Hierarchical Bayesian Inference + Probabilistic Graphical Models

Daniela Huppenkothen

DIRAC Institute, University of Washington

Nature is complex!

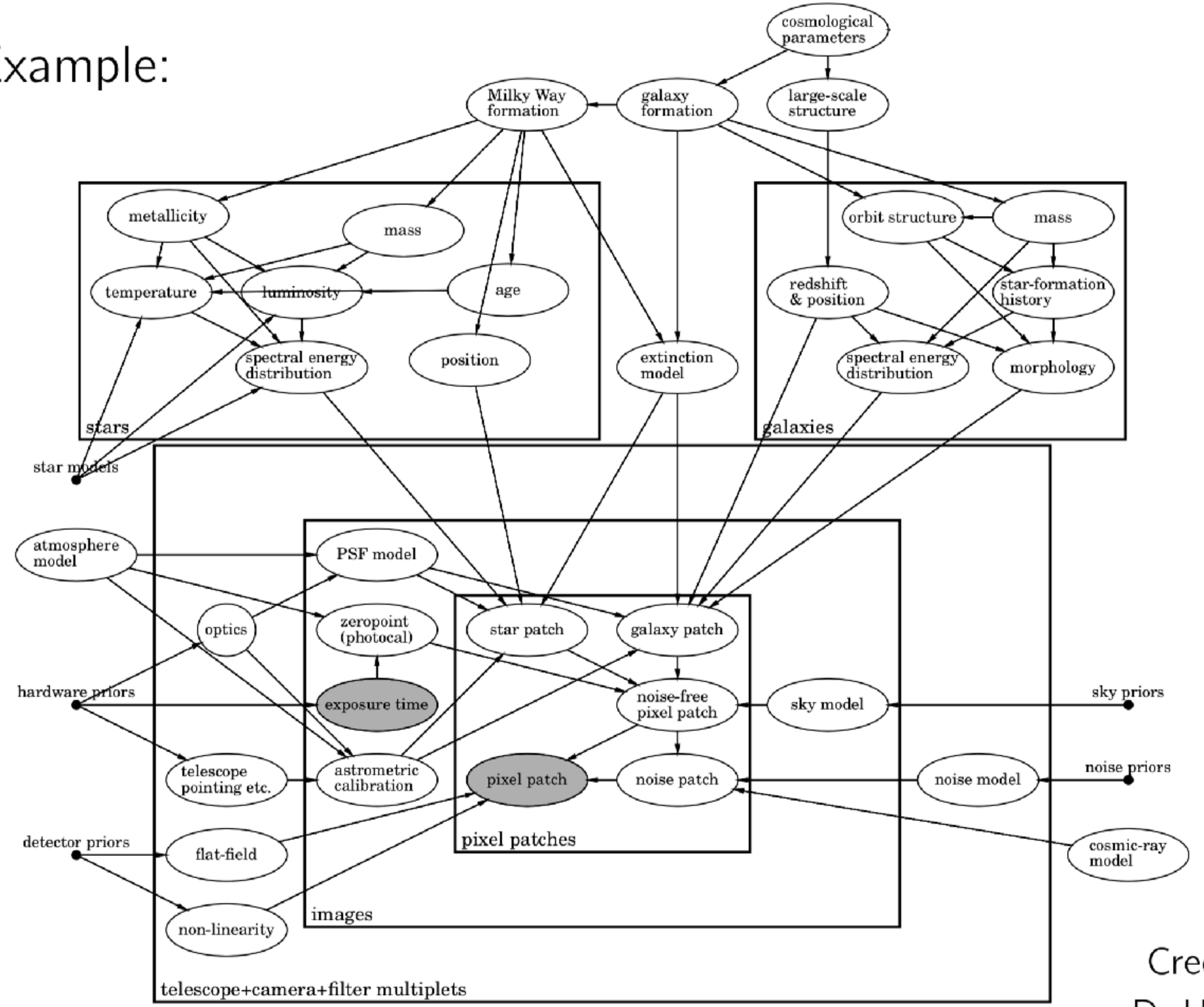
https://github.com/dhuppenkothen/cargese2018_tutorials

... so is our data (collection)!

All models are **wrong**,
but some are **useful**.

— George Box

Example:



Credit:
D. Hogg

A quick census …

A quick census …

- I have heard of Bayes theorem

A quick census …

- I have heard of Bayes theorem
- I have used Bayes theorem in research

A quick census ...

- I have heard of Bayes theorem
- I have used Bayes theorem in research
- I have used Bayesian hierarchical models in research

A quick census …

- I have heard of Bayes theorem
- I have used Bayes theorem in research
- I have used Bayesian hierarchical models in research
- I have heard of machine learning

A quick census …

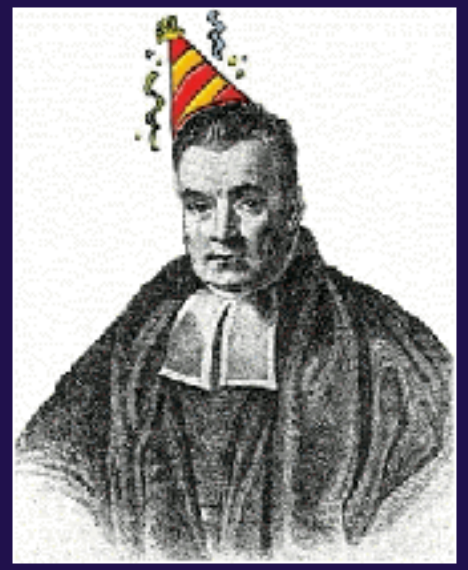
- I have heard of Bayes theorem
- I have used Bayes theorem in research
- I have used Bayesian hierarchical models in research
- I have heard of machine learning
- I have used machine learning in research

A quick census ...

- I have heard of Bayes theorem
- I have used Bayes theorem in research
- I have used Bayesian hierarchical models in research
- I have heard of machine learning
- I have used machine learning in research
- I have written code in Python before

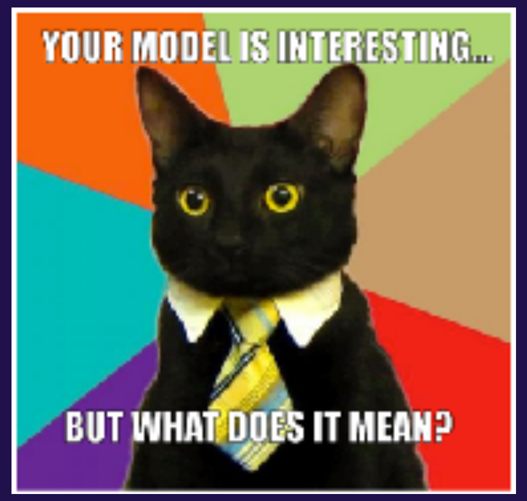
This week

(1) Fun With Bayes(ian Hierarchical Models)



(2) Machine Learning

(3) Statistical Machine Learning



Is it sunny today?

p(☀️)

Is it sunny today?

	☀️	🌧️
🏝️	0.98	0.01
🏔️	0.004	0.006

1) p(☀️,🏝️) = ???
2) p(☀️|🏝️) = ???
3) p(🏔️) = ???
4) p(🌧️|🏔️) = ???

Is it sunny today?

	☀️	🌧️
🏝️	0.98	0.01
🏔️	0.004	0.006

1) p(☀️,🏝️) =
2) p(☀️|🏝️) =
3) p(🏔️) =
4) p(🌧️|🏔️) =

Is it sunny today?

	☀️	🌧️
🏝️	0.98	0.01
🏔️	0.004	0.006

1) p(☀️,🏝️) = 0.98
2) p(☀️|🏝️) =
3) p(🏔️) =
4) p(🌧️|🏔️) =

Is it sunny today?

	☀️	🌧️
🏝️	0.98	0.01
🏔️	0.004	0.006

1) p(☀️,🏝️) = 0.98
2) p(☀️|🏝️) = 0.9898
3) p(🏔️) =
4) p(🌧️|🏔️) =

Is it sunny today?

	☀️	🌧️
🏝️	0.98	0.01
🏔️	0.004	0.006

1) p(☀️,🏝️) = 0.98
2) p(☀️|🏝️) = 0.9898
3) p(🏔️) = 0.01
4) p(🌧️|🏔️) =

Is it sunny today?

	☀️	🌧️
🏝️	0.98	0.01
🏔️	0.004	0.006

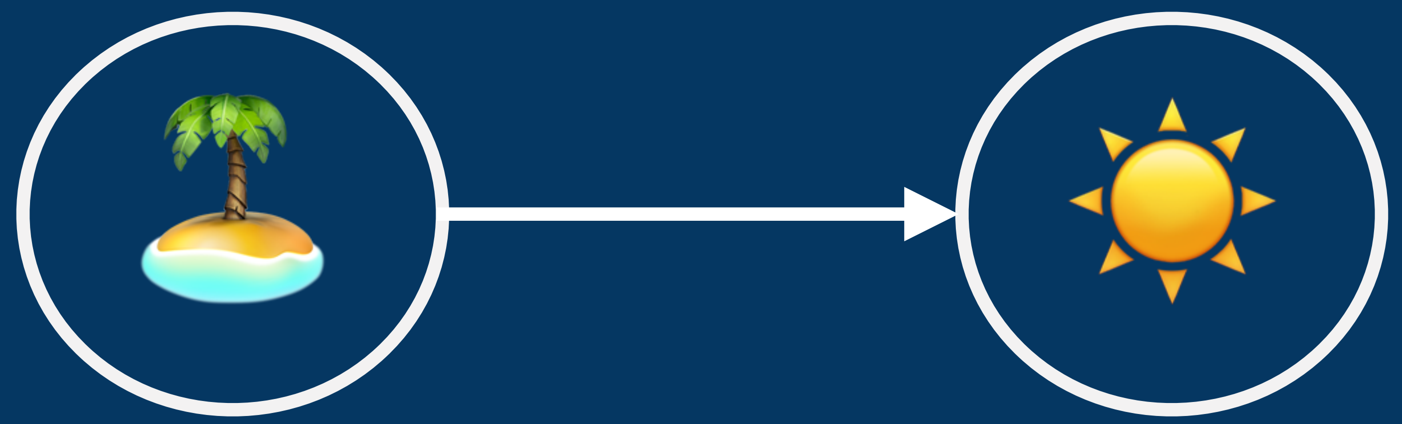
1) p(☀️,🏝️) = 0.98
2) p(☀️|🏝️) = 0.9898
3) p(🏔️) = 0.01
4) p(🌧️|🏔️) = 0.6

“The weather depends on your location”

p(☀️,🏝️) = p(☀️|🏝️)p(🏝️)

"The weather depends on your location"

p(☀️,🏝️) = p(☀️|🏝️)p(🏝️)

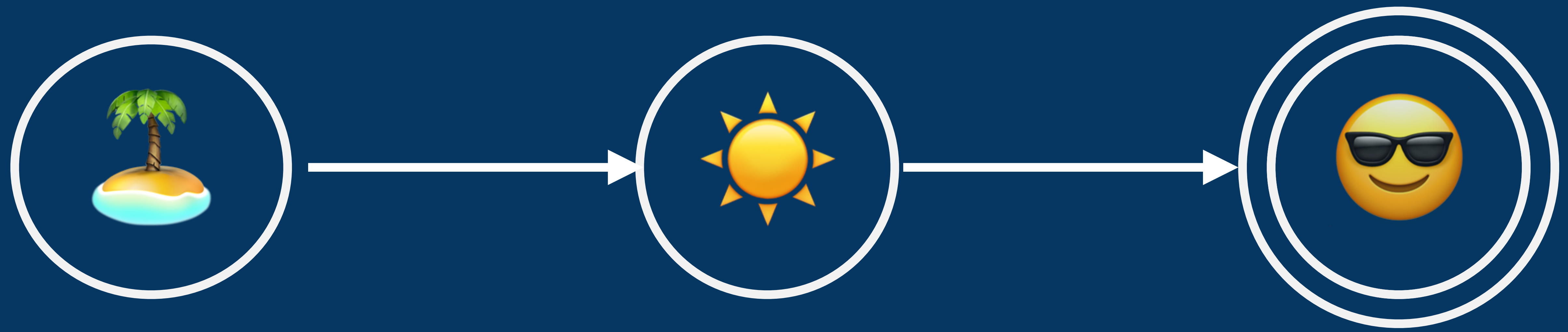


Let's make this more complicated …

p(😎 , ☀️ , 🏝️) = p(😎 | ☀️ , 🏝️)p(☀️ | 🏝️)p(🏝️)

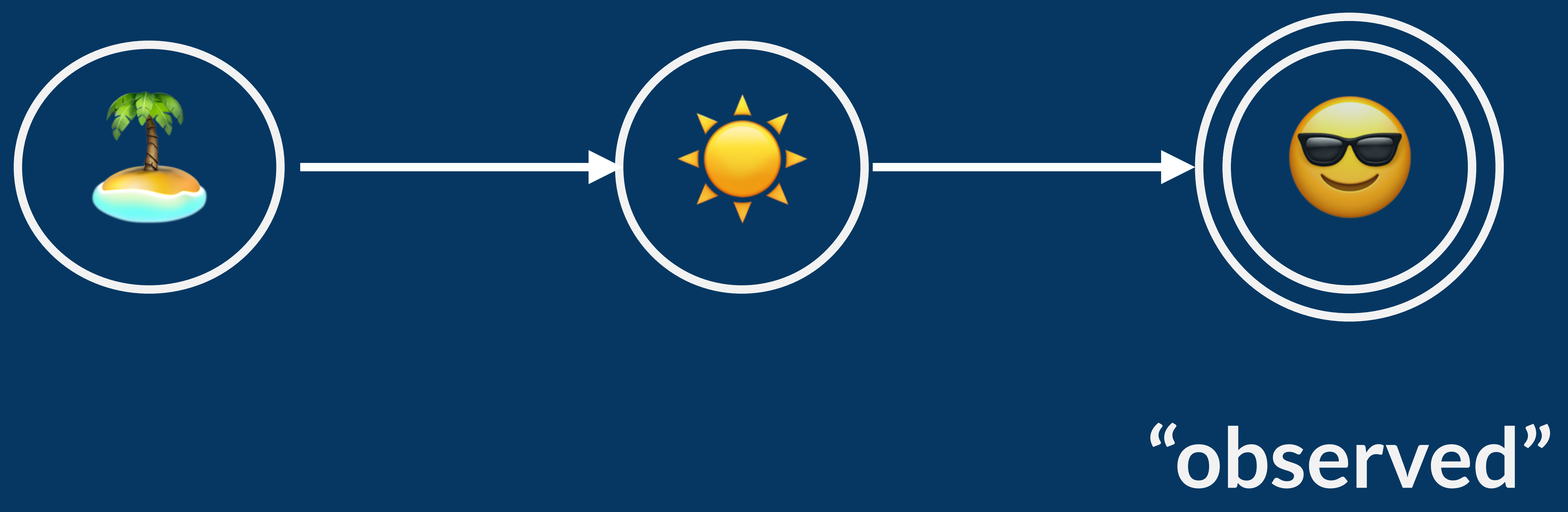
Let's make this more complicated ...

p(😎 , ☀️ , 🏝️) = p(😎 | ☀️ , 🏝️)p(☀️ | 🏝️)p(🏝️)



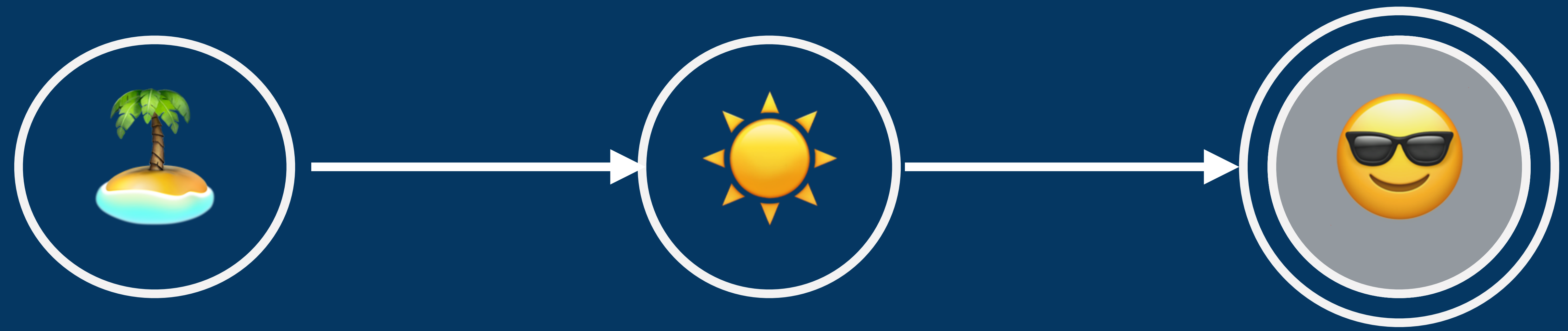
Let’s make this more complicated …

p(😎 , ☀️ , 🏝️) = p(😎 | ☀️ , 🏝️)p(☀️ | 🏝️)p(🏝️)



Let’s make this more complicated …

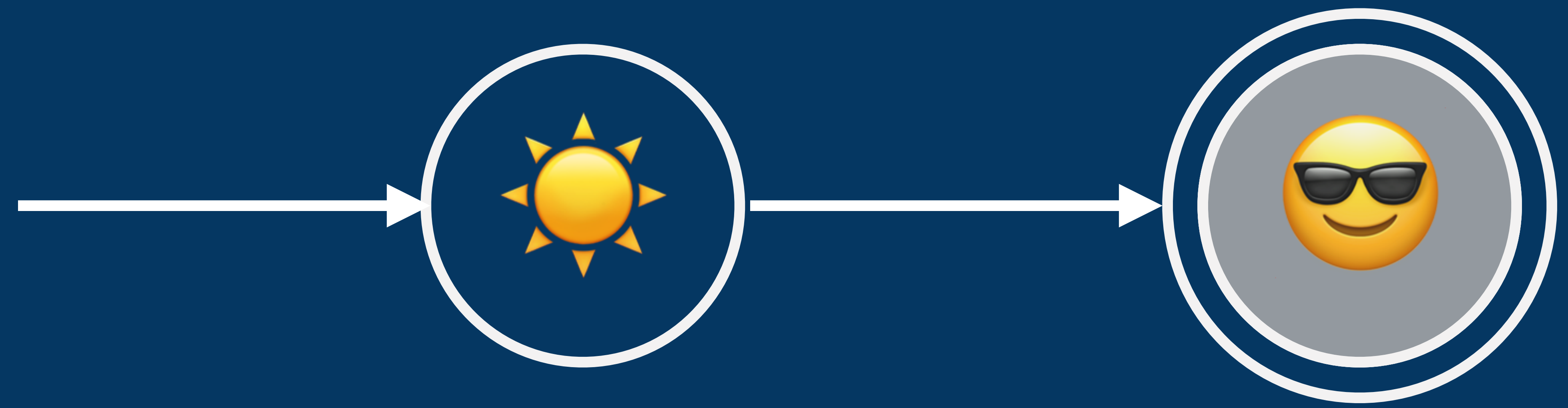
p(😎 , ☀️ , 🏝️) = p(😎 | ☀️ , 🏝️)p(☀️ | 🏝️)p(🏝️)



“observed”

Let’s make this more complicated …

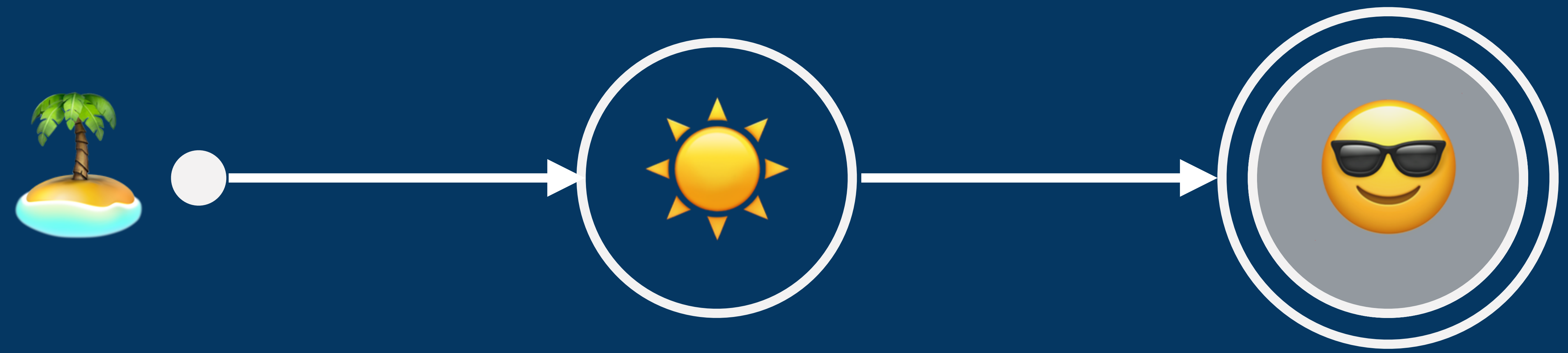
p(😎, ☀️, 🏝️) = p(😎 | ☀️, 🏝️)p(☀️ | 🏝️)p(🏝️)



“observed”

Let’s make this more complicated …

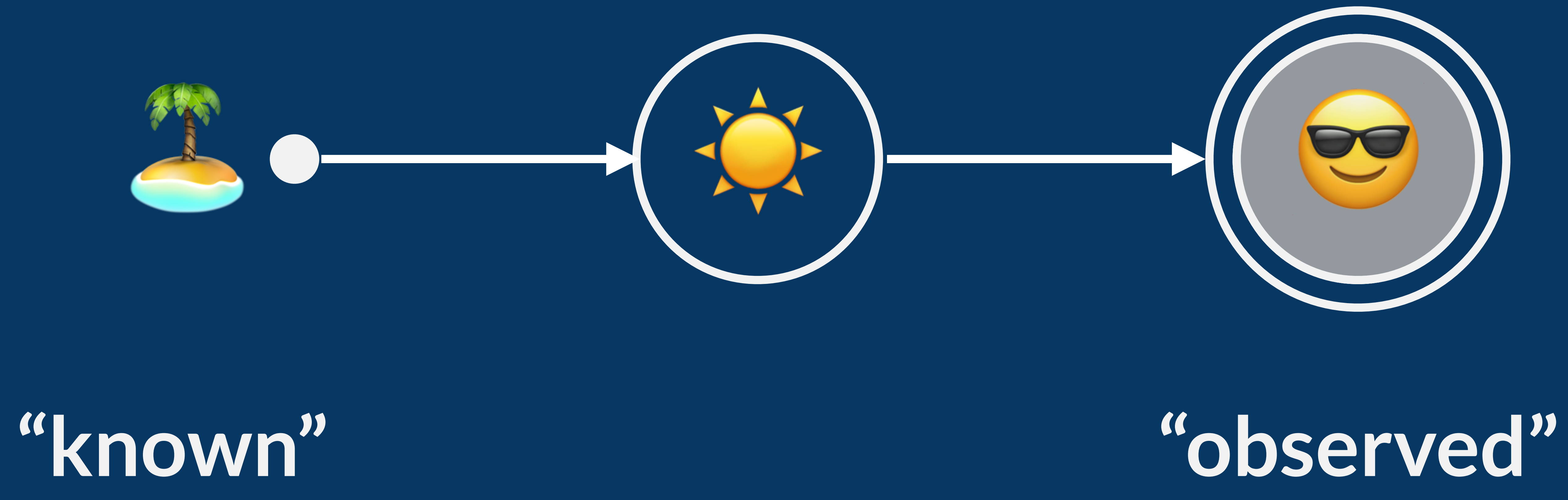
p(😎 , ☀️ , 🏝️) = p(😎 | ☀️ , 🏝️)p(☀️ | 🏝️)p(🏝️)



“observed”

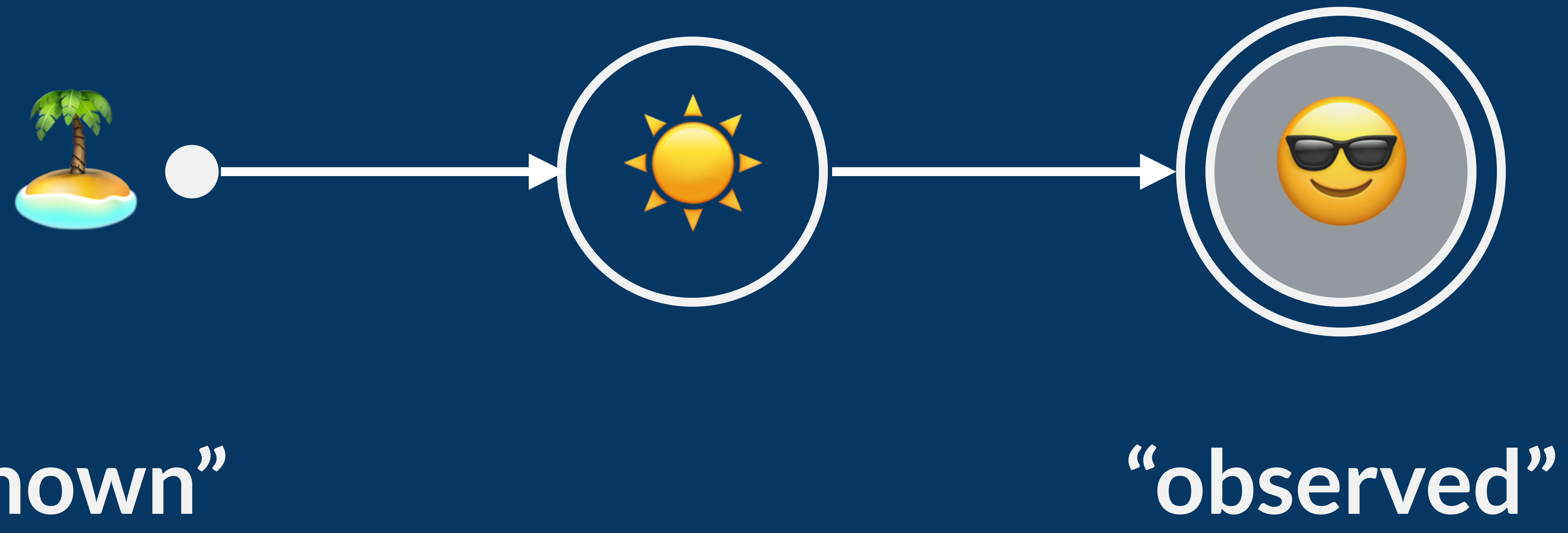
Let’s make this more complicated …

p(😎 , ☀️ , 🏝️) = p(😎 | ☀️ , 🏝️)p(☀️ | 🏝️)p(🏝️)



Let’s make this more complicated …

p(😎 , ☀️ , 🏝️) = p(😎 | ☀️ , 🏝️)p(☀️ | 🏝️)p(~~🏝️~~)



What else does 😎 depend on?

⏰ Is it daytime?
🏡 are you outside?
🌡️ what's the temperature?

Exercise: Add these variables to a graphical network!

What else does 👀 depend on?

⏰ Is it daytime?
🏡 are you outside?
🌡️ what's the temperature?

p(😎 , ☀️ , 🏝️ , ⏰ , 🏡 , 🌡️) = p(😎 | ☀️ , 🏡) p(☀️ | 🏝️ , ⏰)
x p(🏡 | 🌡️) p(🏝️) p(🌡️) p(⏰)

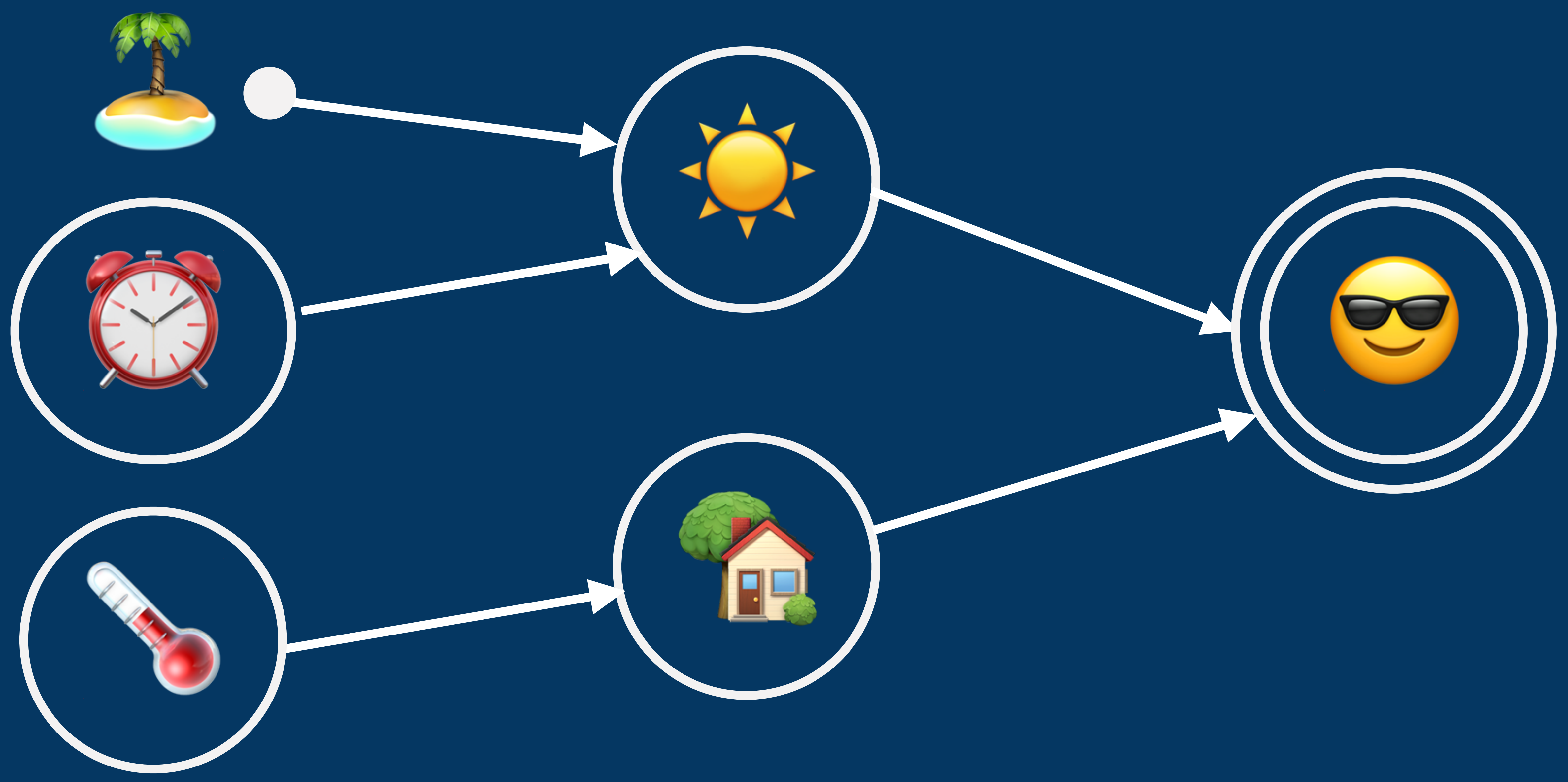
What else does 👀 depend on?

p(😎 , ☀️ , 🏝️ , ⏰ , 🏡 , 🌡️) = p(😎 | ☀️ , 🏡) p(☀️ | 🏝️ , ⏰)
x p(🏡 | 🌡️) p(🌡️) p(⏰)

What else does 👀 depend on?

p(😎, ☀️, 🏝️, ⏰, 🏡, 🌡️) = p(😎 | ☀️, 🏡) p(☀️ | 🏝️, ⏰)

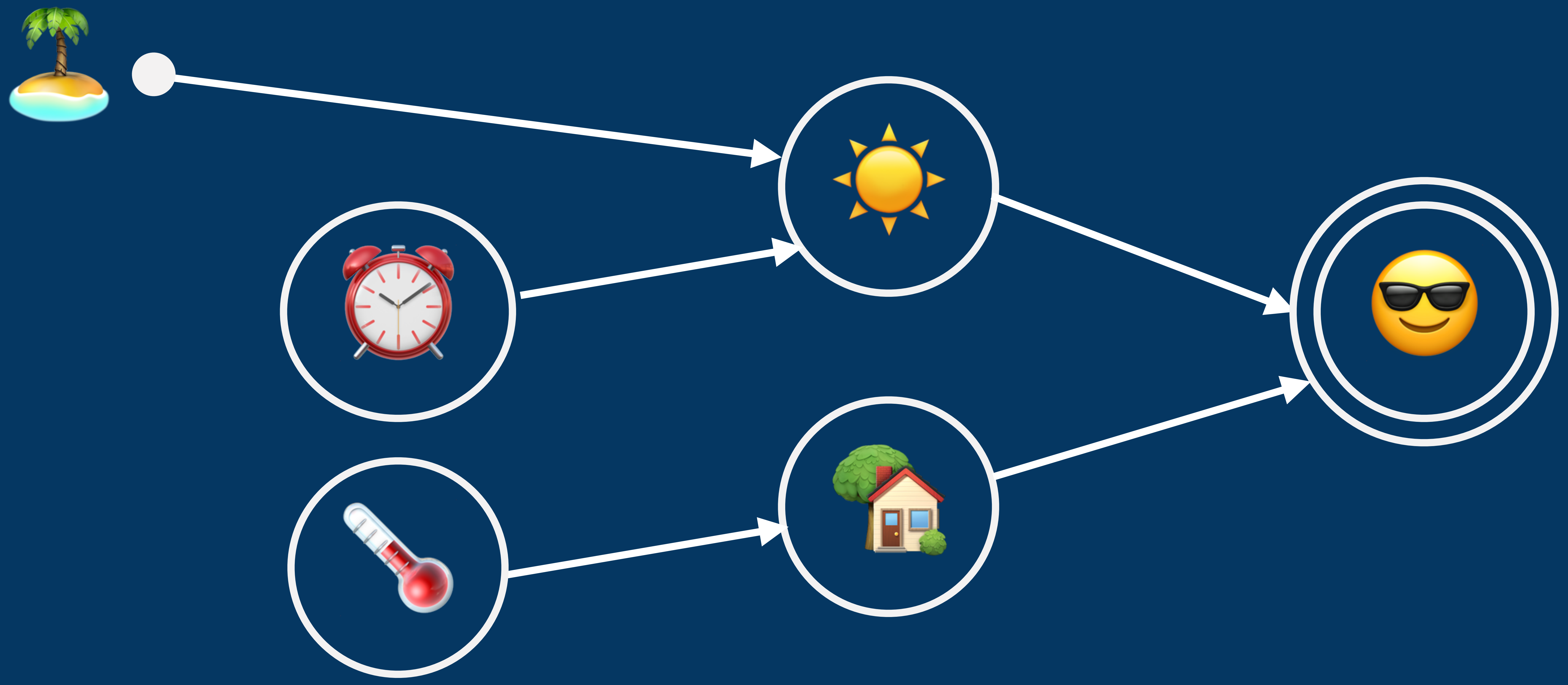
x p(🏡 | 🌡️) p(🌡️) p(⏰)



Repeated Observations

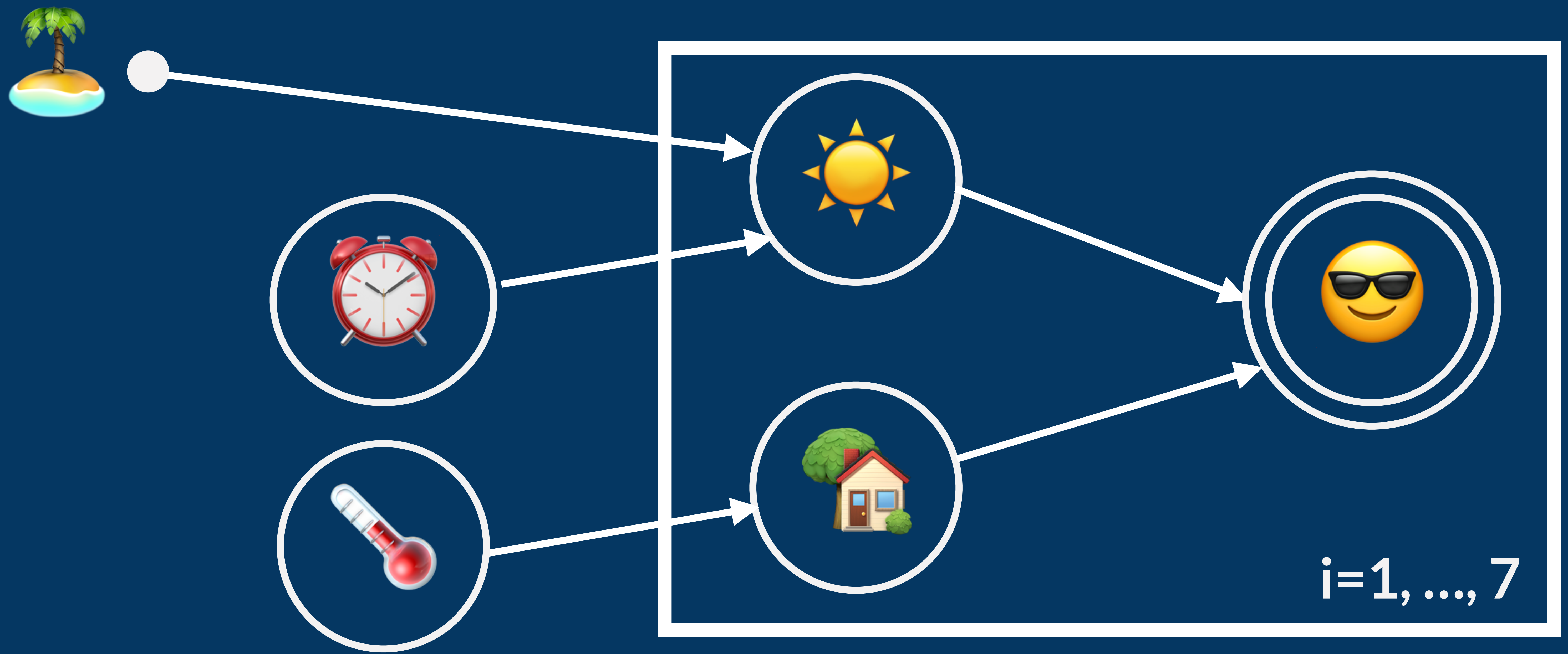
p(😎 , ☀️ , 🏝️ , ⏰ , 🏡 , 🌡️) = $\prod_{i=1}^{7}$ p(😎 | ☀️ , 🏡) p(☀️ | 🏝️ , ⏰)

x p(🏡 | 🌡️) p(🌡️) p(⏰)



Repeated Observations

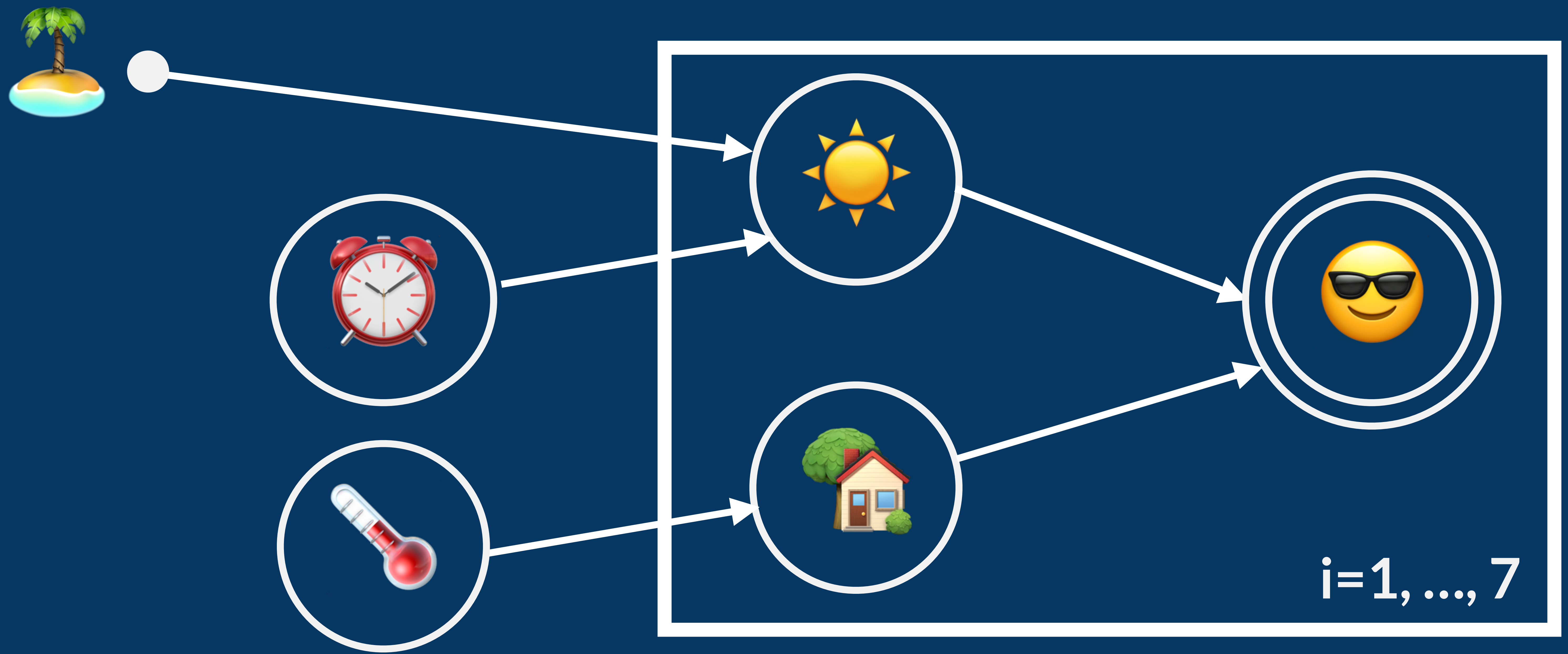
$$p(\text{😎}, \text{☀️}, \text{🏝️}, \text{⏰}, \text{🏡}, \text{🌡️}) = \prod_{i=1}^{7} p(\text{😎} \mid \text{☀️}, \text{🏡})\, p(\text{☀️} \mid \text{🏝️}, \text{⏰}) \times p(\text{🏡} \mid \text{🌡️})\, p(\text{🌡️})\, p(\text{⏰})$$



Repeated Observations

p(😎, ☀️, 🏝️, ⏰, 🏡, 🌡️) = ∏ᵢ₌₁⁷ p(😎 | ☀️, 🏡) p(☀️ | 🏝️, ⏰)

x p(🏡 | 🌡️) p(🌡️) p(⏰)



“repeated variables”

“The weather depends on your location”

p(☀️,🏝️) = p(☀️|🏝️)p(🏝️)

= p(🏝️|☀️)p(☀️)

“The weather depends on your location”

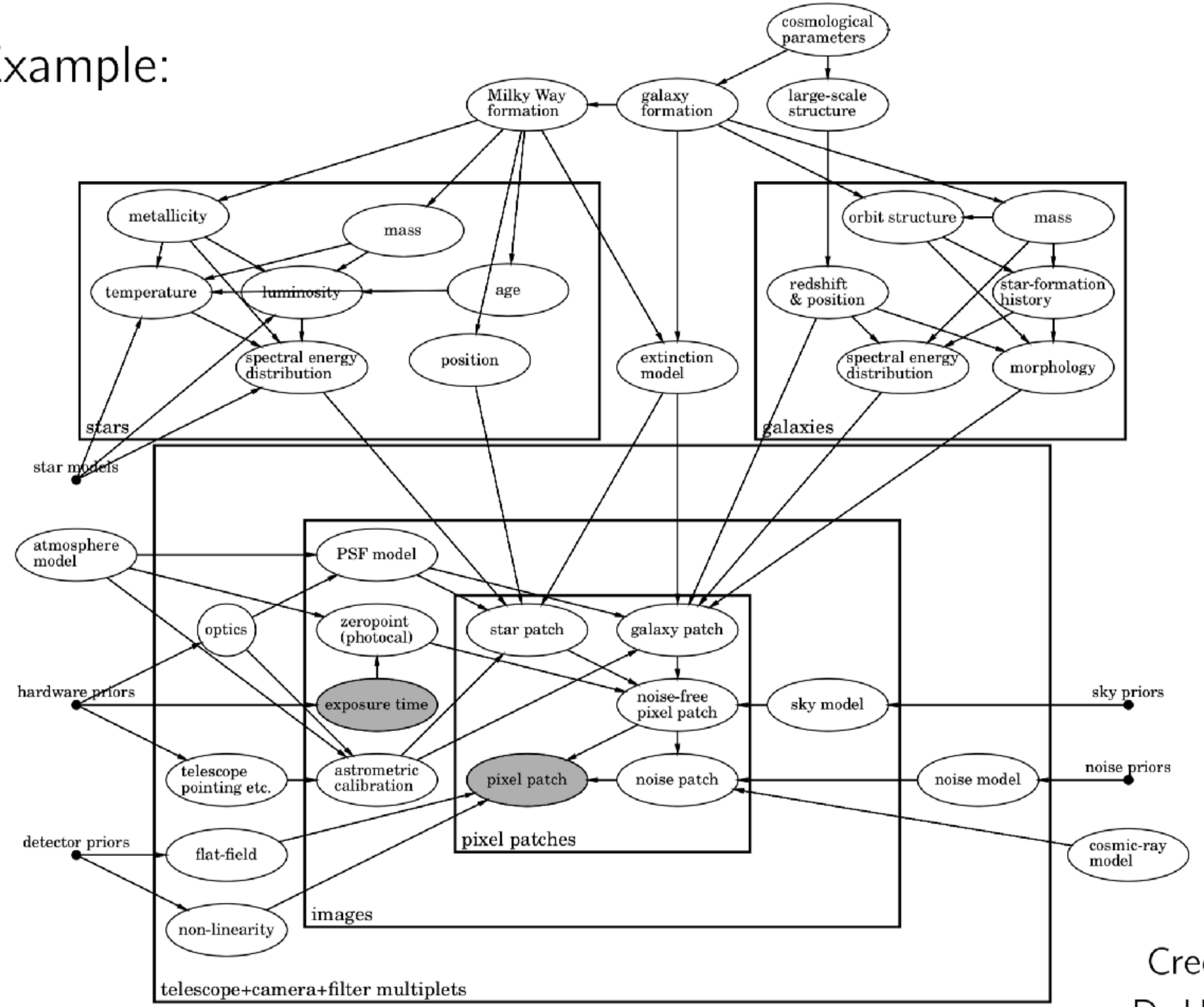
p(☀️,🏝️) = p(☀️|🏝️)p(🏝️)

= p(🏝️|☀️)p(☀️)

p(😎) *p(☀️ , 🏝️ , ⏰ , 🏡 , 🌡️ | 😎) = $\prod_{i=1}^{7}$ p(😎 | ☀️ , 🏡) p(☀️ | 🏝️ , ⏰)

x p(🏡 | 🌡️) p(🌡️) p(⏰)

Example:



Credit:
D. Hogg

Exercise

Write down a graphical model for the toy cosmological parameter inference exercise from the Bayesian statistics session

Alternative Exercise

Write down a graphical model for the probability of catching a cold. What different factors does that probability depend on? What variables should you take into account? What do they in turn depend on?

Bayesian Hierarchical Models

We have shown that we can write down arbitrarily complex probability distributions …

p(😎) p(☀️ , 🏝️ , ⏰ , 🏡 , 🌡️ | 😎) = $\prod_{i=1}^{7}$ p(😎 | ☀️ , 🏡) p(☀️ | 🏝️ , ⏰)

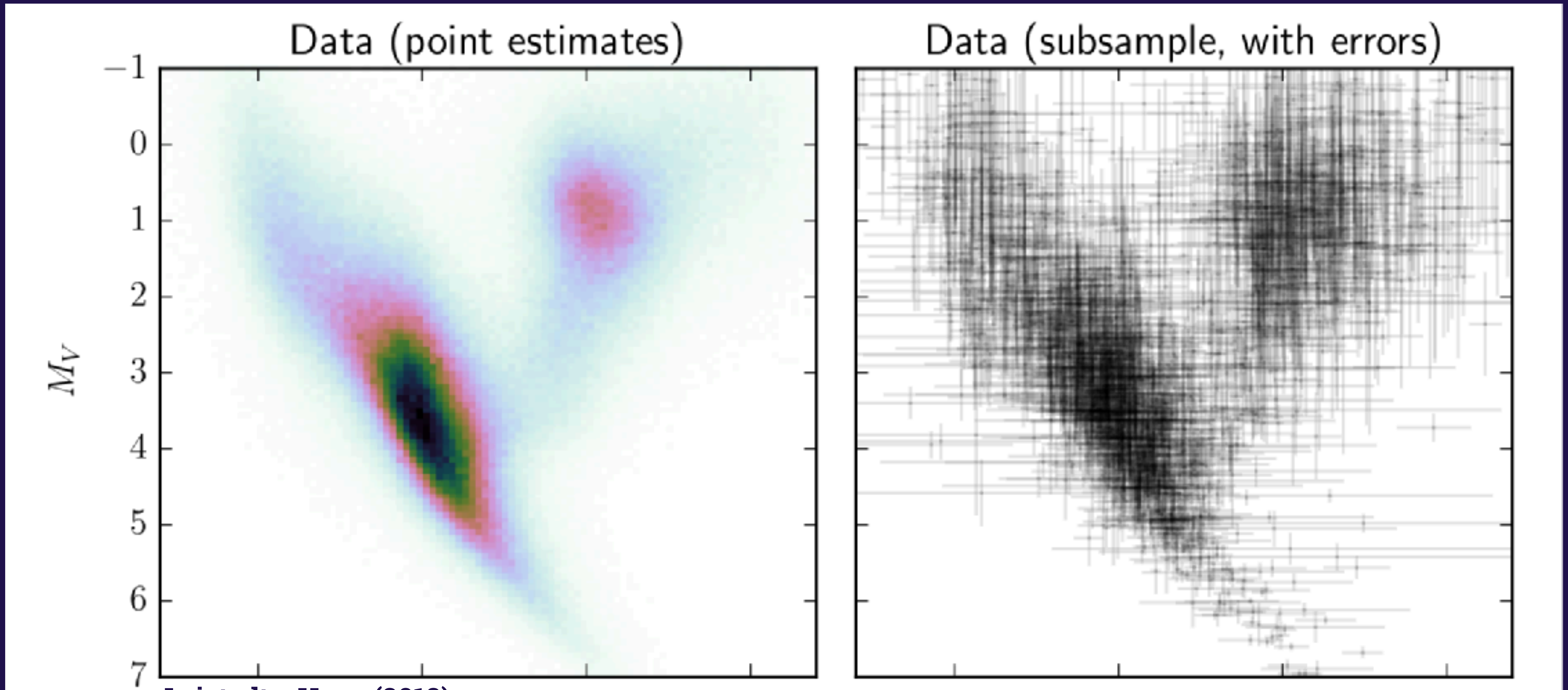
x p(🏡 | 🌡️) p(🌡️) p(⏰)

We have shown that we can write down arbitrarily complex probability distributions …

p(😎) p(☀️ , 🏝️ , ⏰ , 🏡 , 🌡️ | 😎) = $\prod_{i=1}^{7}$ p(😎 | ☀️ , 🏡) p(☀️ | 🏝️ , ⏰)
x p(🏡 | 🌡️) p(🌡️) p(⏰)

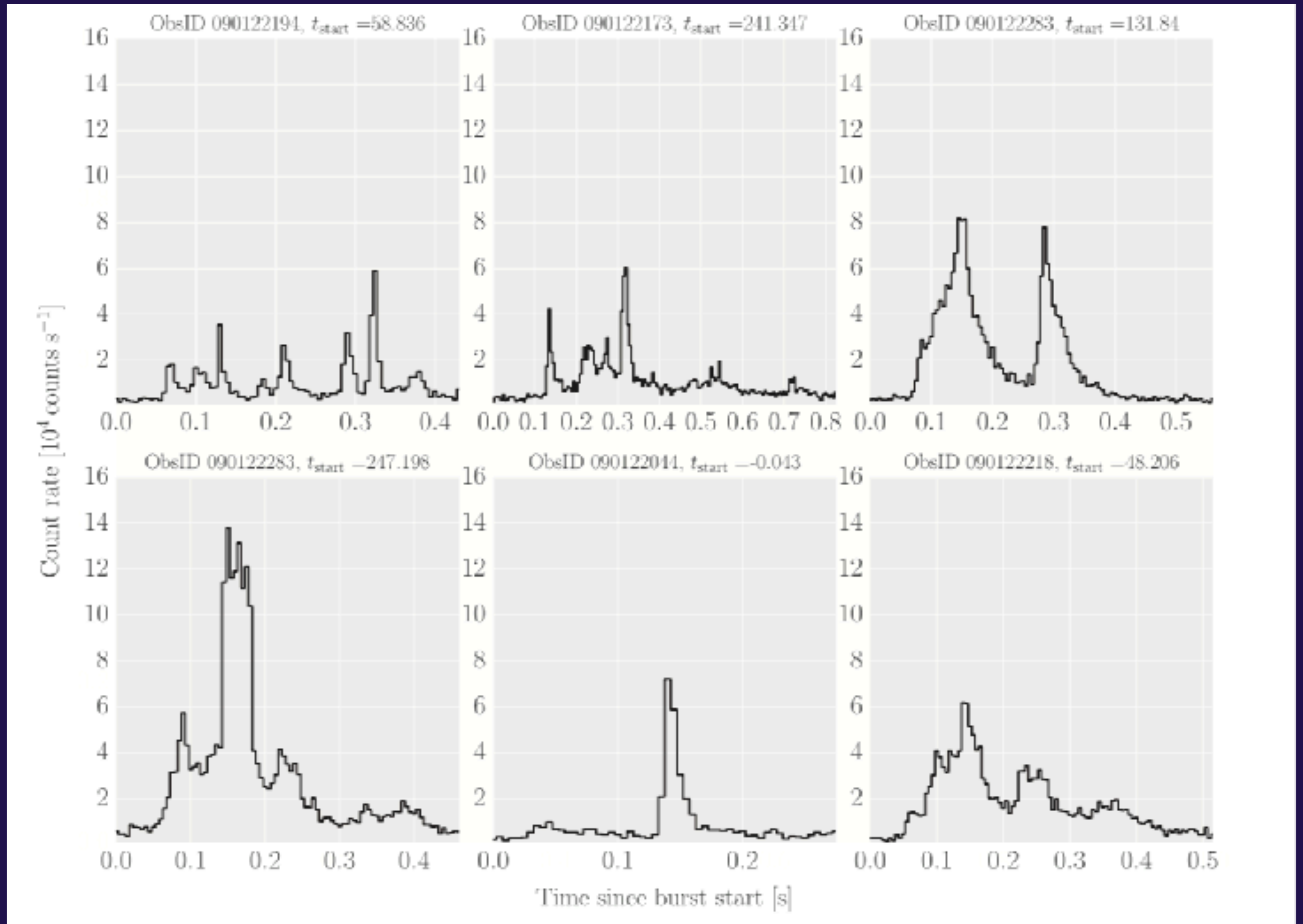
… now what?

Might have many objects ...



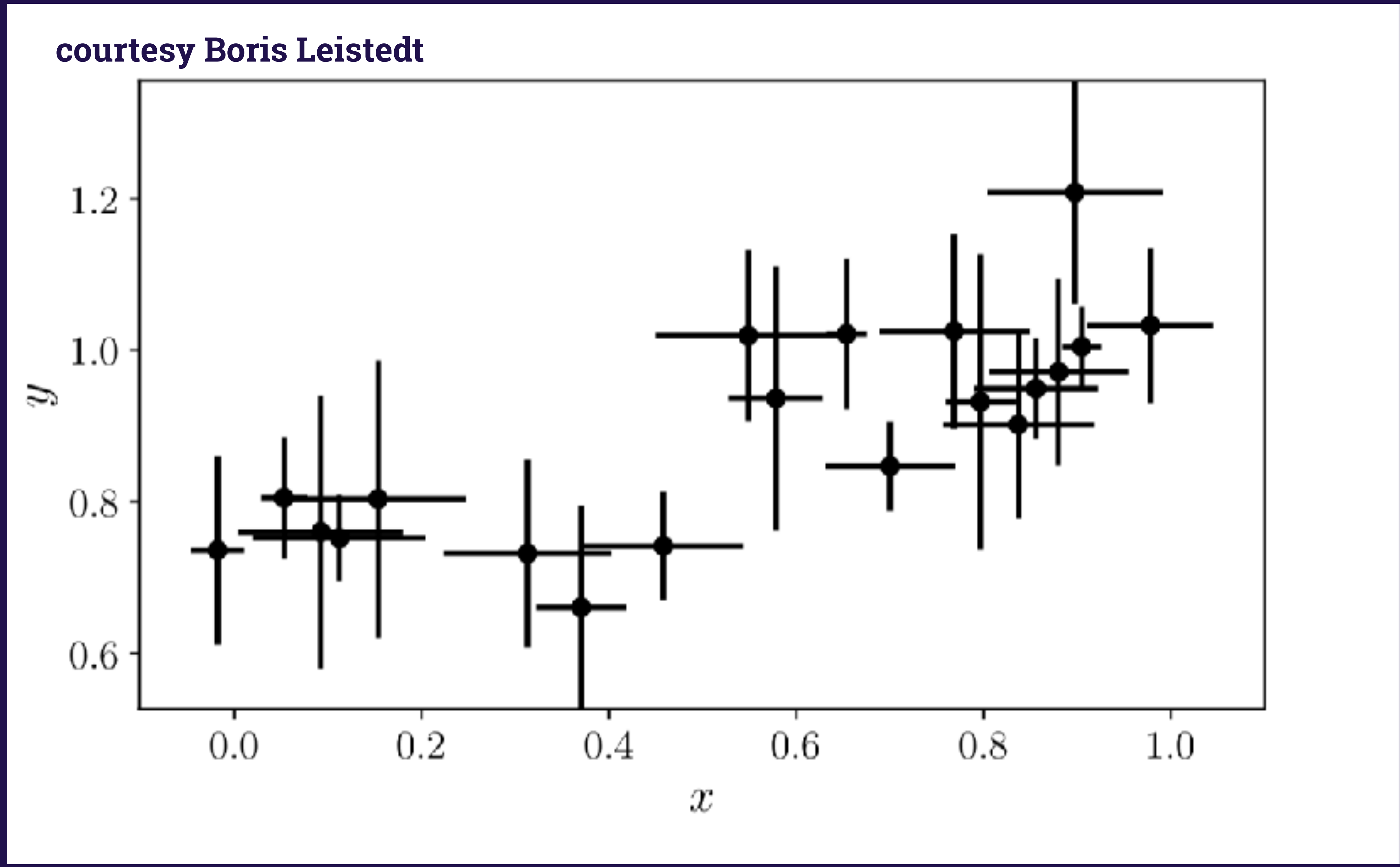
Leistedt + Hogg (2018)

… or many observations per object

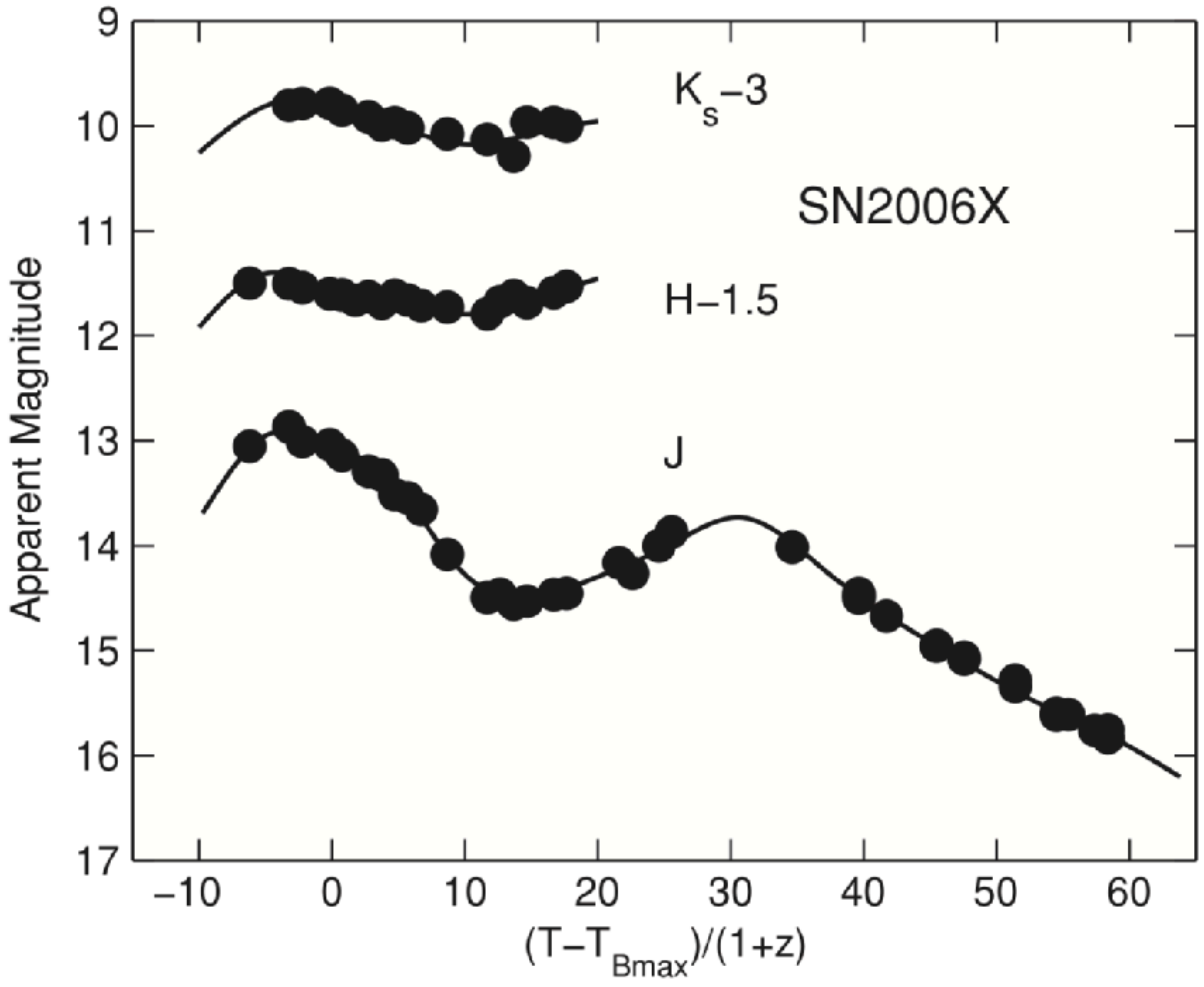


Huppenkothen et al (2015)

... or several types of uncertainties

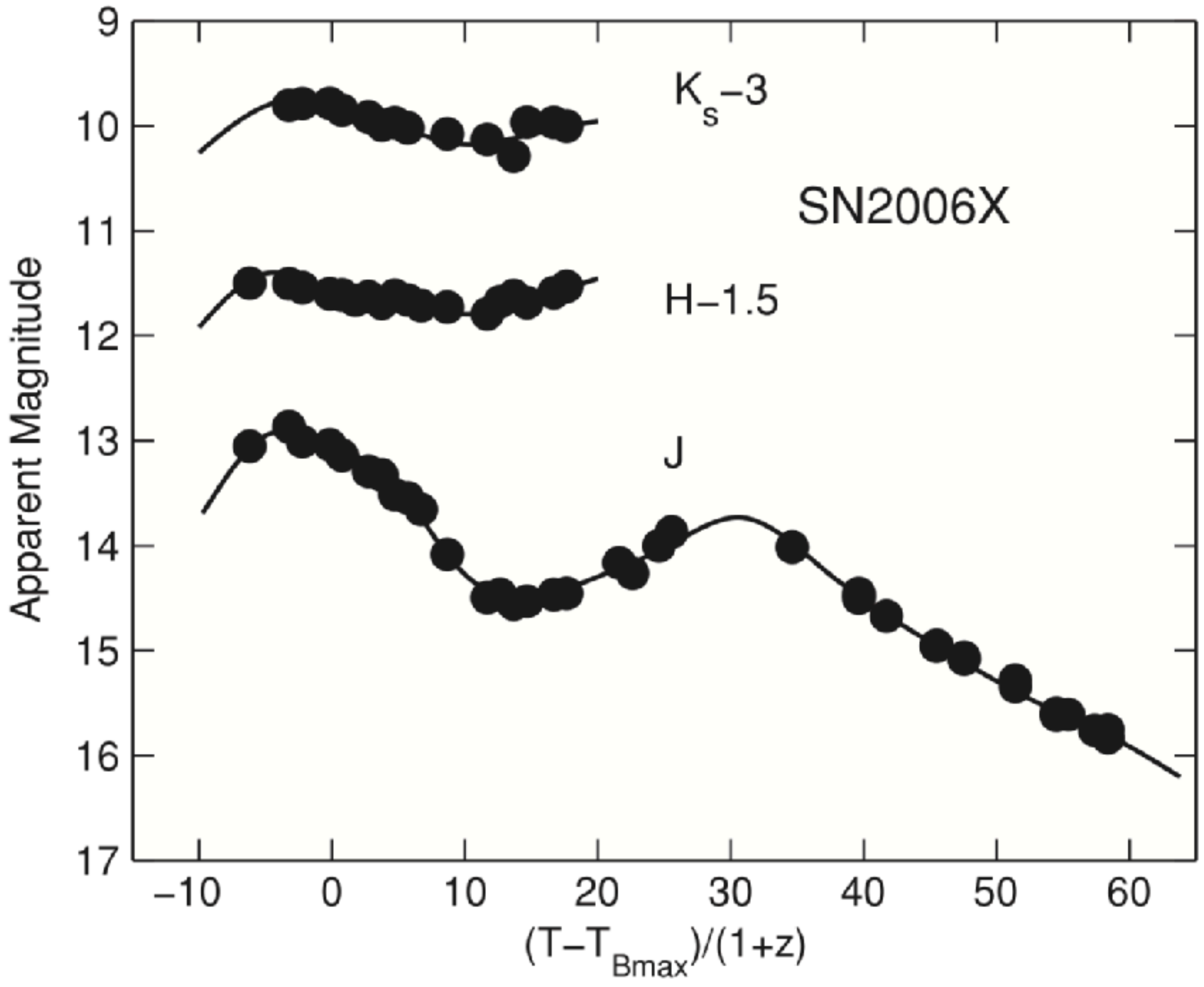


A motivating example ...



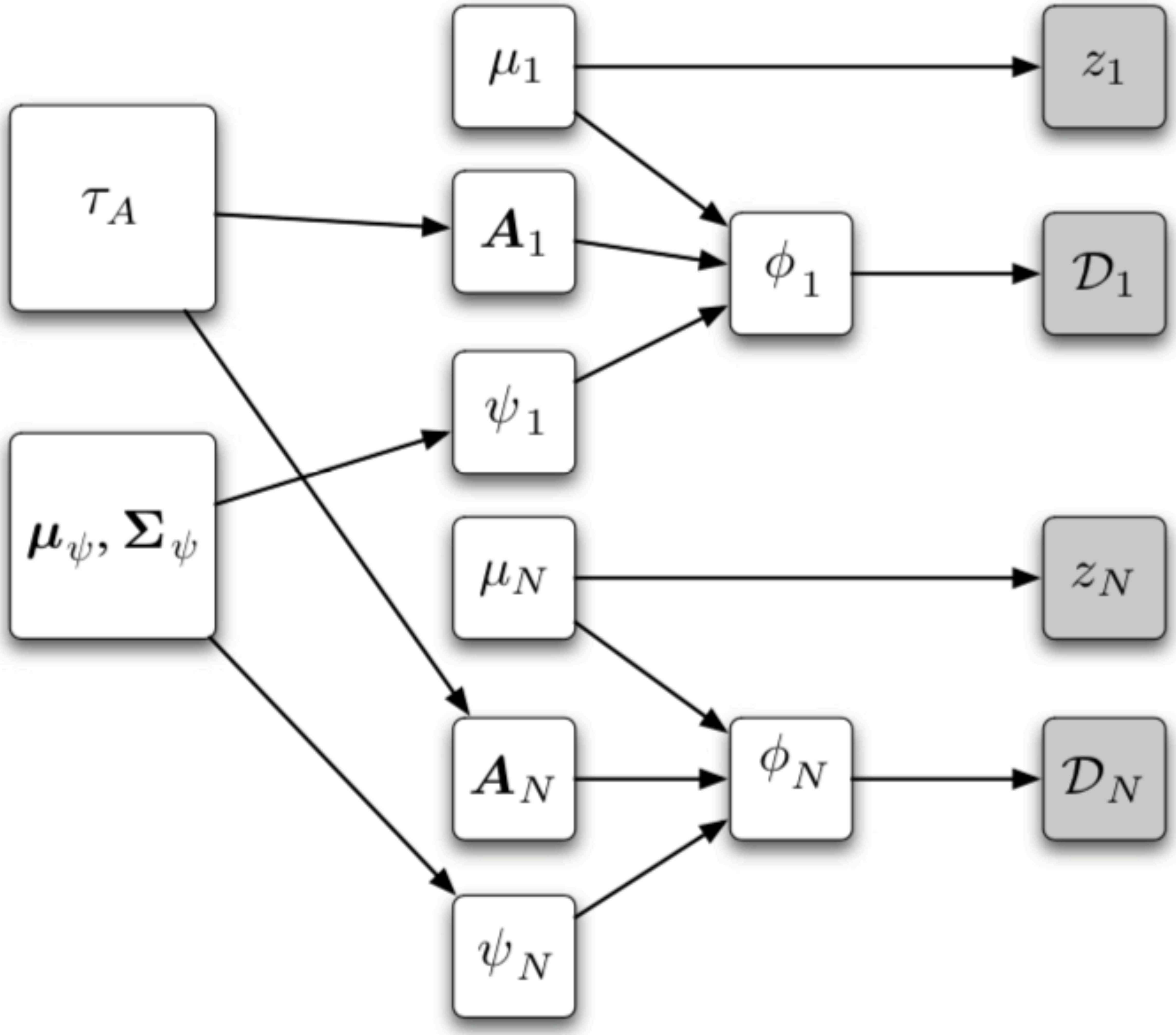
Mandel et al (2009)

SN Ia Light Curves

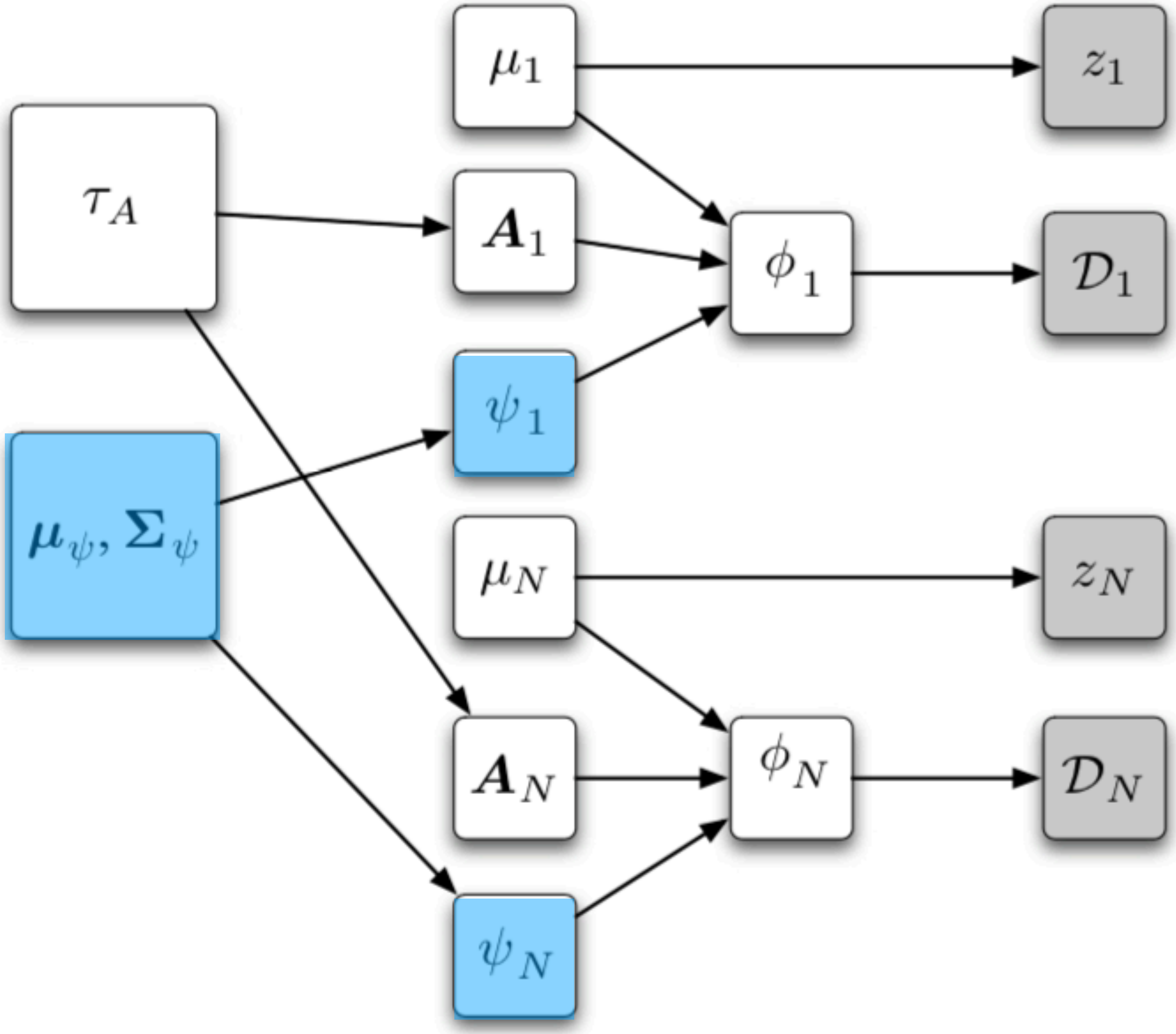


Mandel et al (2009)

What observables and parameters do you have in SN Ia light curves?

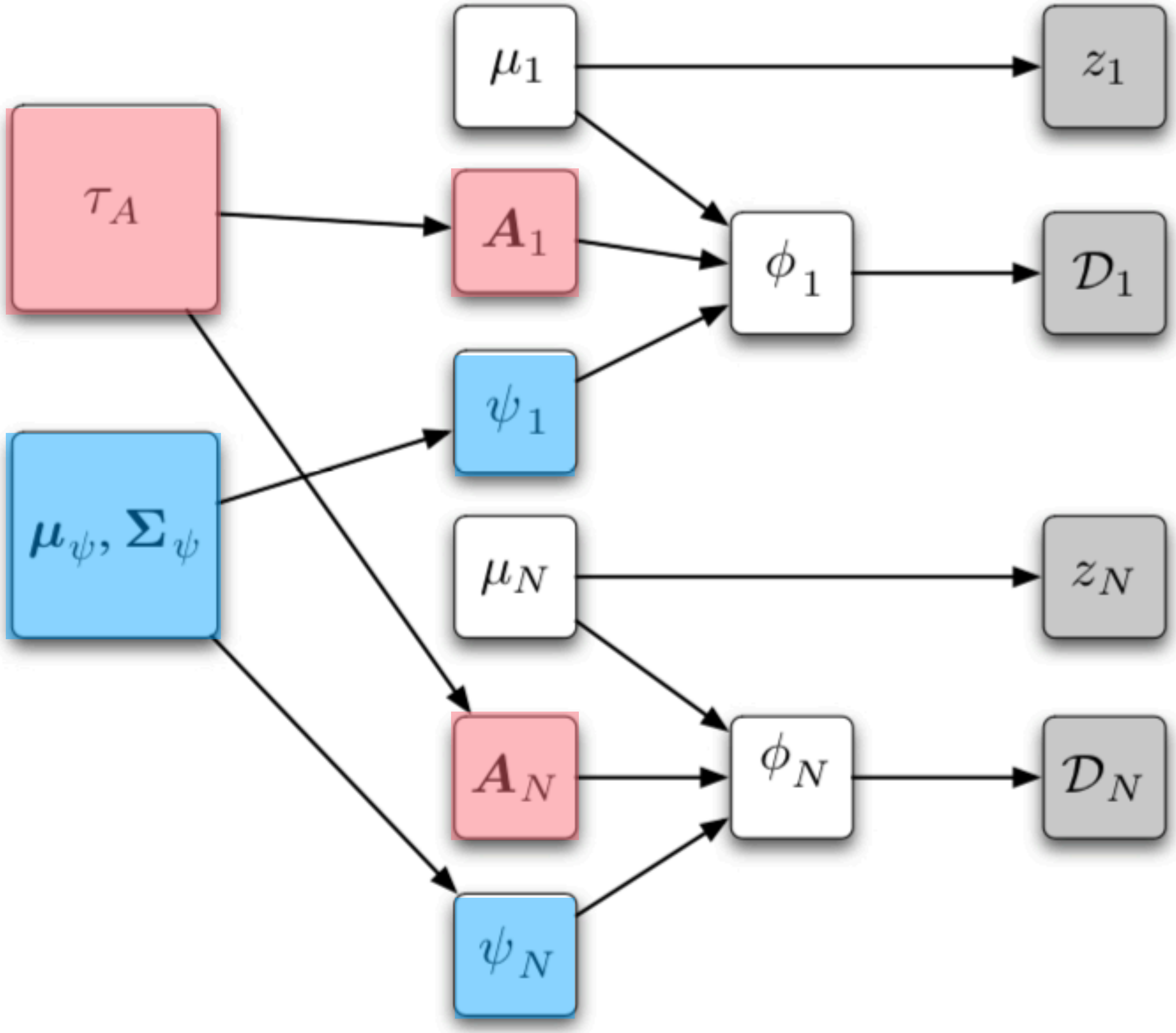


Mandel et al (2009)



supernova physics

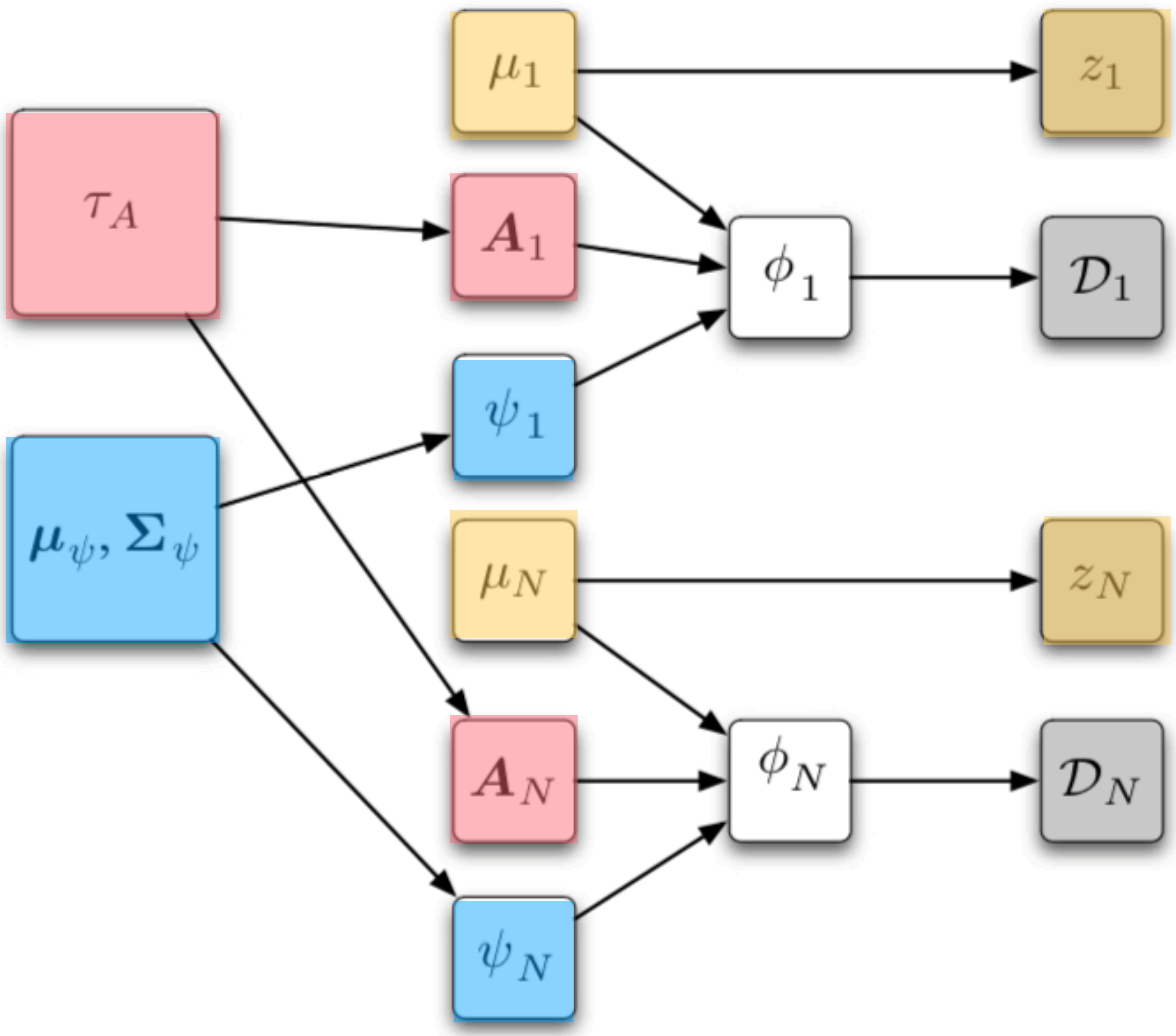
Mandel et al (2009)



supernova physics

dust extinction/reddening

Mandel et al (2009)



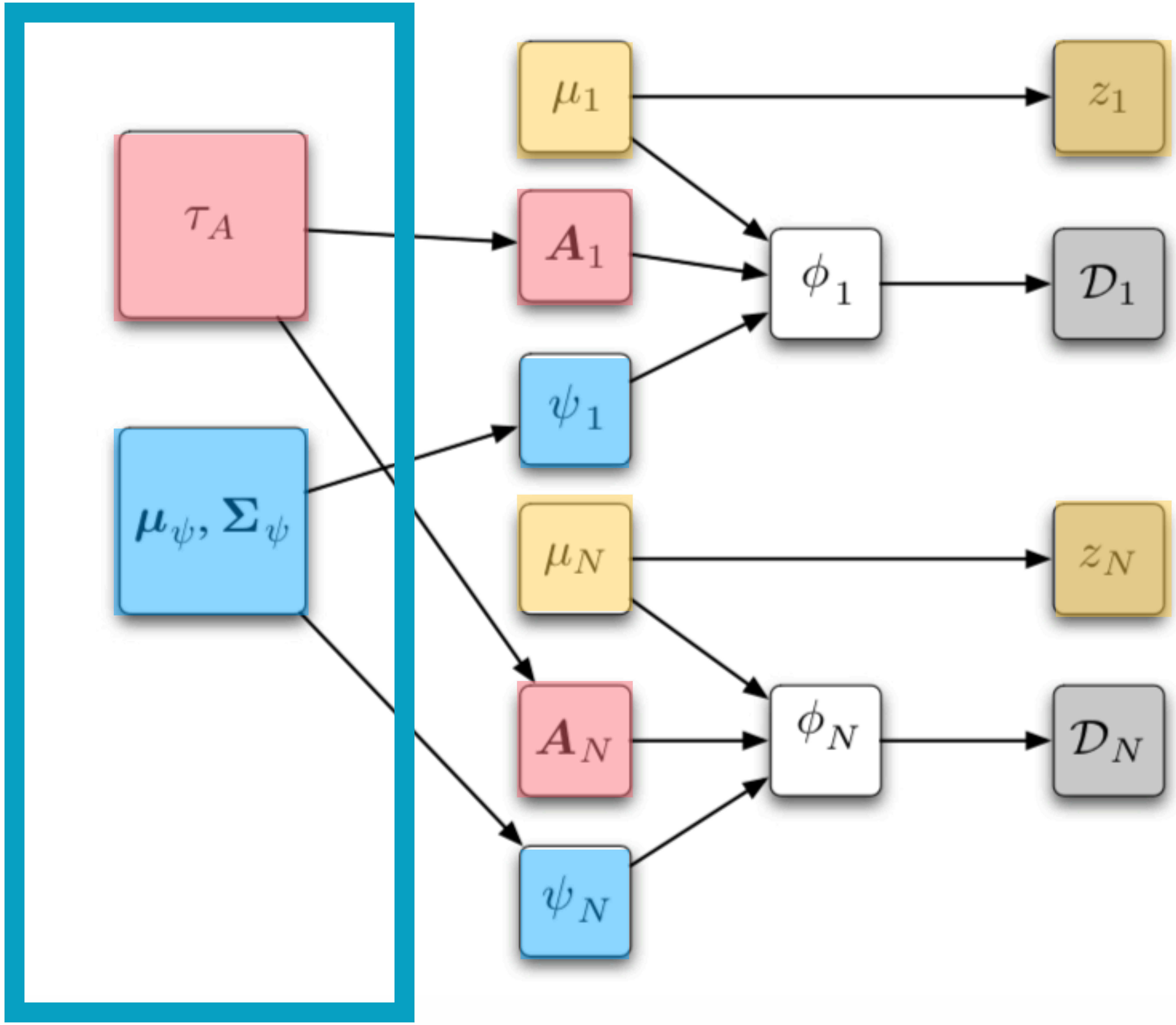
supernova physics

dust extinction/reddening

distance modulus

Mandel et al (2009)

population-level
(global) parameters

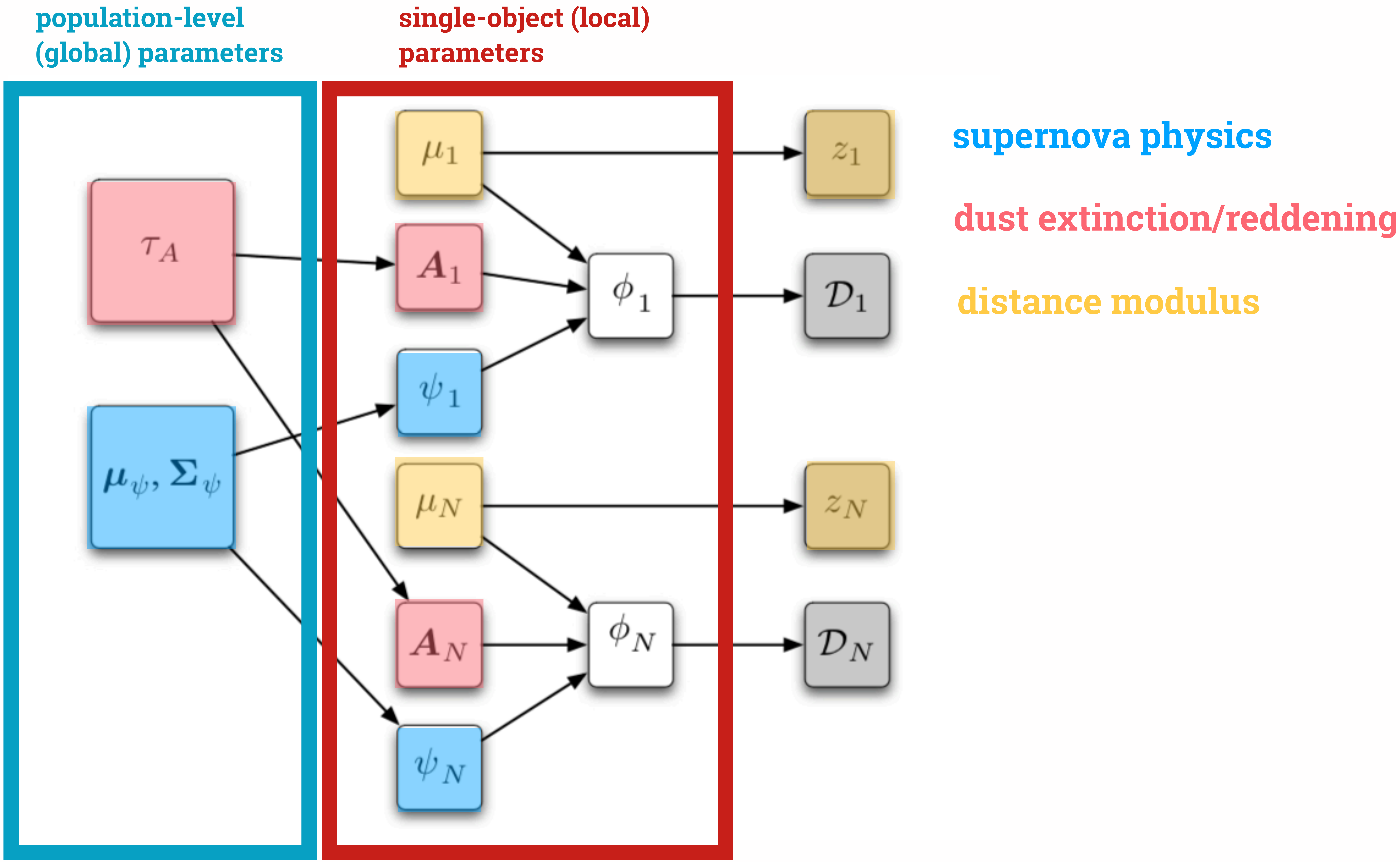


supernova physics

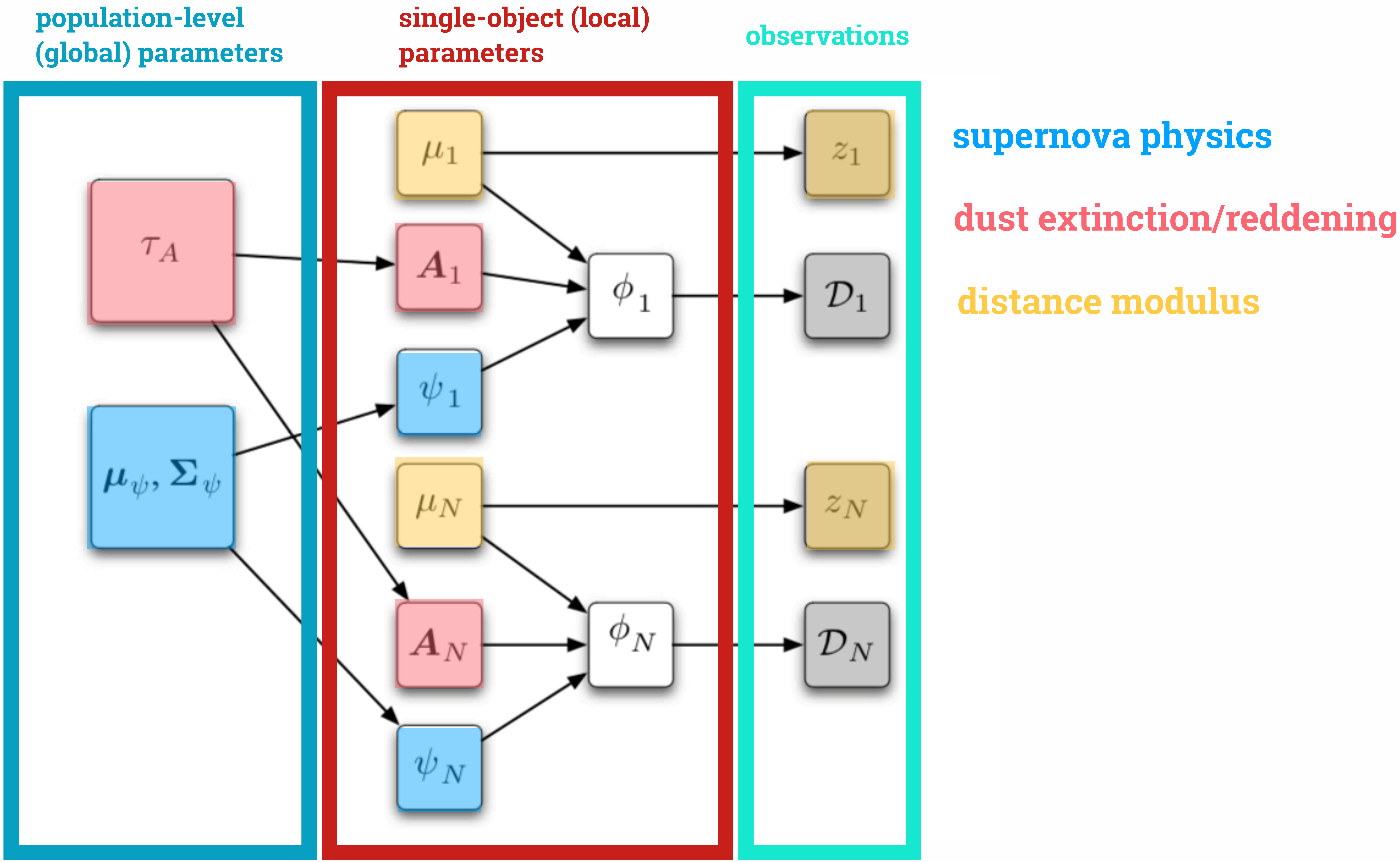
dust extinction/reddening

distance modulus

Mandel et al (2009)



Mandel et al (2009)



Mandel et al (2009)

How is this different?

$$p(\theta|D, I) \propto p(D|\theta, I)p(\theta|\alpha, I)$$

How is this different?

$$p(\theta|D, I) \propto p(D|\theta, I)p(\theta|\alpha, I)$$

assume to be known

How is this different?

$$p(\theta|D, I) \propto p(D|\theta, I)p(\theta|\alpha, I)$$

assume to be known

$$p(\theta, \alpha|D, I) \propto p(D|\theta, I)p(\theta|\alpha, I)p(\alpha|I)$$

How is this different?

$$p(\theta|D, I) \propto p(D|\theta, I)p(\theta|\alpha, I)$$

assume to be known

$$p(\theta,\alpha|D, I) \propto p(D|\theta, I)p(\theta|\alpha, I)p(\alpha|I)$$

infer α along with θ

$$
\begin{aligned}
P\left(\boldsymbol{\phi}_s, \mu_s, A_H^s, R_V^s \,\middle|\, \mathcal{D}_s, z_s; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \alpha_R\right)
&\propto P(\mathcal{D}_s | \boldsymbol{\phi}_s) \times P(\mu_s | z_s) \\
&\times P(\boldsymbol{\psi}_s = \boldsymbol{\phi}_s - \mathbf{v}\mu_s - \mathbf{A}_s | \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi) \\
&\times P\left(A_H^s, R_V^s \,\middle|\, \tau_A, \alpha_R\right).
\end{aligned} \tag{17}
$$

Mandel et al (2009)

Could histogram individual parameters ...

... but how?

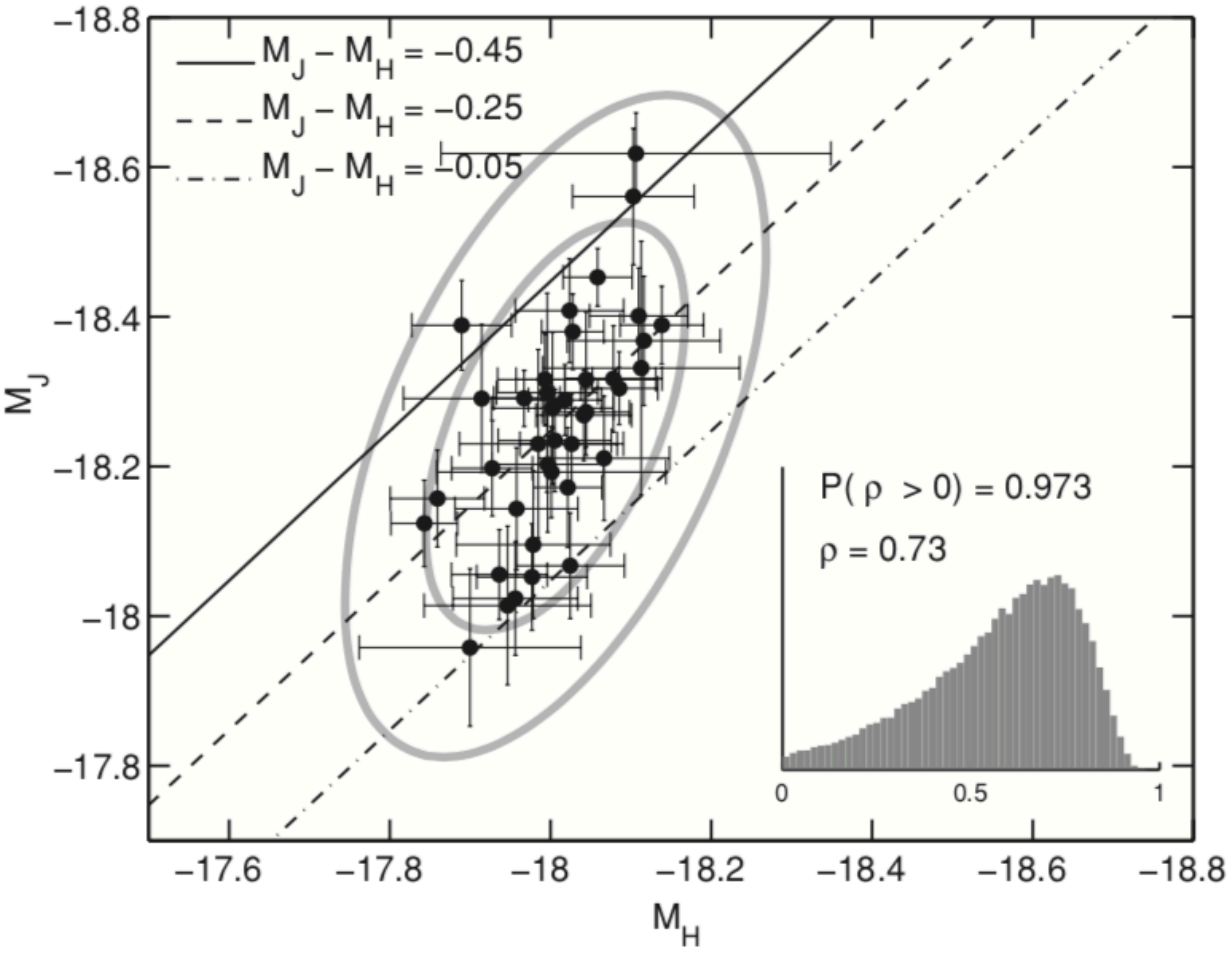
$$
\begin{aligned}
&P\left(\left\{\boldsymbol{\phi}_s, \mu_s, A_H^s, R_V^s\right\}; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \alpha_R \,\middle|\, \mathcal{D}, \mathcal{Z}\right) \\
&\quad \propto \left[\prod_{s=1}^{N_{\mathrm{SN}}} P(\boldsymbol{\phi}_s, \mu_s, A_H^s, R_V^s \,|\, \mathcal{D}_s, z_s; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \alpha_R)\right] \qquad (18) \\
&\quad \times P\left(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi\right) \times P(\tau_A, \alpha_R).
\end{aligned}
$$

Mandel et al (2009)

Why?

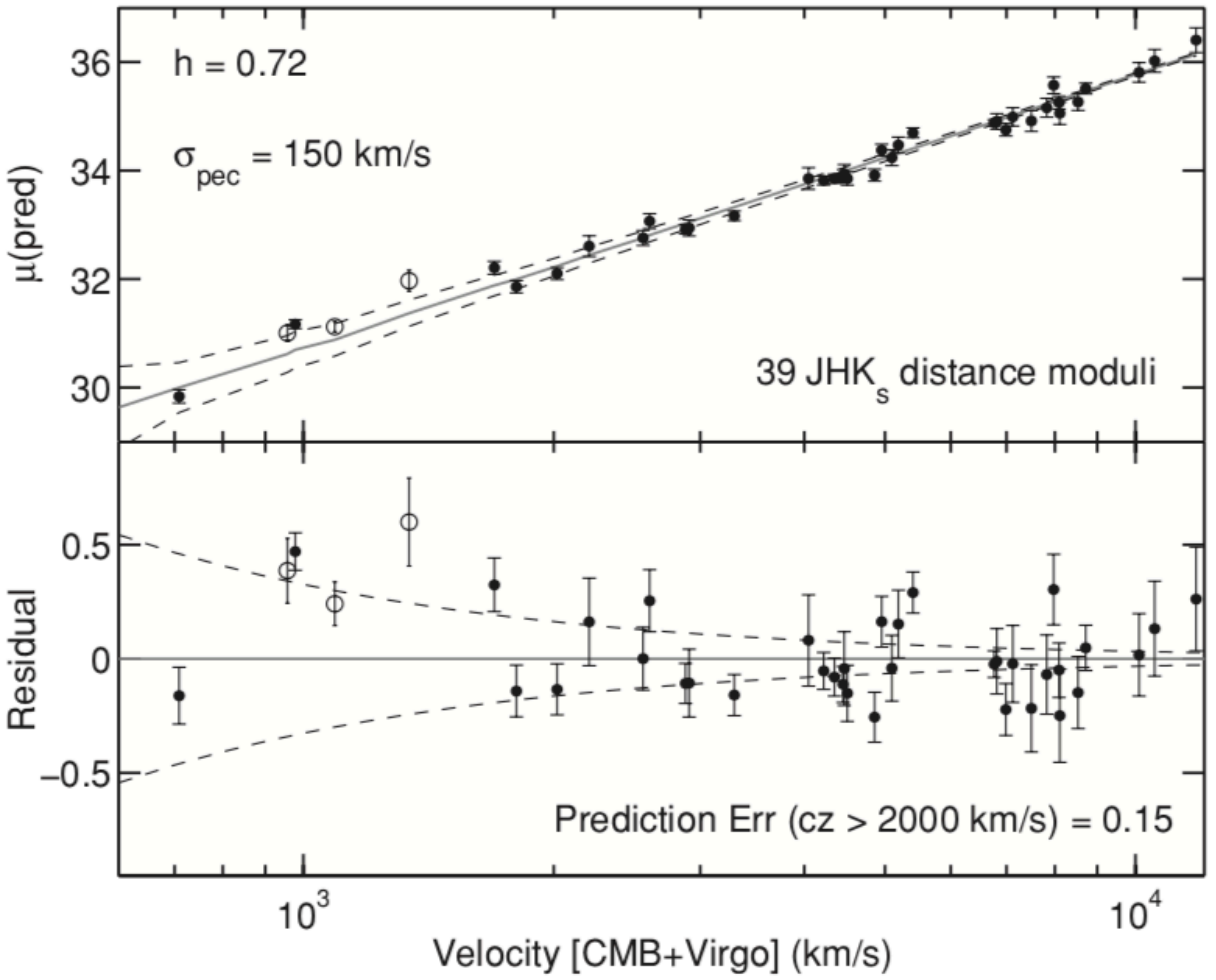
- Learn population parameters
- Improve inferences on individual population members
- self-consistent constraints on the physics
- can deal with large measurement uncertainties, systematic uncertainties and upper limits
- enables direct, probabilistic relationships between theory and observations

Mandel et al (2009)



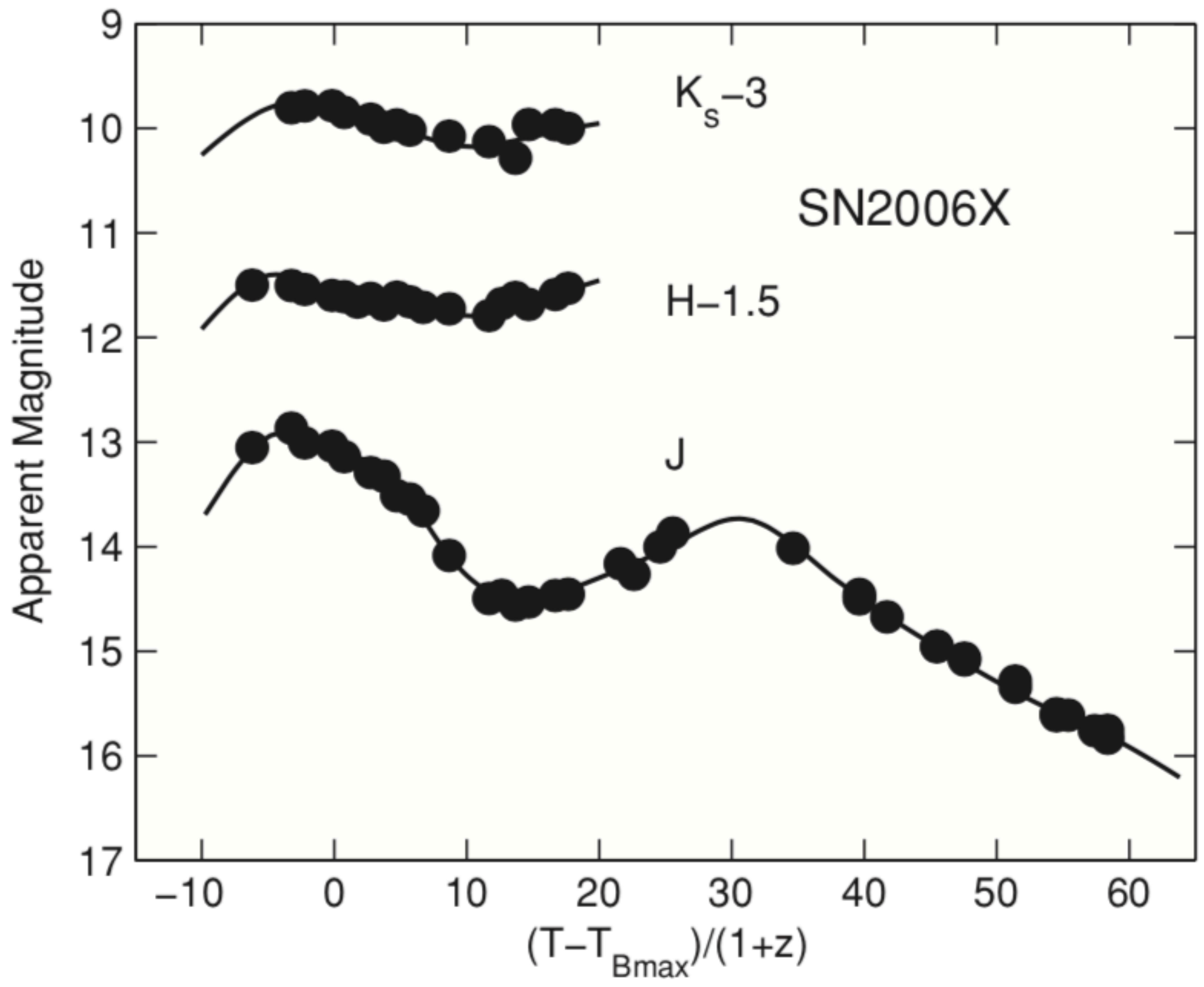
Population Parameters

Mandel et al (2009)



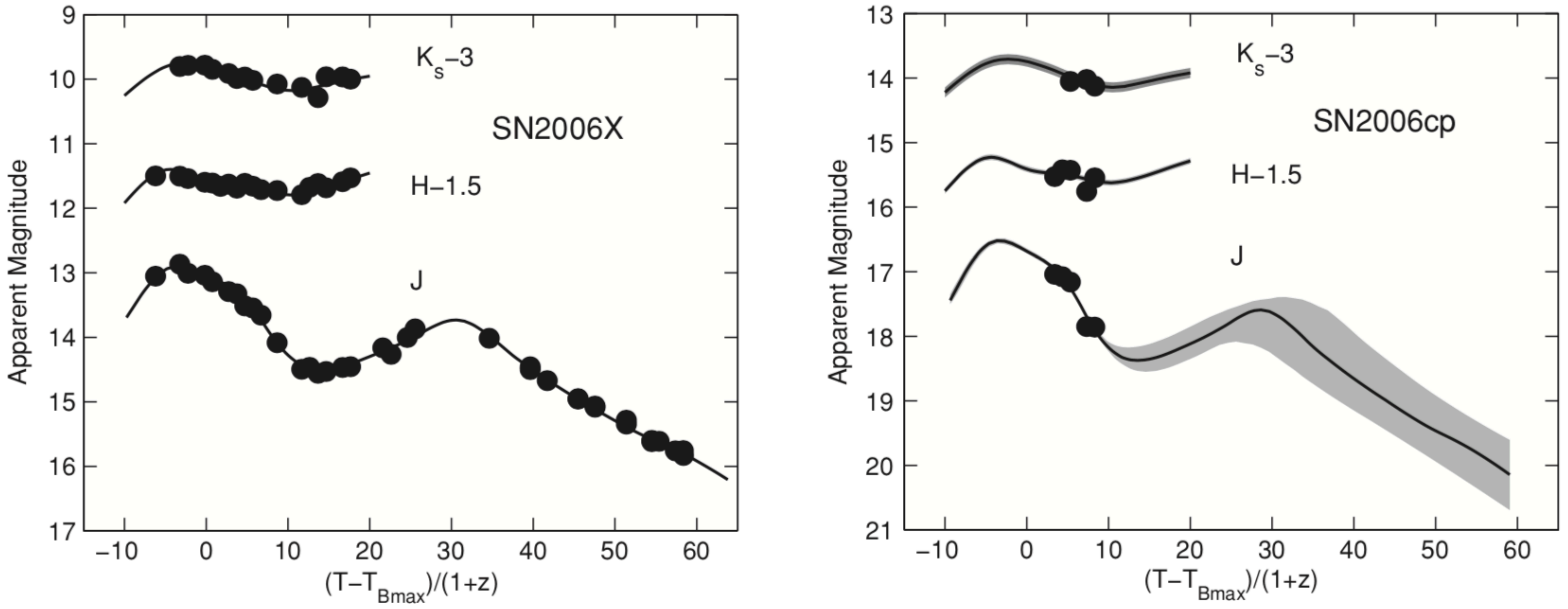
Population Parameters

Mandel et al (2009)



Bayesian Shrinkage

Mandel et al (2009)



Bayesian Shrinkage

Short Aside: Gibbs Sampling

- hierarchical models can be hard to sample with regular MCMC
- sample each variable (or group of variables) in turn, keep everything else constant
- exploit conditional independence in joint probability density to make sampling much more efficient

Exercises: See Notebook!

Nature is complex!

Click to add title

Click to add subtitle

Click to add text

Click to add title

Click to add subtitle

Click to add title

Click to add subtitle

Click to add title

Click to add text

Click to add text

Click to add title

Click to add text

Click to add text

Click to add title

Click to add text

Click to add title

Click to add title

Click to add subtitle

Click to add text

Click to add title

Click to add subtitle

Click to add text

Click to add text

xx %

xx %

Click to add title

Click to add title

Click to add title

Click to add subtitle