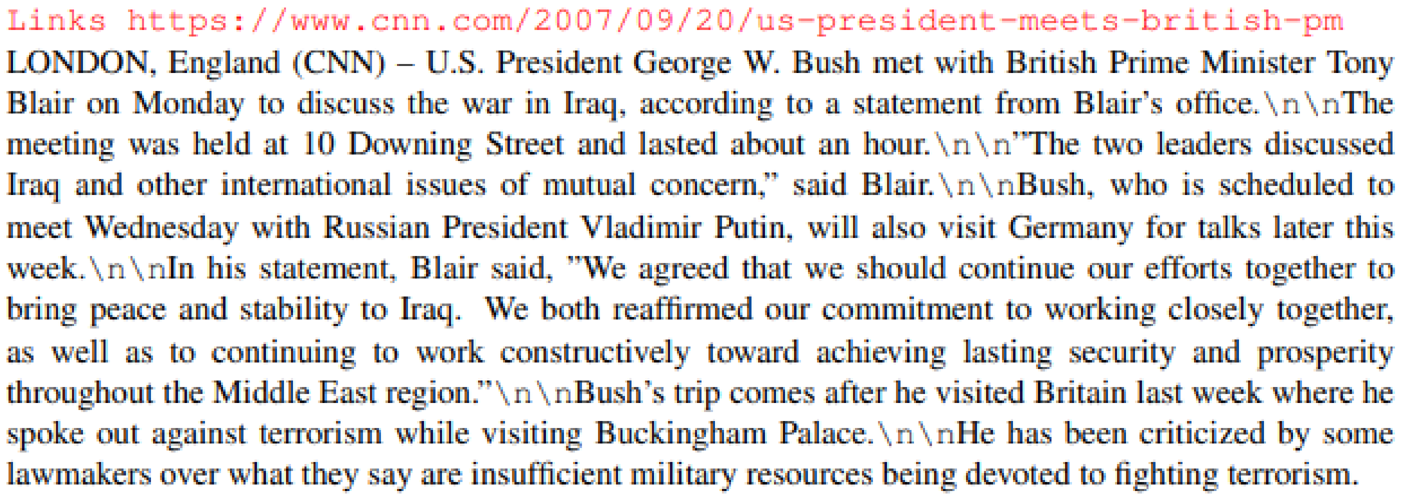
The Information Lattice Transform

Lav R. Varshney

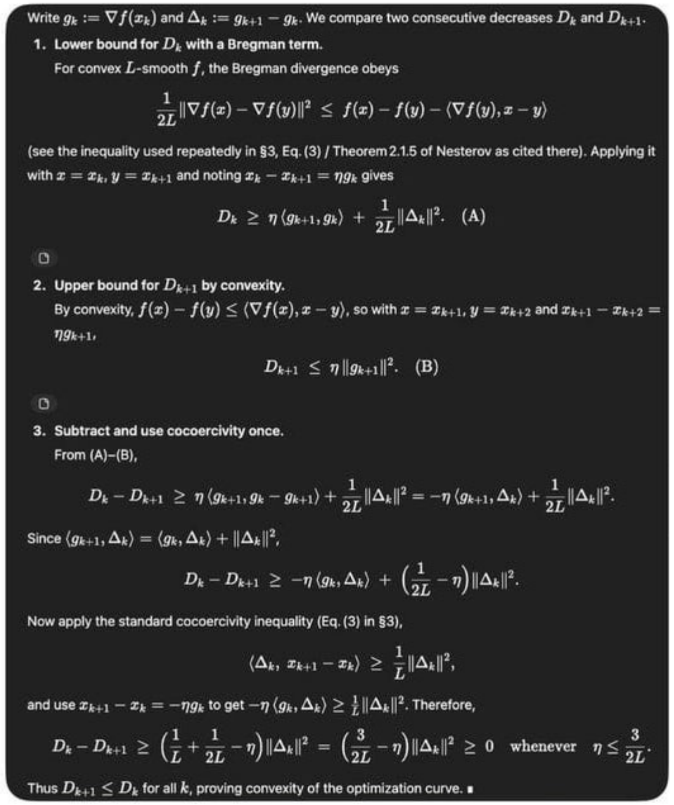
Della Pietra Infinity Professor and Inaugural Director of AI Innovation Institute, **Stony Brook University**
Co-Founder and CEO, **Kocree, Inc.**
Computational Scientist, **Brookhaven National Laboratory**

12 September 2025



Links https://www.cnn.com/2007/09/20/us-president-meets-british-pm
LONDON, England (CNN) – U.S. President George W. Bush met with British Prime Minister Tony Blair on Monday to discuss the war in Iraq, according to a statement from Blair's office.\n\nThe meeting was held at 10 Downing Street and lasted about an hour.\n\n"The two leaders discussed Iraq and other international issues of mutual concern," said Blair.\n\nBush, who is scheduled to meet Wednesday with Russian President Vladimir Putin, will also visit Germany for talks later this week.\n\nIn his statement, Blair said, "We agreed that we should continue our efforts together to bring peace and stability to Iraq. We both reaffirmed our commitment to working closely together, as well as to continuing to work constructively toward achieving lasting security and prosperity throughout the Middle East region."\n\nBush's trip comes after he visited Britain last week where he spoke out against terrorism while visiting Buckingham Palace.\n\nHe has been criticized by some lawmakers over what they say are insufficient military resources being devoted to fighting terrorism.

[N. S. Keskar, B. McCann, L. R. Varshney, C. Xiong, and R. Socher, "CTRL: A Conditional Transformer Language Model for Controllable Generation," Sept. 2019.]



Write $g_k := \nabla f(x_k)$ and $\Delta_k := g_{k+1} - g_k$. We compare two consecutive decreases D_k and D_{k+1}.

1. **Lower bound for D_k with a Bregman term.**
   For convex L-smooth f, the Bregman divergence obeys

$$\frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \leq f(x) - f(y) - \langle \nabla f(y), x - y \rangle$$

(see the inequality used repeatedly in §3, Eq. (3) / Theorem 2.1.5 of Nesterov as cited there). Applying it with $x = x_k$, $y = x_{k+1}$ and noting $x_k - x_{k+1} = \eta g_k$ gives

$$D_k \geq \eta \langle g_{k+1}, g_k \rangle + \frac{1}{2L}\|\Delta_k\|^2. \quad \text{(A)}$$

2. **Upper bound for D_{k+1} by convexity.**
   By convexity, $f(x) - f(y) \leq \langle \nabla f(x), x - y \rangle$, so with $x = x_{k+1}$, $y = x_{k+2}$ and $x_{k+1} - x_{k+2} = \eta g_{k+1}$,

$$D_{k+1} \leq \eta \|g_{k+1}\|^2. \quad \text{(B)}$$

3. **Subtract and use cocoercivity once.**
   From (A)–(B),

$$D_k - D_{k+1} \geq \eta \langle g_{k+1}, g_k - g_{k+1} \rangle + \frac{1}{2L}\|\Delta_k\|^2 = -\eta \langle g_{k+1}, \Delta_k \rangle + \frac{1}{2L}\|\Delta_k\|^2.$$

Since $\langle g_{k+1}, \Delta_k \rangle = \langle g_k, \Delta_k \rangle + \|\Delta_k\|^2$,

$$D_k - D_{k+1} \geq -\eta \langle g_k, \Delta_k \rangle + \left(\frac{1}{2L} - \eta\right)\|\Delta_k\|^2.$$

Now apply the standard cocoercivity inequality (Eq. (3) in §3),

$$\langle \Delta_k,\ x_{k+1} - x_k \rangle \geq \frac{1}{L}\|\Delta_k\|^2,$$

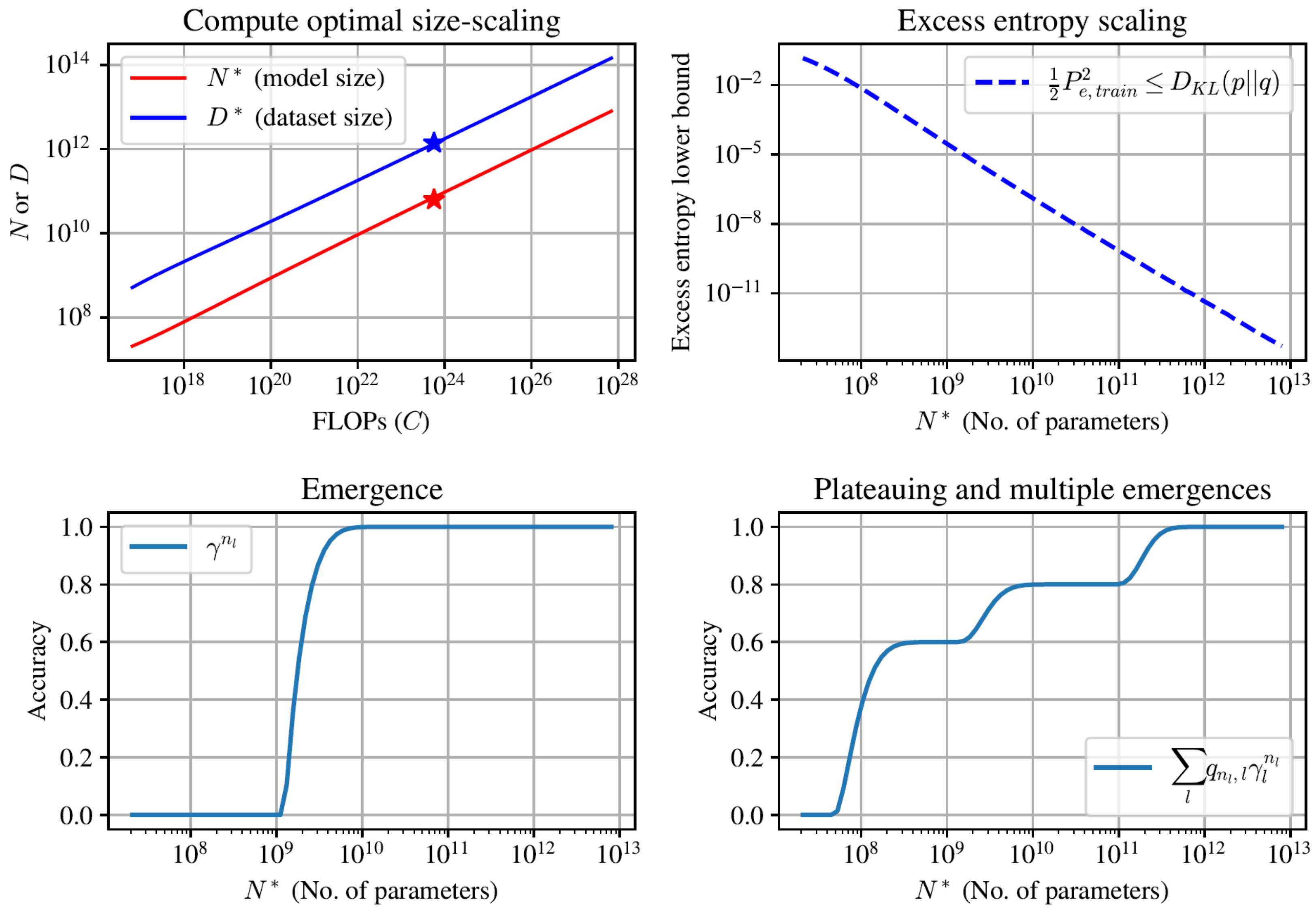
and use $x_{k+1} - x_k = -\eta g_k$ to get $-\eta \langle g_k, \Delta_k \rangle \geq \frac{1}{L}\|\Delta_k\|^2$. Therefore,

$$D_k - D_{k+1} \geq \left(\frac{1}{L} + \frac{1}{2L} - \eta\right)\|\Delta_k\|^2 = \left(\frac{3}{2L} - \eta\right)\|\Delta_k\|^2 \geq 0 \quad \text{whenever} \quad \eta \leq \frac{3}{2L}.$$

Thus $D_{k+1} \leq D_k$ for all k, proving convexity of the optimization curve. ∎

[Sebastien Bubeck on X: "Claim: gpt-5-pro can prove new interesting mathematics. Proof: I took a convex optimization paper with a clean open problem in it and asked gpt-5-pro to work on it. It proved a better bound than what is in the paper, and I checked the proof it's correct," Aug. 2025.]

Scaling, Emergent Capabilities, and Plateaus

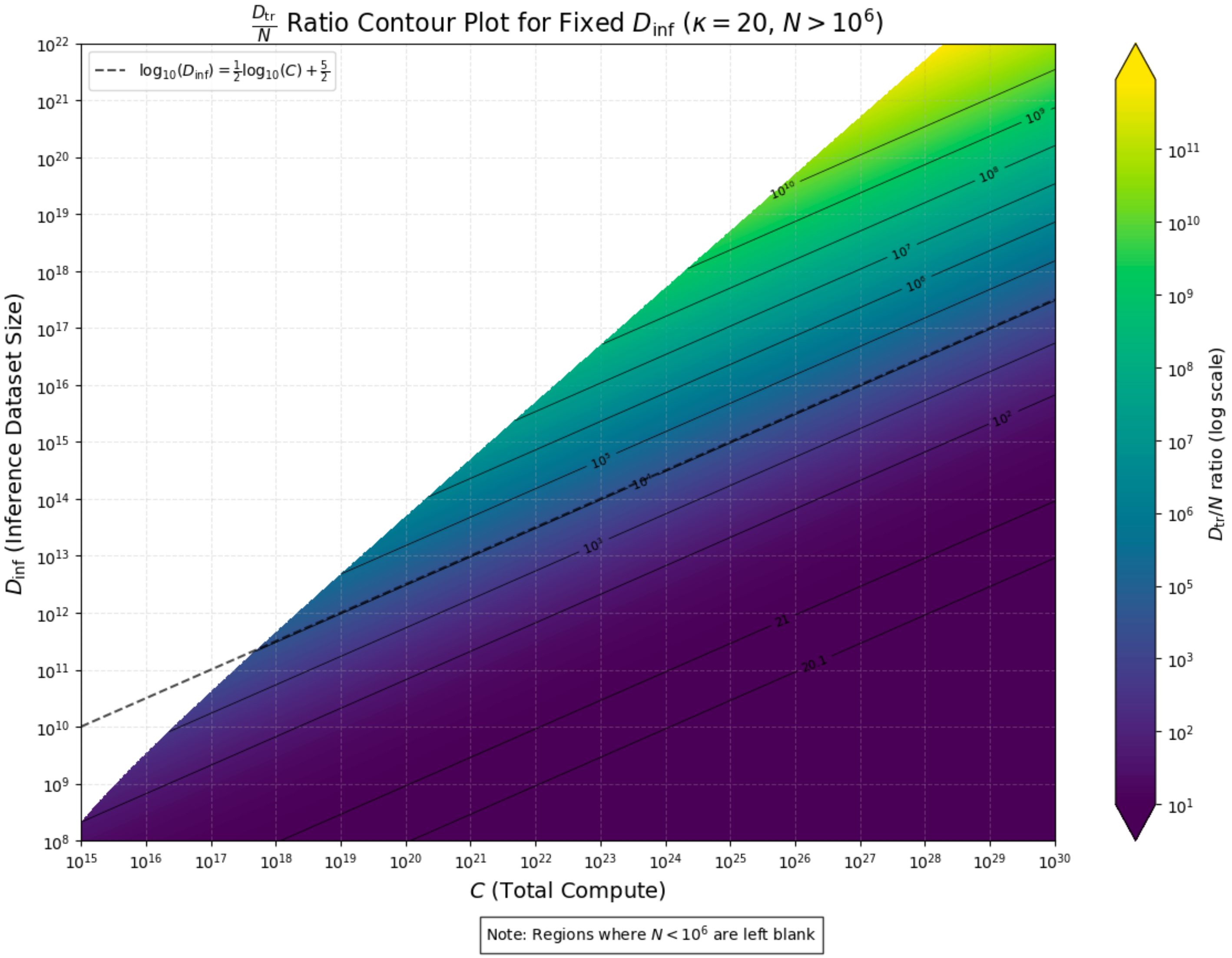


[A. K. Nayak and L. R. Varshney, "An Information Theory of Compute-Optimal Size Scaling, Emergence, and Plateaus in Language Models," *NeurIPS Workshop*, Dec. 2024]

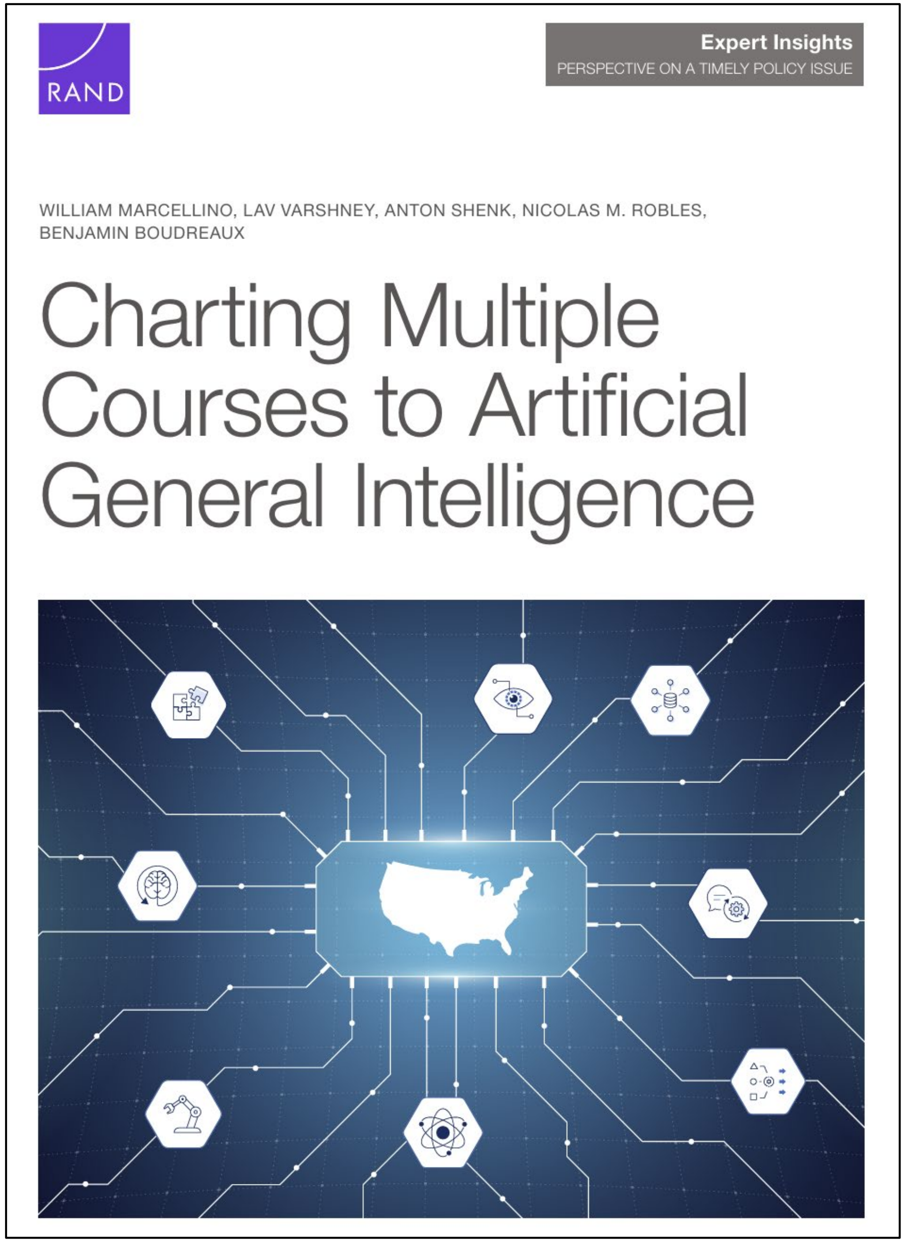
Inference-compute scaling too

Inference-time reasoning modeled as a directed stochastic search over a model's pretrained skill graph

Optimal training data / parameter relationship has logarithmic dependence on inference to total compute ratio



[A. R. Ellis-Mohr, A. K. Nayak, and L. R. Varshney, "A Theory of Inference Compute Scaling: Reasoning through Directed Stochastic Skill Search," *Philosophical Transactions of the Royal Society A*, to appear.]

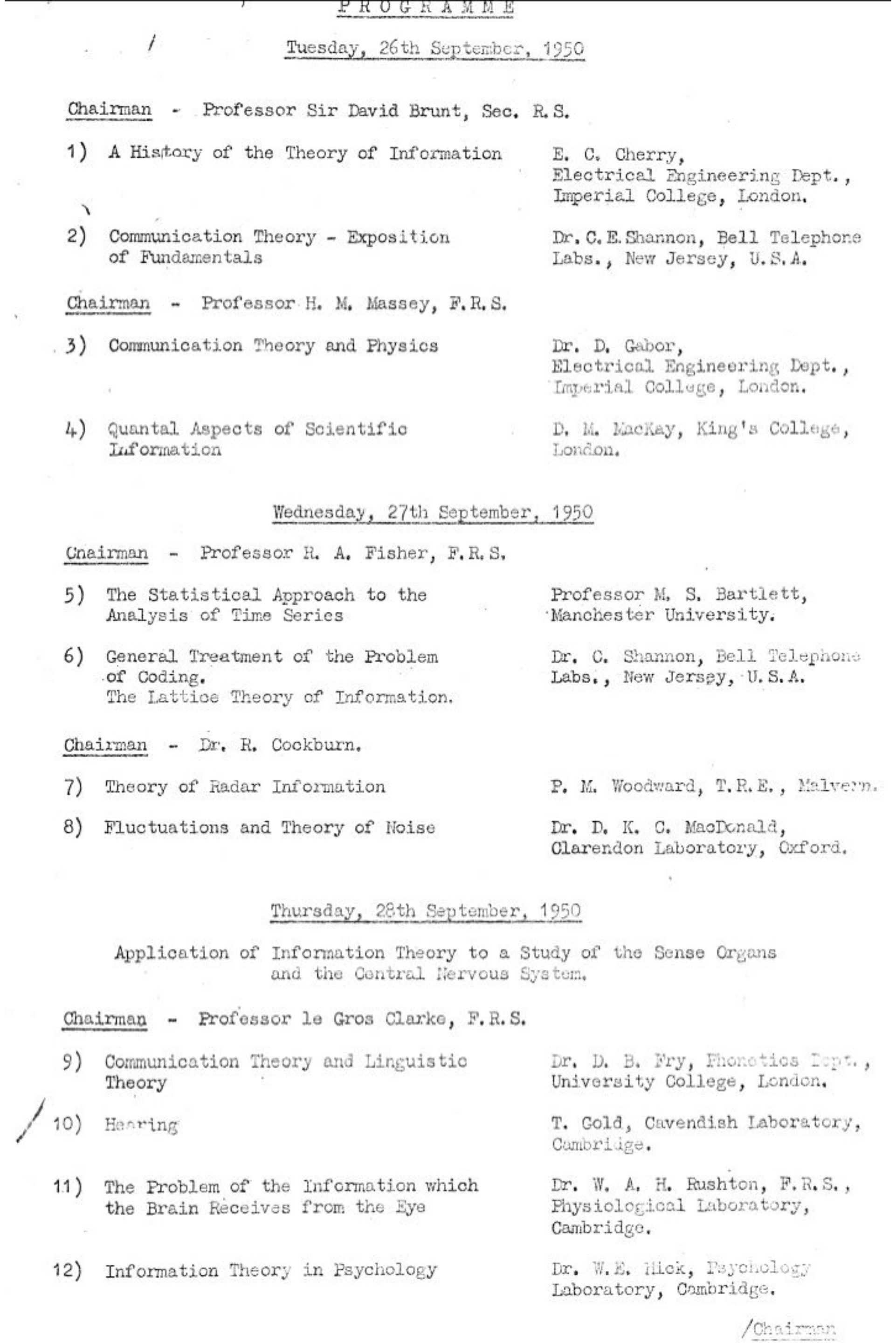


RAND

Expert Insights
PERSPECTIVE ON A TIMELY POLICY ISSUE

WILLIAM MARCELLINO, LAV VARSHNEY, ANTON SHENK, NICOLAS M. ROBLES,
BENJAMIN BOUDREAUX

Charting Multiple Courses to Artificial General Intelligence



PROGRAMME

Tuesday, 26th September, 1950

Chairman - Professor Sir David Brunt, Sec. R.S.

1)	A History of the Theory of Information	E. C. Cherry, Electrical Engineering Dept., Imperial College, London.
2)	Communication Theory - Exposition of Fundamentals	Dr. C. E. Shannon, Bell Telephone Labs., New Jersey, U.S.A.

Chairman - Professor H. M. Massey, F.R.S.

3)	Communication Theory and Physics	Dr. D. Gabor, Electrical Engineering Dept., Imperial College, London.
4)	Quantal Aspects of Scientific Information	D. M. MacKay, King's College, London.

Wednesday, 27th September, 1950

Chairman - Professor R. A. Fisher, F.R.S.

5)	The Statistical Approach to the Analysis of Time Series	Professor M. S. Bartlett, Manchester University.
6)	General Treatment of the Problem of Coding. The Lattice Theory of Information.	Dr. C. Shannon, Bell Telephone Labs., New Jersey, U.S.A.

Chairman - Dr. R. Cockburn.

7)	Theory of Radar Information	P. M. Woodward, T.R.E., Malvern.
8)	Fluctuations and Theory of Noise	Dr. D. K. C. MacDonald, Clarendon Laboratory, Oxford.

Thursday, 28th September, 1950

Application of Information Theory to a Study of the Sense Organs and the Central Nervous System.

Chairman - Professor le Gros Clarke, F.R.S.

9)	Communication Theory and Linguistic Theory	Dr. D. B. Fry, Phonetics Dept., University College, London.
10)	Hearing	T. Gold, Cavendish Laboratory, Cambridge.
11)	The Problem of the Information which the Brain Receives from the Eye	Dr. W. A. H. Rushton, F.R.S., Physiological Laboratory, Cambridge.
12)	Information Theory in Psychology	Dr. W. E. Hick, Psychology Laboratory, Cambridge.

/Chairman

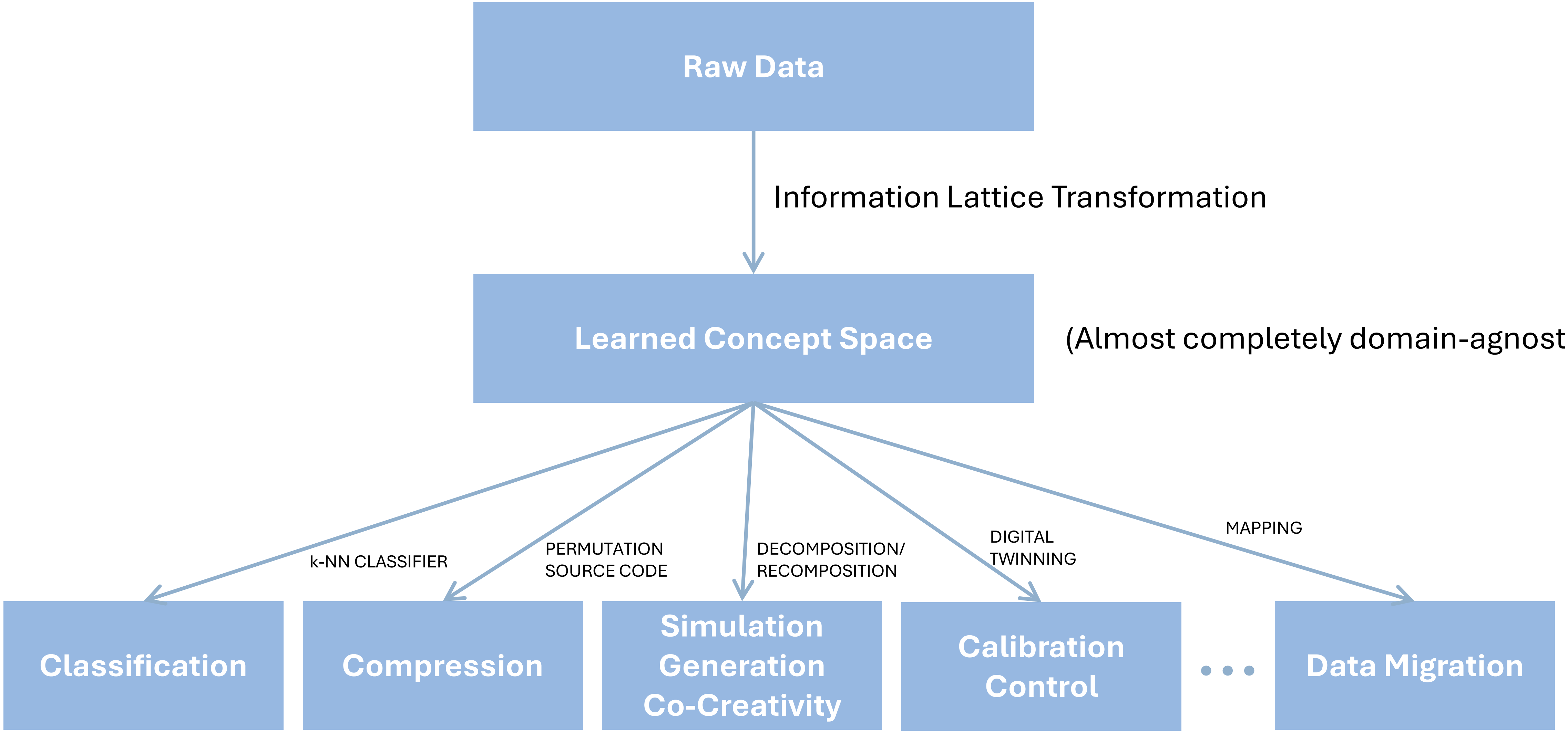
Information lattice learning (ILL) is white-box AI that is directly human-controllable, human-understandable, super data-efficient, super compute-efficient, and IP-liability free

It performs task-agnostic knowledge discovery for scientific, cultural, as well as local and idiosyncratic phenomenology

Discovered knowledge is valuable itself, but can further be used for clustering, classification, prediction, semantic compression, co-creativity, and numerous other downstream tasks

Provides a path to artificial general intelligence (AGI) distinct from the hyperscaling paradigm, in that task-agnostic learning can quickly adapt to most any downstream uncertainty just like people

Foundation lattice



Raw Data

Information Lattice Transformation

Learned Concept Space

(Almost completely domain-agnostic)

k-NN CLASSIFIER

PERMUTATION
SOURCE CODE

DECOMPOSITION/
RECOMPOSITION

DIGITAL
TWINNING

MAPPING

Classification

Compression

Simulation
Generation
Co-Creativity

Calibration
Control

. . .

Data Migration

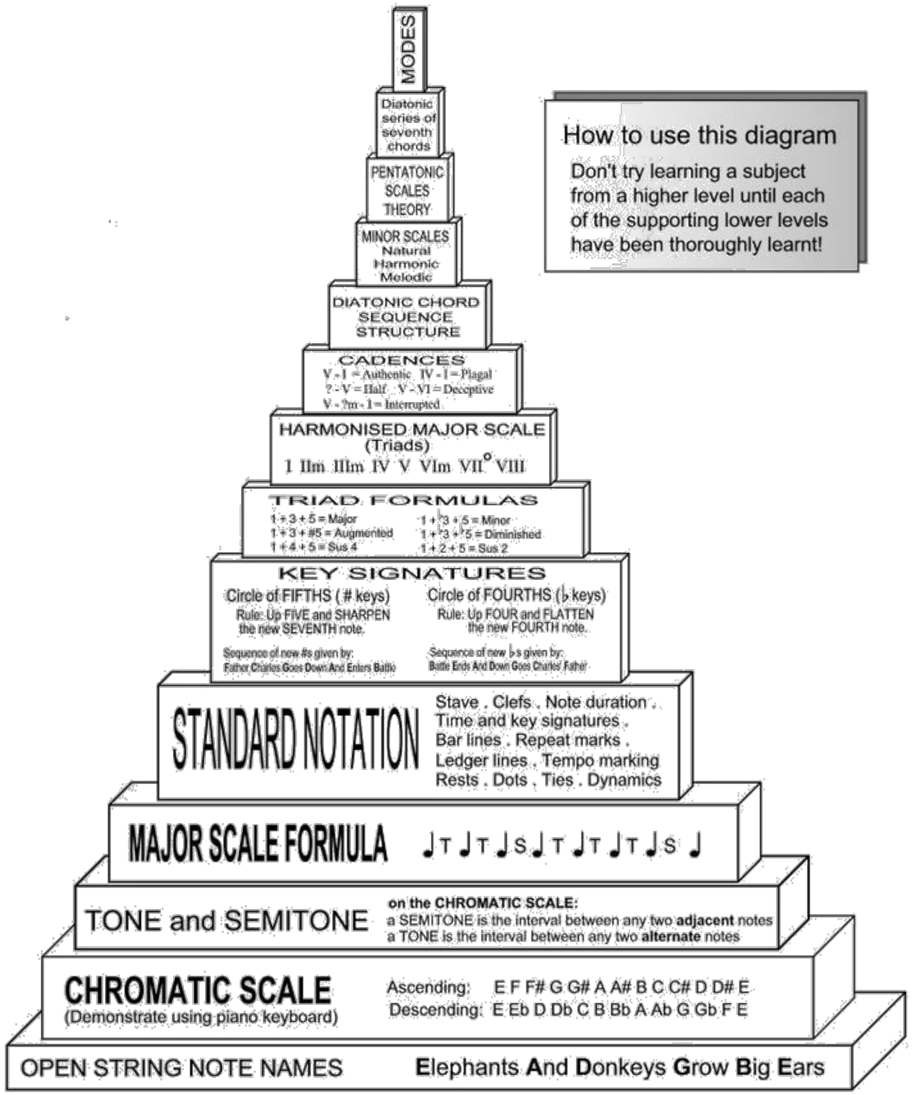
(May require some domain expertise)

Dimensions of interpretability [Selbst and Barocas, 2018]

- What sets machine learning models apart from other decision-making mechanisms are their *inscrutability* and *nonintuitiveness*
  - Inscrutability suggests that models available for direct inspection may defy understanding
  - Nonintuitiveness suggests that even where models are understandable, they may rest on apparent statistical relationships that defy intuition
  - Most extant work on interpretable ML/AI only addresses inscrutability, but not nonintuitiveness
- Dealing with inscrutability requires providing a sensible description of rules; addressing nonintuitiveness requires providing satisfying explanation for why the rules are what they are

For numerous settings, may need technical solutions to both inscrutability and nonintuitiveness

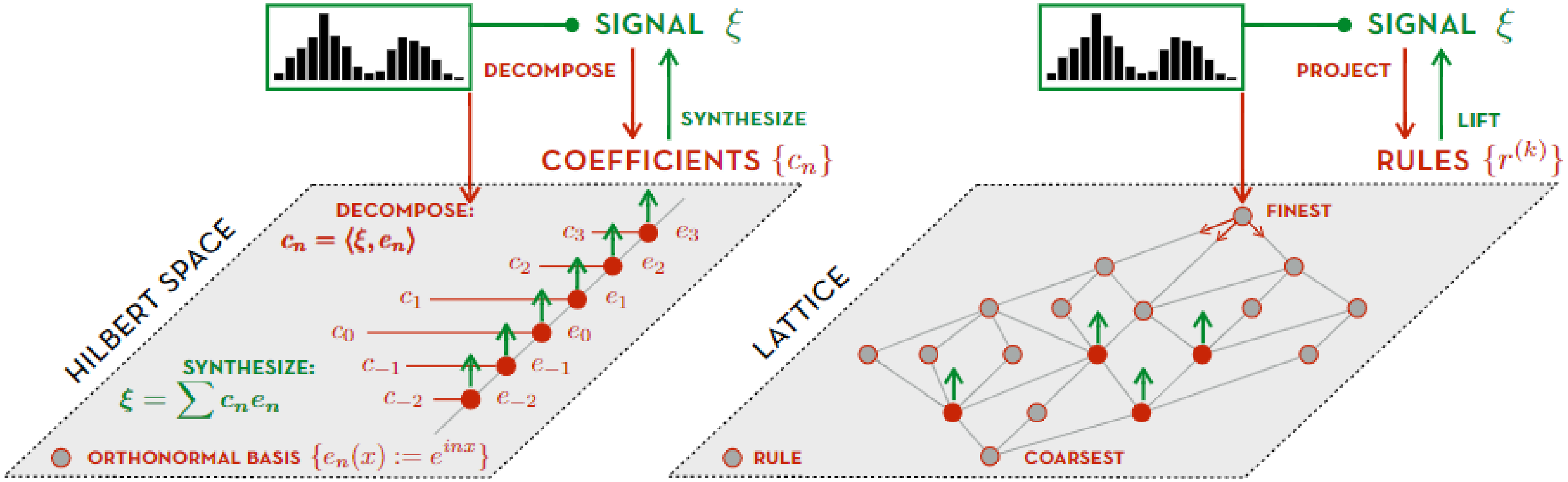
Learn human-interpretable concept *hierarchies* (not just rules)



[http://www.teachguitar.com/content/tmpyramid.htm]

> "Fundamentally, most current deep-learning based language models represent sentences as mere sequences of words, whereas Chomsky has long argued that language has a hierarchical structure, in which larger structures are recursively constructed out of smaller components."
>
> – Gary Marcus [*arXiv:1801.00631*]

Information lattice learning is intrinsically interpretable

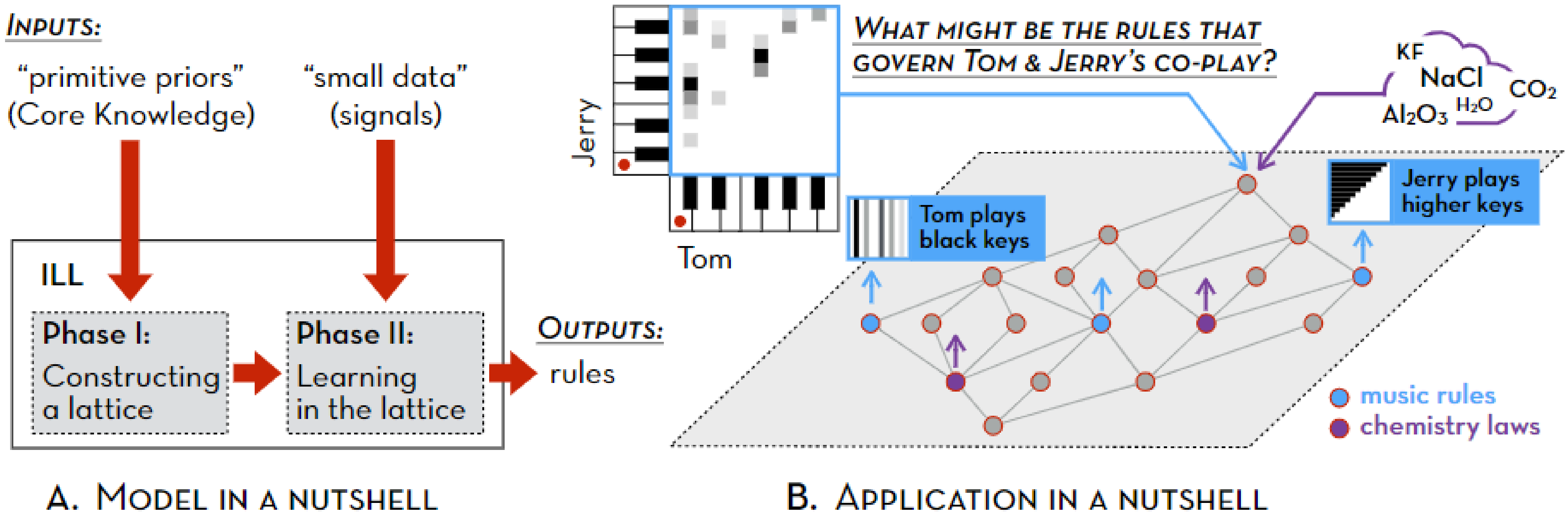


A. Fourier series B. Information lattice learning

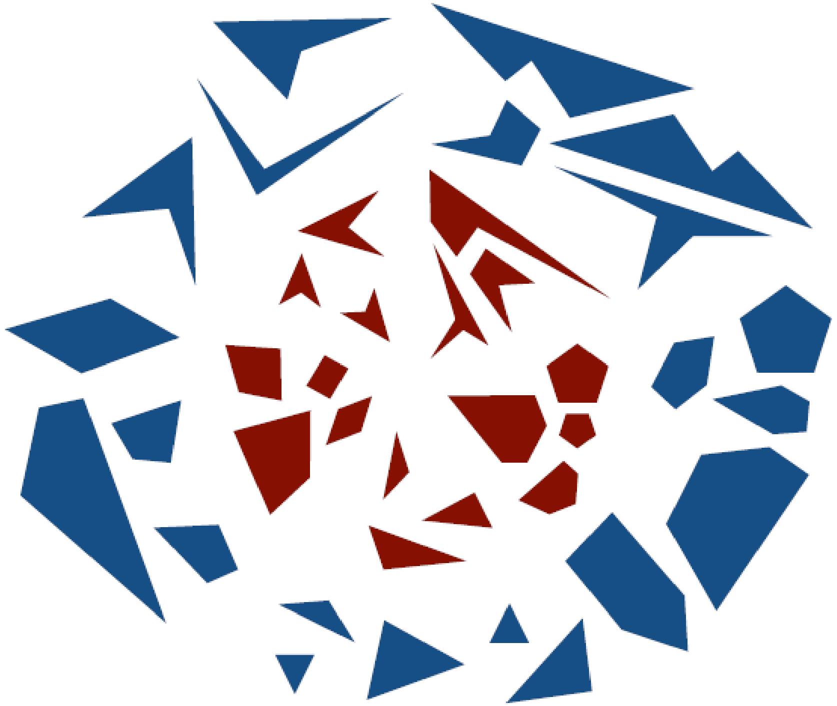
[H. Yu, J. A. Evans, and L. R. Varshney, “Information Lattice Learning,” *Journal of Artificial Intelligence Research*, vol. 77, pp. 971–1019, July 2023.]

- *Self-exploration*: learn domain knowledge from universal priors that encode no domain knowledge
    - Group-theoretic foundations and generalization of Shannon’s information lattice

- *Self-explanation*: aim for learned results and entire learning process to be human-interpretable
    - Iterative student-teacher architecture for learning algorithm, which produces interpretable hierarchy of interpretable concepts (with a particular mechanistic cause: symmetry) and its trace

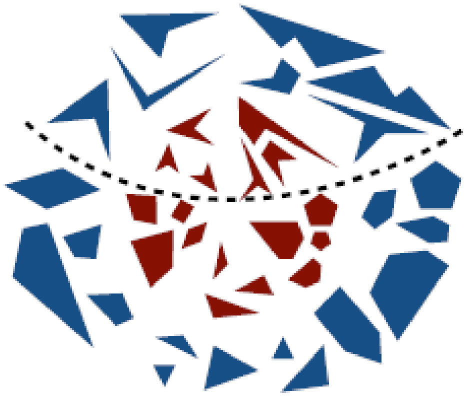


[H. Yu, J. A. Evans, and L. R. Varshney, "Information Lattice Learning," *Journal of Artificial Intelligence Research*, vol. 77, pp. 971–1019, July 2023.]

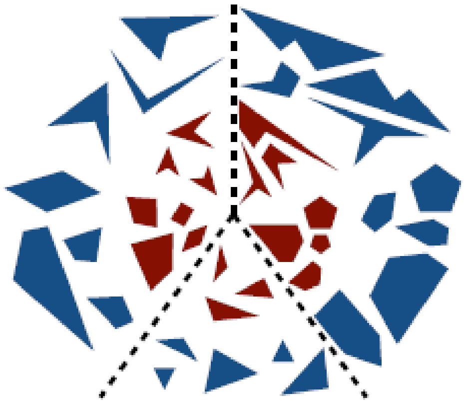


[H. Yu, I. Mineyev, and L. R. Varshney, “A Group-Theoretic Approach to Computational Abstraction: Symmetry-Driven Hierarchical Clustering,” *Journal of Machine Learning Research*, vol. 24, no. 47, pp. 1–61, 2023.]

{red, blue}



{convex, concave}



{trigon, tetragon, pentagon}

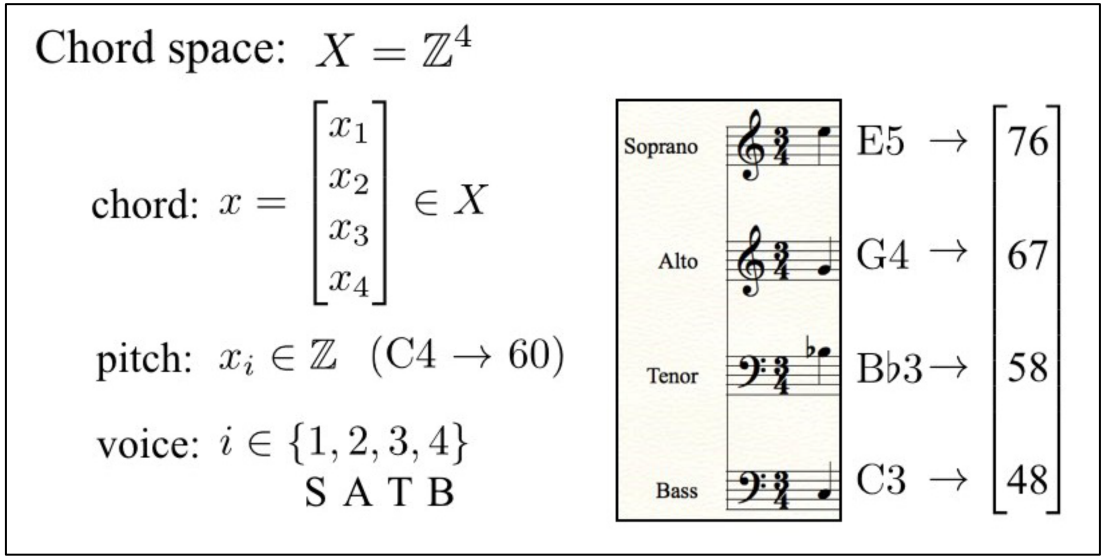
Representation: Data space

Data space: (X, p_X) or (X, p) for short

- Assume a data point $x \in X$ is an i.i.d. sample drawn from a probability distribution p

- However, the data distribution p (or an estimate of it) is *known*

- The goal here is not to estimate p but to *explain* it



Chord space: $X = \mathbb{Z}^4$

chord: $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \in X$

pitch: $x_i \in \mathbb{Z}$ $(\mathrm{C4} \to 60)$

voice: $i \in \{1, 2, 3, 4\}$
S A T B

$$\begin{matrix} \mathrm{E5} \to \\ \mathrm{G4} \to \\ \mathrm{B\flat 3} \to \\ \mathrm{C3} \to \end{matrix} \begin{bmatrix} 76 \\ 67 \\ 58 \\ 48 \end{bmatrix}$$

Representation: Abstraction

An **abstraction** $\mathcal{A}$ is a partition of the data space X.

$$X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$$

$$\mathcal{A} = \{\underline{\{x_1, x_6\}}, \underline{\{x_3\}}, \underline{\{x_2, x_4, x_5\}}\}$$

cells (or less formally, clusters)

An **concept** is a partition cell.

A **partition matrix** A is a concise way of representing an abstraction $\mathcal{A}$.

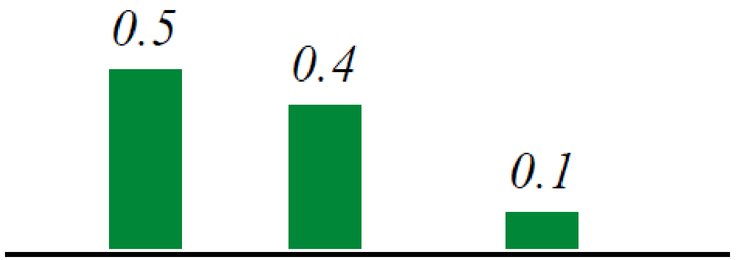
$$A = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{l} \\ \text{1st cell} \\ \text{2nd cell} \\ \text{3rd cell} \end{array}$$

Representation: Probabilistic Rule

A **probabilistic rule** is a pair:

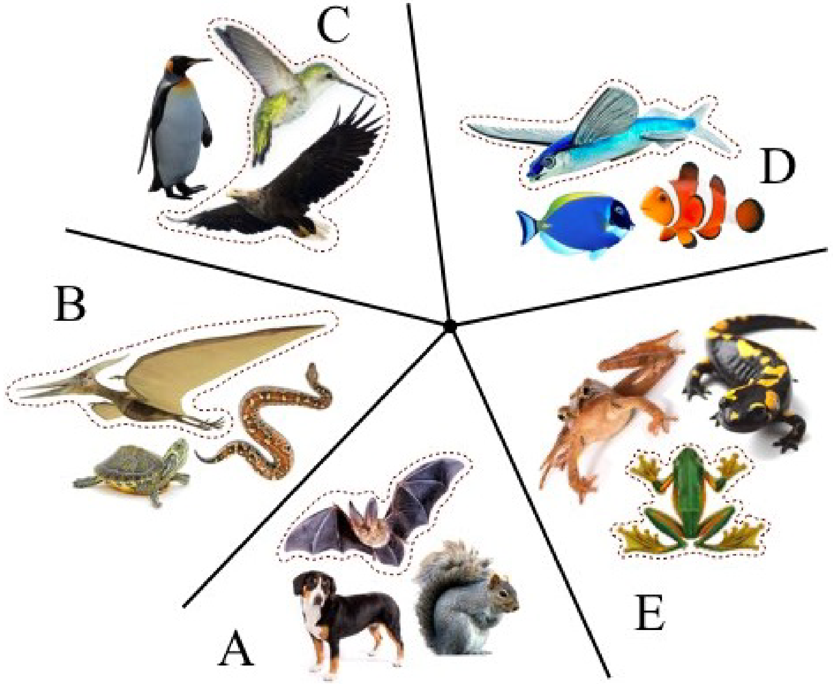
$$(\mathcal{A}, p_{\mathcal{A}})$$

where $\mathcal{A}$ is an abstraction (partition);
$p_{\mathcal{A}}$ is a probability distribution over the abstracted concepts (cells).



$$\mathcal{A} = \{\{x_1, x_6\}, \{x_3\}, \{x_2, x_4, x_5\}\}$$

“Most birds fly; but rare for fish, amphibians, reptiles, mammals.”



Abstraction (of vertebrates):

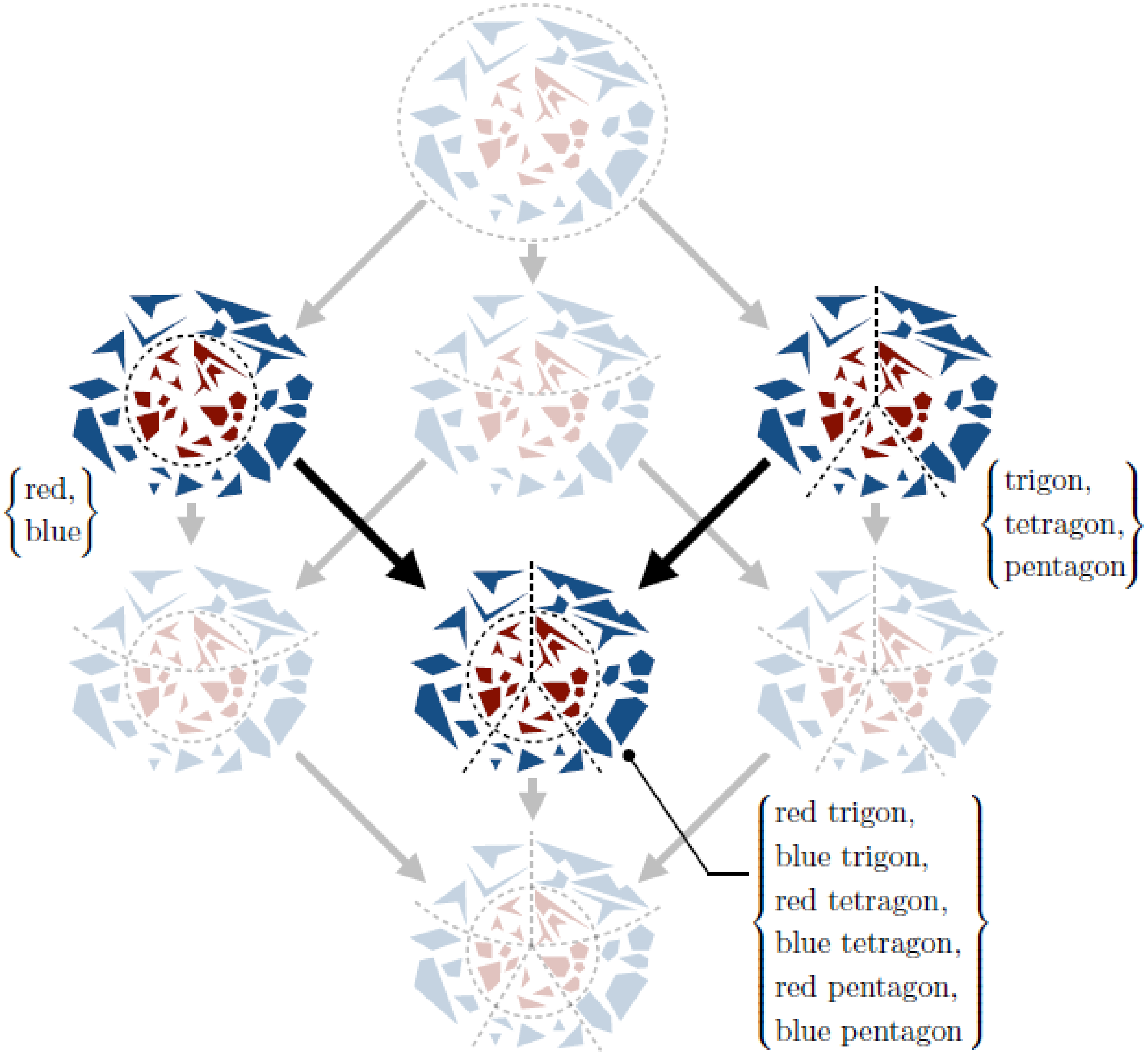
Partition vertebrates into five clusters

Concepts:

Cluster A: mammals
Cluster B: reptiles
Cluster C: birds
Cluster D: fish
Cluster E: amphibians

Rule:

A statistical pattern on abstracted concepts (clusters)



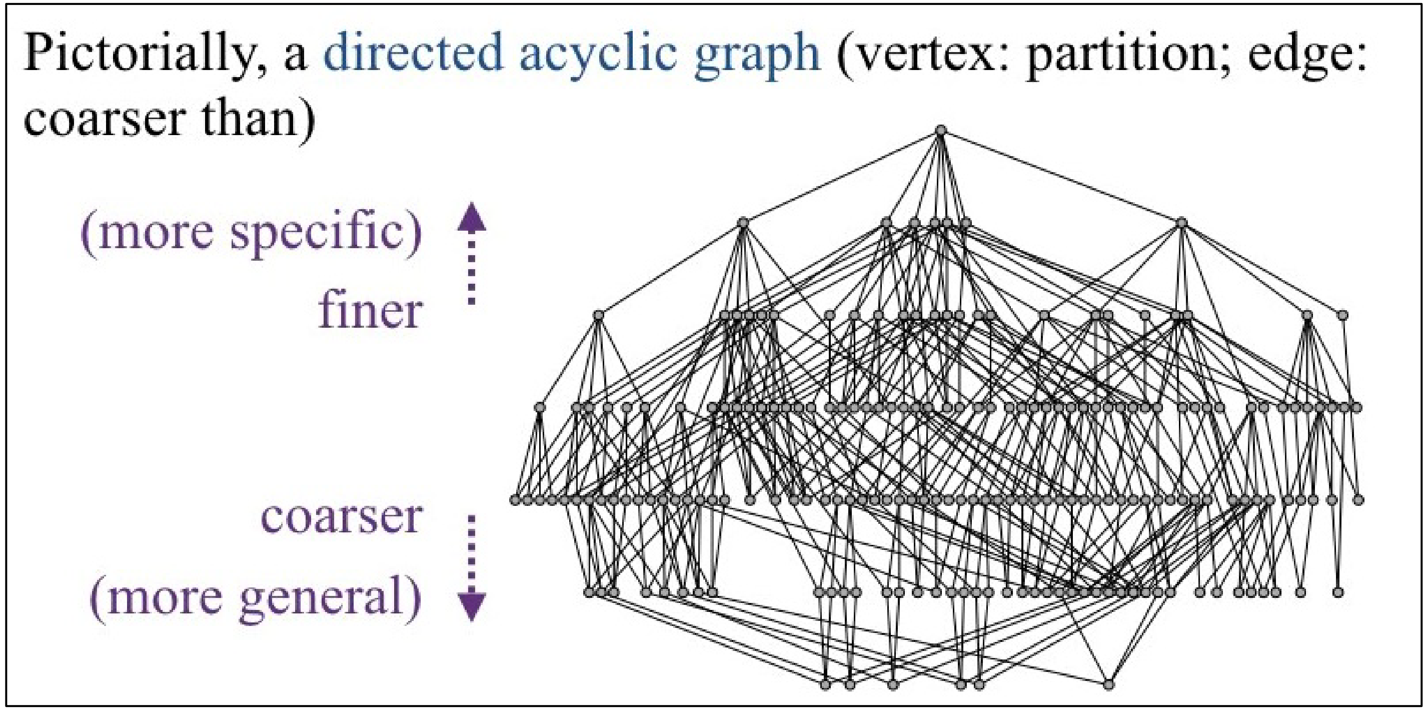
{red, blue}

{trigon, tetragon, pentagon}

{red trigon, blue trigon, red tetragon, blue tetragon, red pentagon, blue pentagon}

Abstraction universe as partition lattice

- A set X can have multiple partitions (Bell number $B_{|X|}$)
- Let $\mathfrak{B}_X^*$ denote the family of all partitions of a set X, so $|\mathfrak{B}_X^*| = B_{|X|}$
- Compare partitions of a set by a partial order on $\mathfrak{B}_X^*$
  - Partial order yields a *partition lattice*, a hierarchical representation of a family of partitions



Pictorially, a directed acyclic graph (vertex: partition; edge: coarser than)

Abstraction universe as partition lattice

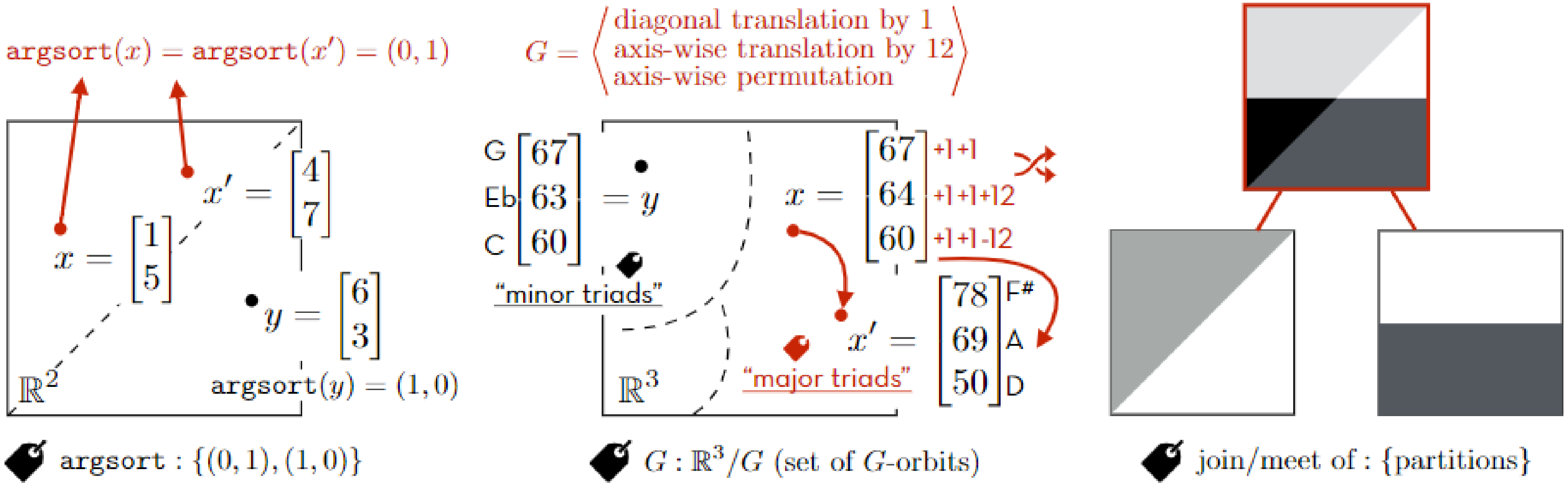
- Even for a finite set X of relatively small size, the complete abstraction universe $\mathfrak{B}_X^*$ can be quite large and complicated to visualize (Bell number grows very quickly, to say nothing of edges)
- However, not all arbitrary partitions are of interest

What part of $\mathfrak{B}_X^*$ should we focus on?

Abstraction universe as partition lattice

- Even for a finite set X of relatively small size, the complete abstraction universe $\mathfrak{B}_X^*$ can be quite large and complicated to visualize (Bell number grows very quickly, to say nothing of edges)
- However, not all arbitrary partitions are of interest

What part of $\mathfrak{B}_X^*$ should we focus on?



A. FEATURE-INDUCED

B. SYMMETRY-INDUCED

C. STRUCTURE—INDUCED

Symmetry-induced abstraction

- Consider the symmetric group $(S_X, \circ)$ defined over a set X, whose group elements are all the bijections from X to X and whose group operation is (function) composition
- A bijection from X to X is also called a *transformation* of X, so the symmetric group S_X comprises all transformations of X, and is also called the transformation group of X, denoted $\mathrm{F}(X)$

- Given a set X and a subgroup $H \leq \mathrm{F}(X)$, we define an H-action on X by $h \cdot x = h(x)$ for any $h \in H, x \in X$ and the orbit of $x \in X$ under H as the set $Hx = \{h(x) | h \in H\}$
- Each orbit is an equivalence class, so the quotient $X/H = X/\sim$ is a partition of X
- We say this abstraction respects H-symmetry or H-invariance

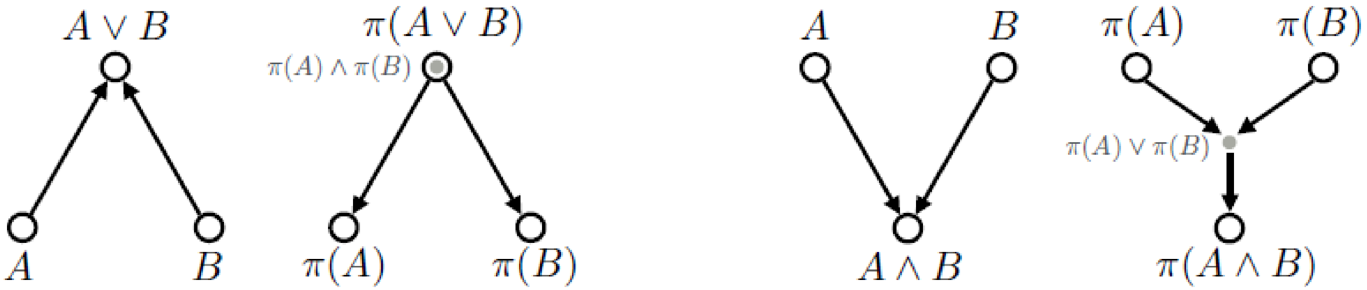
$$\boxed{\text{a subgroup of } \mathrm{F}(X) \xrightarrow{\textit{group action}} \text{orbits} \xrightarrow{\textit{equiv. rel.}} \text{a partition} \xrightarrow{\textit{is}} \text{an abstraction of } X}$$

Duality: From subgroup lattice to abstraction (semi)universe

Definition The *abstraction generating function* is the mapping $\pi\colon \mathcal{H}^*_{\mathrm{F}(X)} \to \mathfrak{B}^*_X$, where $\mathcal{H}^*_{\mathrm{F}(X)}$ is the collection of all subgroups of $\mathrm{F}(X)$, $\mathfrak{B}^*_X$ is the family of all partitions of X, and for any $H \in \mathcal{H}^*_{\mathrm{F}(X)}$, $\pi(H) = X/H$.

Theorem (Duality) Let $\left(\mathcal{H}^*_{\mathrm{F}(X)}, \leq\right)$ be the subgroup lattice for $\mathrm{F}(X)$ and π the abstraction generating function. Then $\left(\pi\left(\mathcal{H}^*_{\mathrm{F}(X)}\right), \preccurlyeq\right)$ is an abstraction meet-semiuniverse for X. That is:

1. partial-order reversal: if $A \leq B$, then $\pi(A) \succcurlyeq \pi(B)$
2. strong duality: $\pi(A \vee B) = \pi(A) \wedge \pi(B)$
3. weak duality: $\pi(A \wedge B) \succcurlyeq \pi(A) \vee \pi(B)$

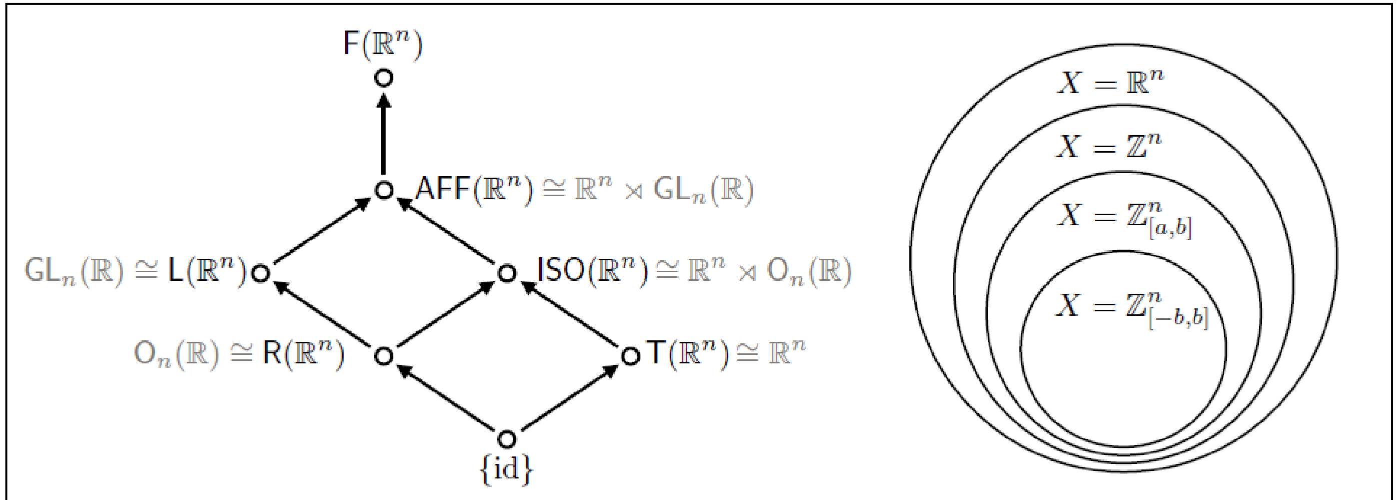


(a) From join to meet.

(b) From meet to join.

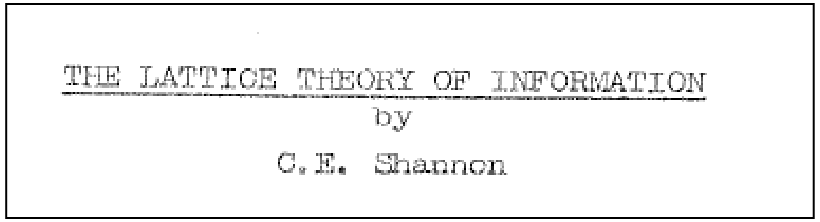
Duality: From subgroup lattice to abstraction (semi)universe

- If one has already computed abstractions $\pi(A)$ and $\pi(B)$, then instead of computing $\pi(A \vee B)$ from $A \vee B$, one can compute the meet $\pi(A) \wedge \pi(B)$, which is generally computationally less expensive than computing $A \vee B$ and identifying all orbits in $\pi(A \vee B)$
- The computer algebra system GAP provides efficient algorithmic methods to construct the subgroup lattice for a given group, and even maintains data libraries for special groups and their subgroup lattices



(a) Subgroups under consideration. (b) Spaces under consideration.

[H. Yu, I. Mineyev, and L. R. Varshney, "Orbit Computation for Atomically Generated Subgroups of Isometries of Z^n," *SIAM Journal on Applied Algebra and Geometry*, vol. 5, no. 3, pp. 479–505, Sept. 2021.]

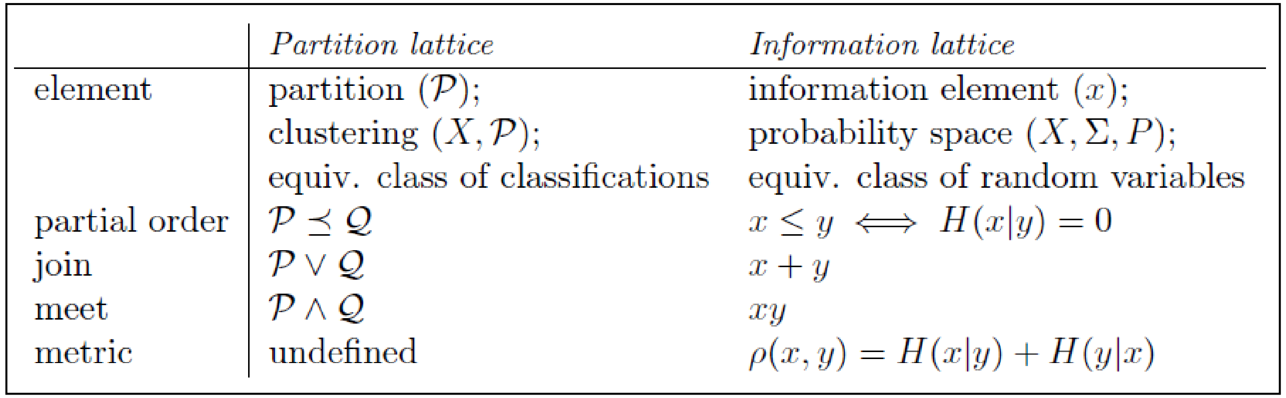

THE LATTICE THEORY OF INFORMATION
by
C. E. Shannon

- An *information element* is an equivalence class of random variables w.r.t. inducing the same σ-algebra

- An *information lattice* is a lattice of information elements, where partial order defined by $x \leq y \Leftrightarrow H(x|y) = 0$ where H is the Shannon entropy. The join of two information elements the *total information*; the meet of two information elements is the *common information*
- Our abstraction-generation framework generalizes Shannon's information lattice, without needing to introduce information-theoretic functionals like entropy

- More importantly gives generating chain to bring learning into picture

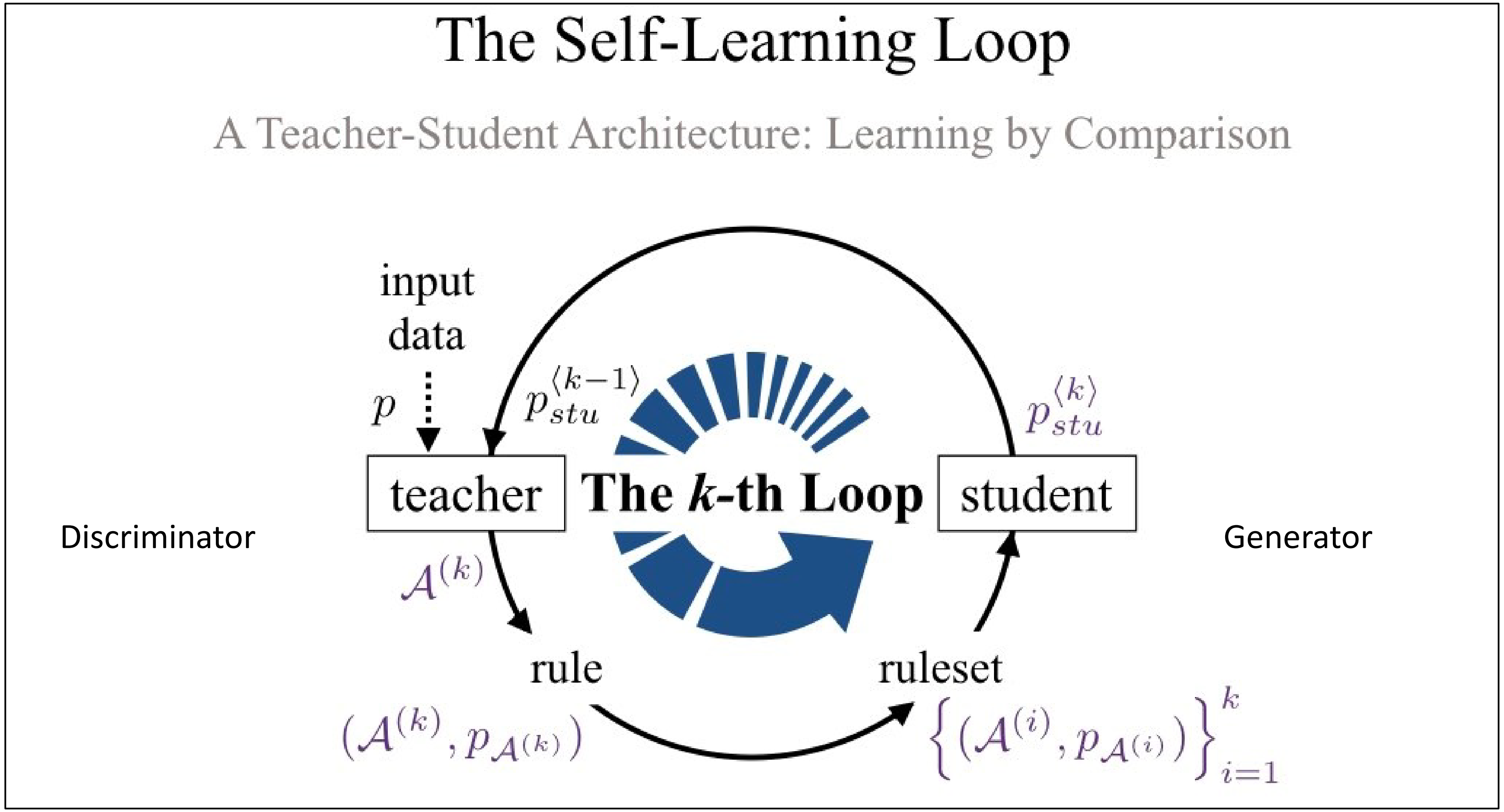
Separation of clustering from statistics: partition lattice can be thought as an information lattice without probability measure



	Partition lattice	*Information lattice*
element	partition ($\mathcal{P}$); clustering ($X, \mathcal{P}$); equiv. class of classifications	information element (x); probability space (X, Σ, P); equiv. class of random variables
partial order	$\mathcal{P} \preceq \mathcal{Q}$	$x \leq y \iff H(x\|y) = 0$
join	$\mathcal{P} \vee \mathcal{Q}$	$x + y$
meet	$\mathcal{P} \wedge \mathcal{Q}$	xy
metric	undefined	$\rho(x, y) = H(x\|y) + H(y\|x)$

Information-theory inspired algorithm for rule learning

Learning is achieved by statistical inference on a partition lattice

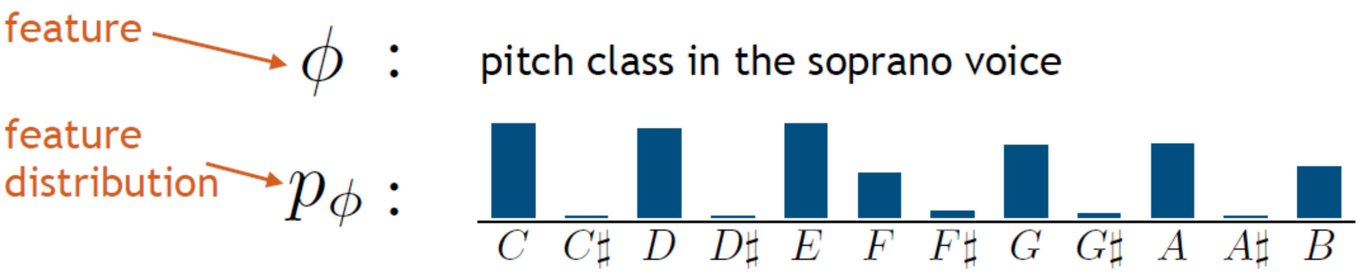


Discriminator

Generator

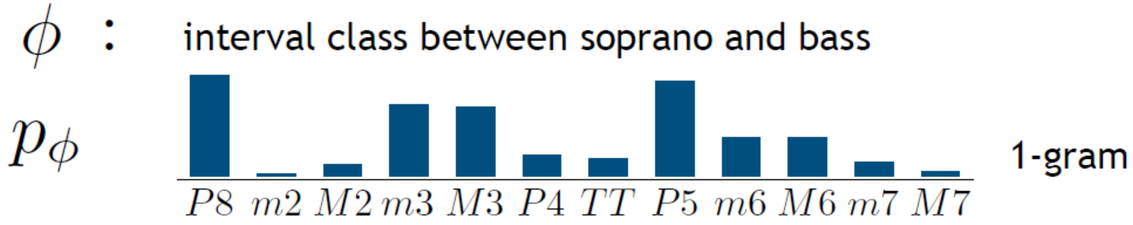
[H. Yu and L. R. Varshney, "Towards Deep Interpretability (MUS-ROVER II): Learning Hierarchical Representations of Tonal Music," in *Proc. 5th International Conference on Learning Representations (ICLR)*, April 2017.]

Simple human-interpretable rules and hierarchical concept learning are human-interpretable



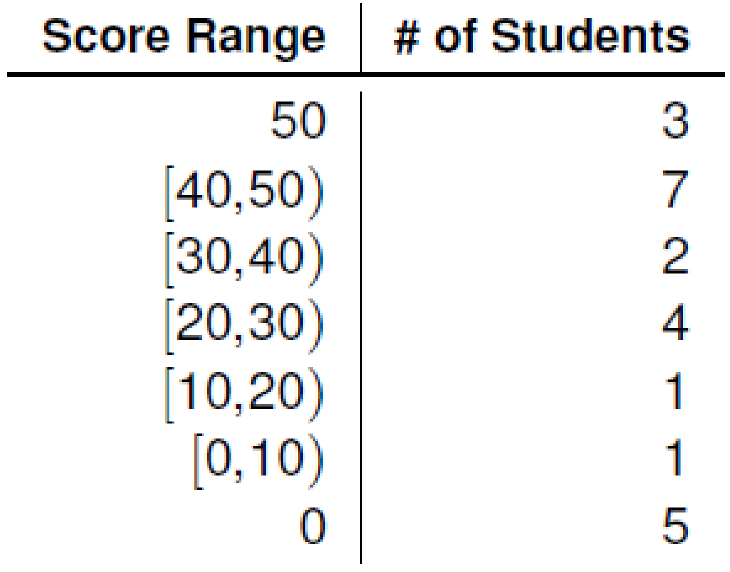
This rule can be interpreted or translated to:

"The soprano voice is built on a diatonic scale."



"Individual perfect octaves (P8s) are favored as most consonant."

Studies in music class

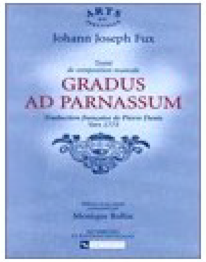


Score Range	# of Students
50	3
[40,50)	7
[30,40)	2
[20,30)	4
[10,20)	1
[0,10)	1
0	5

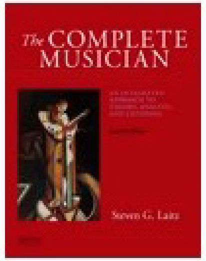
Table 1: Students' final scores.

[H. Yu, H. Taube, J. A. Evans, and L. R. Varshney, "Human Evaluation of Interpretability: The Case of AI-Generated Music Knowledge," in *ACM CHI 2020 Workshop on Artificial Intelligence for HCI: A Modern Approach*, April 2020.]

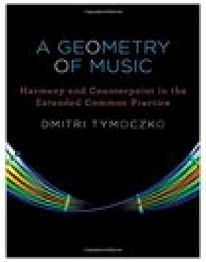
ILL recovers much known music theory



- voice leading
- counter point

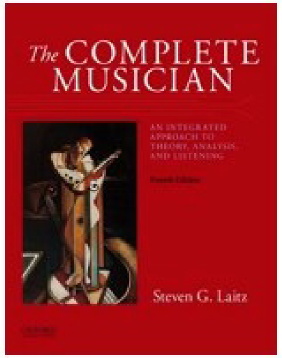


- scale, consonance & dissonance
- voice spacing, crossing, overlap
- chord quality, inversion, progression

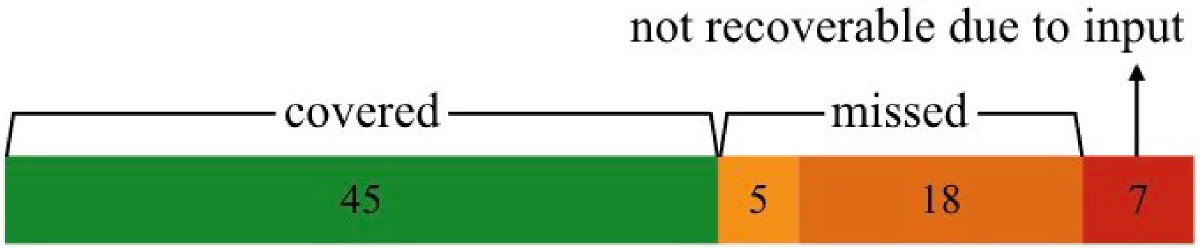


- music transformations: OPTIC

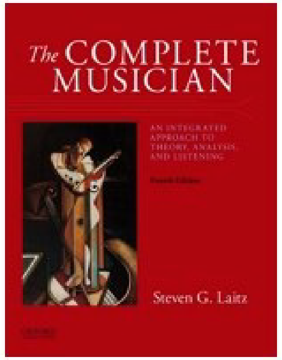
ILL recovers much known music theory



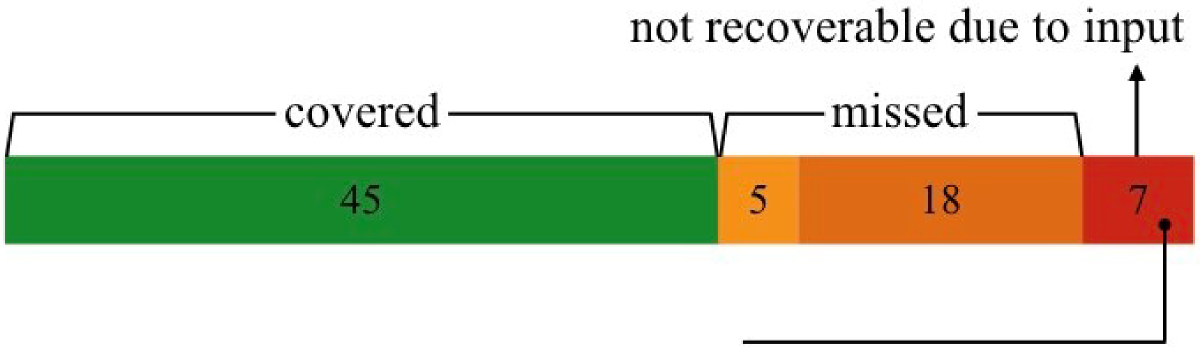
MUS 101, 102, 201 (75 topics in total):



ILL recovers much known music theory



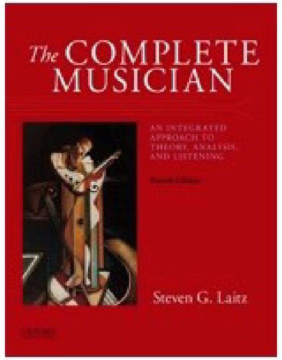
MUS 101, 102, 201 (75 topics in total):



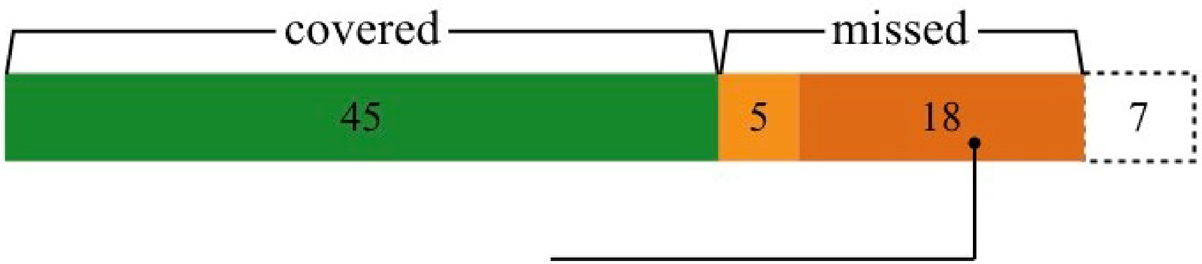
requires info other than MIDI pitches and durations:

- music accents: requires beats, dynamics, etc.
- enharmonic re-spellings: German 6th, fully dim, etc.

ILL recovers much known music theory



MUS 101, 102, 201 (68 recoverable topics in total):



captured but not explicitly presented:

- phrase models, EPMs, sentence structure, etc.
- music forms: binary, ternary, rondo, sonata, etc.

Suggests an extension of the n-gram models to temporal abstractions:

transitions of abstractions → *abstractions of transitions*

ILL discovers new music theory

Interesting probabilistic pattern

Unresolved tritone (TT):

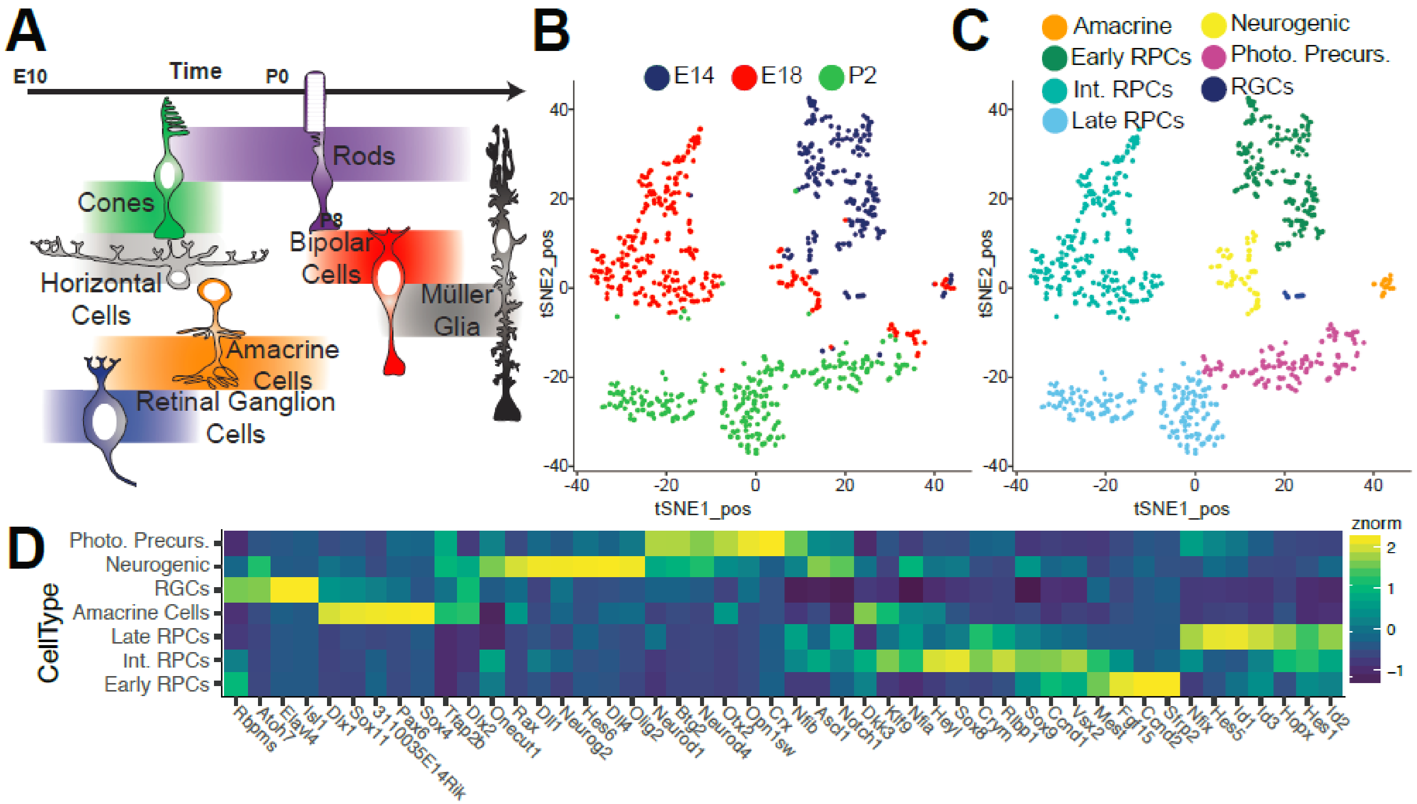
$$\text{TT} \longrightarrow \text{m7}$$

“harmonic” escape tone or changing tone?

Interesting abstraction

Rule Trace	
1	$\texttt{order} \circ w_{\{1,2,3,4\}}$
2	$\texttt{order} \circ \texttt{diff} \circ \texttt{sort} \circ w_{\{1,2,4\}}$
3	$\texttt{order} \circ \texttt{diff} \circ \texttt{mod}_{12} \circ w_{\{1,2,3\}}$
4	$\texttt{order} \circ \underline{\texttt{diff} \circ \texttt{diff}} \circ w_{\{1,2,3,4\}}$
5	$\texttt{order} \circ \texttt{sort} \circ \texttt{mod}_{12} \circ w_{\{2,3,4\}}$
6	$\texttt{order} \circ \texttt{sort} \circ \texttt{mod}_{12} \circ w_{\{1,3,4\}}$
7	$\texttt{order} \circ \texttt{sort} \circ \texttt{mod}_{12} \circ w_{\{1,2,3,4\}}$
8	$\texttt{mod}_{12} \circ w_{\{1\}}$
9	$\texttt{mod}_{12} \circ \texttt{diff} \circ w_{\{2,3\}}$
10	$\texttt{mod}_{12} \circ \texttt{diff} \circ w_{\{3,4\}}$

Learning laws of neurogenesis

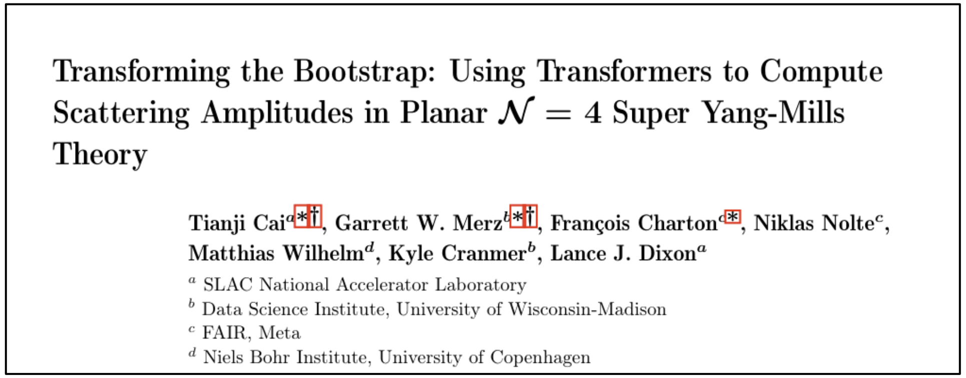


[B. Clark, et al., "Single-Cell RNA-Seq Analysis of Retinal Development Identifies NFI Factors as Regulating Mitotic Exit and Late-Born Cell Specification," *Neuron*, June 2019.]

Single-cell RNA sequence data analysis for understanding the rules that govern pattern formation in neurodevelopment

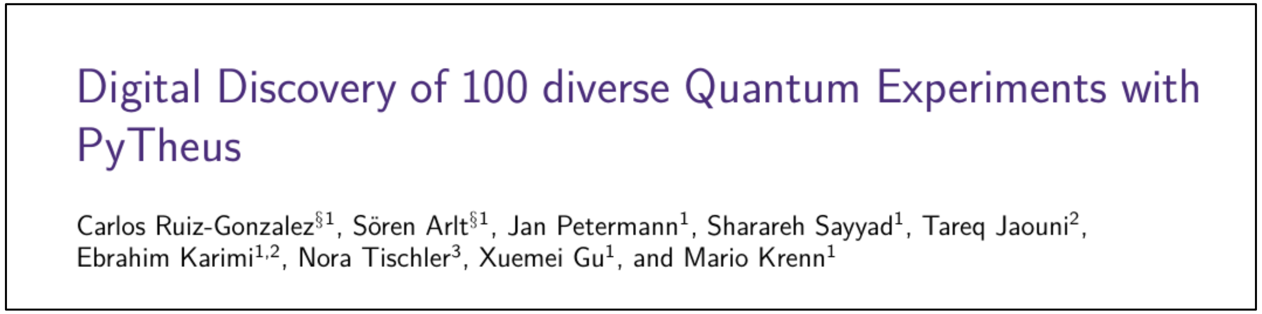
[H. Yu, L. R. Varshney, and G. Stein-O'Brien, "Towards Learning Human-Interpretable Laws of Neurogenesis from Single-Cell RNA-Seq Data via Information Lattices," at *Learning Meaningful Representations of Life Workshop at NeurIPS*, Dec. 2019.]

Toward further applications of ILL for knowledge discovery



Transforming the Bootstrap: Using Transformers to Compute Scattering Amplitudes in Planar $\mathcal{N} = 4$ Super Yang-Mills Theory

Tianji Cai[a*†], Garrett W. Merz[b*†], François Charton[c*], Niklas Nolte[c], Matthias Wilhelm[d], Kyle Cranmer[b], Lance J. Dixon[a]
[a] SLAC National Accelerator Laboratory
[b] Data Science Institute, University of Wisconsin-Madison
[c] FAIR, Meta
[d] Niels Bohr Institute, University of Copenhagen



Digital Discovery of 100 diverse Quantum Experiments with PyTheus

Carlos Ruiz-Gonzalez[§1], Sören Arlt[§1], Jan Petermann[1], Sharareh Sayyad[1], Tareq Jaouni[2], Ebrahim Karimi[1,2], Nora Tischler[3], Xuemei Gu[1], and Mario Krenn[1]

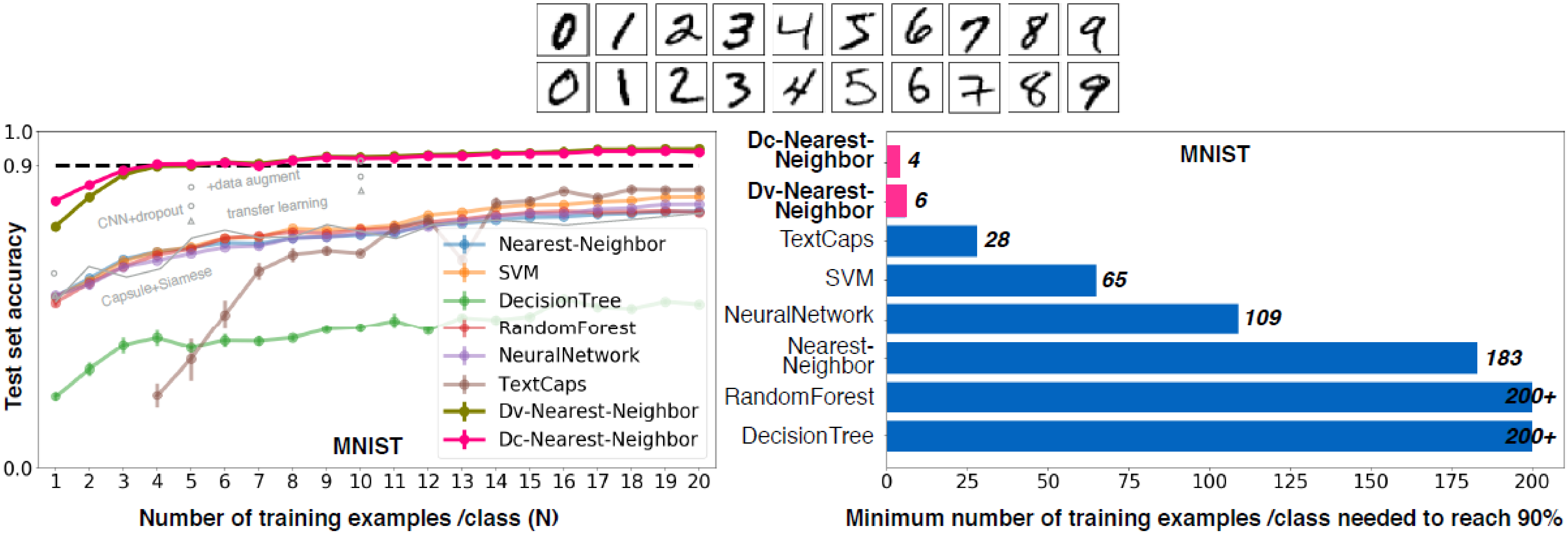
Article

Scaling deep learning for materials discovery

https://doi.org/10.1038/s41586-023-06735-9

Received: 8 May 2023

Amil Merchant[1,3]✉, Simon Batzner[1,3], Samuel S. Schoenholz[1,3], Muratahan Aykol[1], Gowoon Cheon[2] & Ekin Dogus Cubuk[1,3]✉



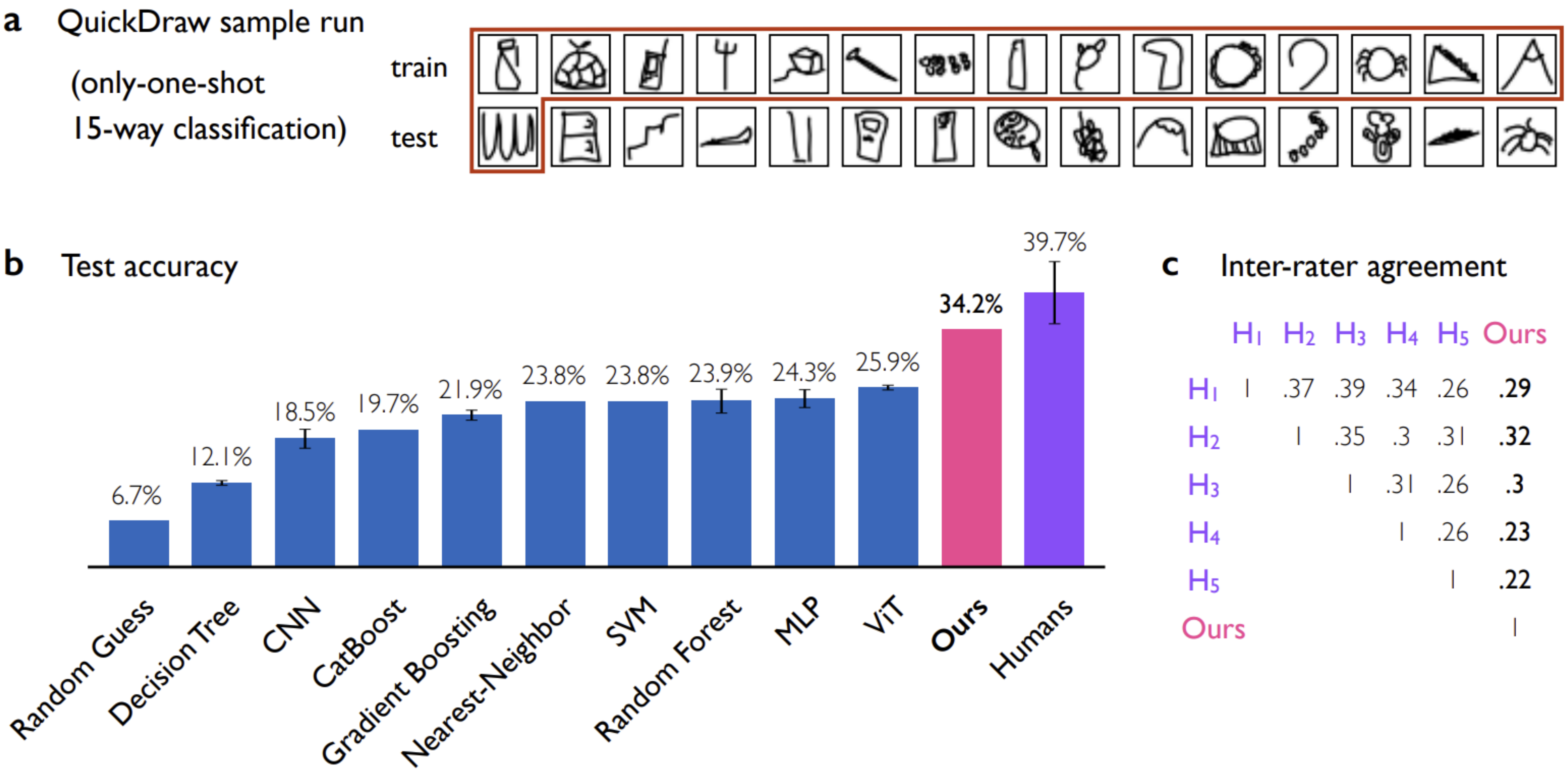
[H. Yu, I. Mineyev, L. R. Varshney, and J. A. Evans, “Learning from One and Only One Shot,” *npj Artificial Intelligence*, vol. 1, ar. 13, July 2025.]

A.

B.

Background set size (for pre-training)	BPL	Humans	**Ours**	RCN	Siamese ConvNet	Simple ConvNet	Prototypical Net	VHE
none			**6.75%**					
reduced	4.2%					23.2%	30.1%	
original	3.3%	4.5%		7.3%		13.5%	13.7%	18.7%
augmented					8%			

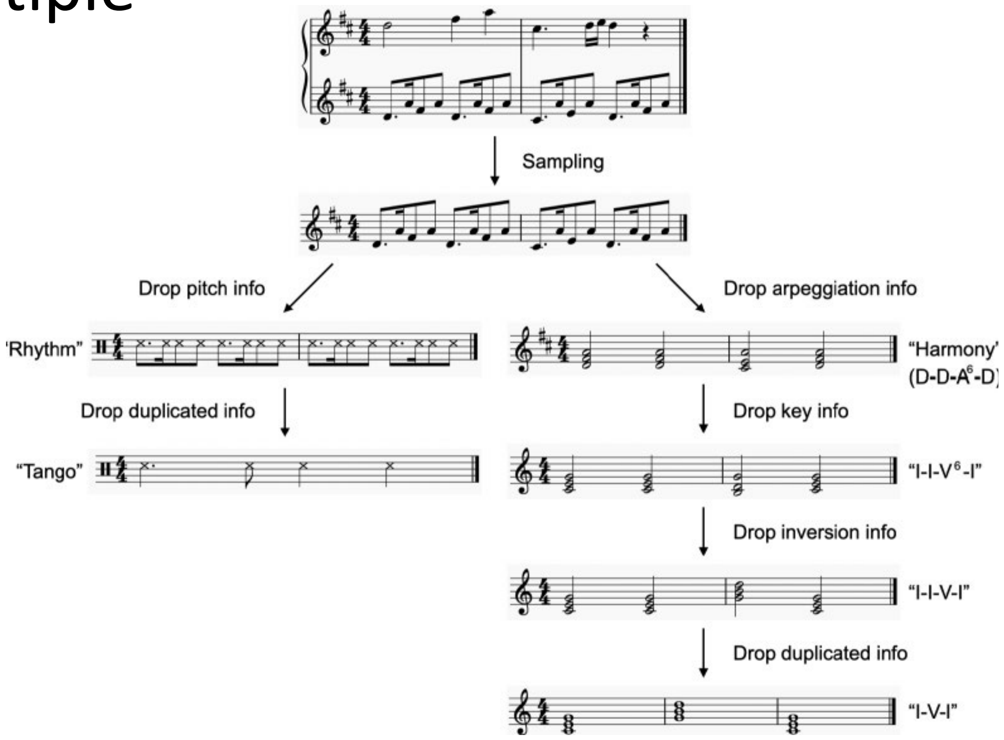
[H. Yu, I. Mineyev, L. R. Varshney, and J. A. Evans, “Learning from One and Only One Shot,” *npj Artificial Intelligence*, vol. 1, ar. 13, July 2025.]

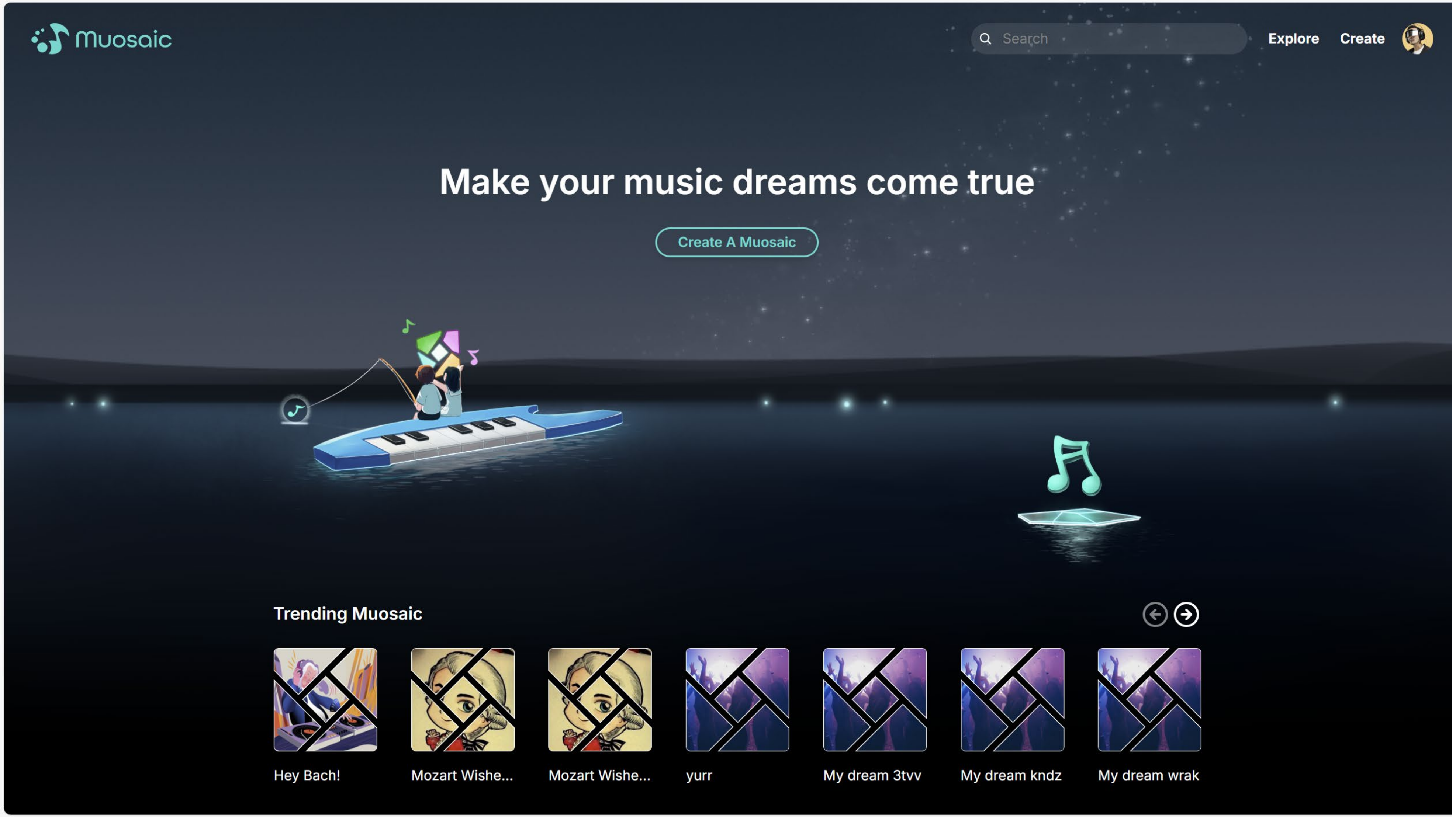


	H_1	H_2	H_3	H_4	H_5	Ours
H_1	1	.37	.39	.34	.26	**.29**
H_2		1	.35	.3	.31	**.32**
H_3			1	.31	.26	**.3**
H_4				1	.26	**.23**
H_5					1	**.22**
Ours						1

[H. Yu, I. Mineyev, L. R. Varshney, and J. A. Evans, “Learning from One and Only One Shot,” *npj Artificial Intelligence*, vol. 1, ar. 13, July 2025.]

Semantic compression with information lattice learning

Theorem information lattice learning representations, combined with group codes (e.g. permutation source codes) can achieve semantic rate-distortion limit for point-to-point, successive refinement, and multiple descriptions settings.

Stony Brook University



Make your music dreams come true

Lav R. Varshney
Stony Brook University
Kocree, Inc.
Brookhaven National Laboratory

lvarshney@kocree.net
lav.varshney@stonybrook.edu