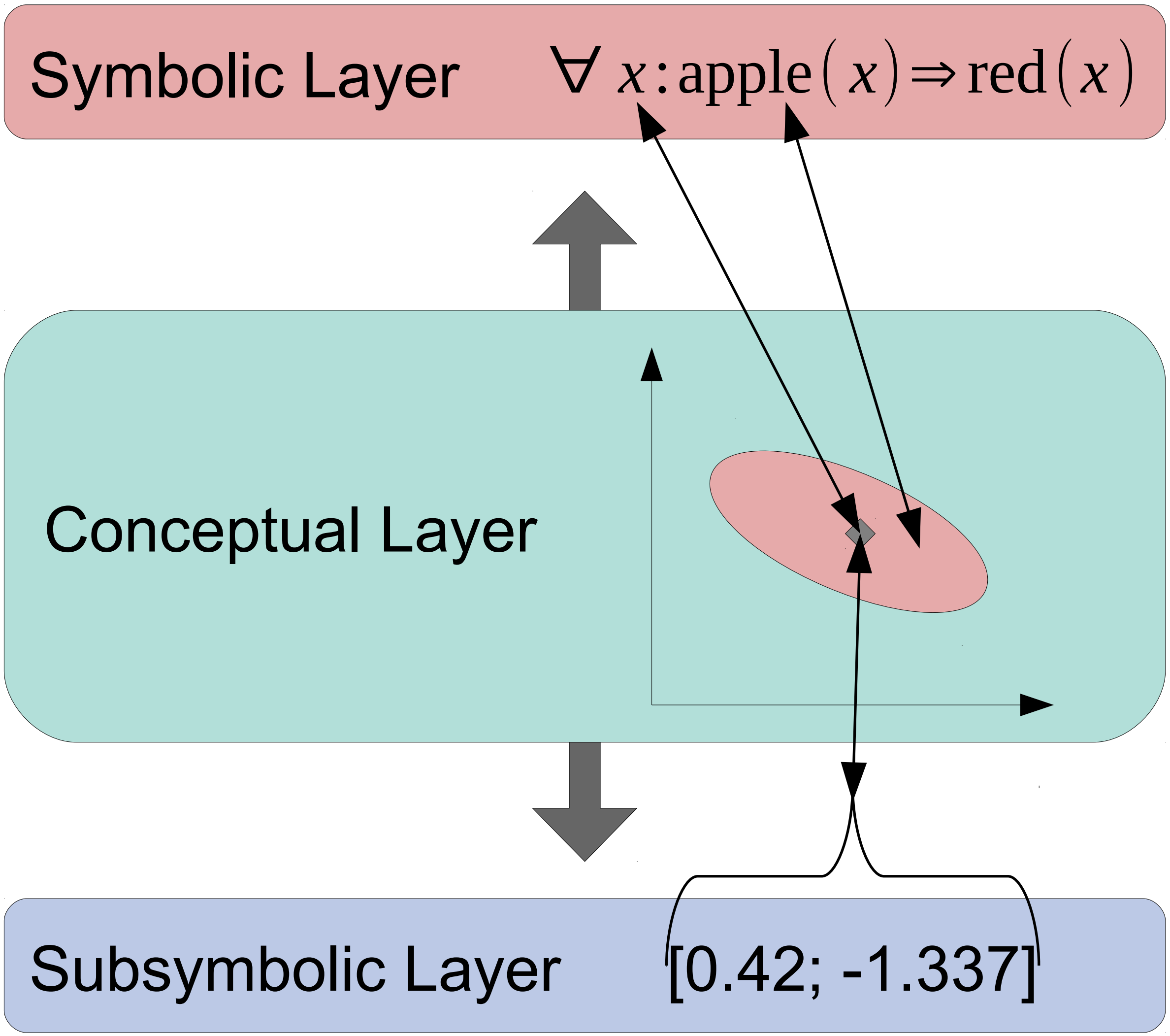
Conceptual Spaces for Artificial Intelligence

Formalization, Domain Grounding, and Concept Formation

Lucas Bechberger

https://www.lucas-bechberger.de

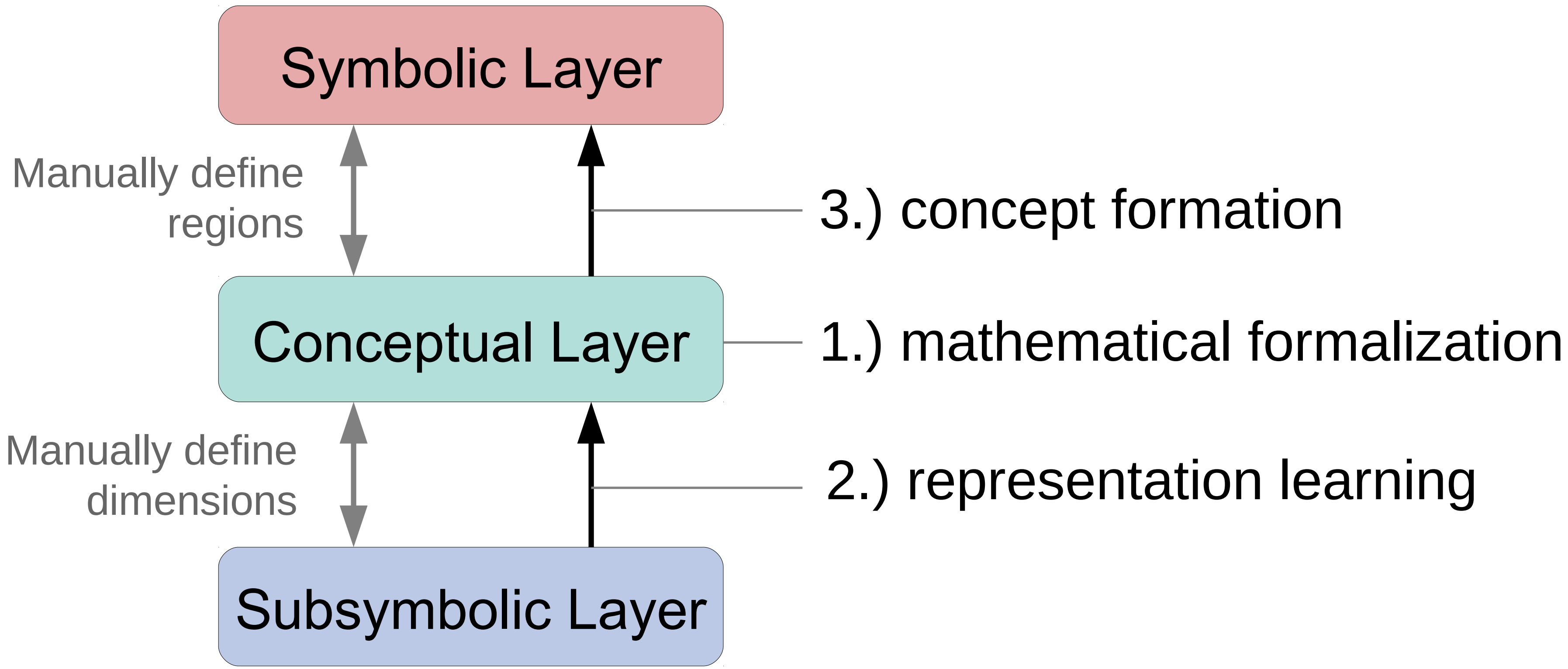
The different layers of representation



Formal Logics

Geometric Representation

Sensor Values, Machine Learning


Symbolic Layer
Manually define regions
3.) concept formation
Conceptual Layer
1.) mathematical formalization
Manually define dimensions
2.) representation learning
Subsymbolic Layer

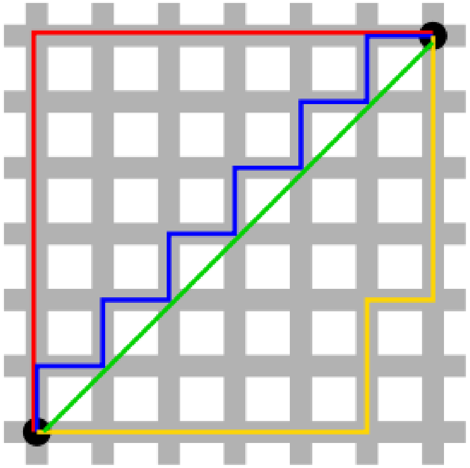
Conceptual Spaces [Gärdenfors2000]

- Quality dimensions
  - Interpretable ways of judging the similarity of two instances
  - E.g., temperature, weight, brightness, pitch
- Domain
  - Set of dimensions that inherently belong together
  - Color: hue, saturation, and brightness
- Distance in this space is inversely related to similarity
  - Within a domain: Euclidean distance
  - Between domains: Manhattan distance
- Concepts
  - Region + correlation information + salience weights

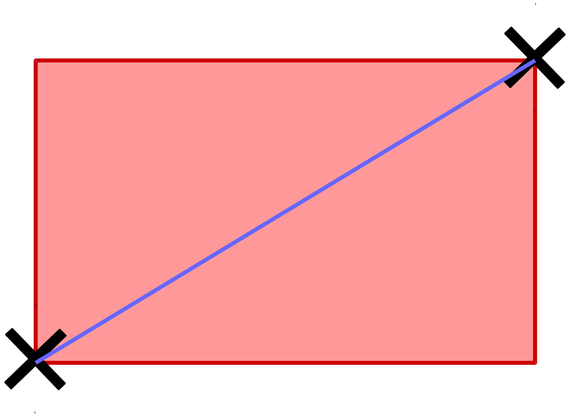
[Gärdenfors 2000] Gärdenfors, P. Conceptual Spaces: The Geometry of Thought. MIT press, 2000.

Betweenness

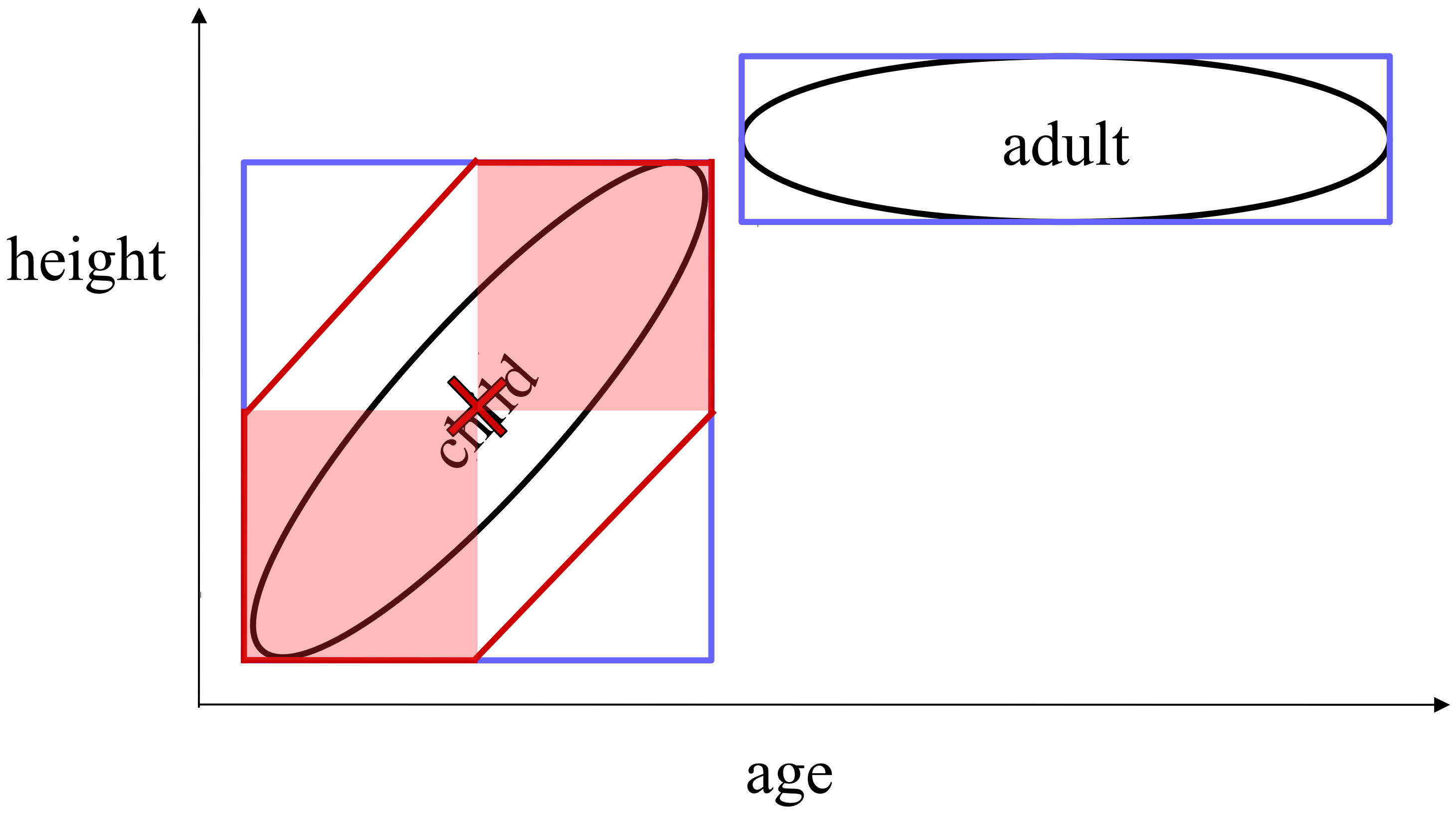
- $B(x,y,z) :\leftrightarrow d(x,y) + d(y,z) = d(x,z)$
- Convex region C: $\forall\ x,z \in C : \forall\ y : B(x,y,z) \Rightarrow y \in C$
- Star-shaped region S: $\exists\, p \in S : \forall\ z \in S : \forall\ y : B(p,y,z) \Rightarrow y \in S$



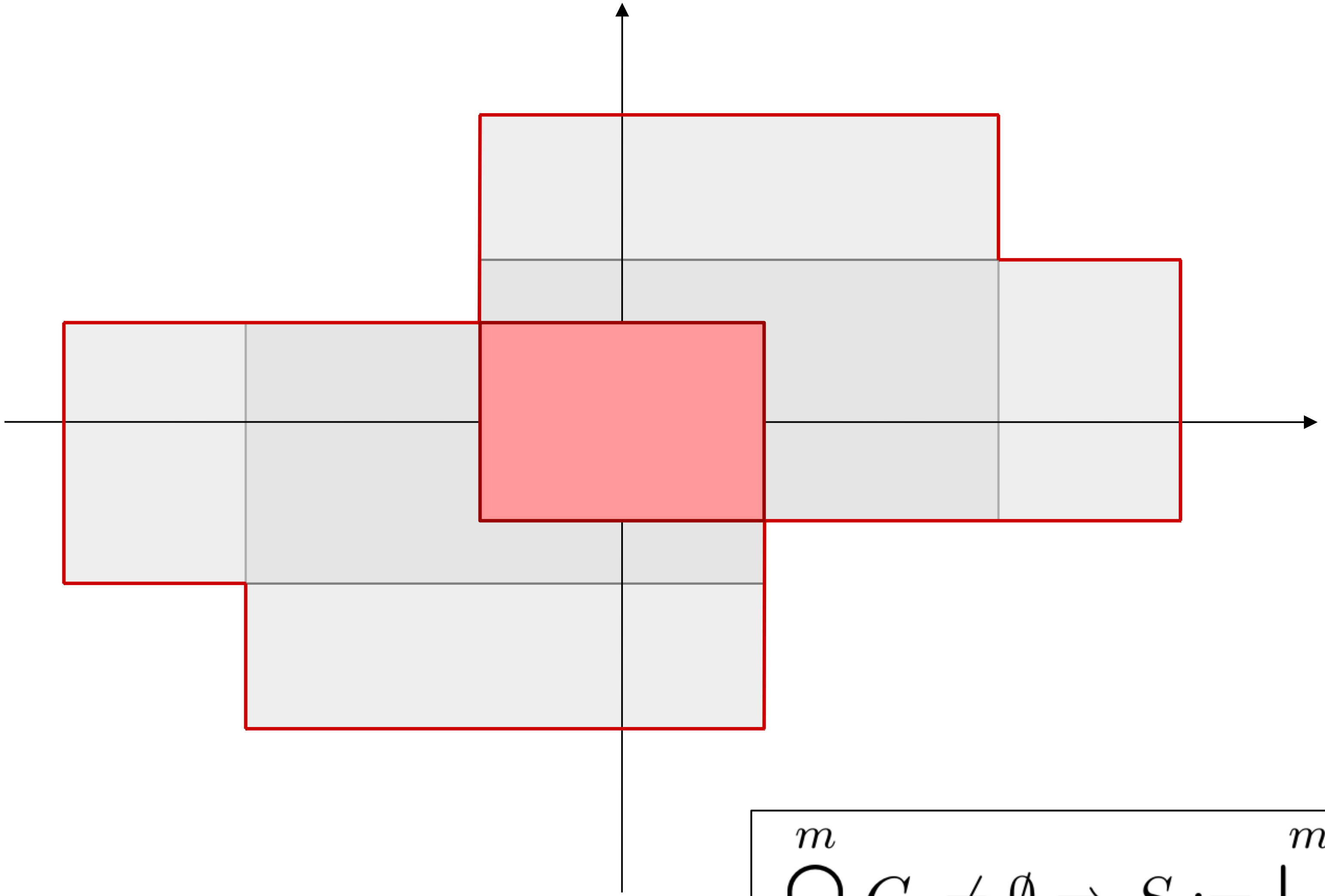
https://en.wikipedia.org/wiki/Taxicab_geometry#/media/File:Manhattan_distance.svg



Convexity and Manhattan distance

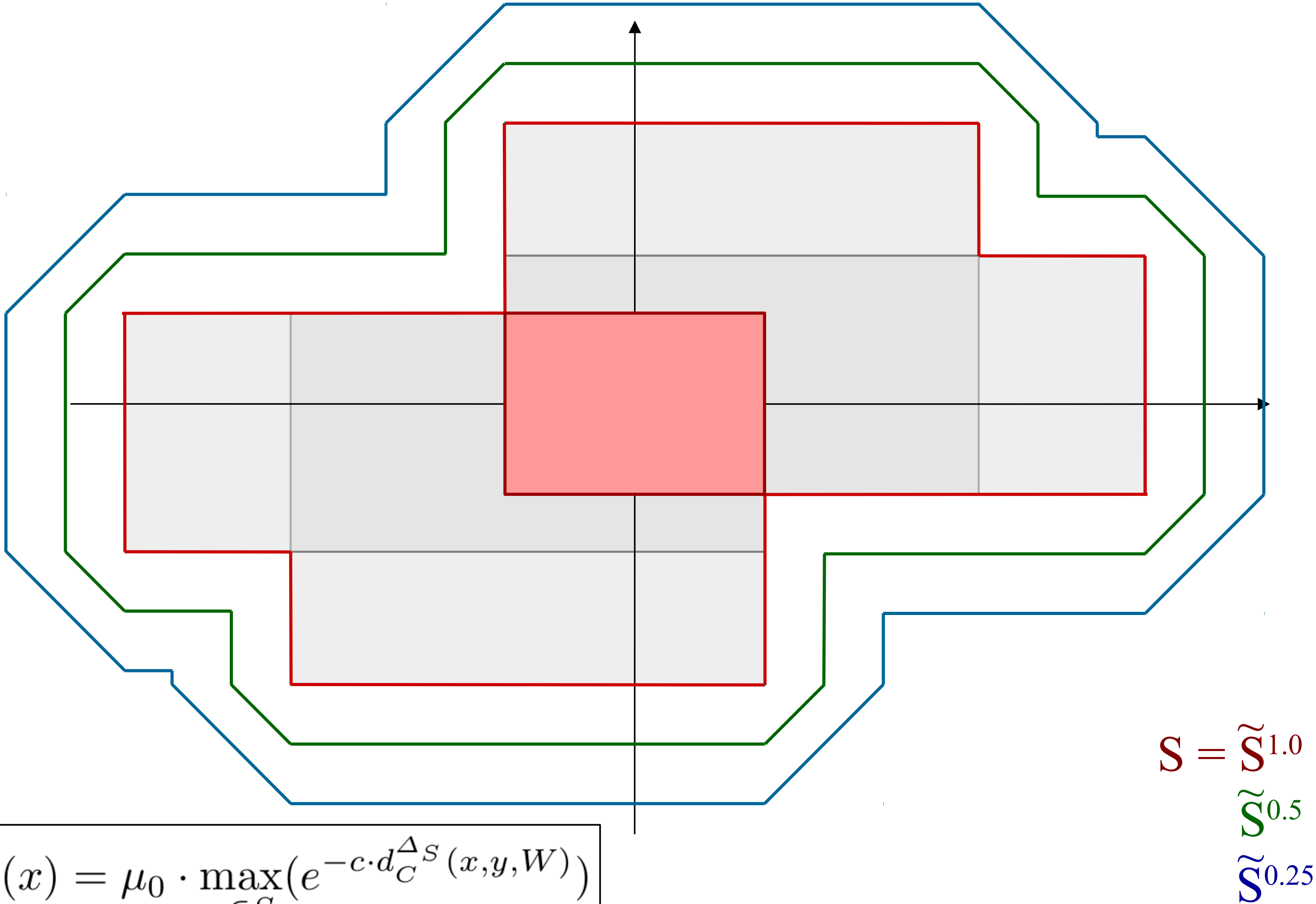


Formalizing Star-Shaped Concepts



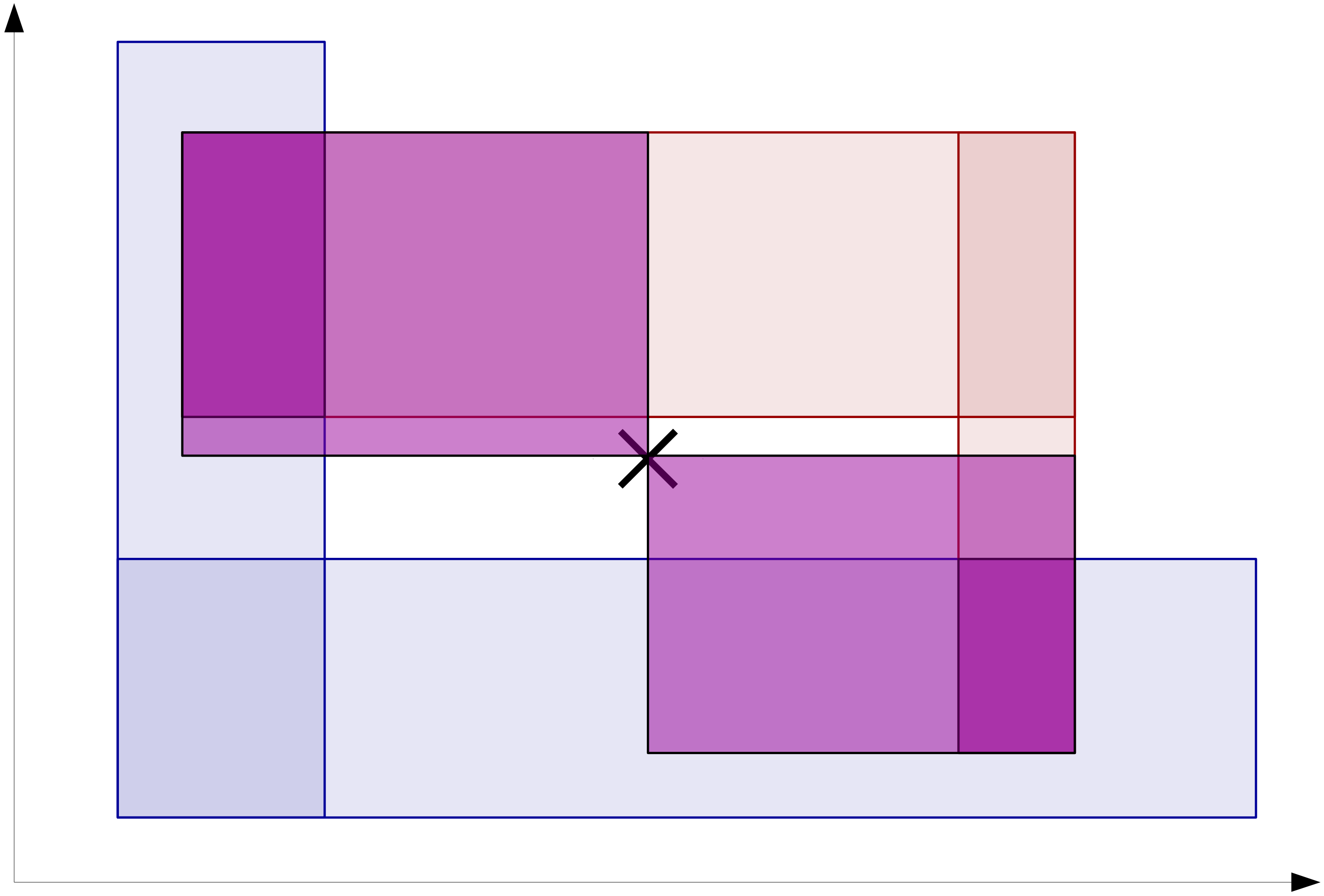
$$\bigcap_{i=1}^{m} C_i \neq \emptyset \Rightarrow S := \bigcup_{i=1}^{m} C_i$$

Formalizing Star-Shaped Concepts

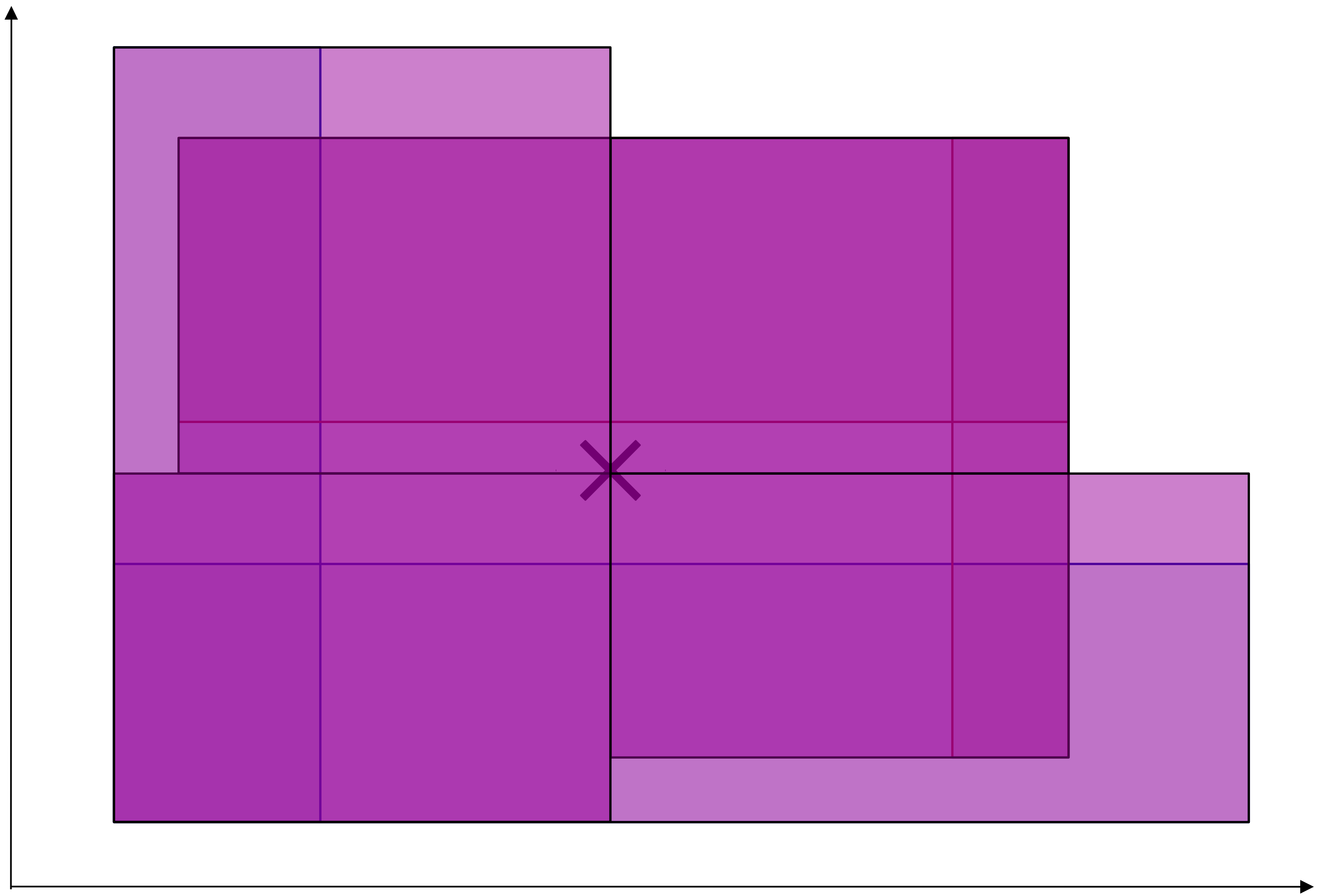


$$\mu_{\widetilde{S}}(x) = \mu_0 \cdot \max_{y \in S}\left(e^{-c \cdot d_C^{\Delta_S}(x,y,W)}\right)$$

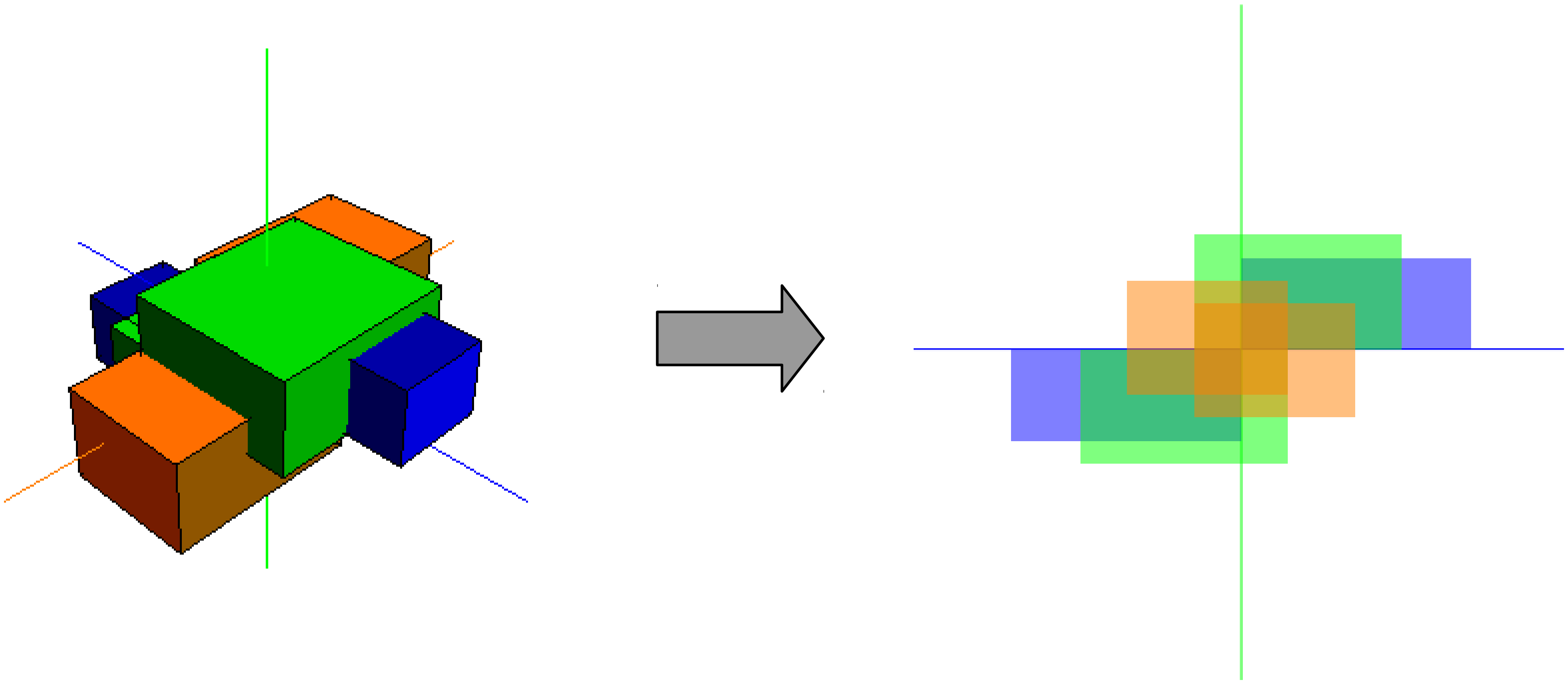
Intersection of Two Concepts



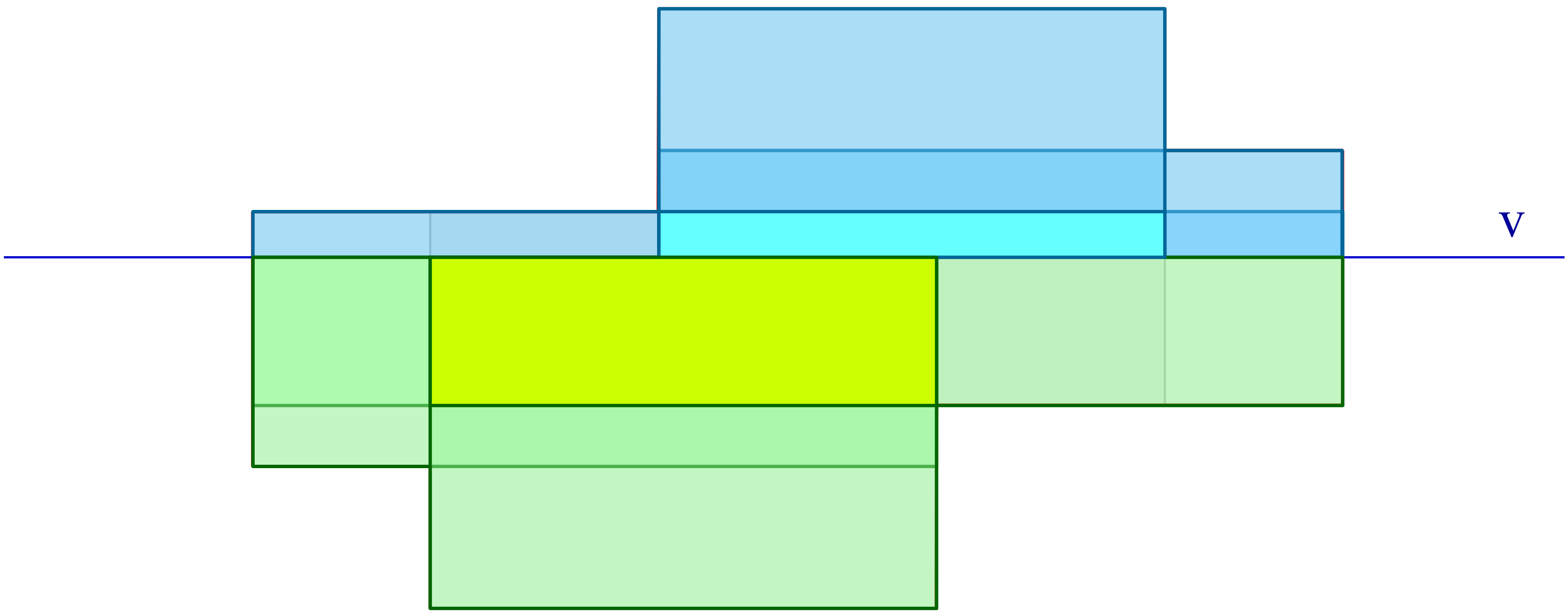
Unification of Two Concepts



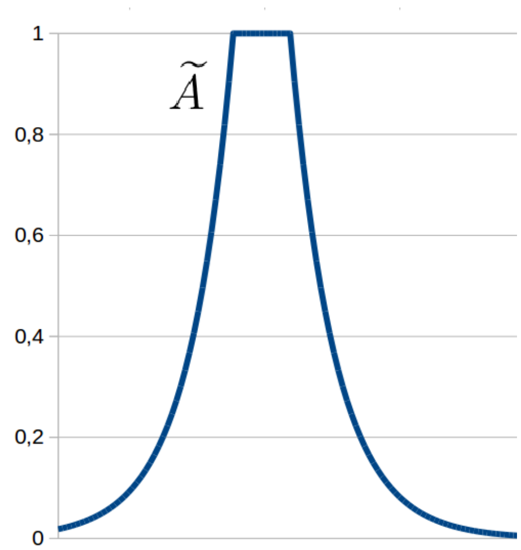
Projection of a Concept

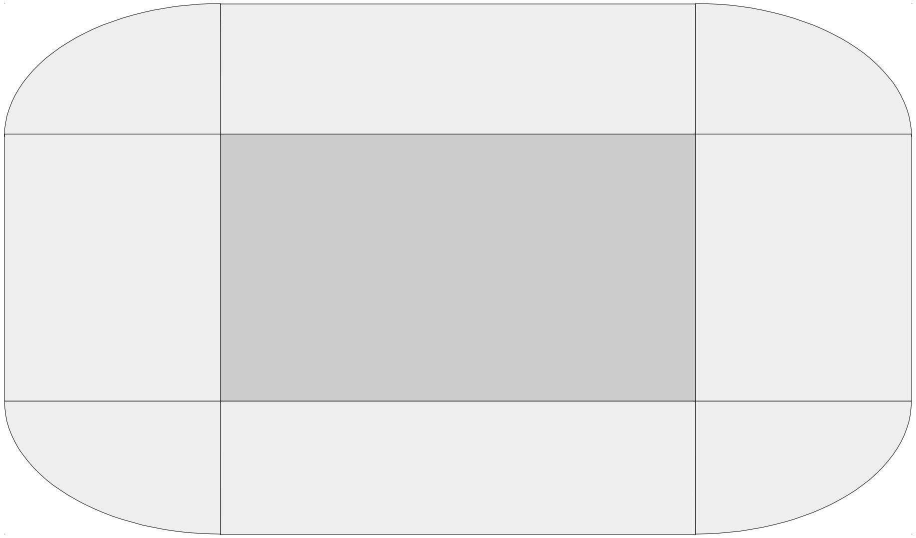


Splitting up a Concept



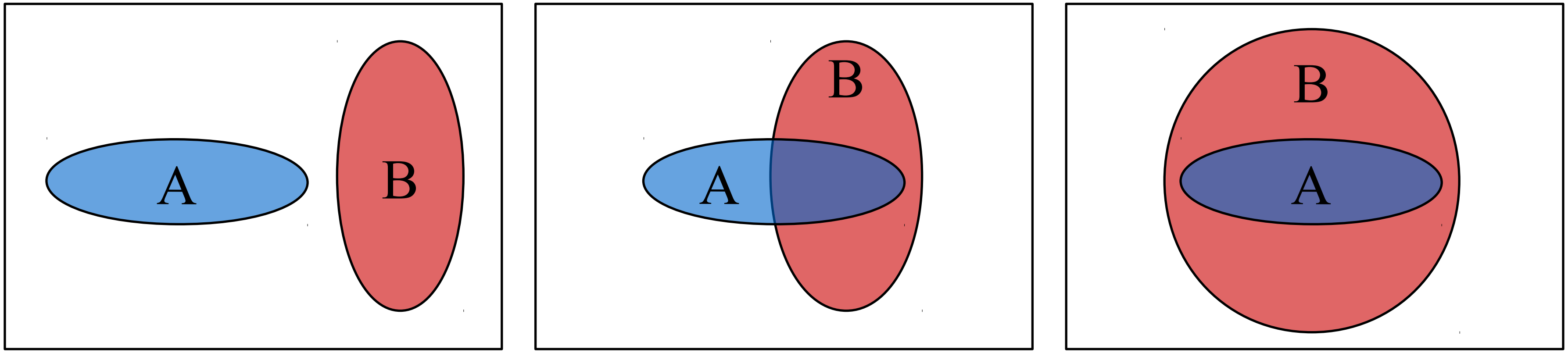
Measuring the Size of a Concept





$$M(\widetilde{C}) = \frac{\mu_0}{c^n \prod_{d\in D} w_{\delta(d)}\sqrt{w_d}} \sum_{i=0}^{n} \left(\sum_{\substack{\{d_1,\dots,d_i\}\\ \subseteq D}} \left(\prod_{\substack{d\in\\ D\setminus\{d_1,\dots,d_i\}}} a_d \right) \cdot \prod_{\substack{\delta\in\\ \Delta_{\{d_1,\dots,d_i\}}}} \left(n_\delta! \cdot \frac{\pi^{\frac{n_\delta}{2}}}{\Gamma(\frac{n_\delta}{2}+1)} \right) \right)$$

Degree of Subsethood & Implication



$sub(A, B) = 0$ $\quad$ $sub(A, B) \approx 0.5$ $\quad$ $sub(A, B) = 1$

$$sub(A, B) = \frac{M(A \cap B)}{M(A)}$$

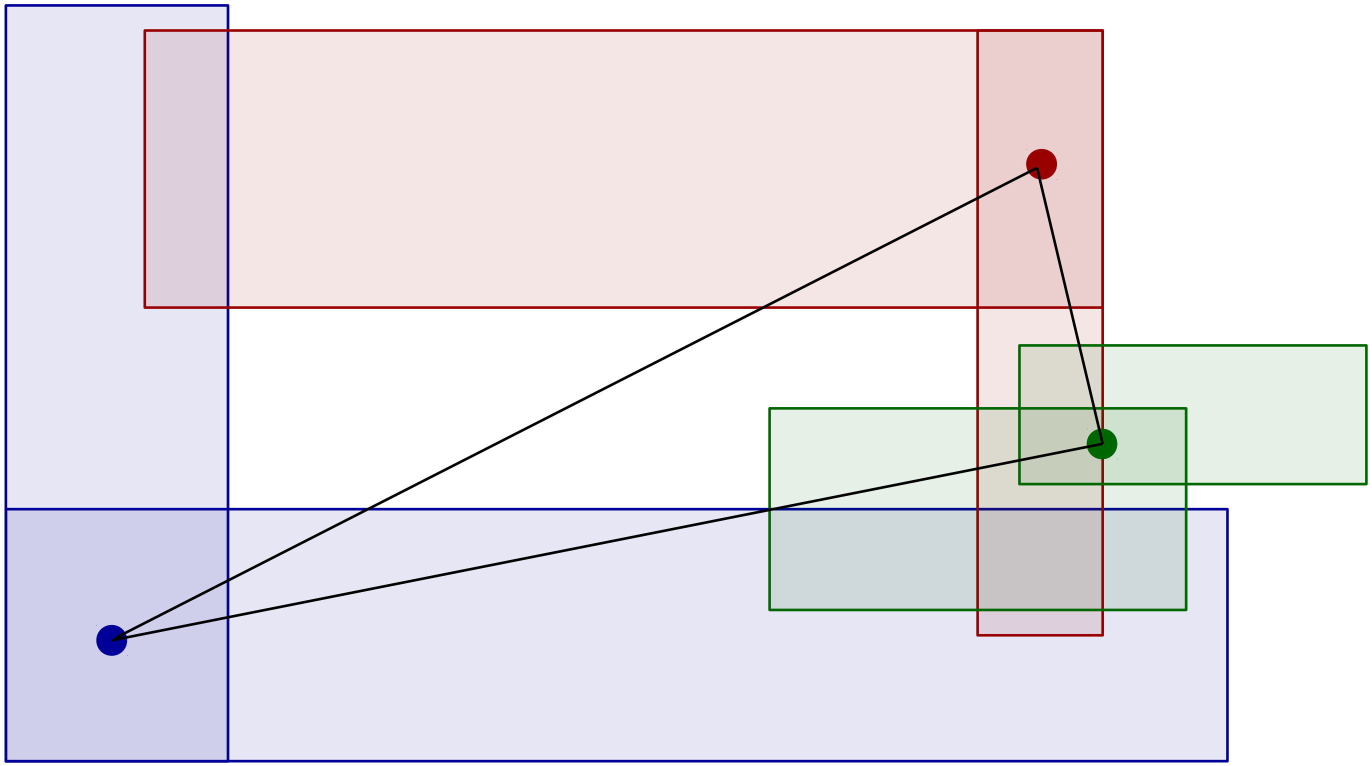
Kosko, B. Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence Prentice Hall, 1992

- “apple” implies “red” to the degree to which “apple” is a subset of “red”

Similarity and Betweenness

- Use point-wise definitions for now [Derrac2014]

$$Sim(a,b) = e^{-c \cdot d(a,b)} \qquad Btw(a,b,c) = \frac{d(a,c)}{d(a,b) + d(b,c)}$$



[Derrac2014] Joaquín Derrac and Steven Schockaert. Enriching Taxonomies of Place Types Using Flickr. FoIKS 2014.

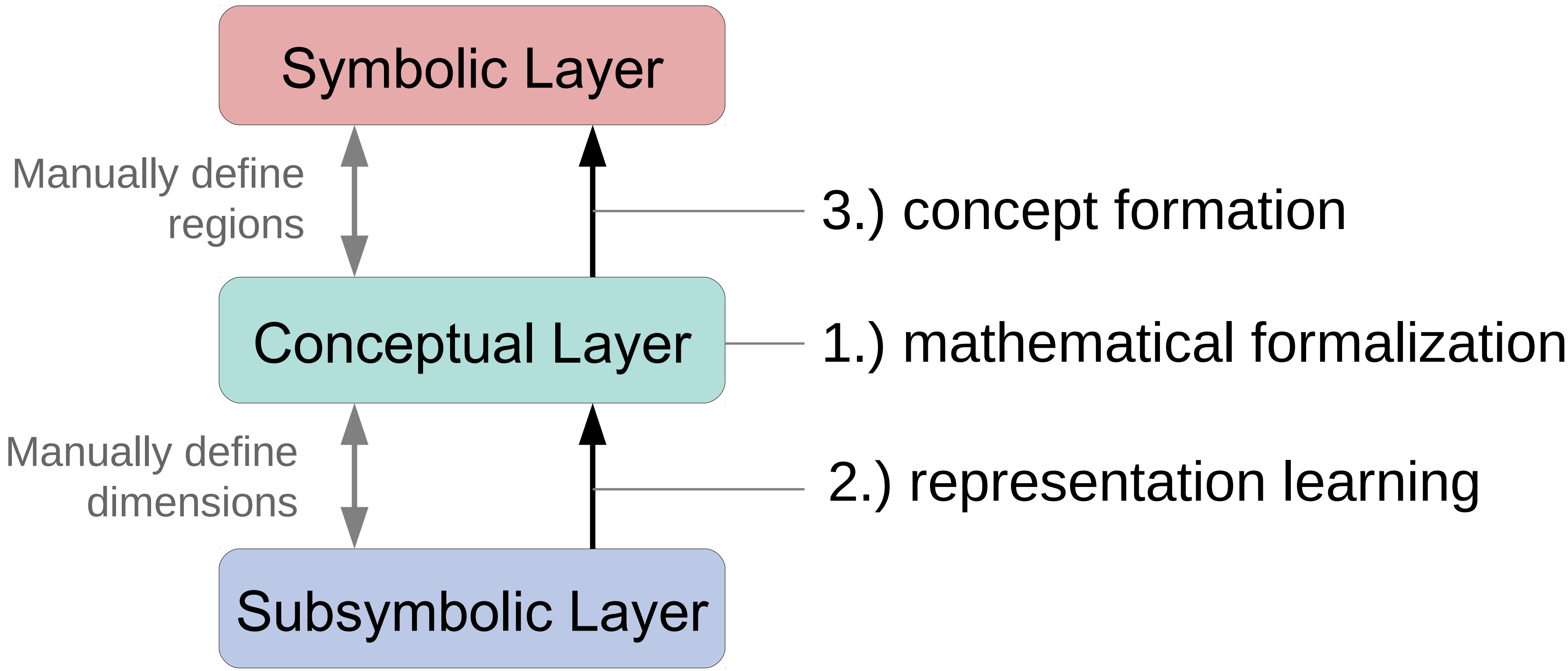
Formalization – Summary

- We can encode correlations in a geometric way
  - Most prior formalizations completely ignore this important aspect
  - [Rickard2006] considers correlations, but not in a geometric way
- Quite straightforward to implement
  - Represent each cuboid by two support points
  - Single constraint: cuboids must intersect
  - https://github.com/lbechberger/ConceptualSpaces
- Comprehensive list of supported operations:
  - Set membership
  - Intersection, Union, Projection, Cut
  - Size, Subsethood/Implication, Similarity, Betweenness

[Rickard2006] Rickard, J. T. A Concept Geometry for Conceptual Spaces. Fuzzy Optimization and Decision Making, 2006

DEMO TIME!

Conceptual Spaces for AI

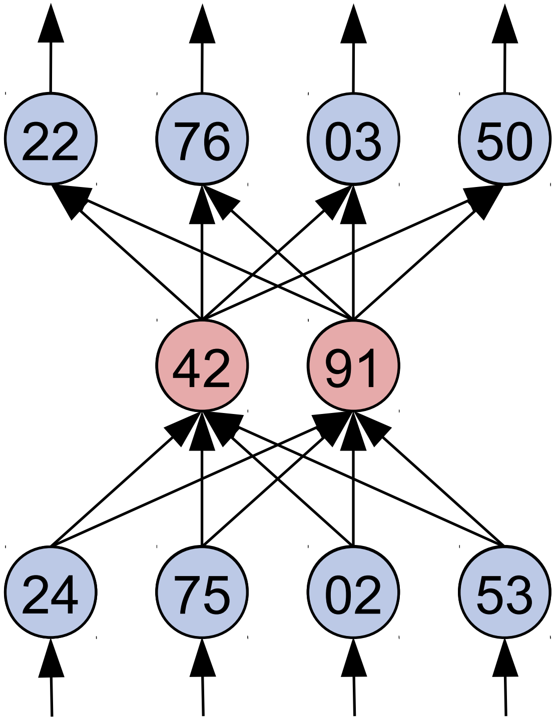


Why do we need domain grounding?

- Isn't it quite obvious which dimensions we need?
  - Color: hue, saturation, brightness
  - Temperature: temperature
  - Emotions: valence, arousal
  - …
- … but what about shape?
  - it's surprisingly hard to define this domain with a handful of dimensions
  - Roundness, convexity, number of corners?
  - But how to extract those from images?
- Idea: **learn** the dimensions of a given domain with ANNs

(Deep) Representation Learning

- Autoencoder (e.g., β-VAE): compress and reconstruct input

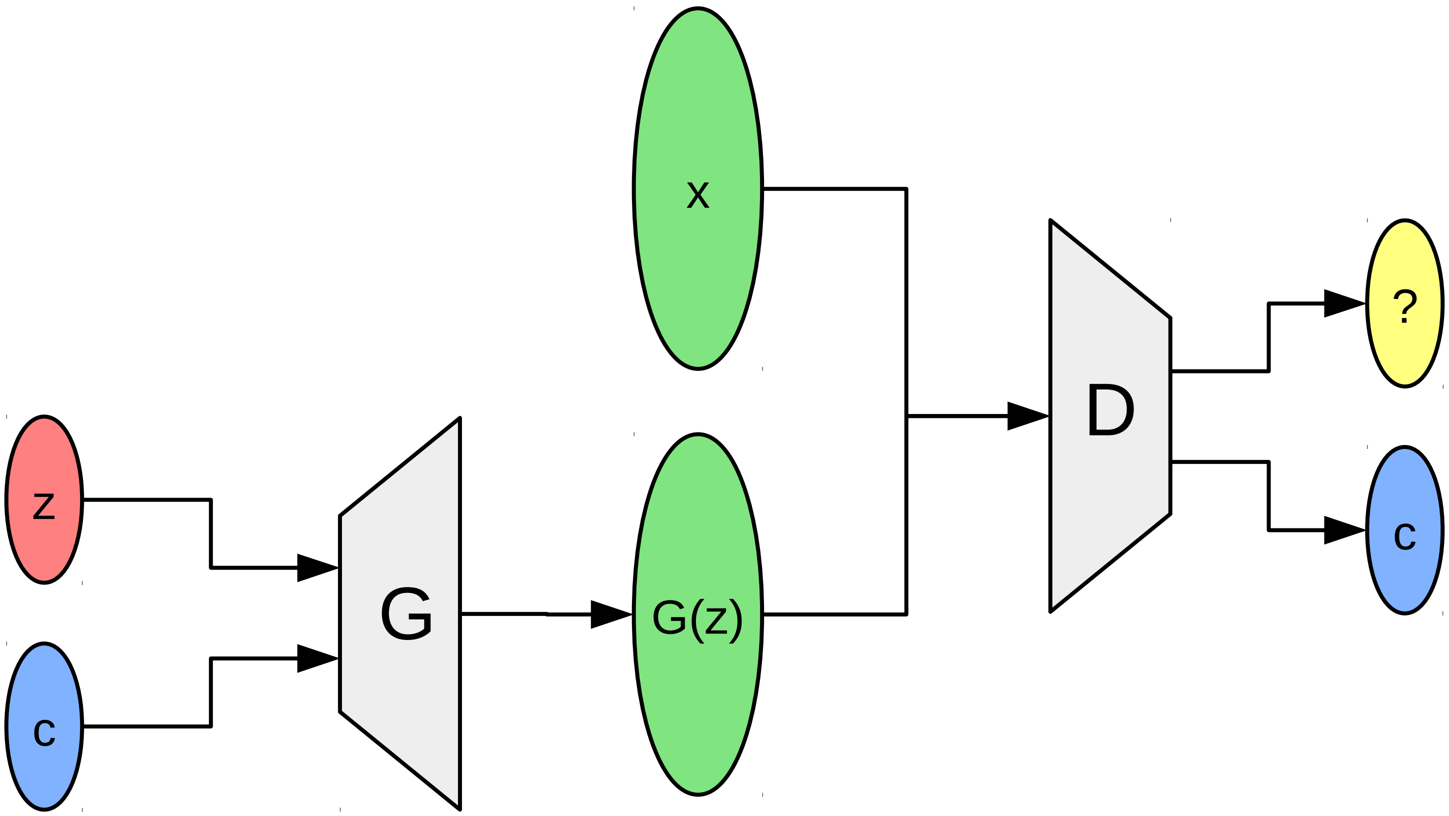


- Hidden neurons = dimensions in our conceptual space

Higgins, I.; Matthey, L.; Pal, A.; Burgess, C.; Glorot, X.; Botvinick, M.; Mohamed, S. & Lerchner, A. β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework, ICLR 2017

InfoGAN – Architecture

- Information Maximizing Generative Adversarial Networks



X. Chen et al., “InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets”, Advances in Neural Information Processing Systems, 2016

InfoGAN – MNIST Results

- Three latent variables
  - Categorical (10 classes)
  - Continuous (uniform)
  - Continuous (uniform)

X. Chen et al., "InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets", Advances in Neural Information Processing Systems, 2016

Domains and Latent Spaces

Domains in CS framework

- Interpretable dimensions
- Distance-based notion of semantic similarity
- Geometric betweenness represents semantic betweenness

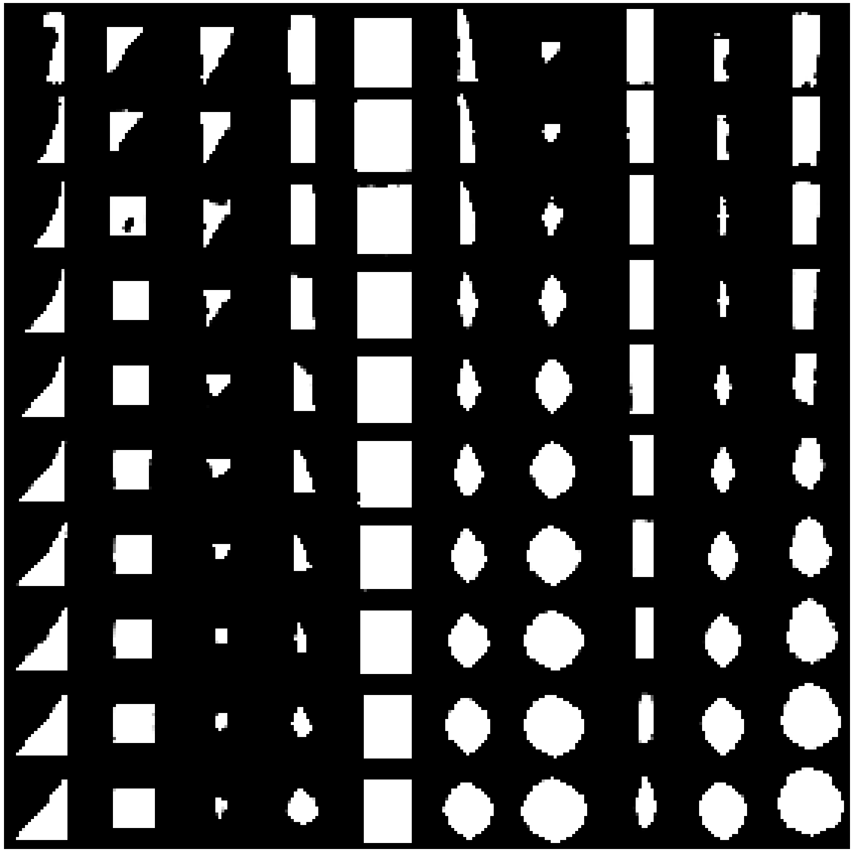
Latent Spaces of InfoGAN and β-VAE

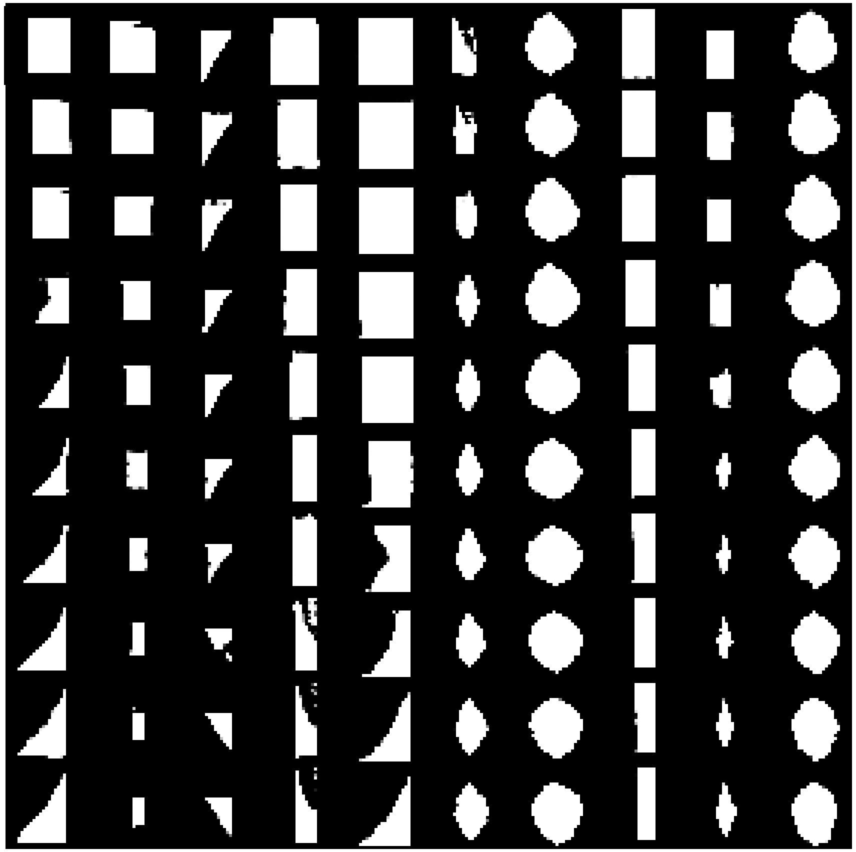
- Tends to be the case
- Smoothness assumption
- Interpolations in latent space describe a meaningful morph

→ **use InfoGAN/β-VAE on a data set of shapes to learn dimensions**

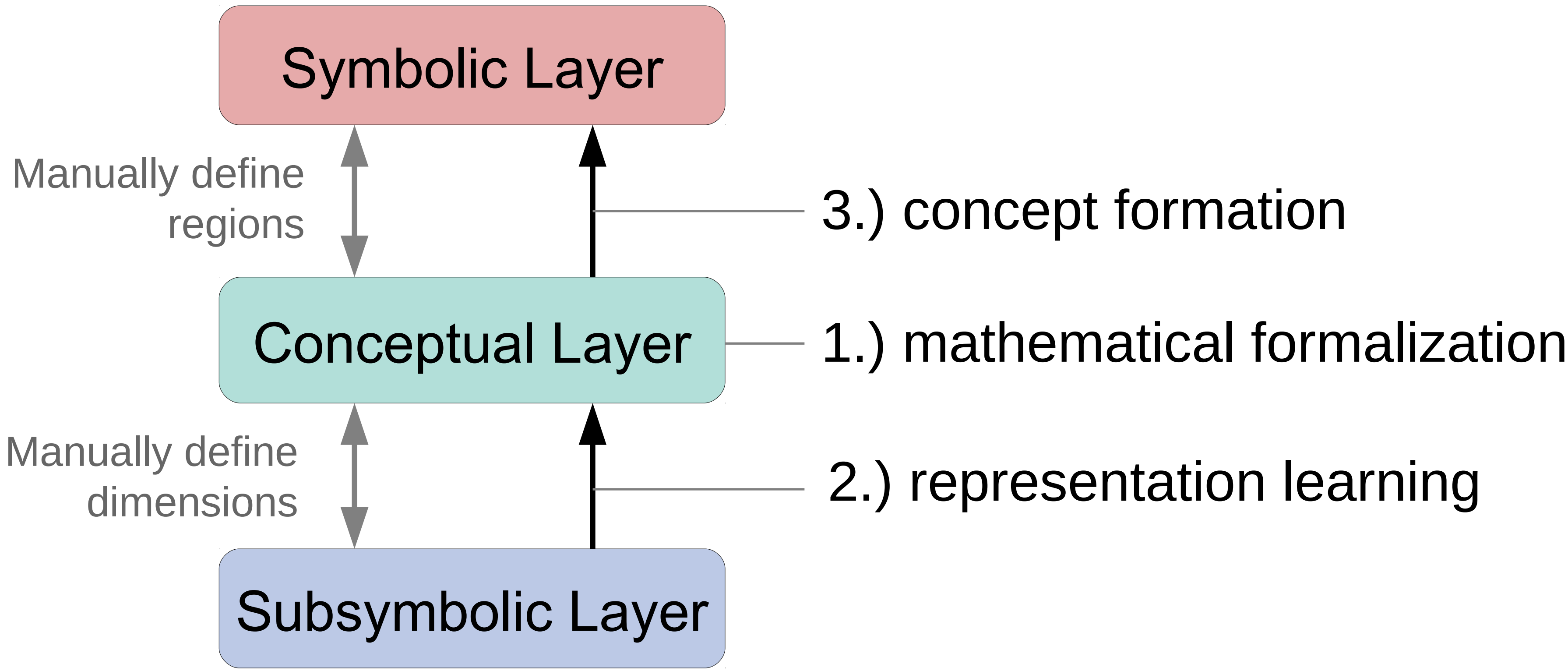
First Preliminary Results (InfoGAN)

- Data set of right-angled triangles, rectangles, and ellipses
- 2 continuous variables (uniform distribution), 500 epochs



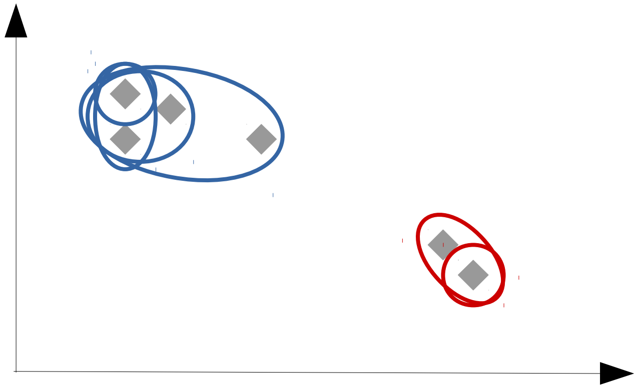


Conceptual Spaces for AI



Concept Formation

- We look for meaningful regions in the conceptual space
  - Concepts = clusters of data points
- Observed objects usually come without class information
  - Unsupervised learning
- Observing one object at a time, limited memory
  - Stream of data points, incremental processing



What do we need?

- Wish list for a clustering algorithm
  - Incremental (stream of observations)
  - Semi-supervised (take into account scarce feedback)
  - Unknown number of clusters
  - Work with my fuzzy formalization of concepts
  - Hierarchical

- Good news: some approaches seem partially fitting
- Bad news: none of them fits perfectly

→ need to combine existing ideas into a new algorithm

What a clustering algorithm needs

- Operations
  - Add clusters
  - Remove clusters
  - Move and resize clusters
  - Change the shape of clusters
  - Merge clusters
  - Split a cluster into sub-clusters

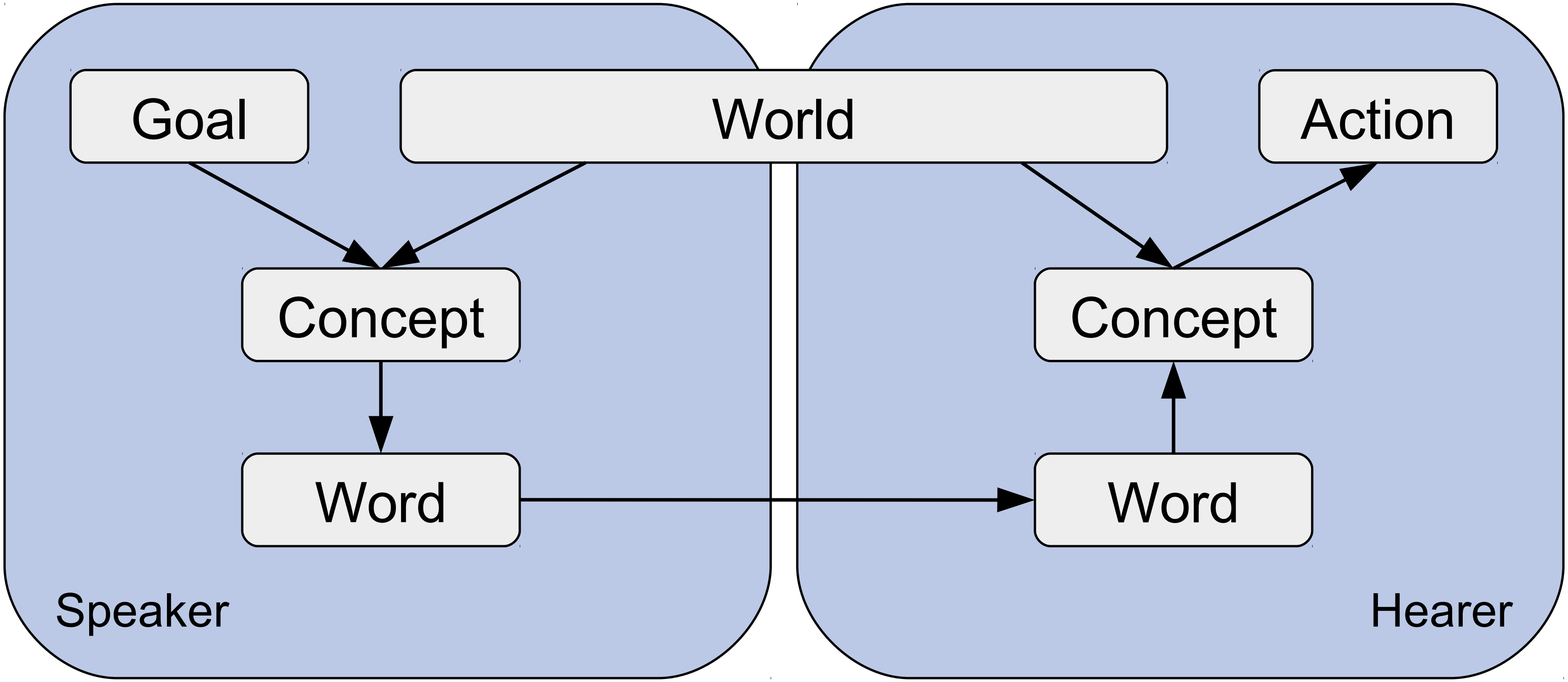
- Information
  - Size of clusters
  - Overlap of clusters
  - Hierarchy of clusters

Is there more grounding needed?

- Concepts are already grounded in perception
- … but there are many ways in which the conceptual space can be divided up into concepts
- Still, humans seem to share their concepts (otherwise we could not communicate)
- Idea: use of concepts in communication gives further constraints

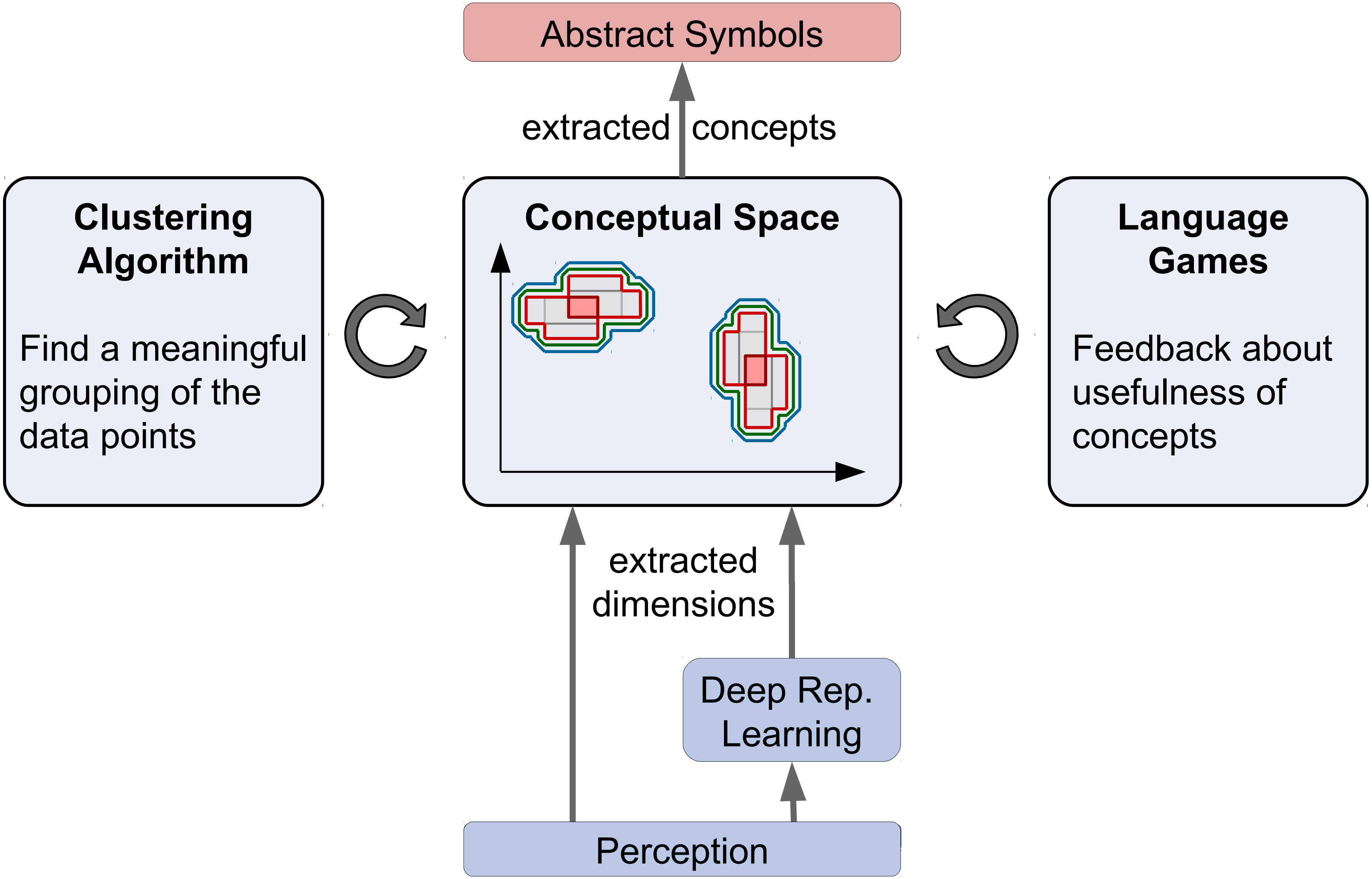
Language games [Steels2015]

https://www.pexels.com/photo/art-artistic-bright-close-up-268435/

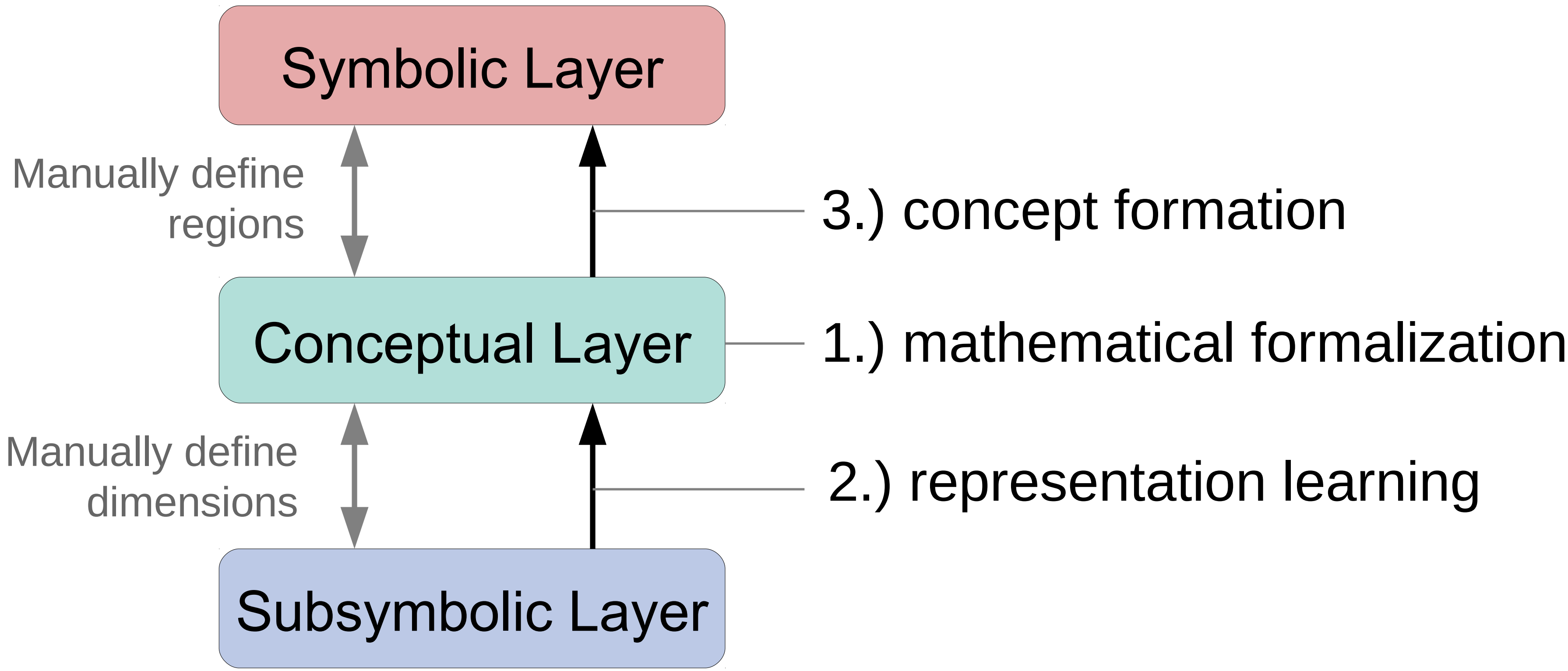


[Steels2015] Luc Steels, „The Talking Heads experiment: Origins of words and meanings“, Language Science Press, 2015

The overall envisioned architecture



Conceptual Spaces for AI



Thank you for your attention!

Questions? Comments? Discussions?

https://www.lucas-bechberger.de

@LucasBechberger