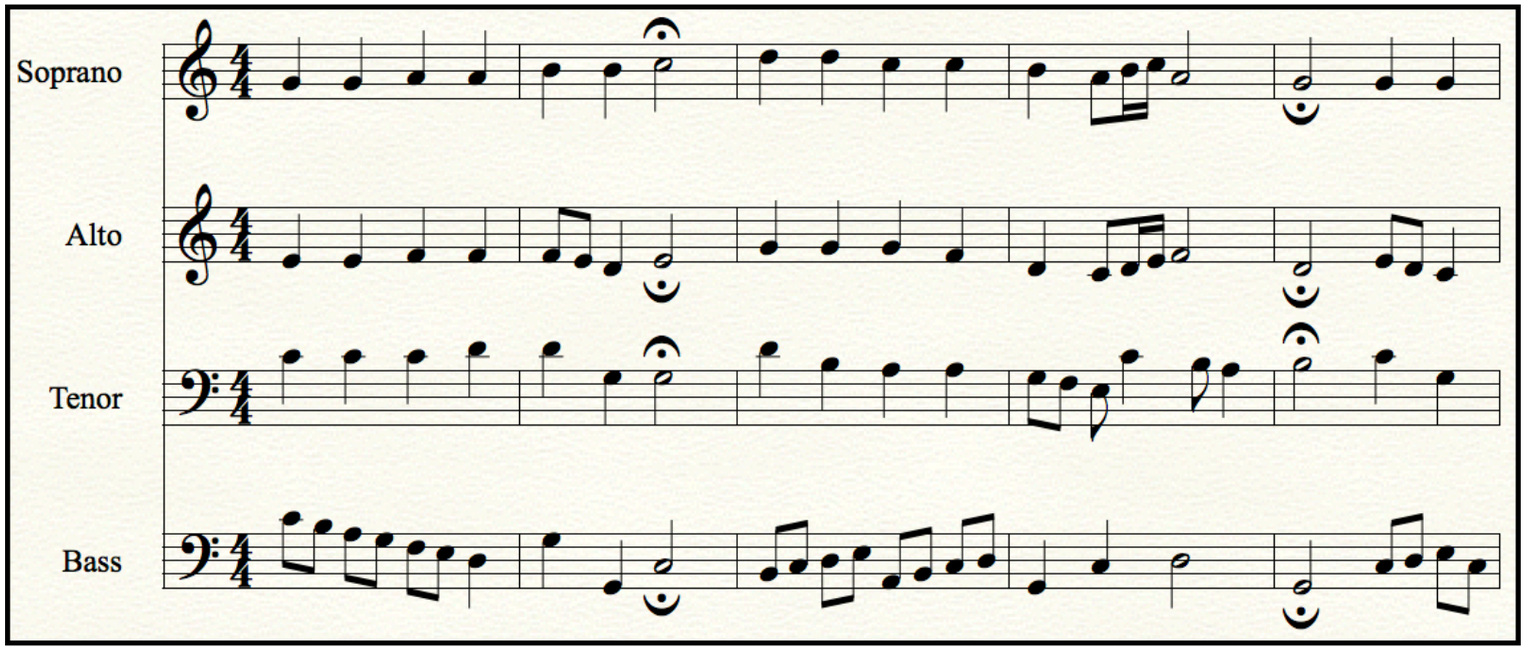
Information Lattice Learning (ILL)

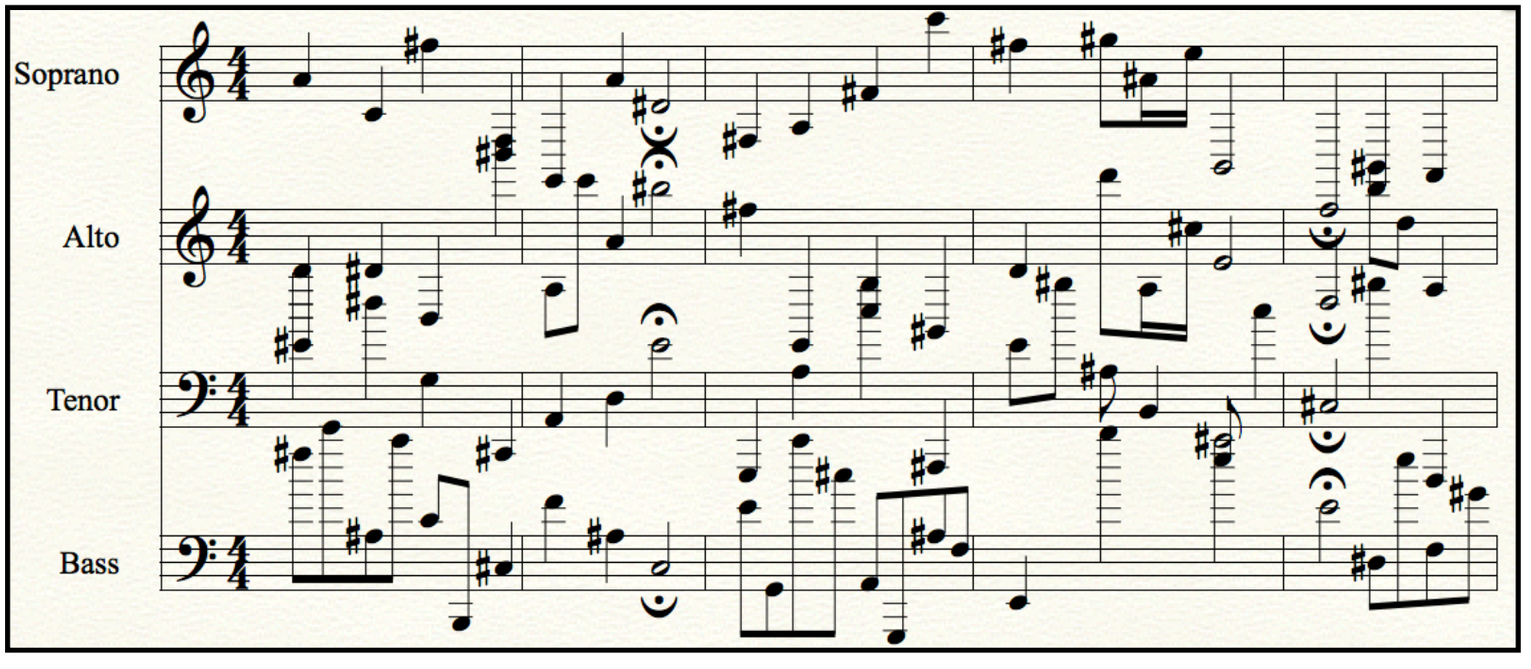
Haizi Yu

Music Quiz 1

?



?

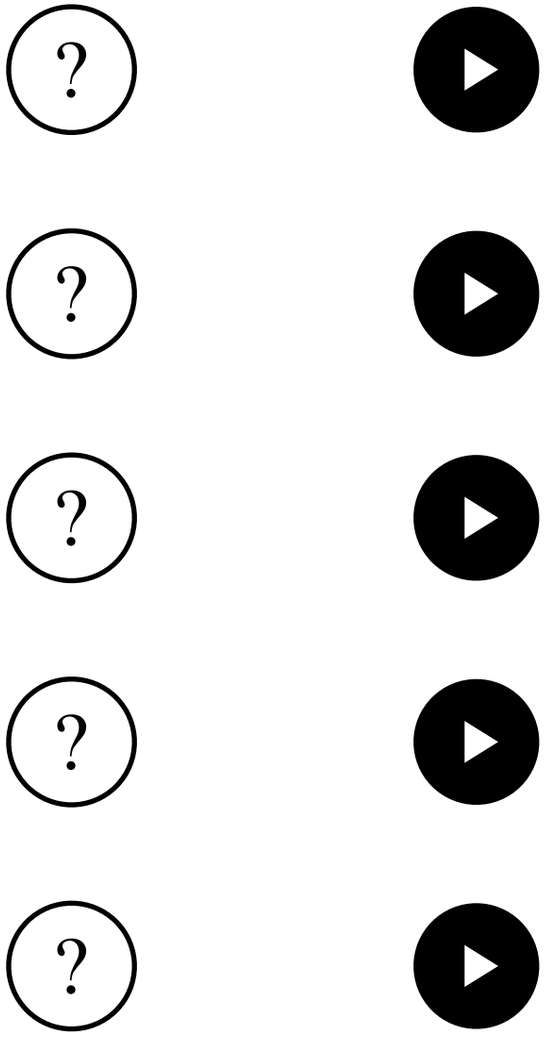


Music Quiz 2

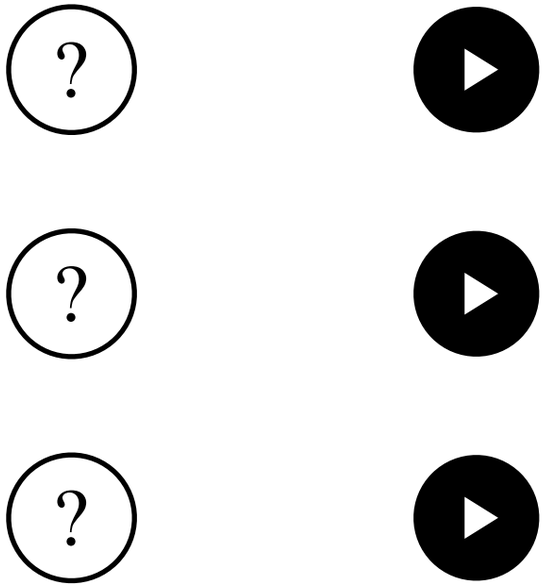
?

?

Music Quiz 3



Music Quiz 4



What makes X X?

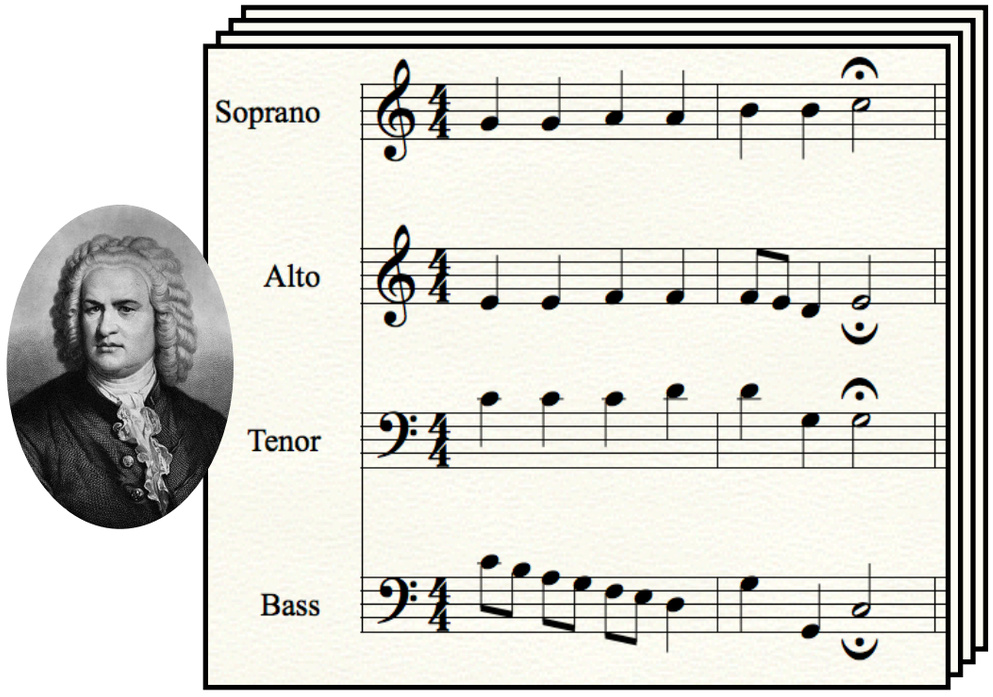
(What makes Bach’s chorales Bach’s chorales?)

How the above question differs from …?

A New Learning Problem

Automatic Concept Learning

Input



Output

Concepts

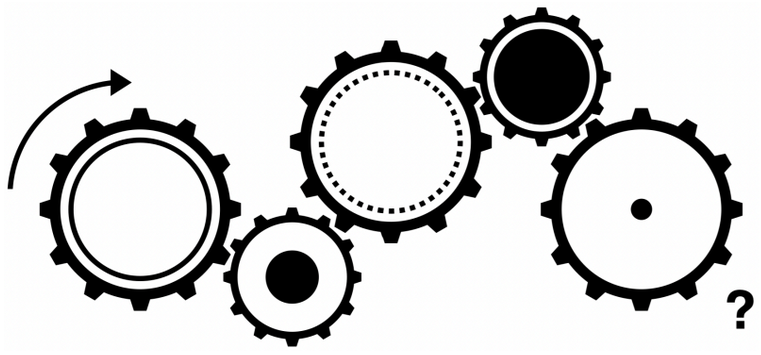
Rules

Laws

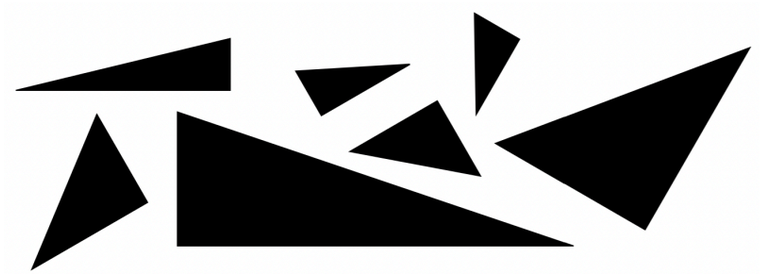
…

Automatic Concept Learning

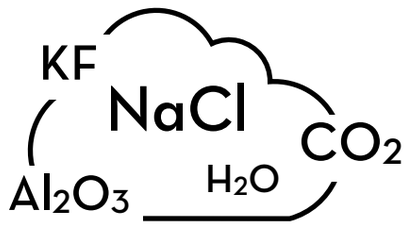
From phenomenology to theory



→ Parity Rule
(clockwise iff odd)



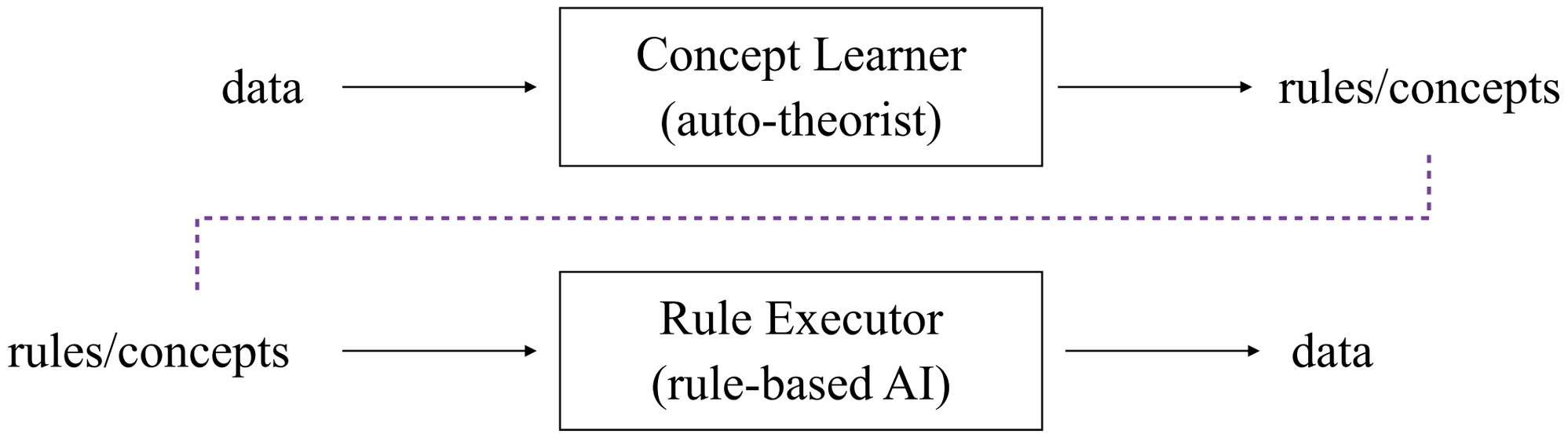
→ Pythagorean Theorem
($a^2 + b^2 = c^2$)



→ Chemical laws on compound formation and nomenclature

Music → Music Theory

Concept Learning in a Nutshell



Information Lattice Learning (ILL)

data ⟶ rules and concepts ⇢ data'

augment human intelligence ▲ augment human creativity ▲

How this differs from an auto-encoder?

A Learning Paradigm

- Representation: Information Lattice (IL)
- Algorithm: Information Lattice Learning (ILL)

Representation: IL

Abstraction, Concept, Rule

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

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

cells (or less formally, clusters)

concepts

A **concept** is a partition cell.

$X = \{\text{every vertebrate}\}$
$\mathcal{A} = \{\text{mammals, birds, fish, amphibians, reptiles}\}$

Abstraction, Concept, Rule

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

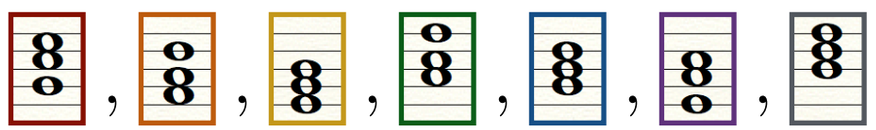
$$X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$$

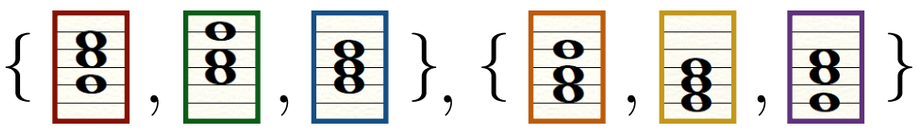
$$\mathcal{A} = \{\{x_1, x_6\}, \{x_3\}, \{x_2, x_4, x_5\}\}$$

cells (or less formally, clusters)

concepts

A **concept** is a partition cell.

$X = \{$  $\}$ (treble clef)

$\mathcal{A} = \{\{$  $\}\}$

major **minor** **diminished**

Abstraction, Concept, Rule

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

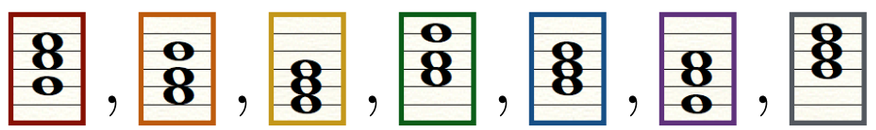
$$X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$$

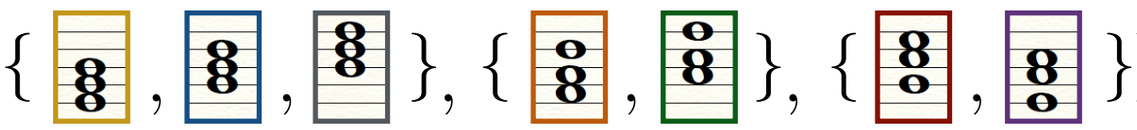
$$\mathcal{A} = \{\{x_1, x_6\}, \{x_3\}, \{x_2, x_4, x_5\}\}$$

cells (or less formally, clusters)

concepts

A **concept** is a partition cell.

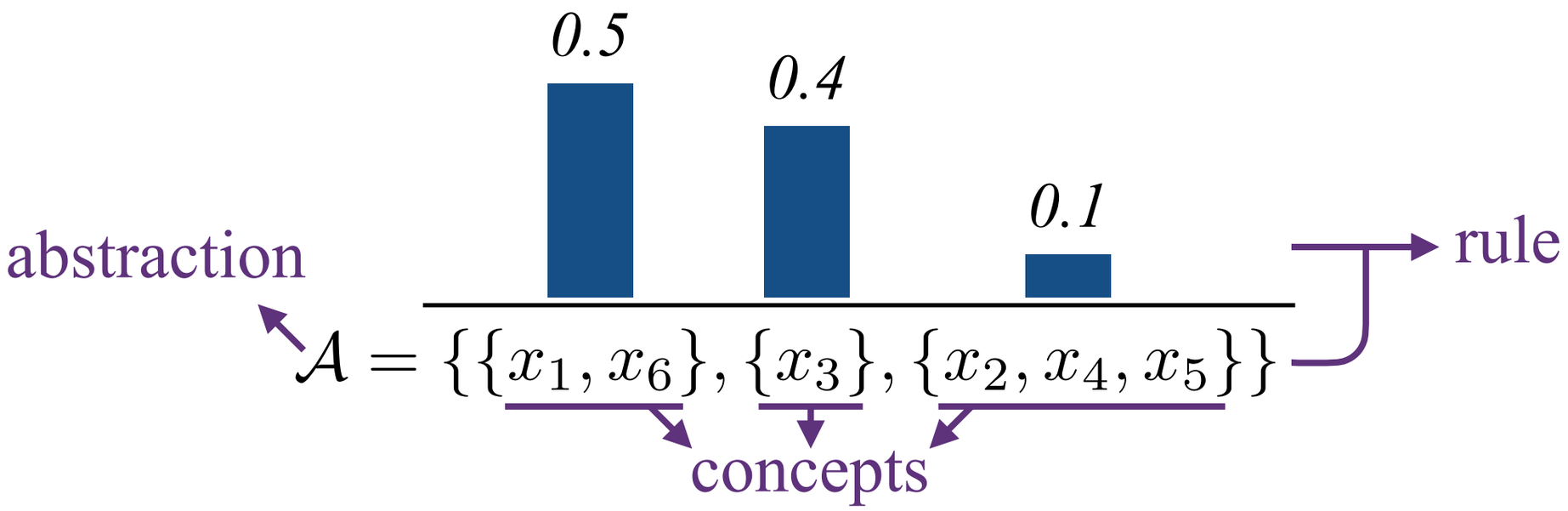
$X = \{$  $\}$ (treble clef)

$\mathcal{A}' = \{\{$  $\}\}$

root position **1st inversion** **2nd inversion**

Abstraction, Concept, Rule

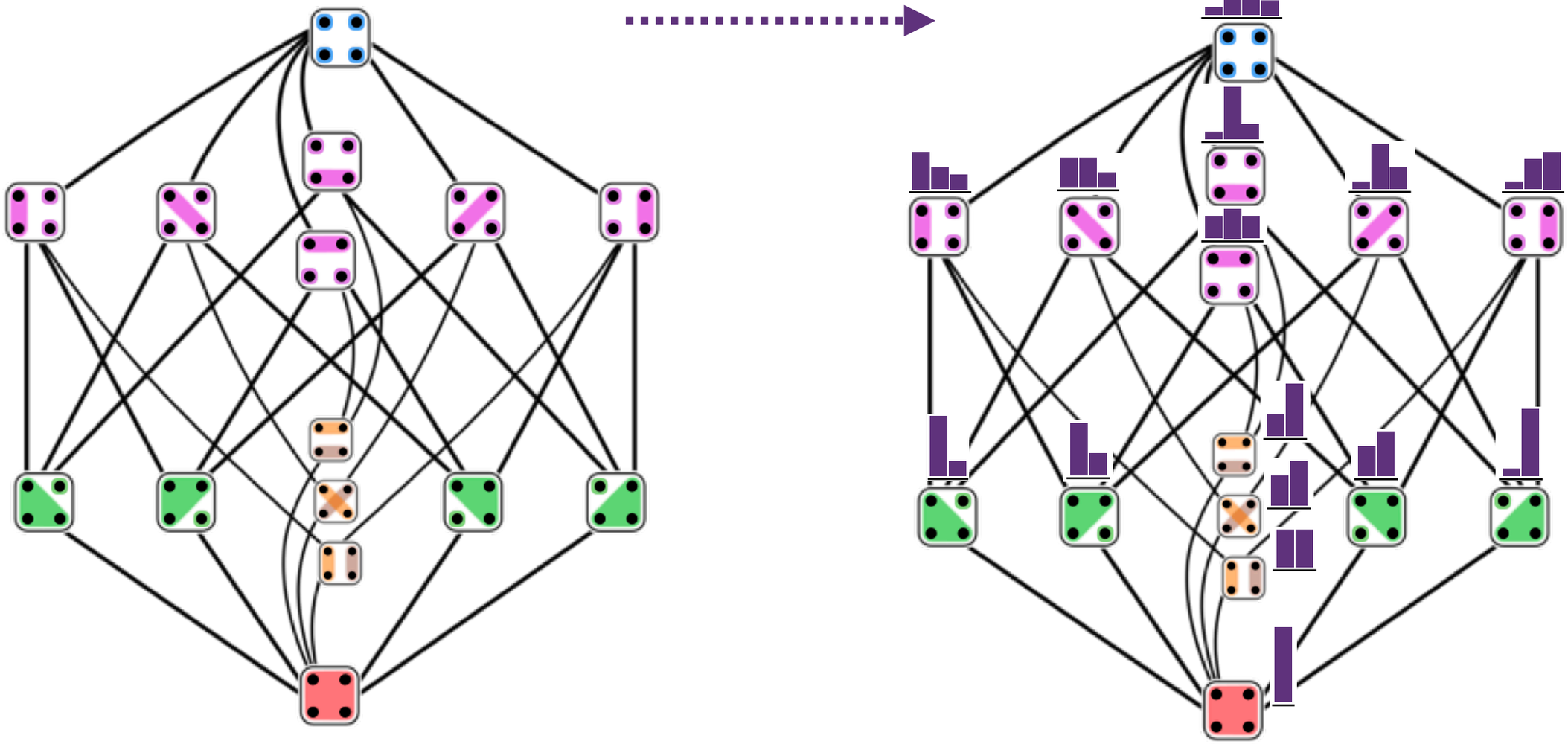
A **(probabilistic) rule** is a probability distribution projected onto a partition



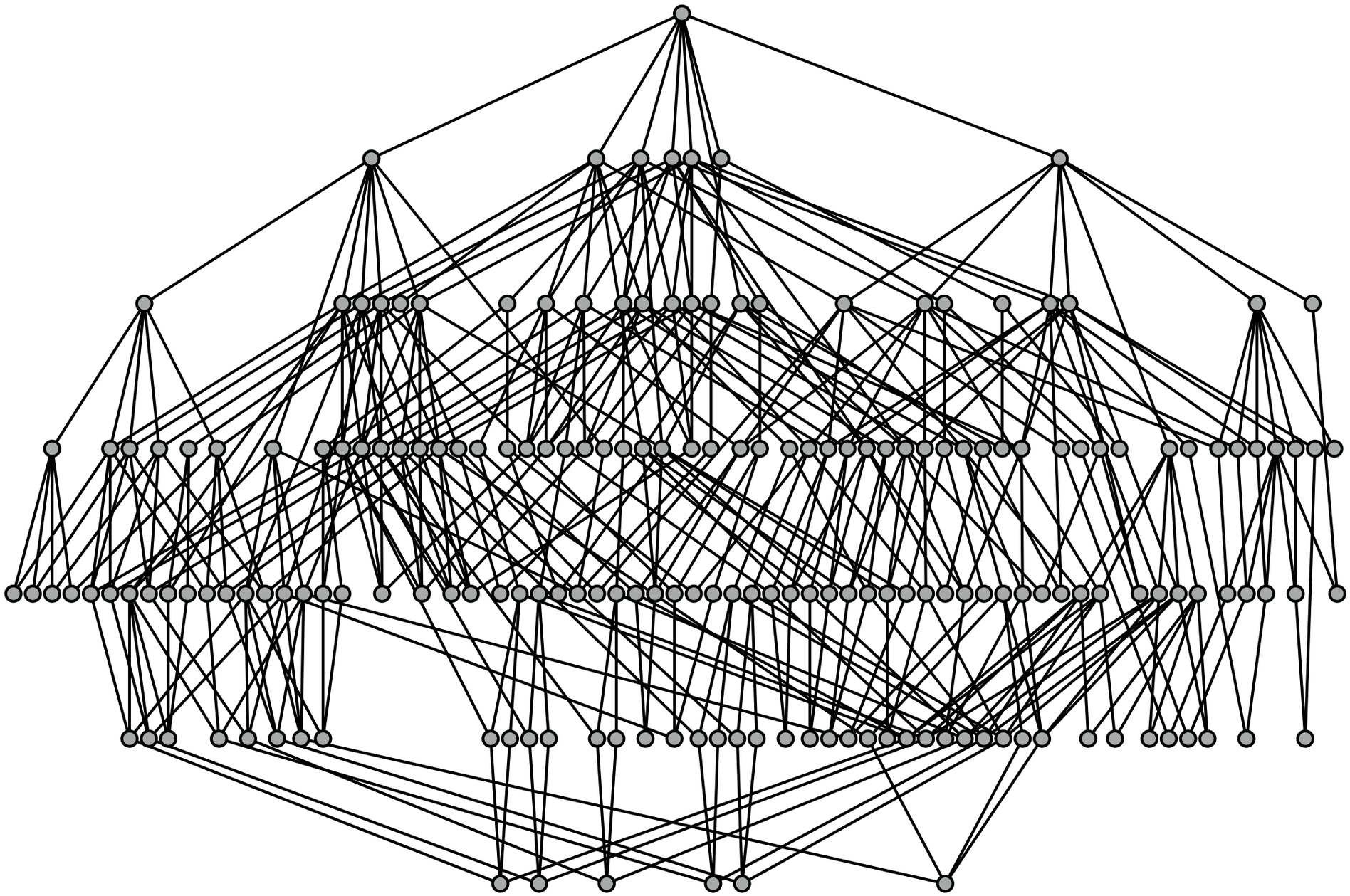
Partition Lattice

Information Lattice

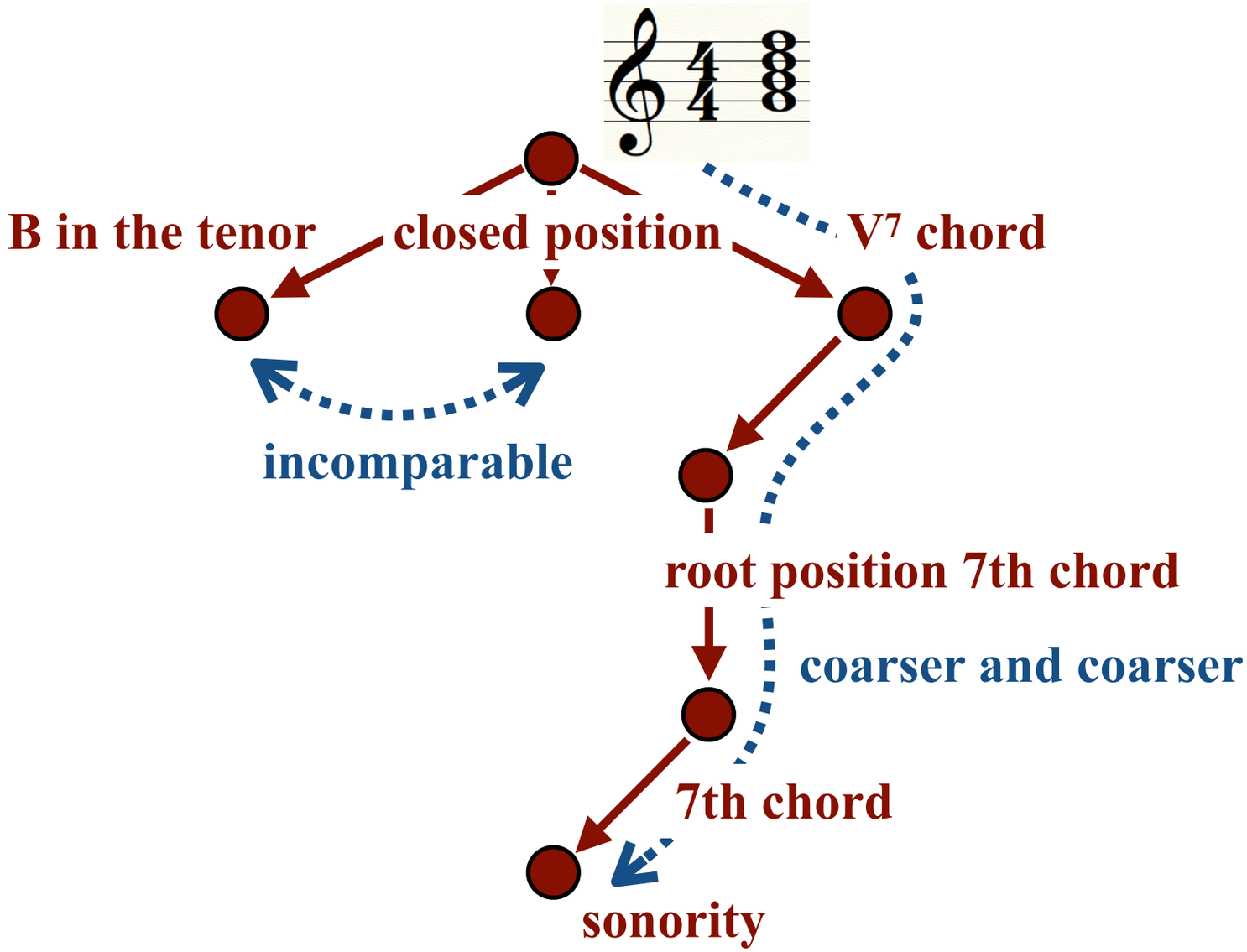
project probability



Bach's Information Lattice



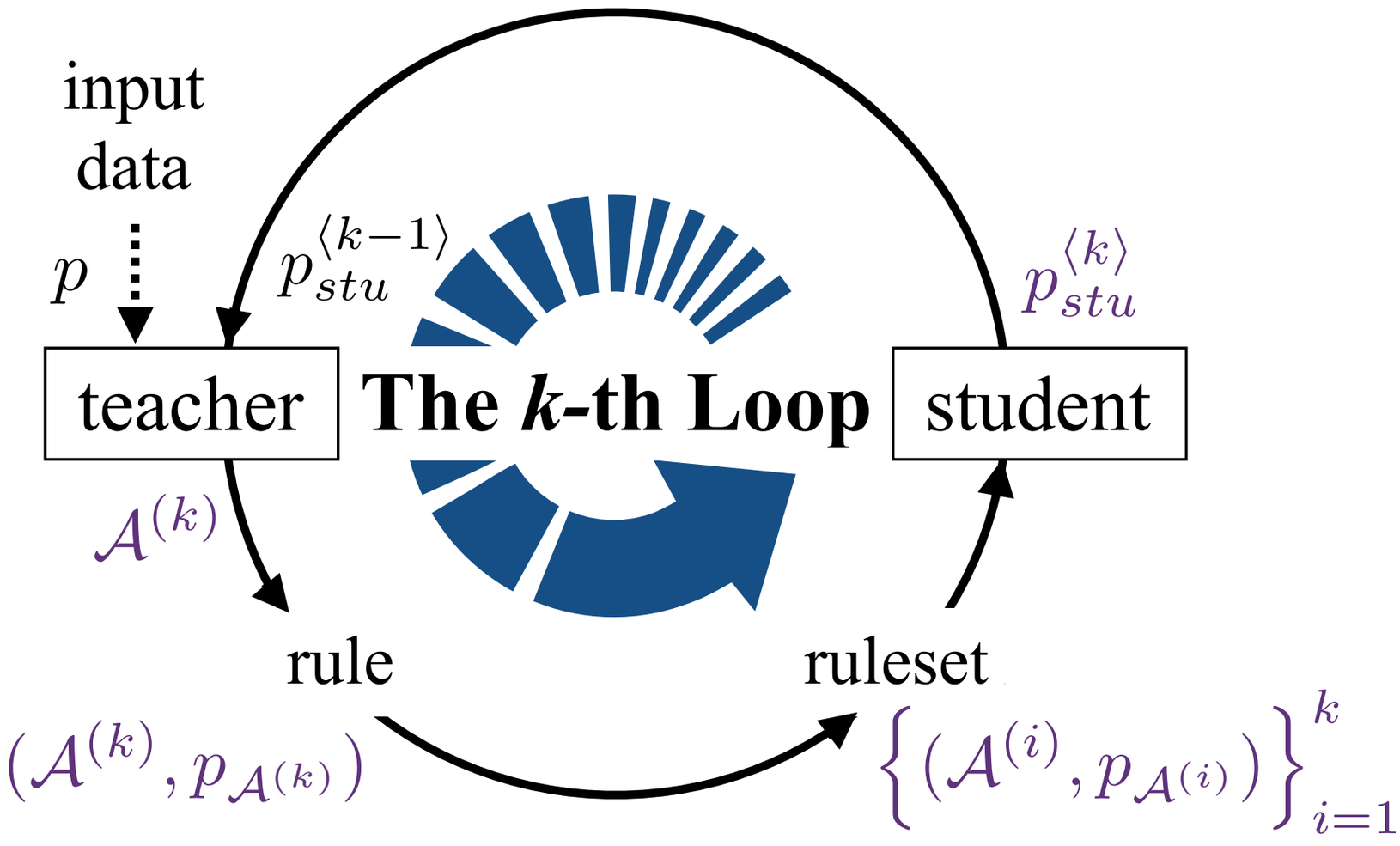
Bach's Concept Lattice



Algorithm: ILL

The Self-Learning Loop

A Teacher-Student Architecture: Learning by Comparison



Teacher: a Discriminative Model

The teacher solves an optimization problem:

$$\begin{aligned} \underset{\mathcal{A} \in \mathfrak{P}_X}{\text{maximize}} \quad & D_{KL}\left(p_{\mathcal{A},stu}^{\langle k-1 \rangle} \parallel p_{\mathcal{A}}\right) \\ \text{subject to} \quad & \mathcal{A} \notin \mathfrak{P}^{\langle k-1 \rangle} \\ & H(p_{\mathcal{A}}) \leq \delta_k \end{aligned}$$

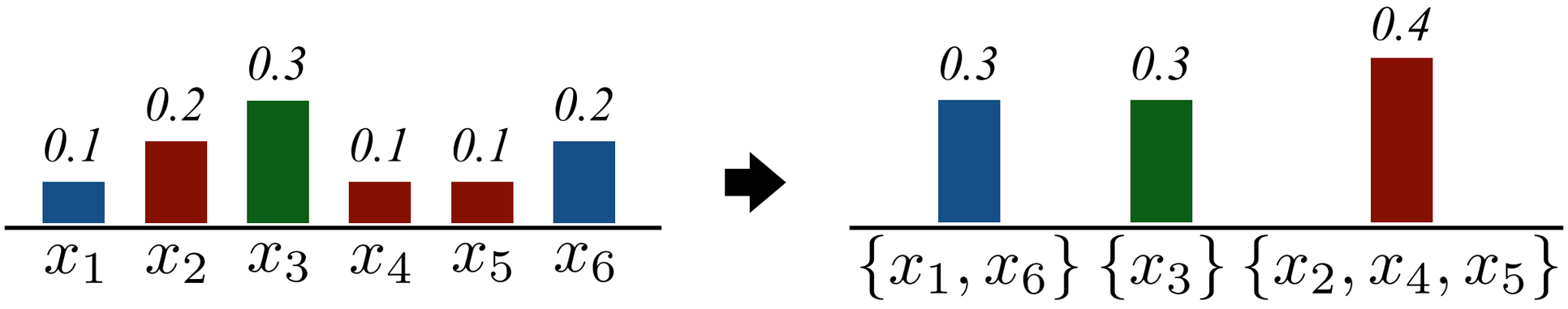
Want to find the abstraction which reveals the largest statistical difference between the student and the input data.

Teacher: a Discriminative Model

The teacher solves an optimization problem:

$$\begin{aligned} \underset{\mathcal{A}\in\mathfrak{P}_X}{\text{maximize}} \quad & D_{KL}\left(p_{\mathcal{A},stu}^{\langle k-1\rangle} \parallel p_{\mathcal{A}}\right) \\ \text{subject to} \quad & \mathcal{A} \notin \mathfrak{P}^{\langle k-1\rangle} \\ & H(p_{\mathcal{A}}) \leq \delta_k \end{aligned}$$

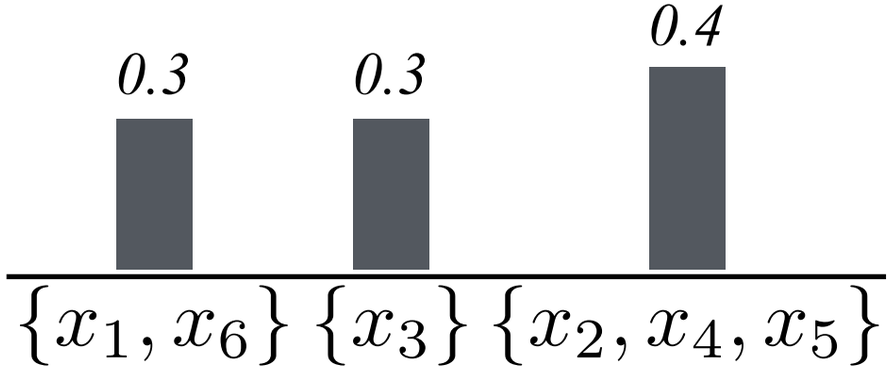
Projection: given p , computing $p_{\mathcal{A}}$ is easy:

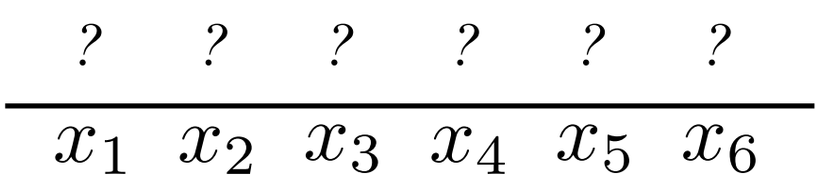


Student: a Generative Model

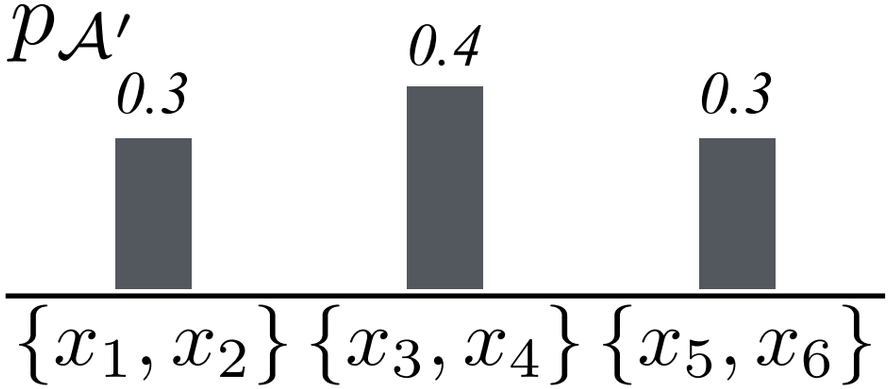
Apply probabilistic rules, which is known as the rule realization problem.

Given $p_{\mathcal{A}}$, compute p:





$p_{\mathcal{A}'}$



…

not necessarily unique

which one do we prefer?

Student: a Generative Model

The student solves a MaxEnt problem:

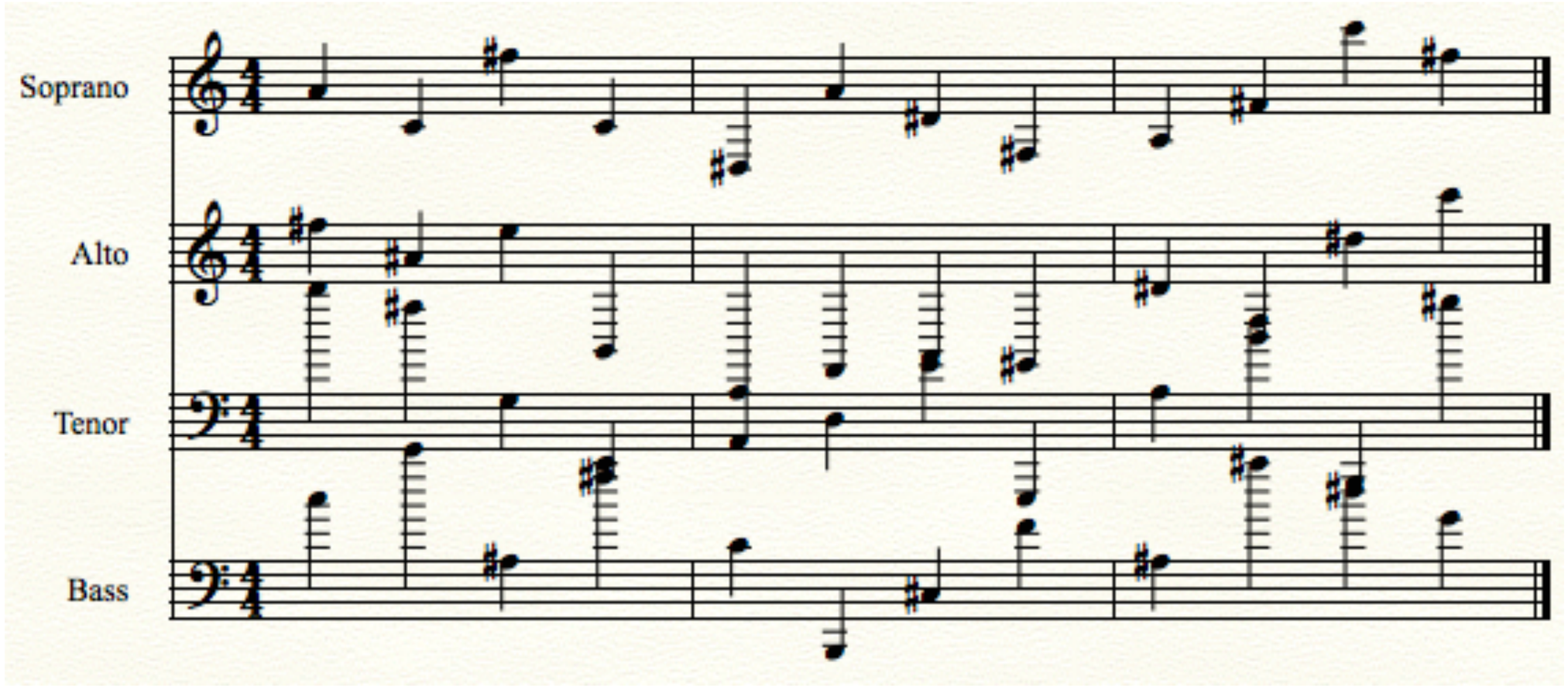
$$\begin{aligned} \underset{p_{stu}^{\langle k \rangle} \in \Delta^{|X|}}{\text{maximize}} \quad & S_q(p_{stu}^{\langle k \rangle}) := (q-1)^{-1}\left(1 - \|p_{stu}^{\langle k \rangle}\|_q^q\right) \\ \text{subject to} \quad & A^{(i)} p_{stu}^{\langle k \rangle} = p_{\mathcal{A}^{(i)}}, \quad i = 1, \ldots, k \end{aligned}$$

Want to find the most creative probabilistic model which satisfies all the rules (quality) while at the same time enables novelty

How MUS-ROVER Self-Evolves?

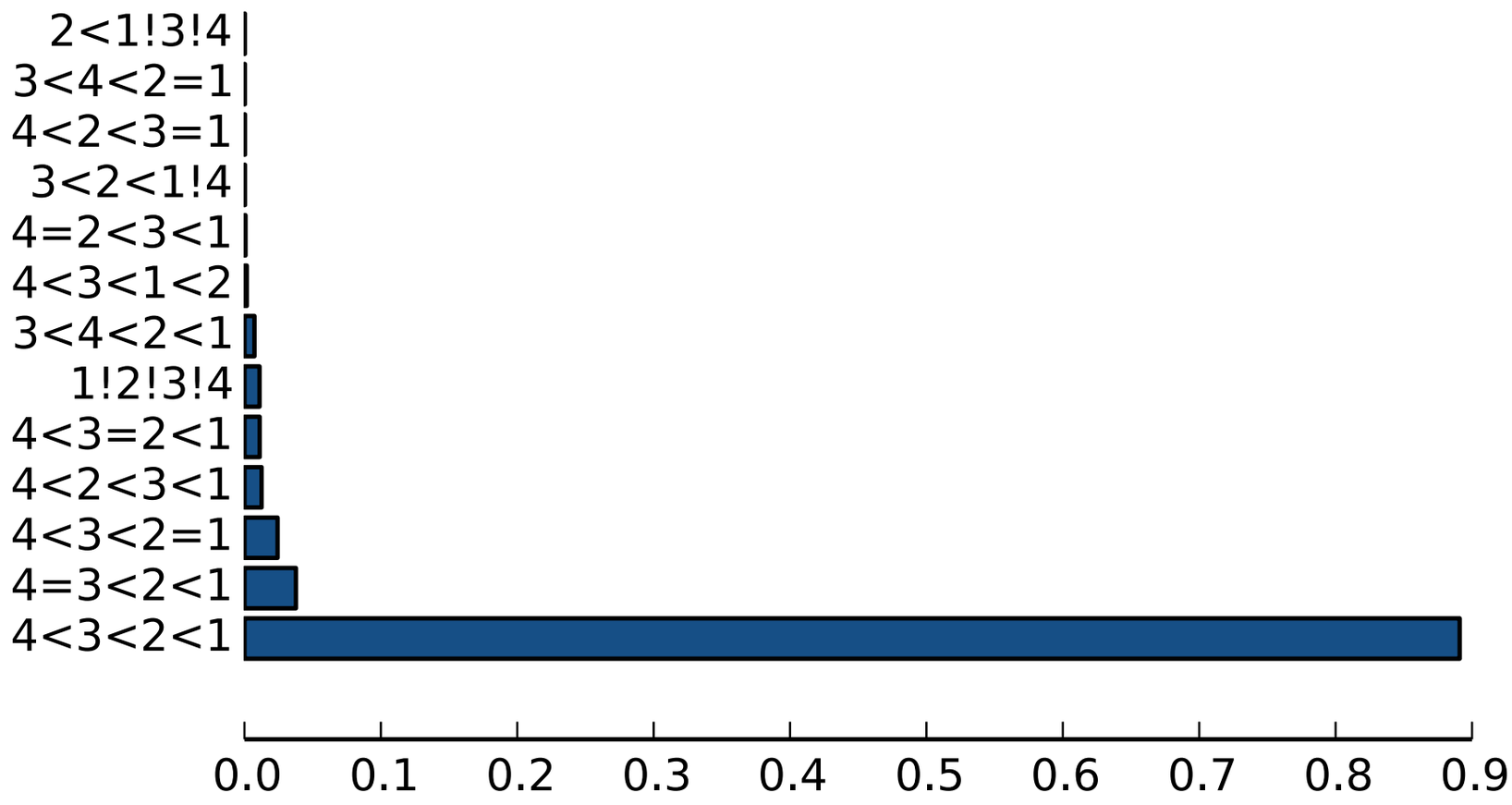
1-gram

Student 0



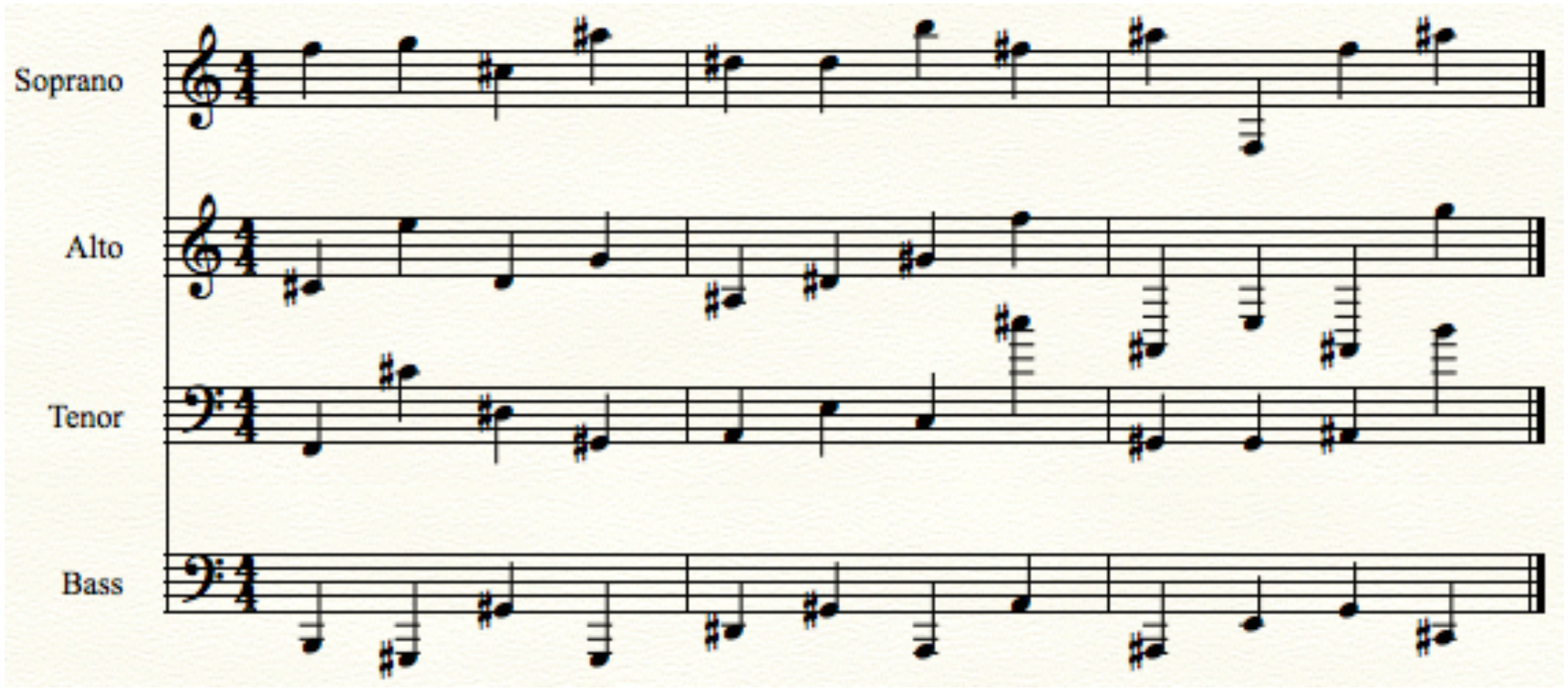
1-gram

Rule 1: `order` $\circ\, w_{1,2,3,4}$



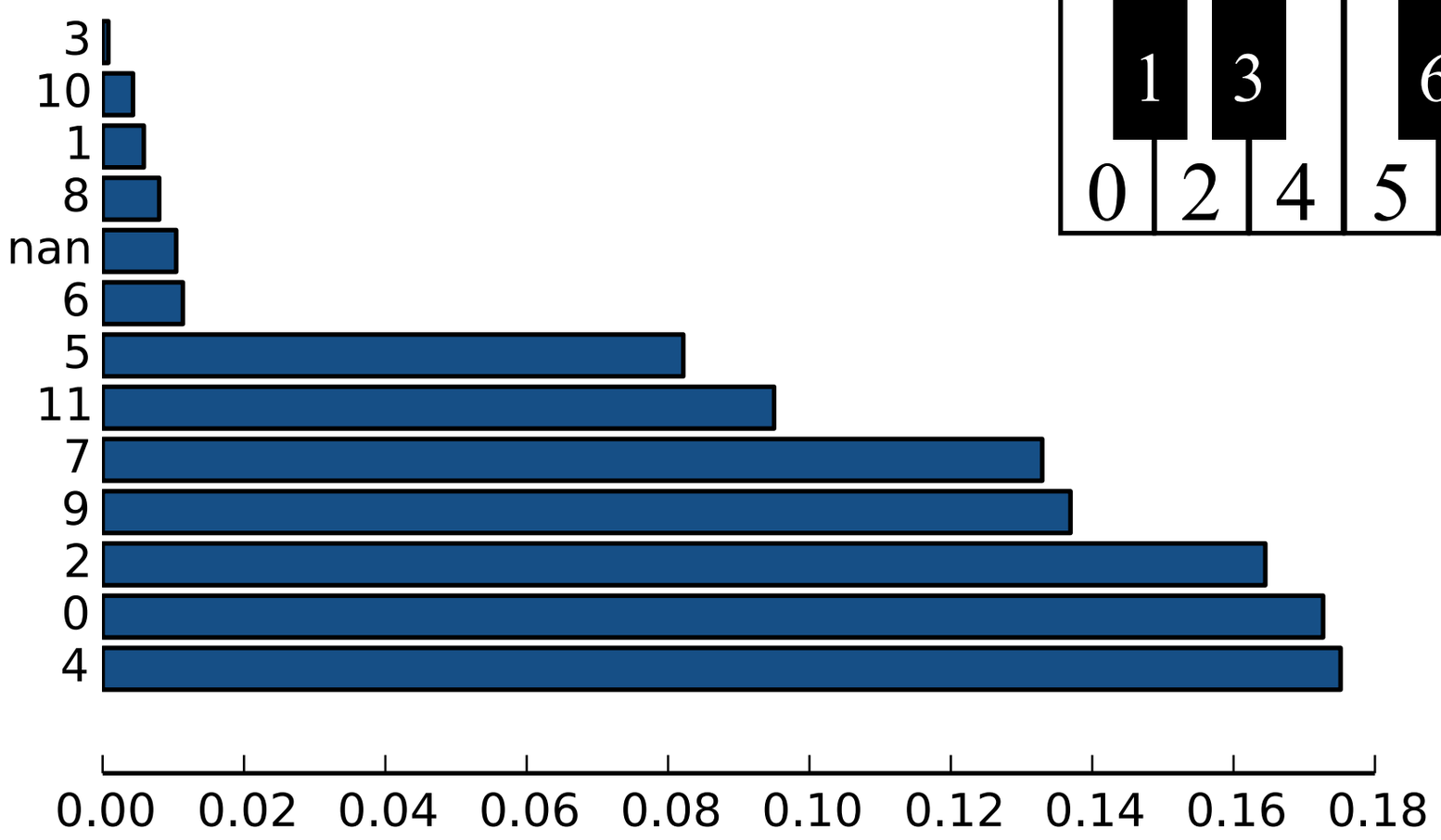
1-gram

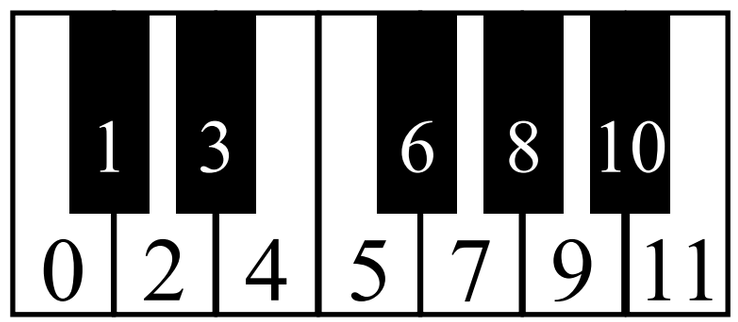
Student 1



1-gram

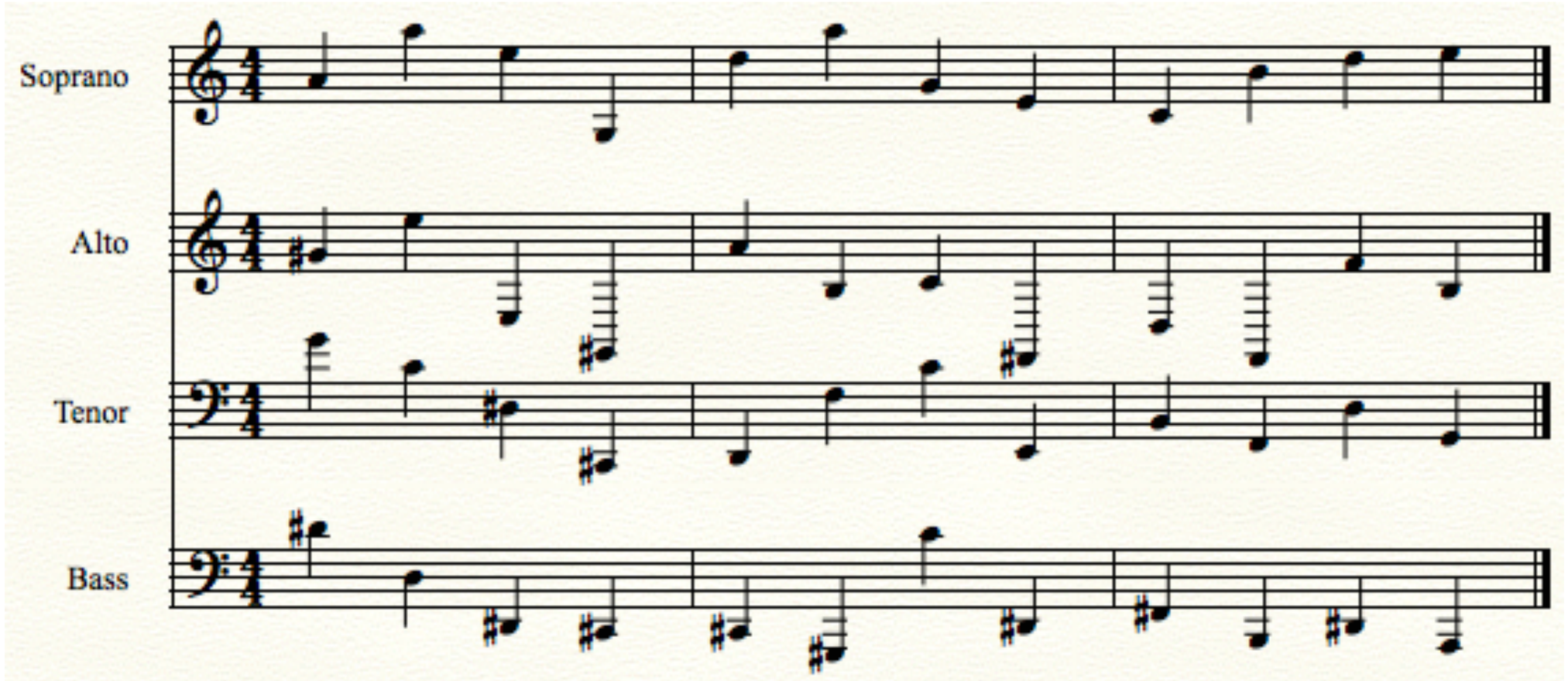
Rule 2: $\texttt{mod}_{12} \circ w_1$





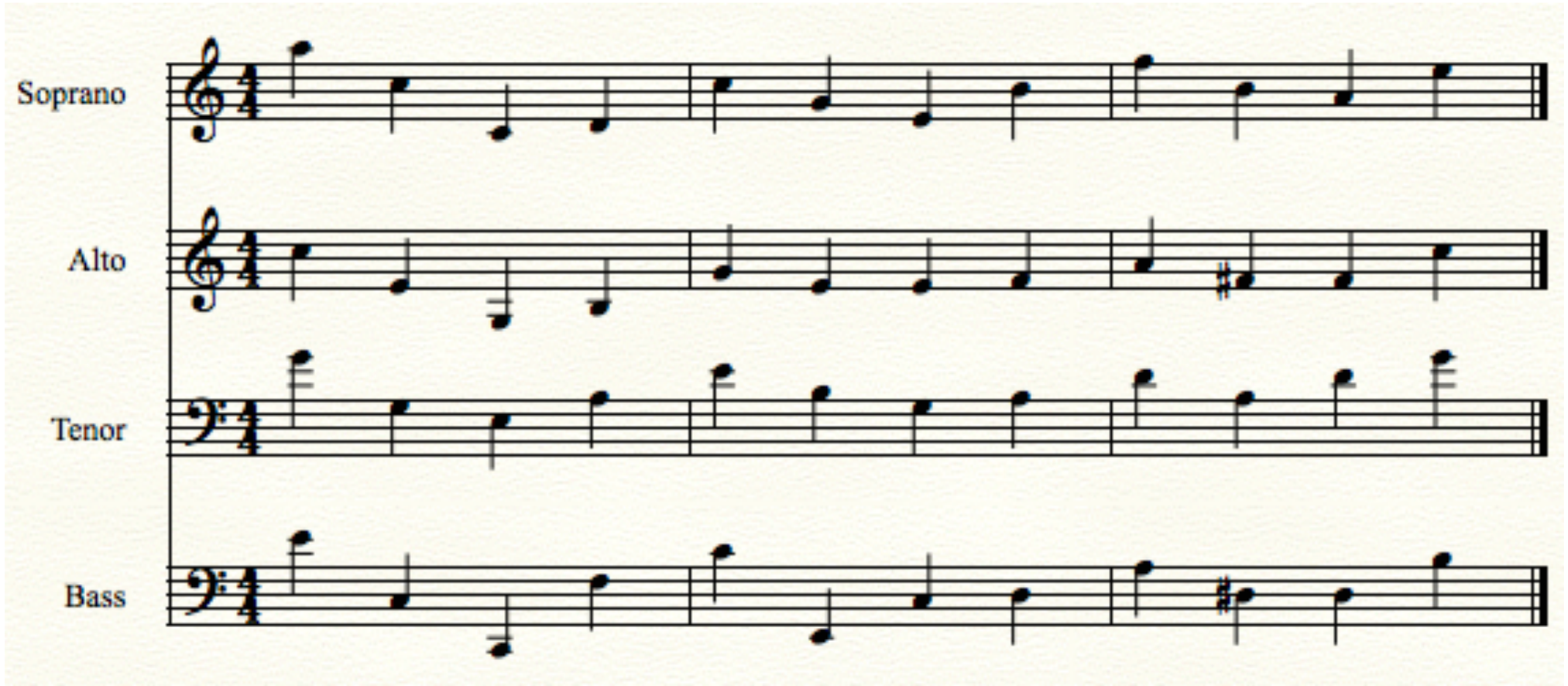
1-gram

Student 2



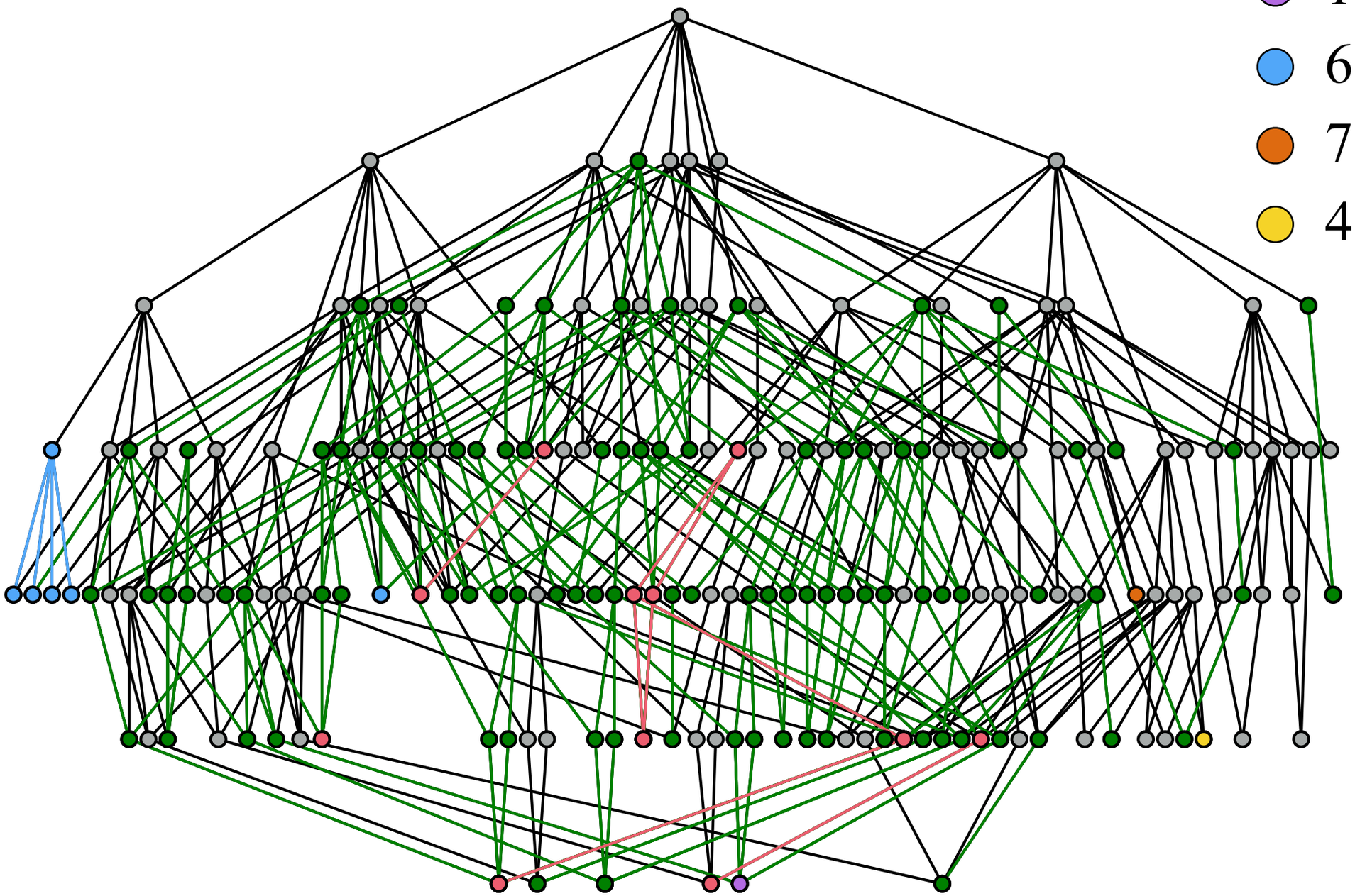
1-gram

Student 22



n-gram

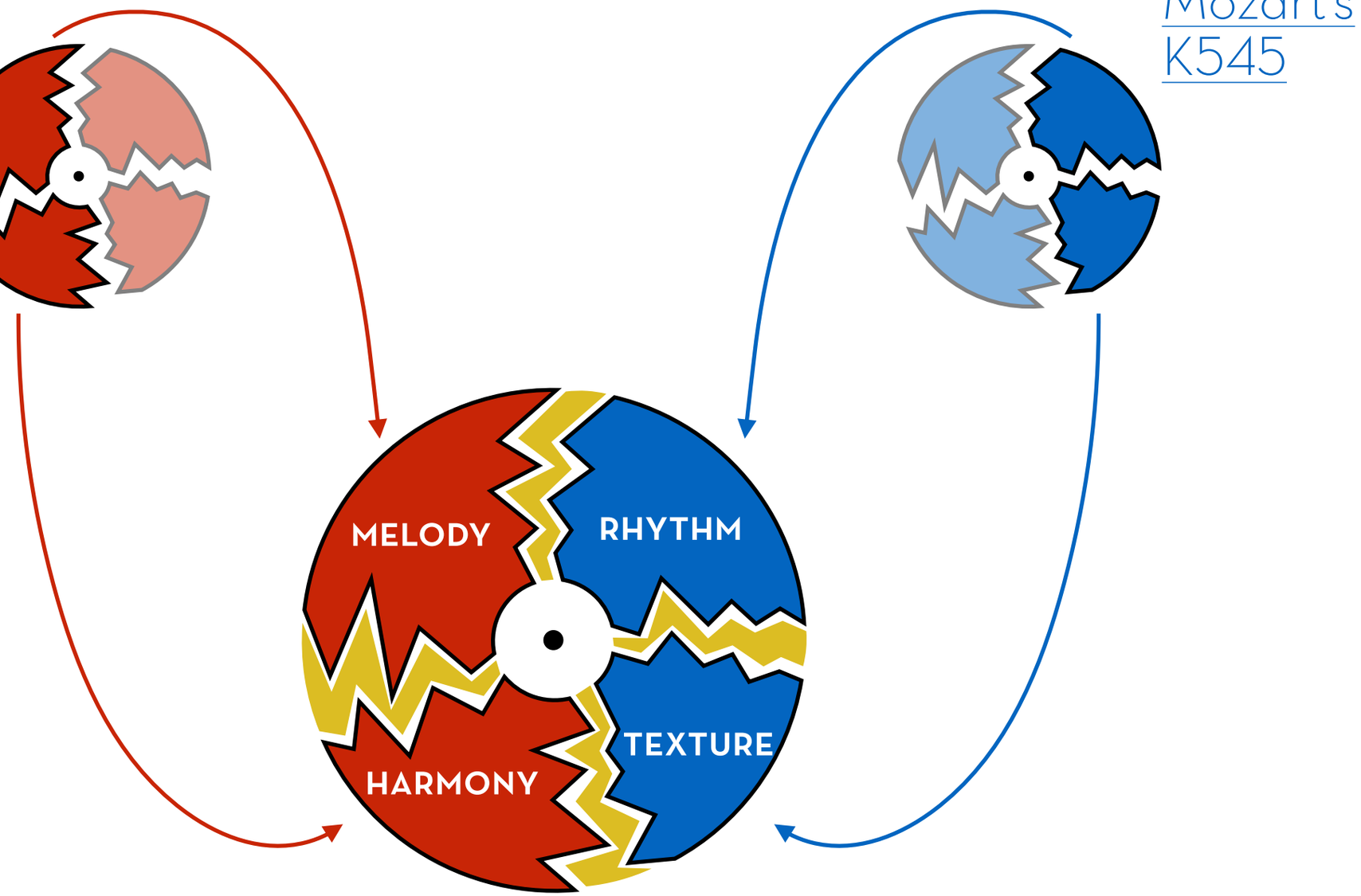
End of Loop 10



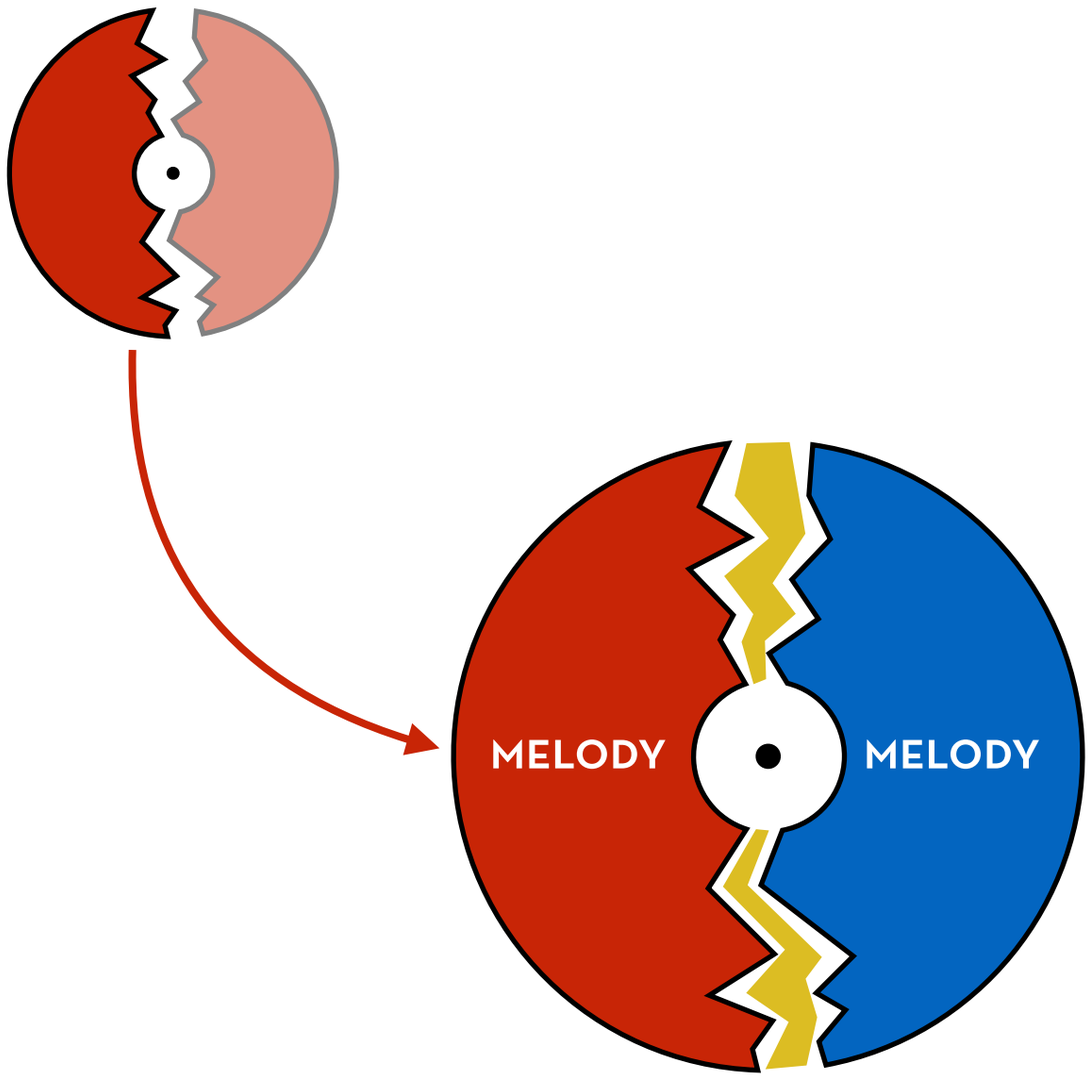
Music Mosaic

Happy Birthday

Mozart's K545



Happy Birthday



Beethoven's 5th Symphony