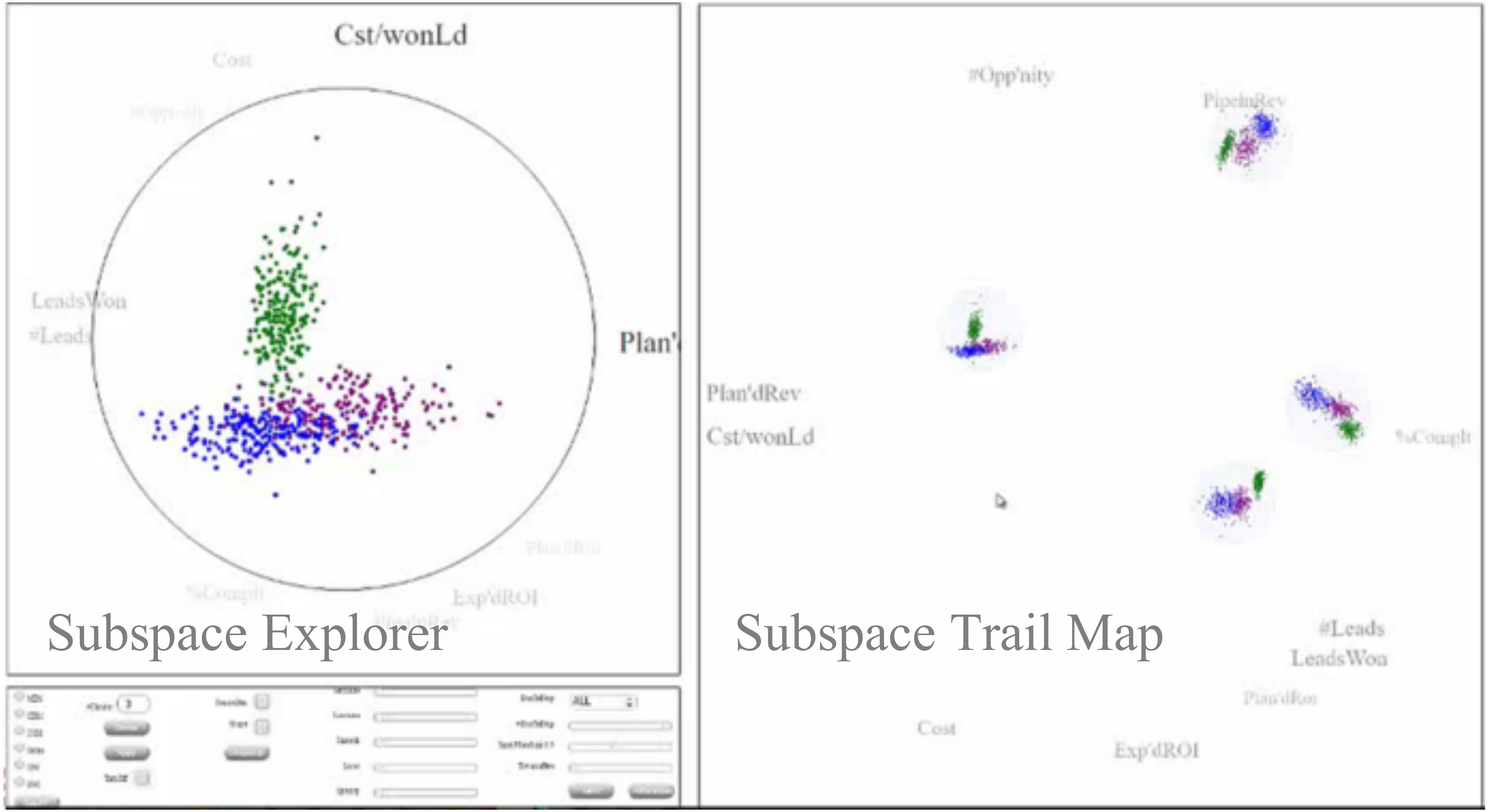
[J. E. Nam & K. Mueller, 2013]

Problems with TripAdvisor-ND

- Have to keep track of two views at once
- …so single window
- Have to move around two points in ND trackpad
- …so trackball interface
- Hard to map axes
- …so direct labeling

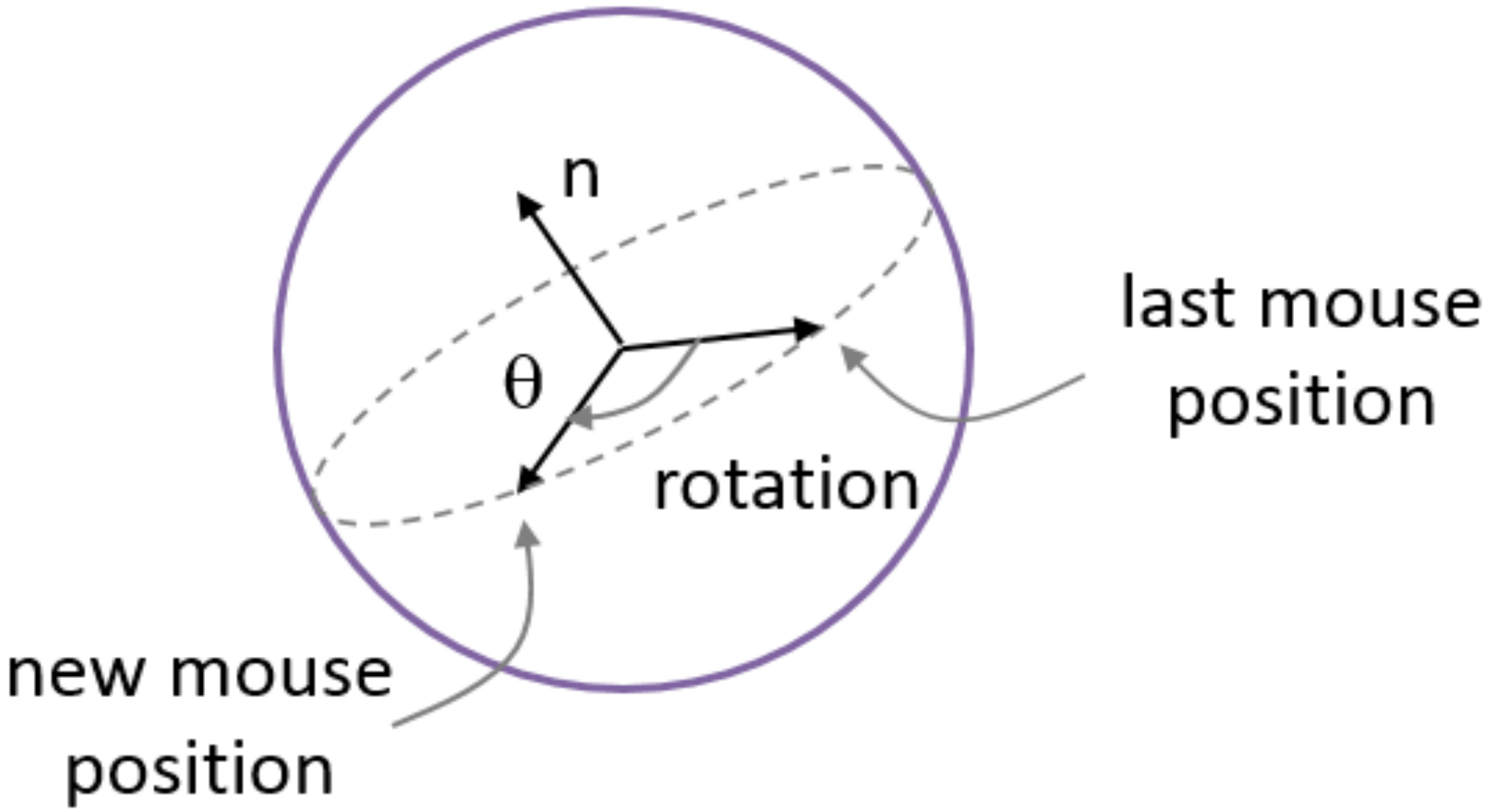
[B. Wang & K. Mueller, 2018]

Subspace Voyager Interface



[B. Wang & K. Mueller, 2018]

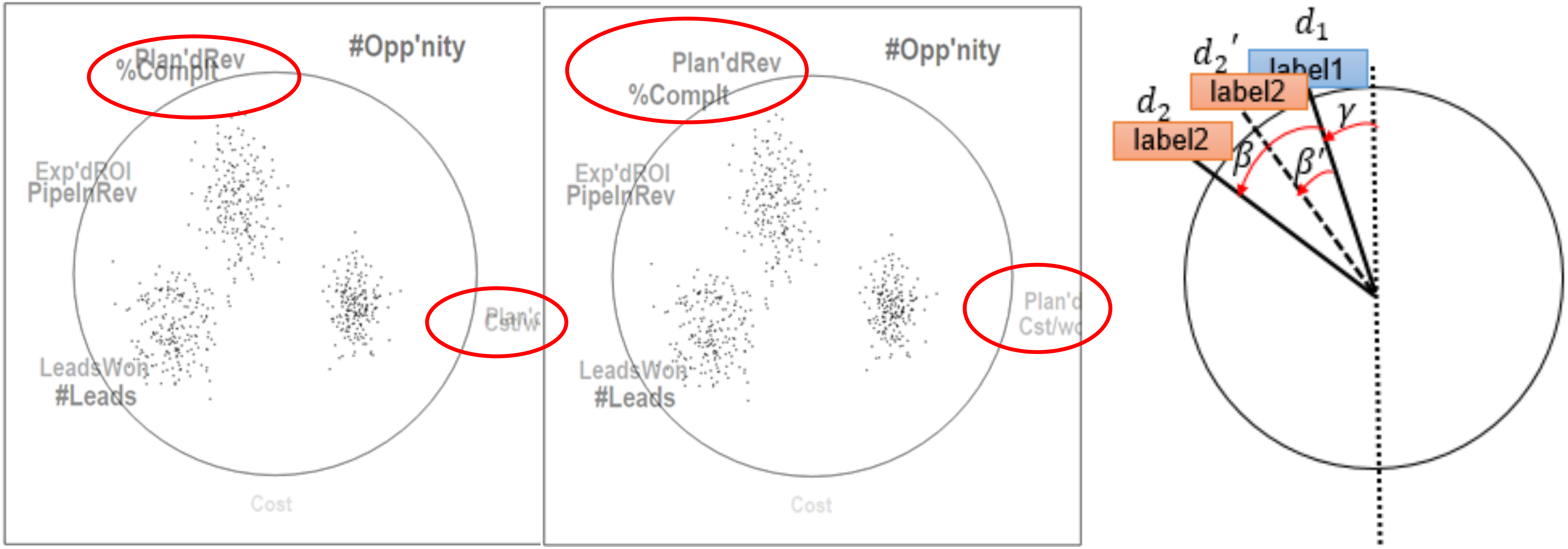
Trackball Interface



- Can use different mouse buttons
- Left: rotation
- Right: transition by changing axis
- Middle: travel along orthogonal vector (a z-axis), can't see until rotation

[B. Wang & K. Mueller, 2018]

Fix Labels



(a) (b) (c)

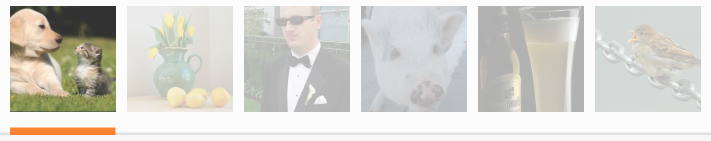
[B. Wang & K. Mueller, 2018]

Next Class: Critique Due

The Building Blocks of Interpretability

Interpretability techniques are normally studied in isolation.
We explore the powerful interfaces that arise when you combine them—
and the rich structure of this combinatorial space.

CHOOSE AN INPUT IMAGE



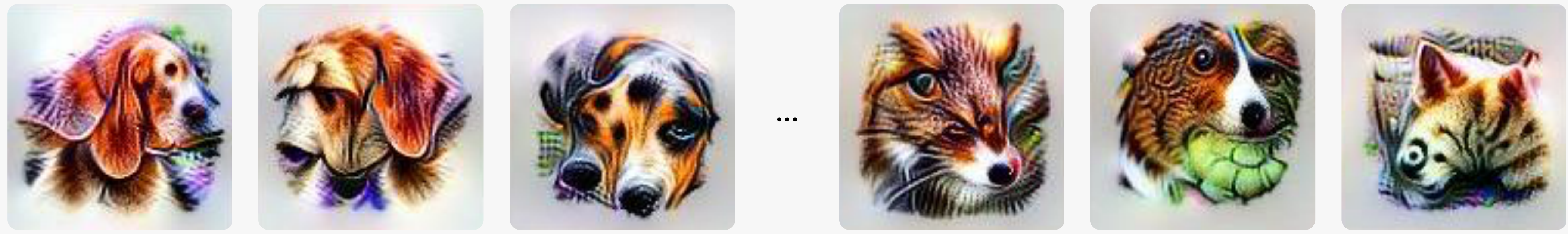
For instance, by combining feature visualization (*what is a neuron looking for?*) with attribution (*how does it affect the output?*), we can explore how the network decides between labels like **Labrador retriever** and **tiger cat**.

Several floppy ear detectors seem to be important when distinguishing dogs, whereas pointy ears are used to classify "tiger cat".

CHANNELS THAT MOST SUPPORT ... LABRADOR RETRIEVER ▾ ← → TIGER CAT ▾

feature visualization of channel

hover for attribution maps →



	LABRADOR RETRIEVER				TIGER CAT		
net evidence	1.63	1.51	1.19	...	1.32	1.54	1.72
for "Labrador retriever"	1.22	1.24	1.32		-0.70	-1.24	-0.43
for "tiger cat"	-0.40	-0.27	0.13		0.62	0.30	1.29

[C. Olah et al.]

Introduction to eXplainable AI (XAI)

Q. V. Liao, M. Singh, Y. Zhang, and R. Bella

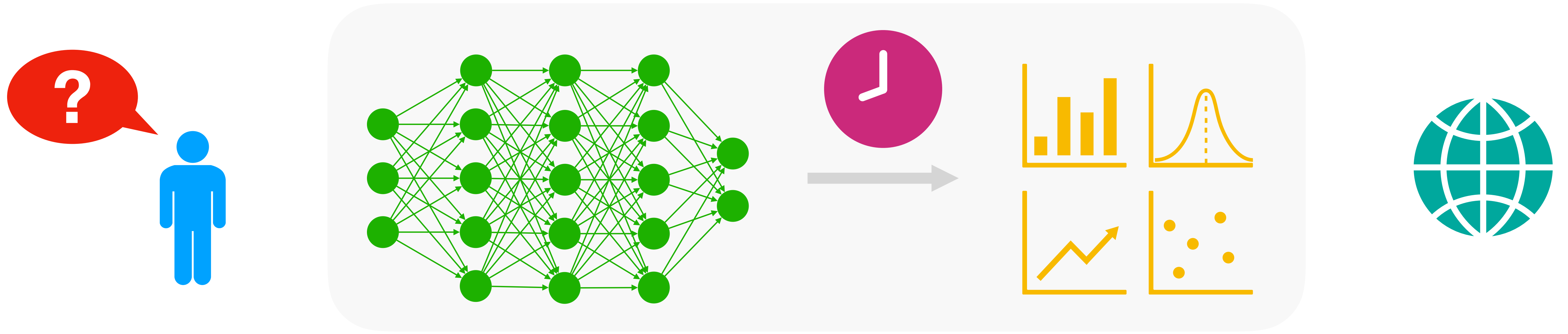
Survey of visualization in deep learning

Visual Analytics in Deep Learning | Interrogative Survey Overview

§4 WHY
Why would one want to use visualization in deep learning?
- Interpretability & Explainability
- Debugging & Improving Models
- Comparing & Selecting Models
- Teaching Deep Learning Concepts

§6 WHAT
What data, features, and relationships in deep learning can be visualized?
- Computational Graph & Network Architecture
- Learned Model Parameters
- Individual Computational Units
- Neurons In High-dimensional Space
- Aggregated Information

§8 WHEN
When in the deep learning process is visualization used?
- During Training
- After Training



§5 WHO
Who would use and benefit from visualizing deep learning?
- Model Developers & Builders
- Model Users
- Non-experts

§7 HOW
How can we visualize deep learning data, features, and relationships?
- Node-link Diagrams for Network Architecture
- Dimensionality Reduction & Scatter Plots
- Line Charts for Temporal Metrics
- Instance-based Analysis & Exploration
- Interactive Experimentation
- Algorithms for Attribution & Feature Visualization

§9 WHERE
Where has deep learning visualization been used?
- Application Domains & Models
- A Vibrant Research Community

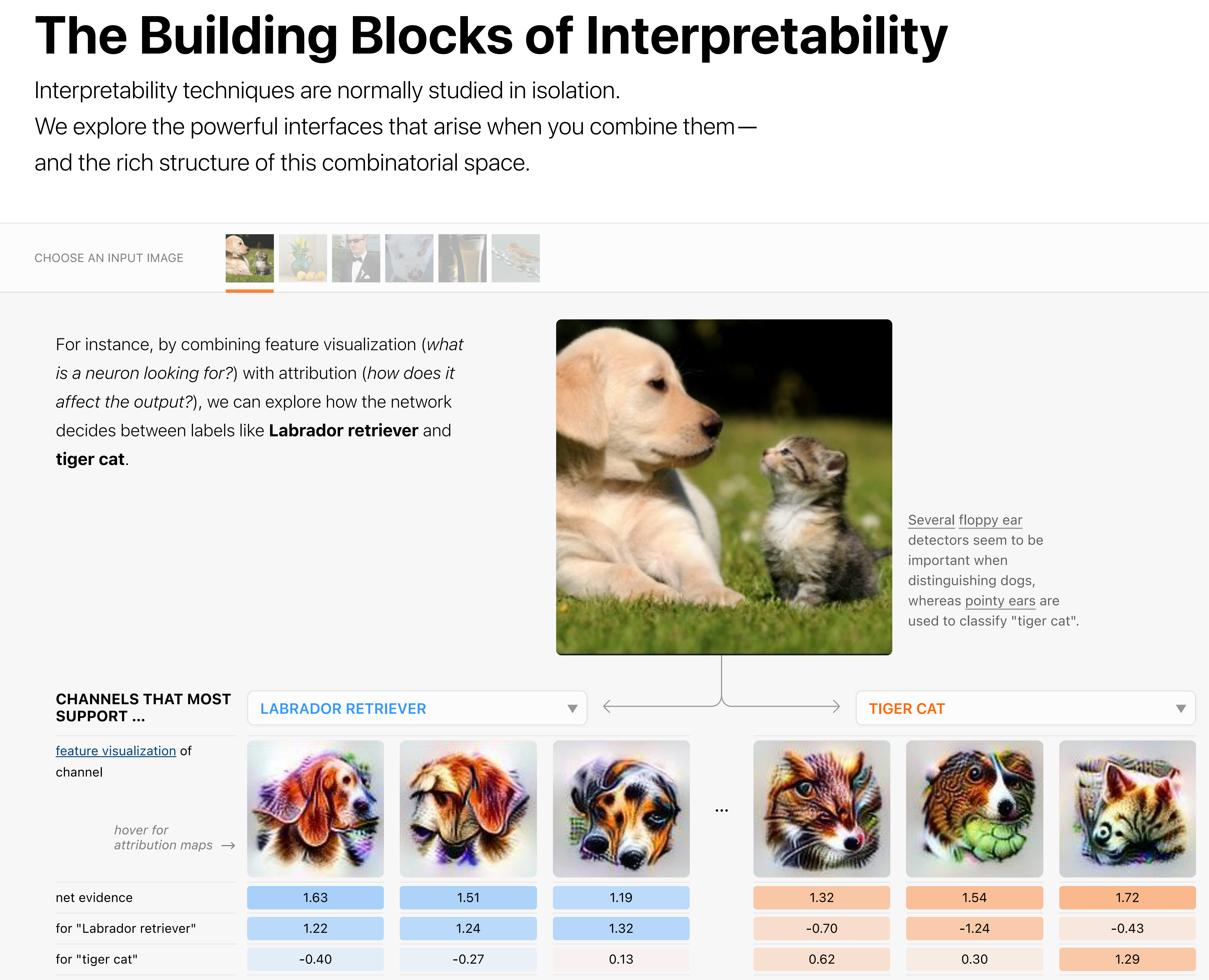
[F. Hohman et al.]

Survey Landscape

	WHY				WHO			WHAT					HOW						WHEN		WHERE
Work	4.1 Interpretability & Explainability	4.2 Debugging & Improving Models	4.3 Comparing & Selecting Models	4.4 Teaching Deep Learning Concepts	5.1 Model Developers & Builders	5.2 Model Users	5.3 Non-experts	6.1 Computational Graph & Network Architecture	6.2 Learned Model Parameters	6.3 Individual Computational Units	6.4 Neurons in High-dimensional Space	6.5 Aggregated Information	7.1 Node-link Diagrams for Network Architecture	7.2 Dimensionality Reduction & Scatter Plots	7.3 Line Charts for Temporal Metrics	7.4 Instance-based Analysis & Exploration	7.5 Interactive Experimentation	7.6 Algorithms for Attribution & Feature Visualization	8.1 During Training	8.2 After Training	9.2 Publication Venue
Abadi, et al., 2016 [27]	●	●	●		●	●						●			●				●	●	arXiv
Bau, et al., 2017 [28]	●		●		●					●						●		●		●	CVPR
Bilal, et al., 2017 [29]	●	●			●					●		●				●		●	●		TVCG
Bojarski, et al., 2016 [30]	●	●			●				●			●				●		●	●	●	arXiv
Bruckner, 2014 [31]	●	●			●			●	●				●			●		●	●		MS Thesis
Carter, et al., 2016 [32]	●			●	●	●	●			●	●	●				●	●			●	Distill
Cashman, et al., 2017 [33]	●	●			●	●			●	●						●				●	VADL

[Hohman et al.]

Next Class: Critique Due



[C. Olah et al.]