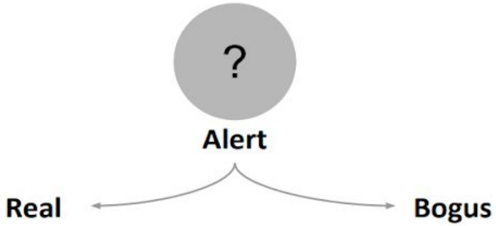
# A Hierarchy of Classification

V. Ashley Villar

Harvard University, Assistant Professor

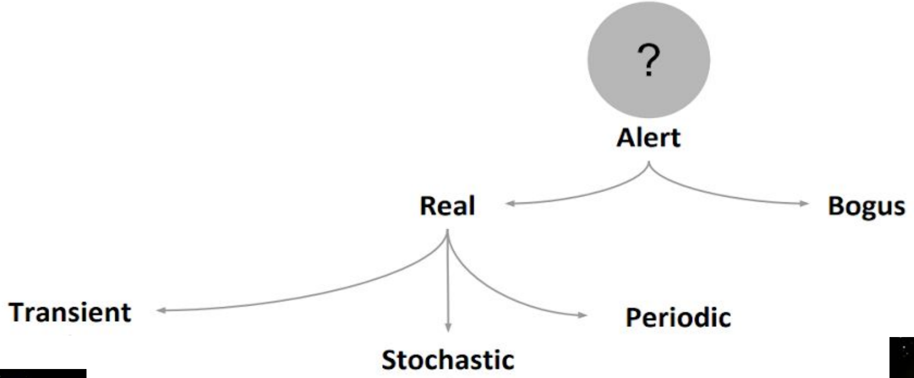
VLT

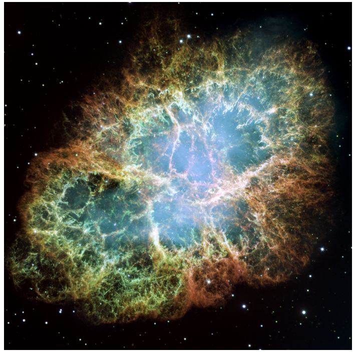
# We must classify a zoo of astronomical events

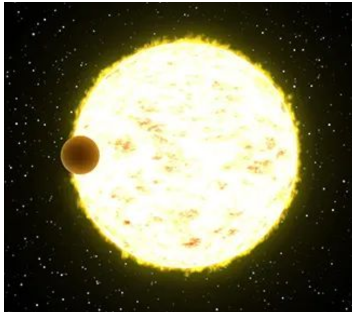



Sánchez-Sáez et al. (2020)
@wandering_astro (illustrations)

# We must classify a zoo of astronomical events

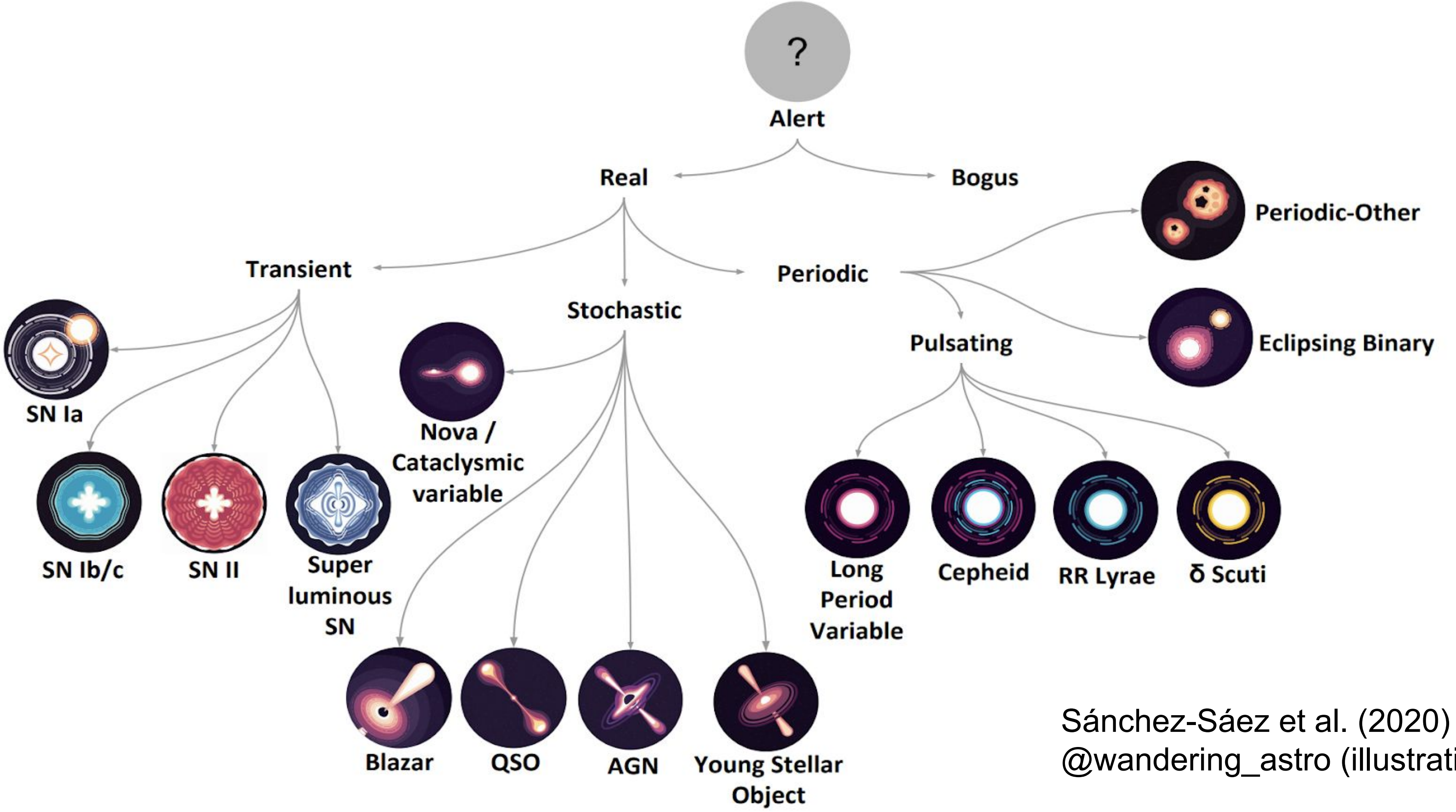




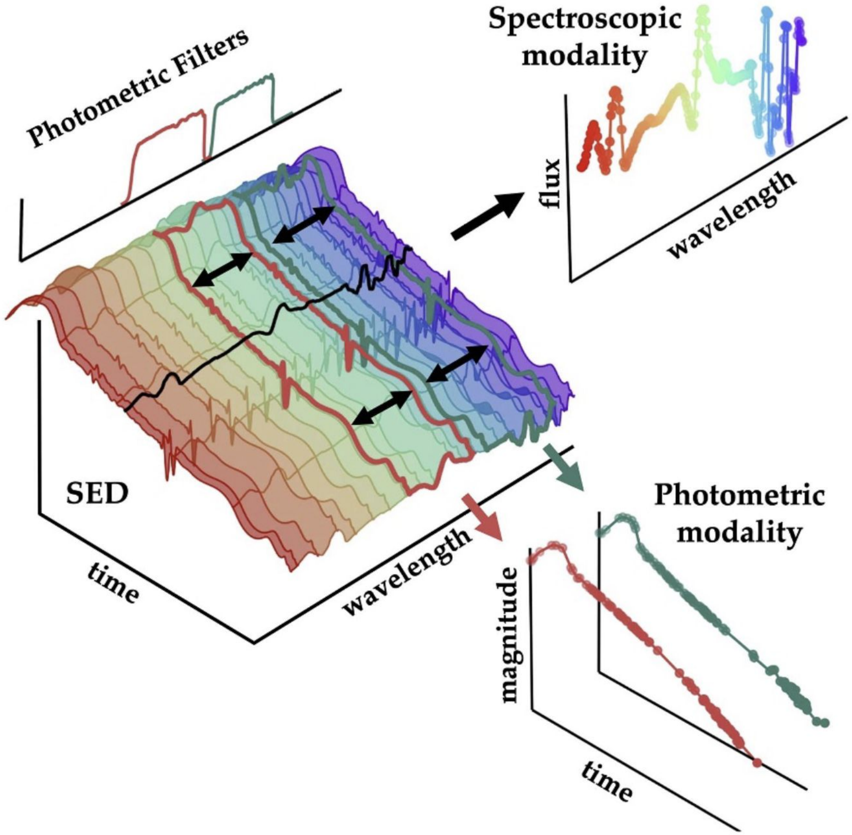
Sánchez-Sáez et al. (2020)
@wandering_astro (illustrations)

# We must classify a zoo of astronomical events



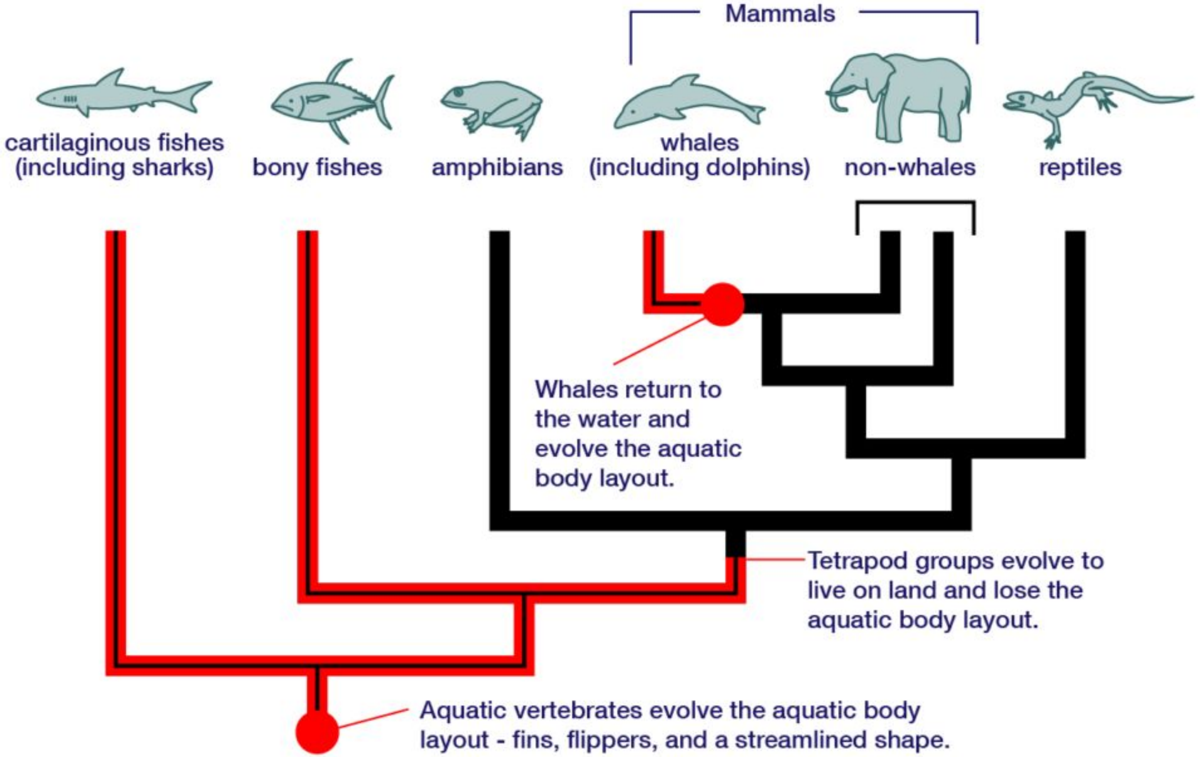

Sánchez-Sáez et al. (2020)
@wandering_astro (illustrations)

# Astrophysical events can be observed in many ways

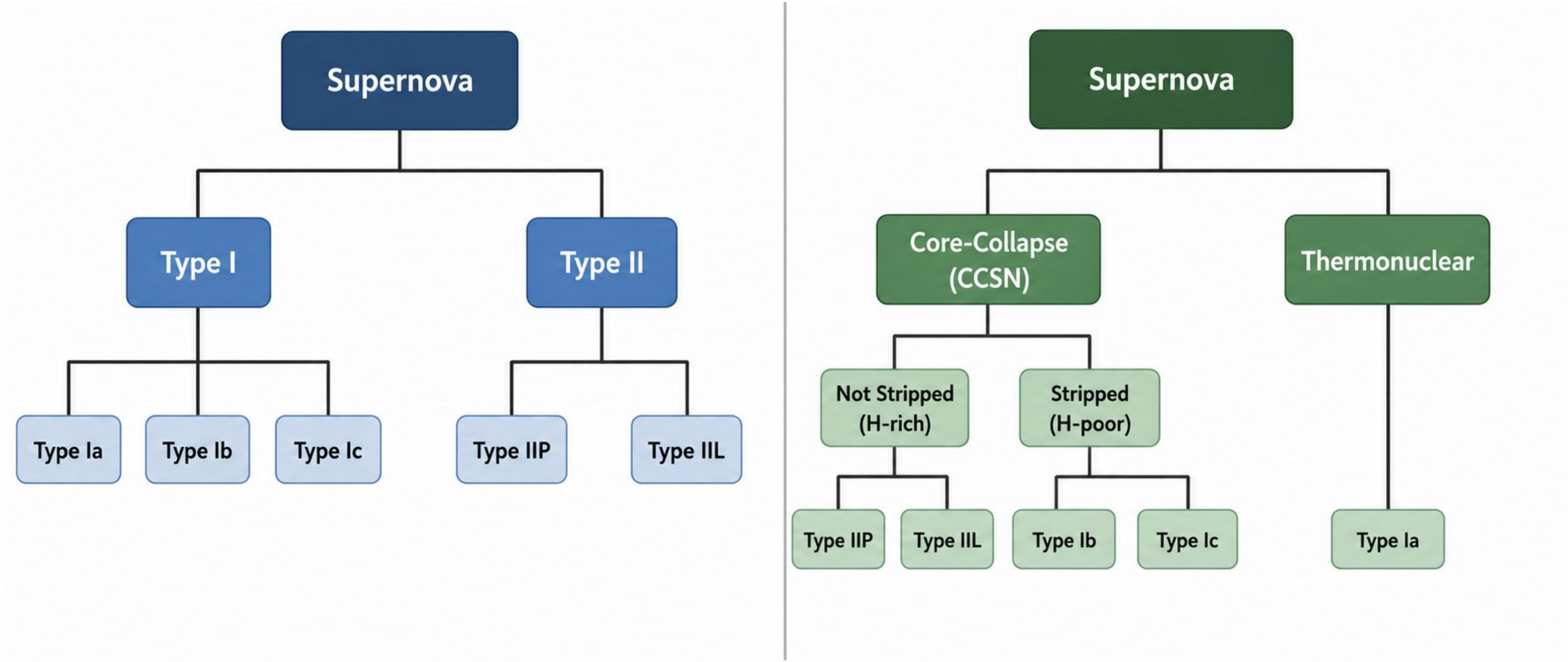


# How does one decide upon a taxonomy anyway?

# How does one decide upon a taxonomy anyway?



# Taxonomy: features or physics?



# Does any of these even matter for classification?

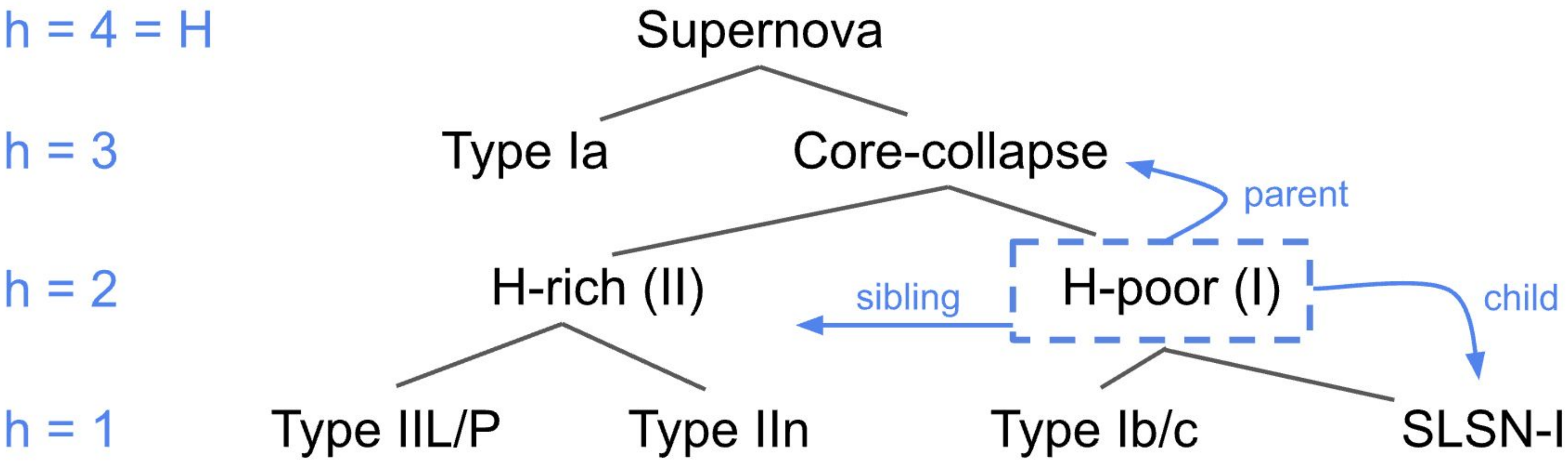
Does any of these even matter for classification?

Traditionally: no!

# Classic classification objective: cross entropy

$$H(P^*|P) = -\sum_i \underbrace{P^*(i)}_{\text{TRUE CLASS DISTIRBUTION}} \log \underbrace{P(i)}_{\text{PREDICTED CLASS DISTIRBUTION}}$$

# But we can incorporate this hierarchy

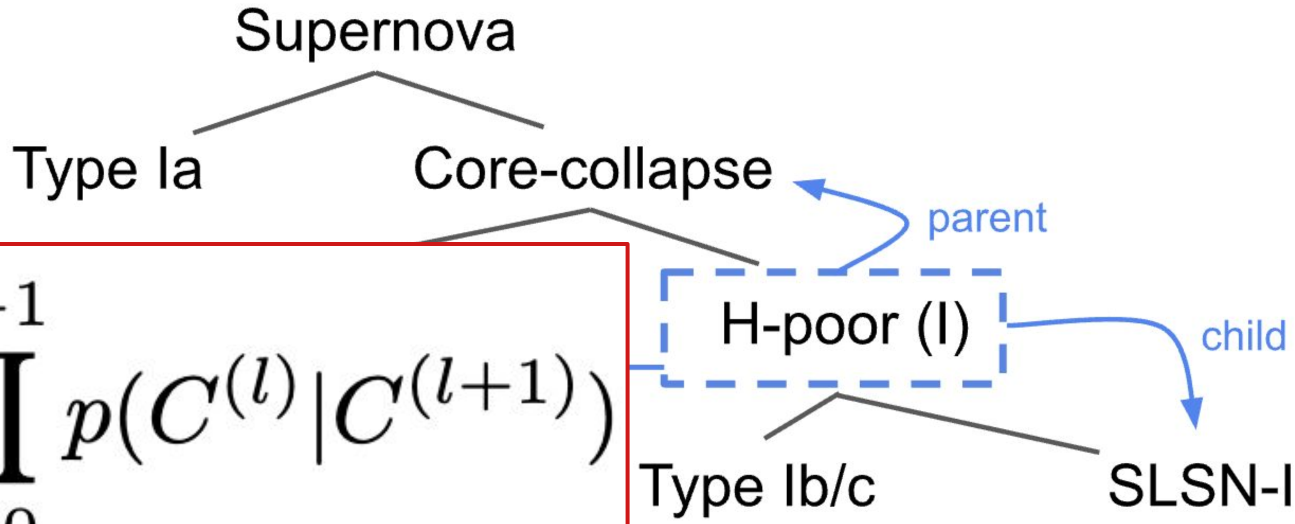



Bertinetto+20, VAV+25

# But we can incorporate this hierarchy

h = 4 = H

h = 3




$$p(C) = \prod_{l=0}^{h-1} p(C^{(l)} | C^{(l+1)})$$

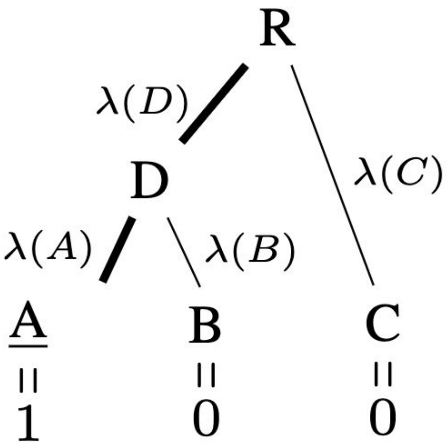
Bertinetto+20, VAV+25

# Hierarchical cross entropy

True edge

$$\mathcal{L}_{\mathrm{HXE}}(p, C) = -\sum_{l=0}^{h-1} \lambda(C^{(l)}) \log p(C^{(l)} | C^{(l+1)}),$$

Predicted class




Bertinetto+20, VAV+25
VAV adds additional weighting to account for class imbalance

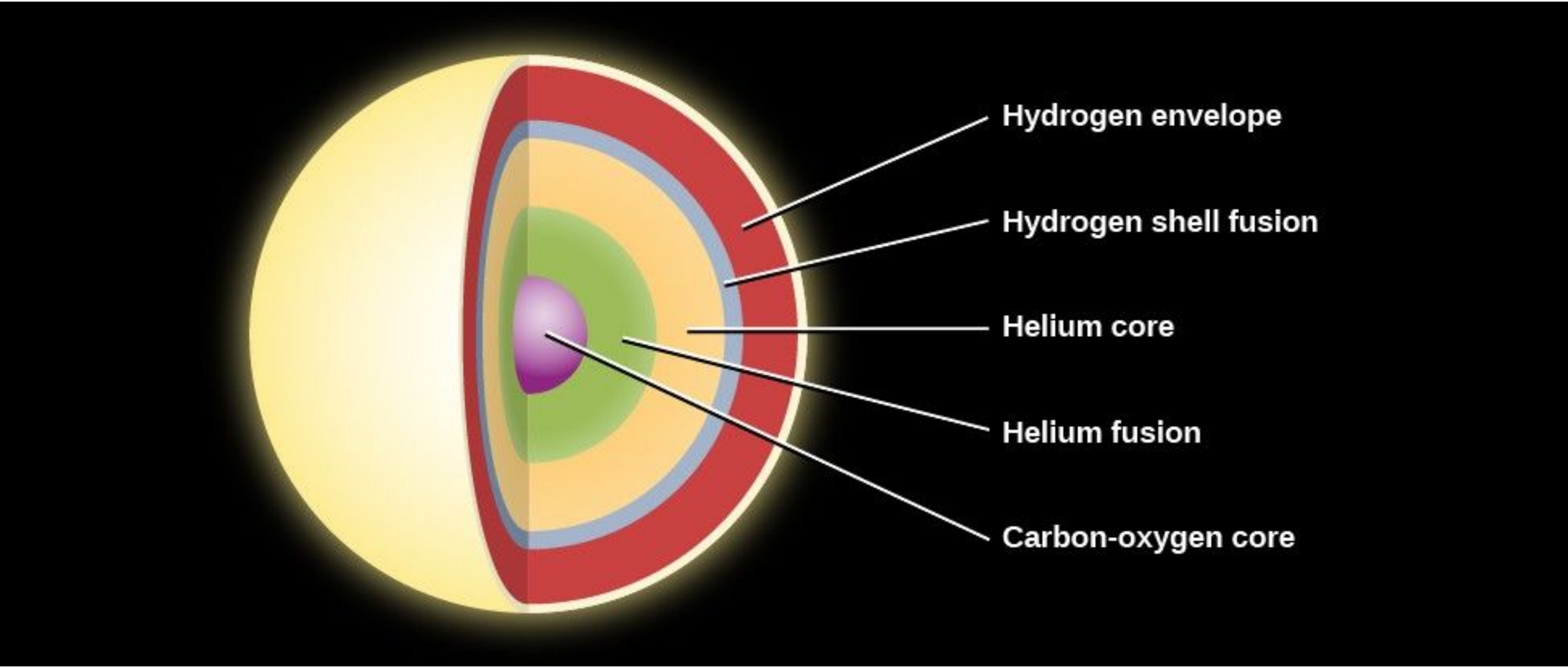
# Hierarchical Cross Entropy...

- Performs as well as flat classifiers, but not necessarily any better
- In theory, makes better ‘mistakes’ by labeling within the correct “family”
  - In practice, this is not necessarily true for the astrophysical case
- Gives notable flexibility in tasks
  - By giving access to a “node-specific” pseudo-probability, one can answer many more classification questions without retraining

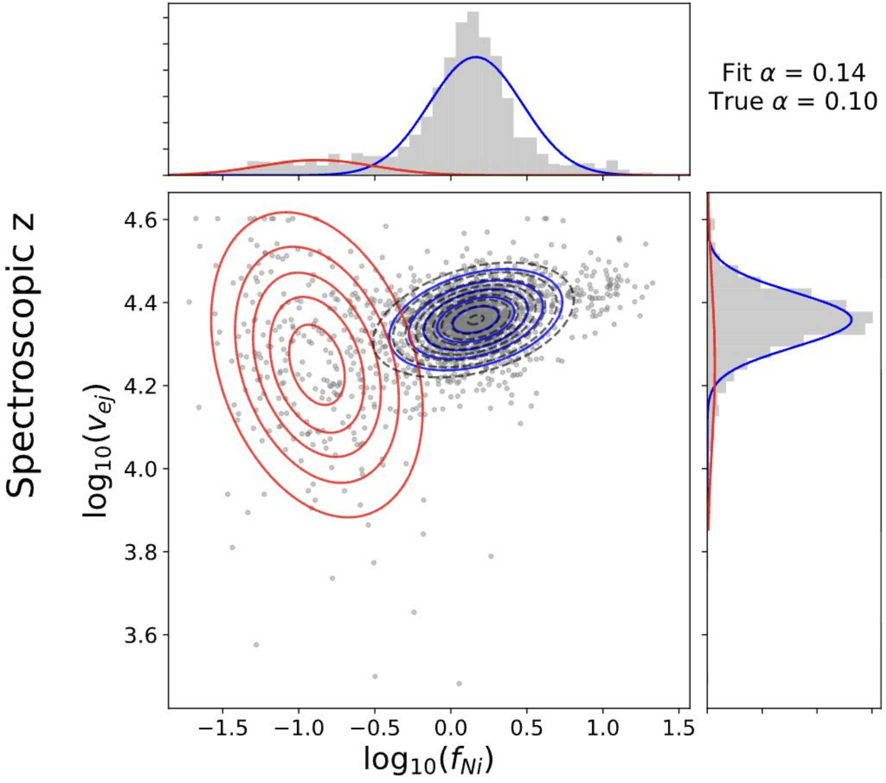
# What if we want a feature-based taxonomy?

# Particularly valuable for fuzzy labels

# E.g., some supernovae classe are clearly "fuzzy"

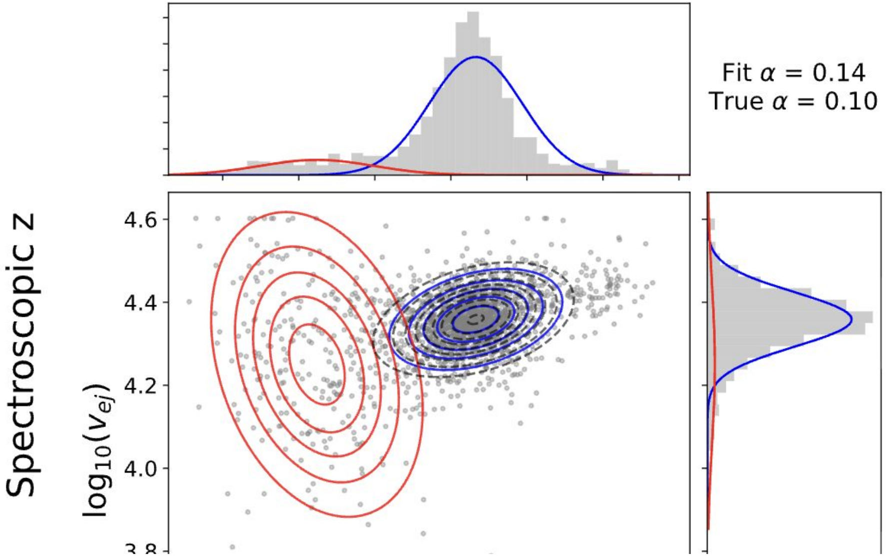


# Classification as a(n unsupervised) mixing problem




Uzsoy+VAV, sub.

# Classification as a(n unsupervised) mixing problem



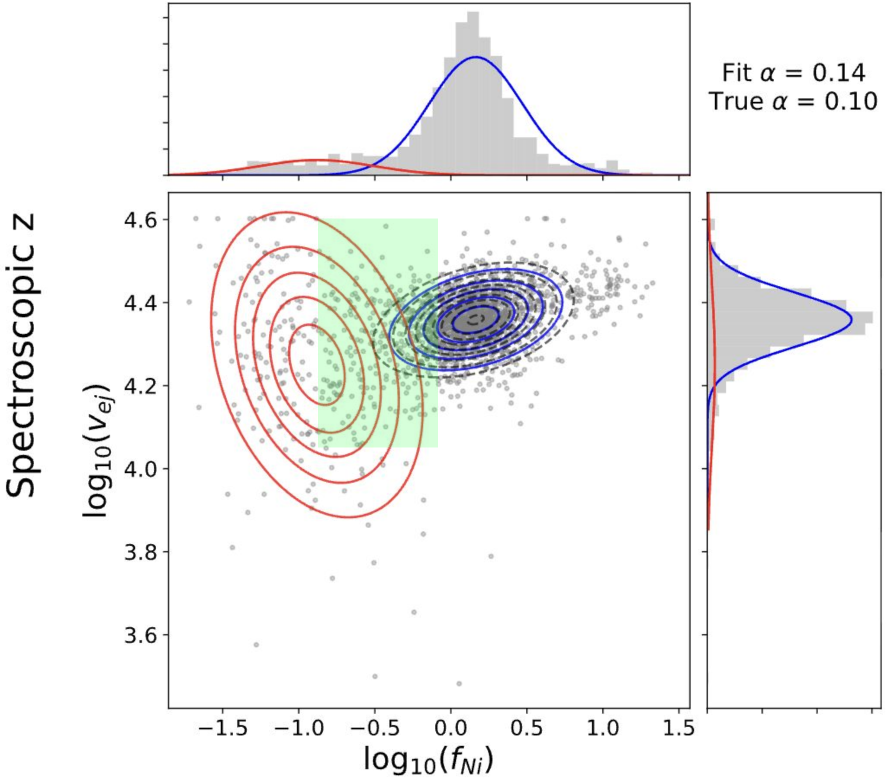

$$\mathbf{x} \sim \alpha\mathcal{N}(\boldsymbol{\mu_{Ibc}}, \boldsymbol{\Sigma}_{\mathbf{Ibc}}) + (1-\alpha)\mathcal{N}(\boldsymbol{\mu_{Ia}}, \boldsymbol{\Sigma}_{\mathbf{Ia}})$$

Uzsoy+VAV, sub.

# Classification as a mixing problem…

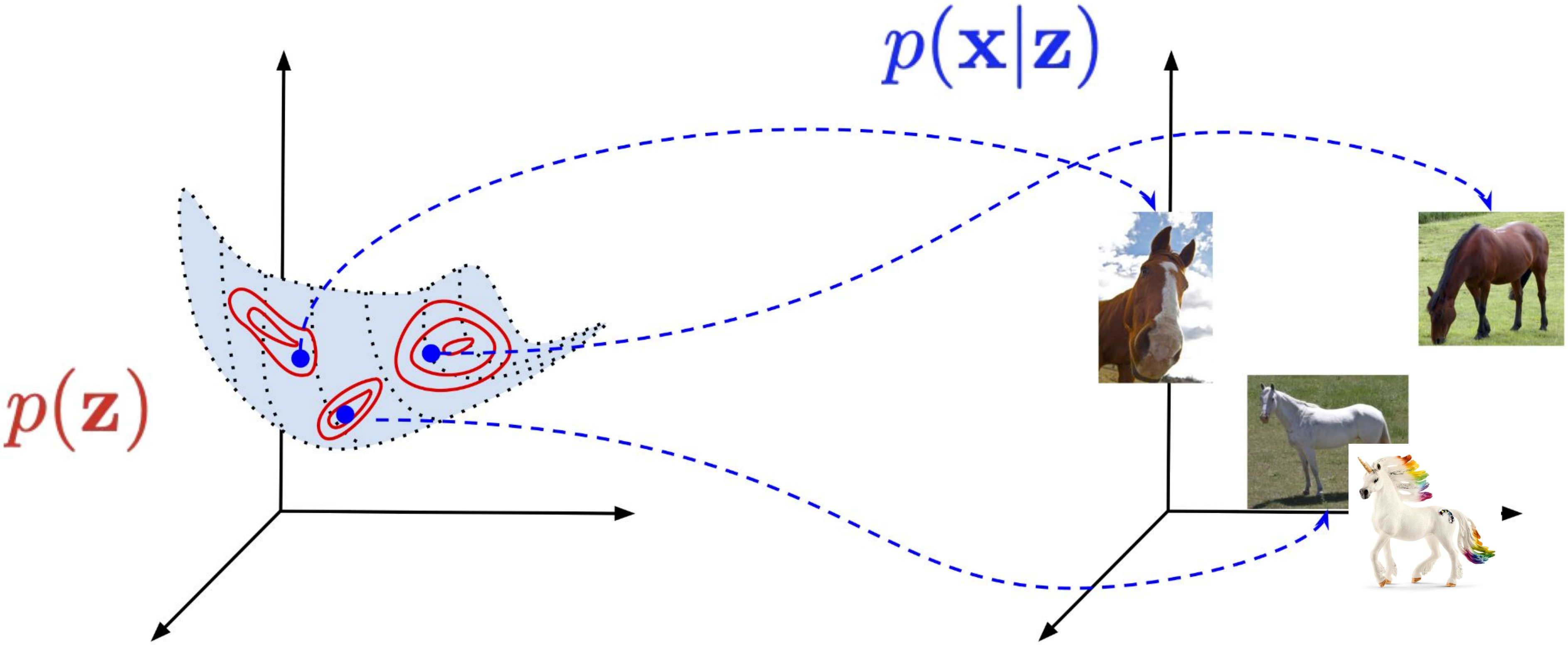
- Performs as well as classic supervised methods, with no need for expensive labeling
- Naturally finds ambiguous events…if your latent space is reasonable

# Classification as a(n unsupervised) mixing problem




Uzsoy+VAV, sub.

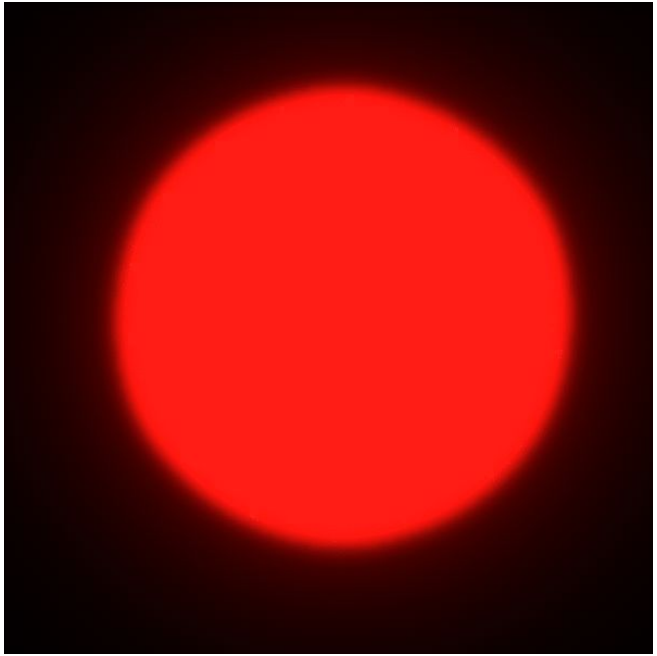
# No guarantee that my feature space captures all



# Many open problems remain

- How do we unify “fuzzy labels” with hierarchical learning in a well-structured way?
- Can we learn the true latent (aka physical) model from a large enough sample?
- How do we calibrate any of these probabilities?

# Thank you!



Title: A hierarchy of classification

Abstract: The taxonomy of physical events in the night sky (erupting stars, stellar explosions, and even gravitational lensing) is inherently hierarchical in nature. Broad physical families subdivide into increasingly specific physics, forming a tree-structured classification problem. Standard classification objectives typically ignore said structure and treat all misclassifications equally, rather than encouraging classifiers to "make better mistakes". In this talk, I will discuss a few approaches to modeling hierarchical class structure within the objective function. I will also discuss open issues of "fuzzy boundaries" within this taxonomy, in which both discrete and continuous subclasses exist.

20 min