



# Hierarchical Bayesian Inference + Probabilistic Graphical Models

Daniela Huppenkothen

*DIRAC Institute, University of Washington*



—

[https://github.com/dhuppenkothen/  
cargese2018\\_tutorials](https://github.com/dhuppenkothen/cargese2018_tutorials)

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

— George Box



# Nature is complex!

—



... so is our data (collection)!

—



# 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 ...

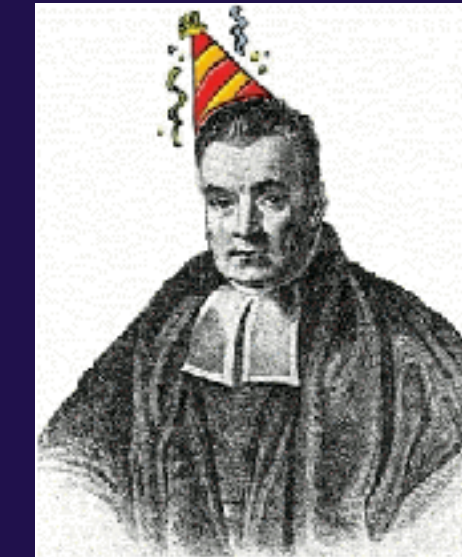
---

- 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

---

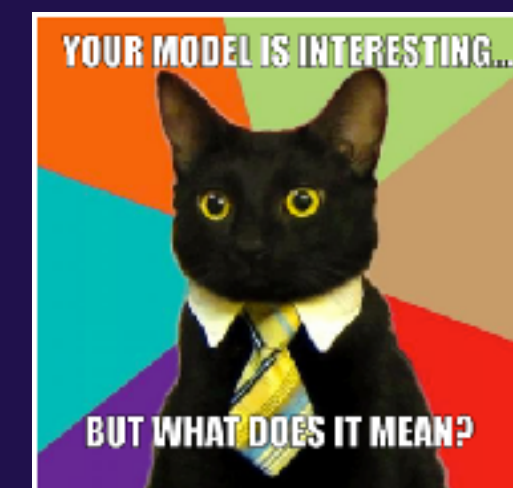
(1) Fun With Bayes(ian Hierarchical Models)



(2) Machine Learning



(3) Statistical Machine Learning



# Is it sunny today?

$p(\text{☀})$

# Is it sunny today?

		
	0.98	0.01
	0.004	0.006

1)  $p(\text{Sun}, \text{Island}) = ???$

2)  $p(\text{Sun} | \text{Island}) = ???$

3)  $p(\text{Mountain}) = ???$

4)  $p(\text{Cloud with rain} | \text{Mountain}) = ???$

# Is it sunny today?

		
	0.98	0.01
	0.004	0.006

$$1) \quad p(\text{Sun}, \text{Tropical Island}) =$$

$$2) \quad p(\text{Sun} | \text{Tropical Island}) =$$

$$3) \quad p(\text{Mountain}) =$$

$$4) \quad p(\text{Cloud with rain} | \text{Mountain}) =$$



# Is it sunny today?

		
	0.98	0.01
	0.004	0.006

$$1) \quad p(\text{Sun}, \text{Tropical Island}) = 0.98$$

$$2) \quad p(\text{Sun} | \text{Tropical Island}) =$$

$$3) \quad p(\text{Mountain}) =$$

$$4) \quad p(\text{Cloud with rain} | \text{Mountain}) =$$

# Is it sunny today?

		
	0.98	0.01
	0.004	0.006

$$1) \quad p(\text{Sun}, \text{Tropical Island}) = 0.98$$

$$2) \quad p(\text{Sun} | \text{Tropical Island}) = 0.9898$$

$$3) \quad p(\text{Mountain}) =$$

$$4) \quad p(\text{Cloud with rain} | \text{Mountain}) =$$

# Is it sunny today?

		
	0.98	0.01
	0.004	0.006

$$1) \quad p(\text{Sun}, \text{Tropical Island}) = 0.98$$

$$2) \quad p(\text{Sun} | \text{Tropical Island}) = 0.9898$$

$$3) \quad p(\text{Mountain}) = 0.01$$

$$4) \quad p(\text{Cloud with rain} | \text{Mountain}) =$$

# Is it sunny today?

		
	0.98	0.01
	0.004	0.006

$$1) \quad p(\text{Sun}, \text{Tropical Island}) = 0.98$$

$$2) \quad p(\text{Sun} | \text{Tropical Island}) = 0.9898$$

$$3) \quad p(\text{Mountain}) = 0.01$$

$$4) \quad p(\text{Cloud with rain} | \text{Mountain}) = 0.6$$

# “The weather depends on your location”

$$p(\text{☀️}, \text{🌴}) = p(\text{☀️} | \text{🌴})p(\text{🌴})$$

# “The weather depends on your location”

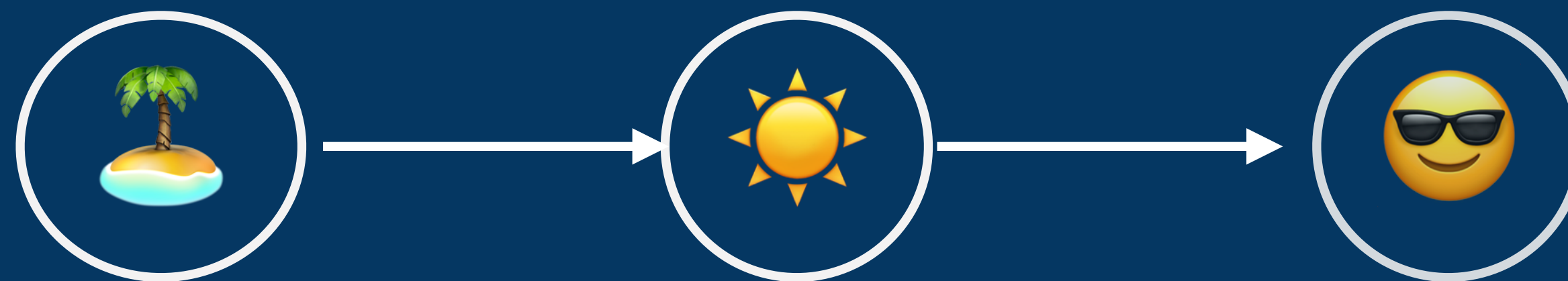
$$p(\text{☀️}, \text{🌴}) = p(\text{☀️} | \text{🌴})p(\text{🌴})$$





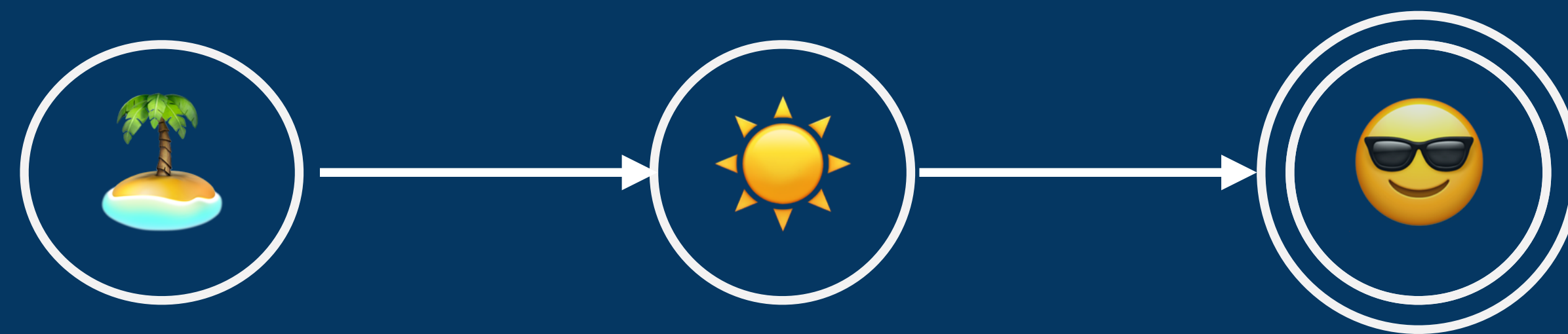
# Let's make this more complicated ...

$$p(\text{😎}, \text{☀️}, \text{🌴}) = p(\text{😎} \mid \text{☀️})p(\text{☀️} \mid \text{🌴})p(\text{🌴})$$



# Let's make this more complicated ...

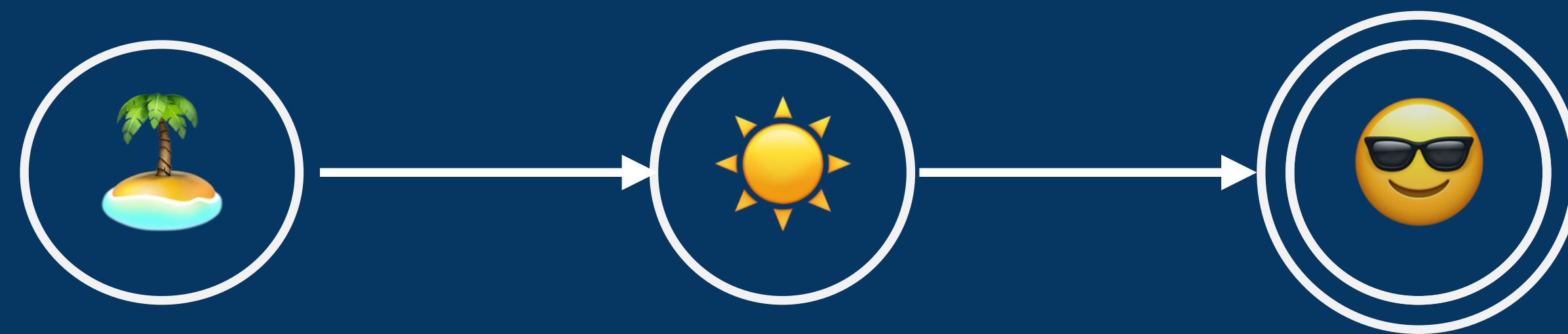
$$p(\text{😎}, \text{☀️}, \text{🌴}) = p(\text{😎} \mid \text{☀️})p(\text{☀️} \mid \text{🌴})p(\text{🌴})$$





# Let's make this more complicated ...

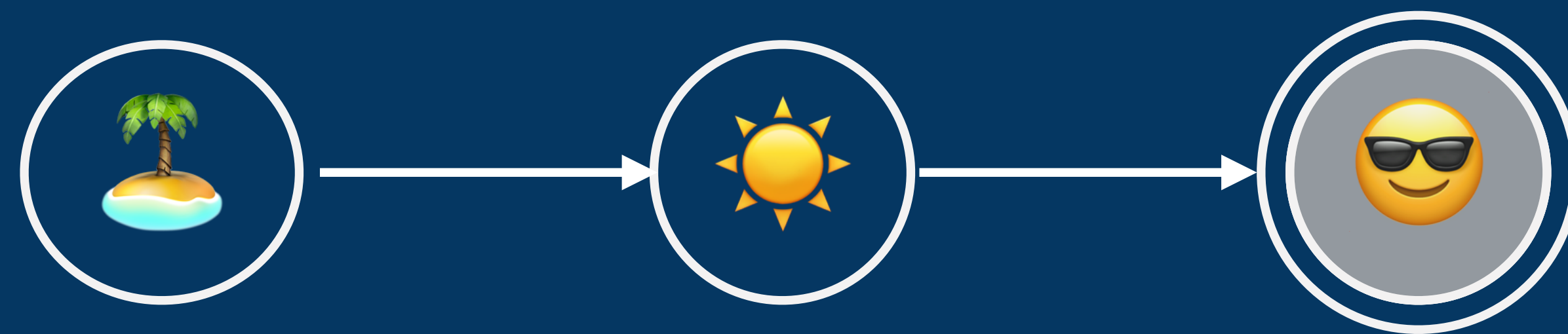
$$p(\text{😎}, \text{☀️}, \text{🌴}) = p(\text{😎} \mid \text{☀️})p(\text{☀️} \mid \text{🌴})p(\text{🌴})$$



“observed”

# Let's make this more complicated ...

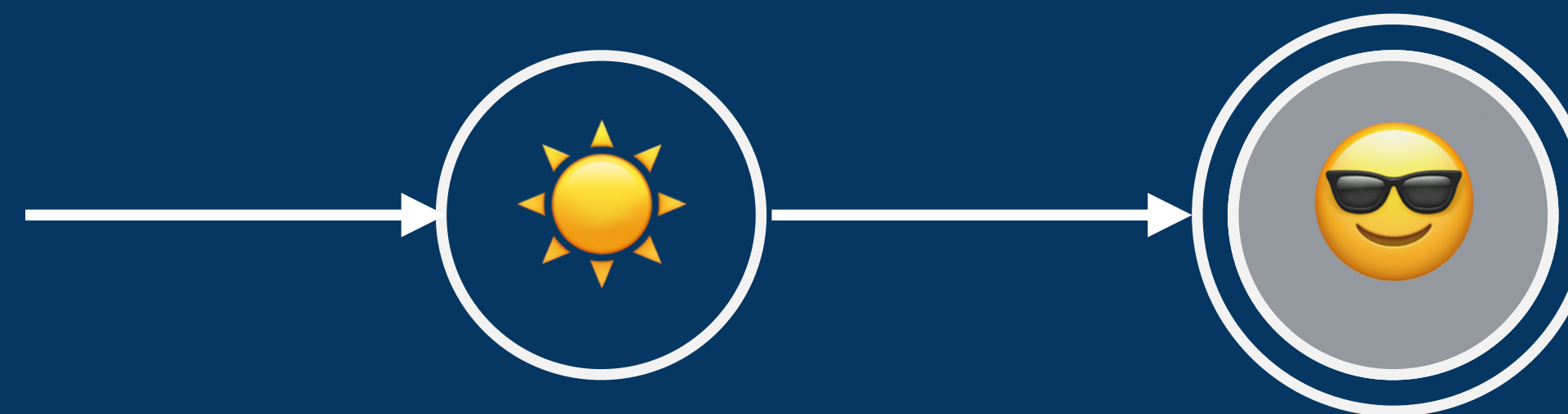
$$p(\text{🕶️}, \text{☀️}, \text{🌴}) = p(\text{🕶️} \mid \text{☀️})p(\text{☀️} \mid \text{🌴})p(\text{🌴})$$



“observed”

# Let's make this more complicated ...

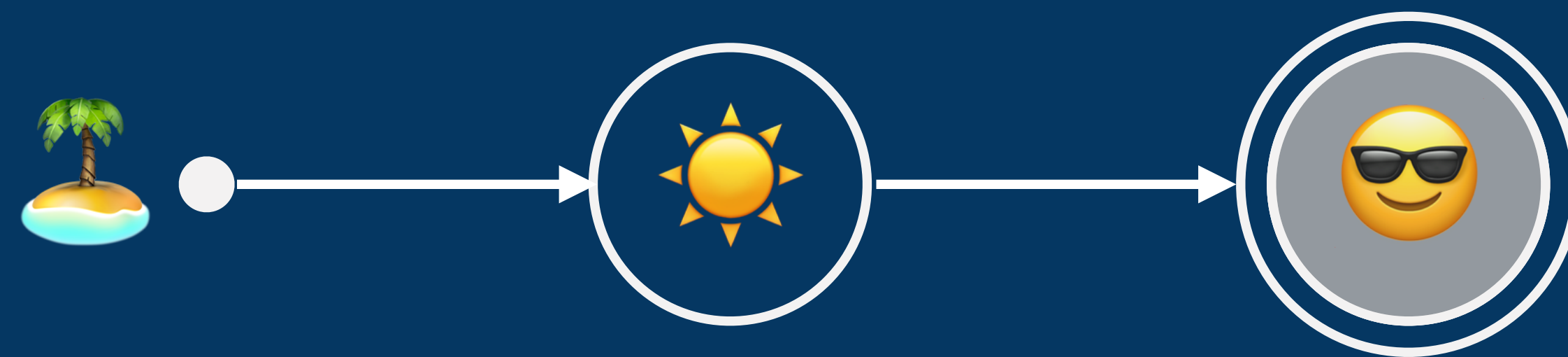
$$p(\text{😎}, \text{☀️}, \text{🌴}) = p(\text{😎} \mid \text{☀️})p(\text{☀️} \mid \text{🌴})p(\text{🌴})$$



“observed”

# Let's make this more complicated ...

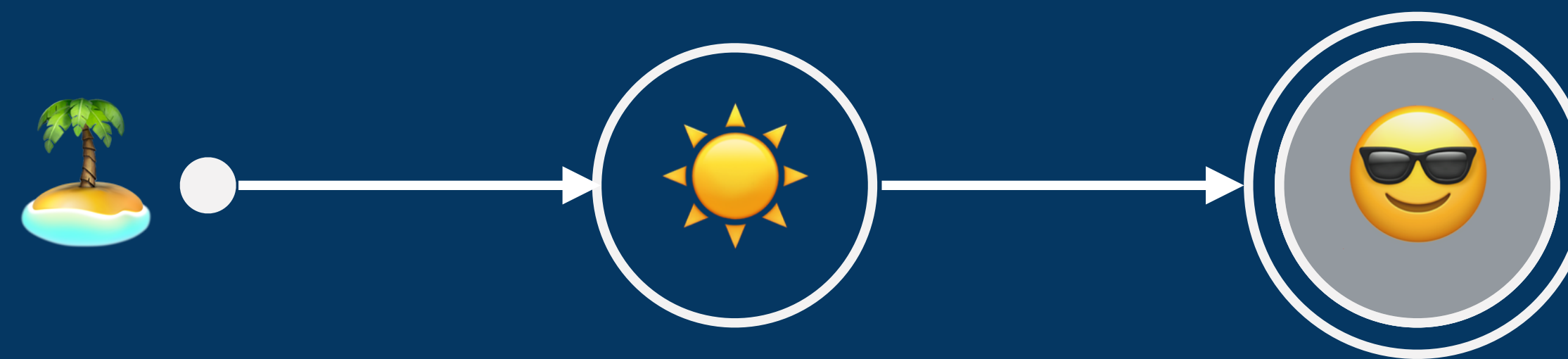
$$p(\text{👓}, \text{☀️}, \text{🌴}) = p(\text{👓} \mid \text{☀️})p(\text{☀️} \mid \text{🌴})p(\text{🌴})$$



“observed”

# Let's make this more complicated ...

$$p(\text{👓}, \text{☀️}, \text{🌴}) = p(\text{👓} \mid \text{☀️})p(\text{☀️} \mid \text{🌴})p(\text{🌴})$$

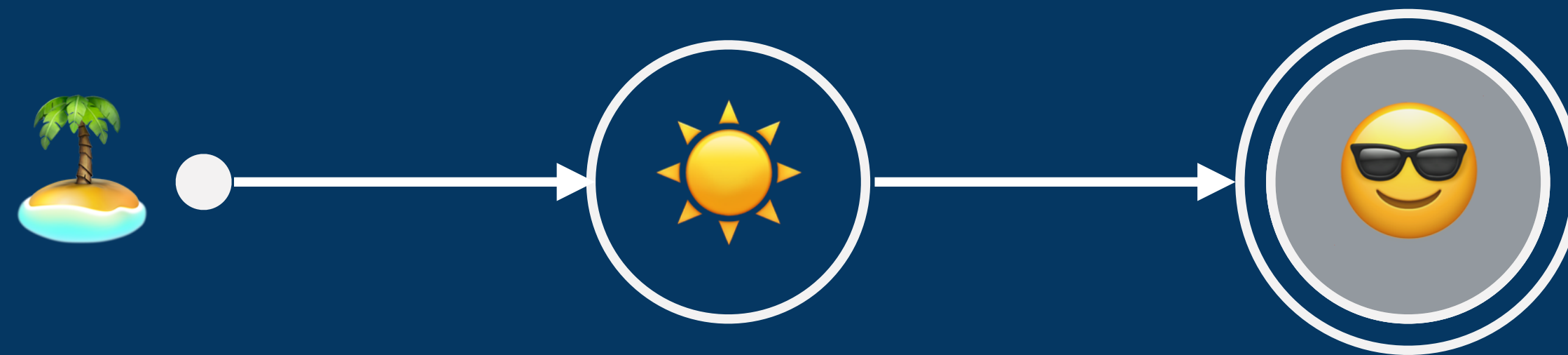


“known”

“observed”

# Let's make this more complicated ...

$$p(\text{👓}, \text{☀️}, \text{🌴}) = p(\text{👓} \mid \text{☀️})p(\text{☀️} \mid \text{🌴})p(\text{🌴})$$



“known”

“observed”



# 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?

$$\begin{aligned}
 p(\text{😎}, \text{☀️}, \text{🌴}, \text{🕒}, \text{🏠}, \text{🌡️}) &= p(\text{😎} \mid \text{☀️}, \text{🏠}) p(\text{☀️} \mid \text{🌴}, \text{🕒}) \\
 &\quad \times p(\text{🏠} \mid \text{🌡️}) p(\text{🌴}) p(\text{🌡️}) p(\text{🕒})
 \end{aligned}$$

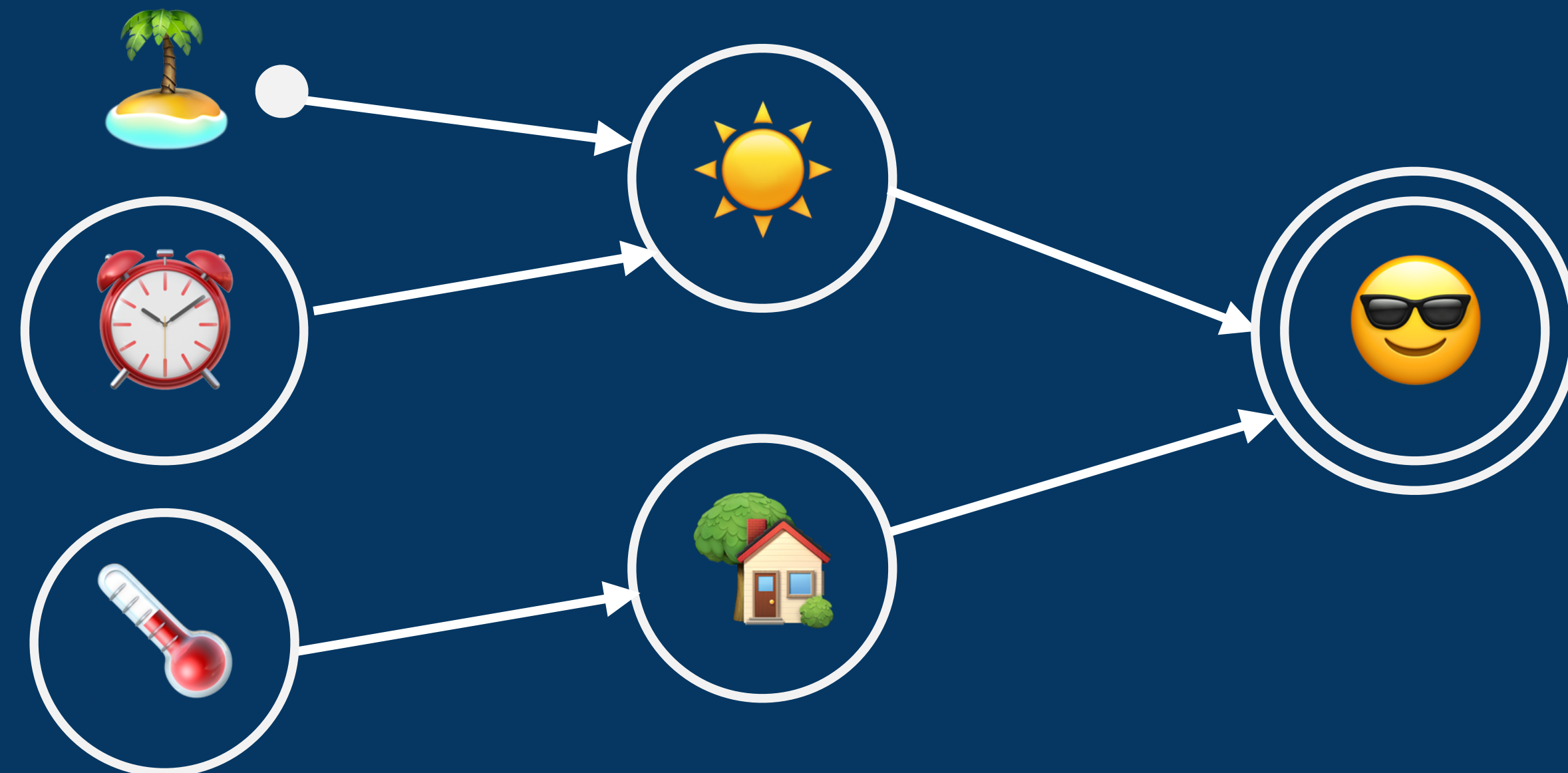


# What else does 🙄 depend on?

$$p(\text{😎}, \text{☀️}, \text{🌴}, \text{🕒}, \text{🏠}, \text{🌡️}) = p(\text{😎} \mid \text{☀️}, \text{🏠}) p(\text{☀️} \mid \text{🌴}, \text{🕒}) \\ \times p(\text{🏠} \mid \text{🌡️}) p(\text{🌡️}) p(\text{🕒})$$

# What else does 🙄 depend on?

$$p(\text{😎}, \text{☀️}, \text{🌴}, \text{🕒}, \text{🏠}, \text{🌡️}) = p(\text{😎} \mid \text{☀️}, \text{🏠}) p(\text{☀️} \mid \text{🌴}, \text{🕒}) \\ \times p(\text{🏠} \mid \text{🌡️}) p(\text{🌡️}) p(\text{🕒})$$



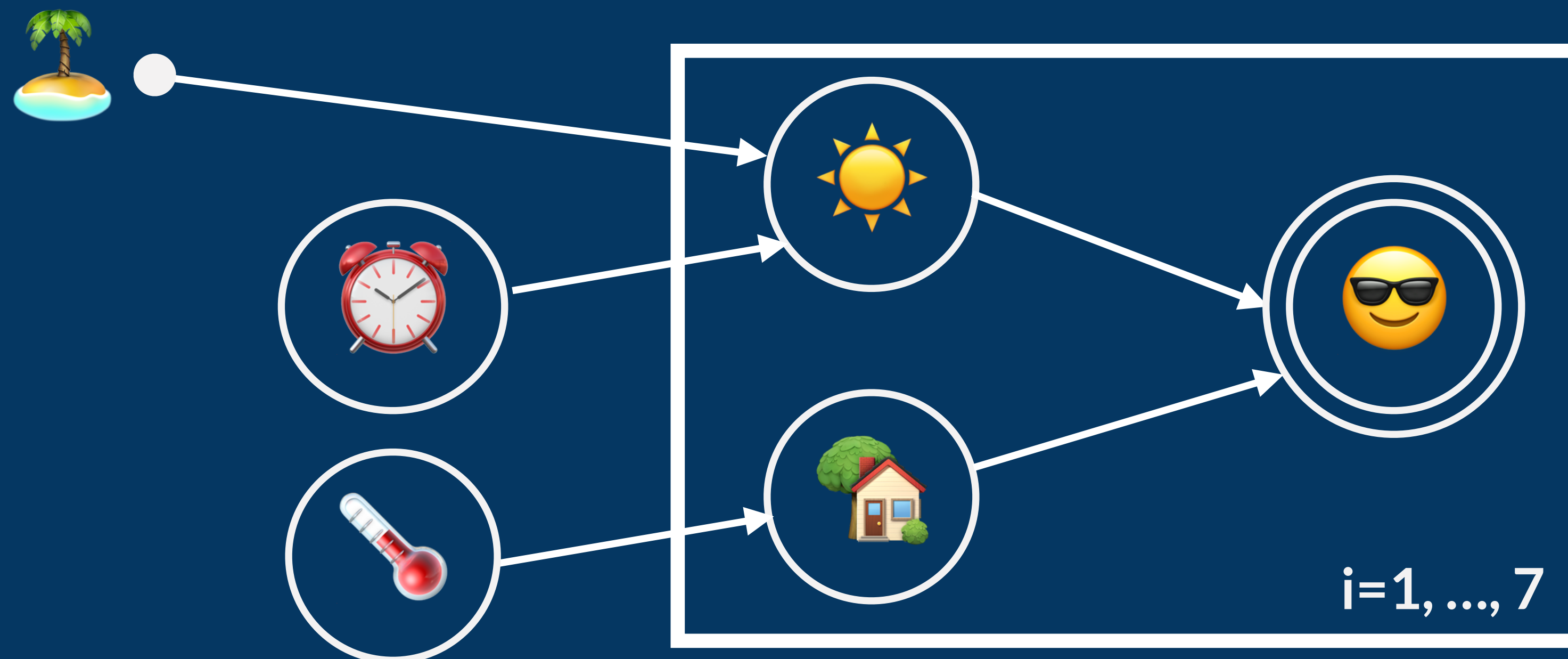
# Repeated Observations

$$p(\text{😎}, \text{☀️}, \text{🌴}, \text{🕒}, \text{🏠}, \text{🌡️}) = \left[ \prod_{i=1}^7 p(\text{😎} \mid \text{☀️}, \text{🏠}) p(\text{☀️} \mid \text{🌴}, \text{🕒}) \right. \\ \left. \times p(\text{🏠} \mid \text{🌡️}) \right] p(\text{🌡️}) p(\text{🕒})$$



# Repeated Observations

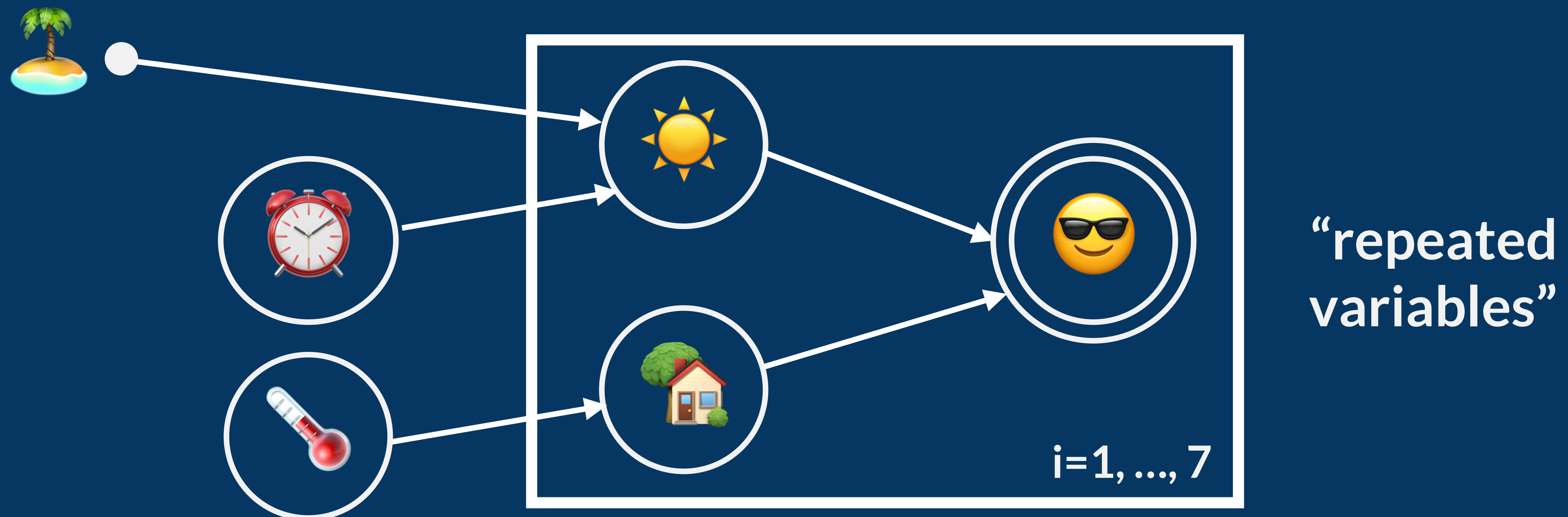
$$p(\text{👤}, \text{☀️}, \text{🌴}, \text{🕒}, \text{🏠}, \text{🌡️}) = \left[ \prod_{i=1}^7 p(\text{👤} \mid \text{☀️}, \text{🏠}) p(\text{☀️} \mid \text{🌴}, \text{🕒}) \right. \\ \left. \times p(\text{🏠} \mid \text{🌡️}) \right] p(\text{🌡️}) p(\text{🕒})$$





# Repeated Observations

$$p(\text{👤}, \text{☀️}, \text{🌴}, \text{🕒}, \text{🏠}, \text{🌡️}) = \left[ \prod_{i=1}^7 p(\text{👤} \mid \text{☀️}, \text{🏠}) p(\text{☀️} \mid \text{🌴}, \text{🕒}) \right. \\ \left. \times p(\text{🏠} \mid \text{🌡️}) \right] p(\text{🌡️}) p(\text{🕒})$$



# “The weather depends on your location”

$$\begin{aligned}
 p(\text{☀️}, \text{🌴}) &= p(\text{☀️} | \text{🌴}) p(\text{🌴}) \\
 &= p(\text{🌴} | \text{☀️}) p(\text{☀️})
 \end{aligned}$$

# “The weather depends on your location”

$$\begin{aligned}
 p(\text{☀️}, \text{🌴}) &= p(\text{☀️} | \text{🌴}) p(\text{🌴}) \\
 &= p(\text{🌴} | \text{☀️}) p(\text{☀️})
 \end{aligned}$$

$$\begin{aligned}
 p(\text{😎}) p(\text{☀️}, \text{🌴}, \text{🕒}, \text{🏠}, \text{🌡️} | \text{😎}) &= \left[ \prod_{i=1}^7 p(\text{😎} | \text{☀️}, \text{🏠}) p(\text{☀️} | \text{🌴}, \text{🕒}) \right. \\
 &\quad \left. \times p(\text{🏠} | \text{🌡️}) \right] p(\text{🌡️}) p(\text{🕒})
 \end{aligned}$$

# 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 ...

$$p(\text{😎}) p(\text{☀️}, \text{🌴}, \text{🕒}, \text{🏠}, \text{🌡️} \mid \text{😎}) = \prod_{i=1}^7 p(\text{😎} \mid \text{☀️}, \text{🏠}) p(\text{☀️} \mid \text{🌴}, \text{🕒}) \\ \times p(\text{🏠} \mid \text{🌡️}) p(\text{🌡️}) p(\text{🕒})$$

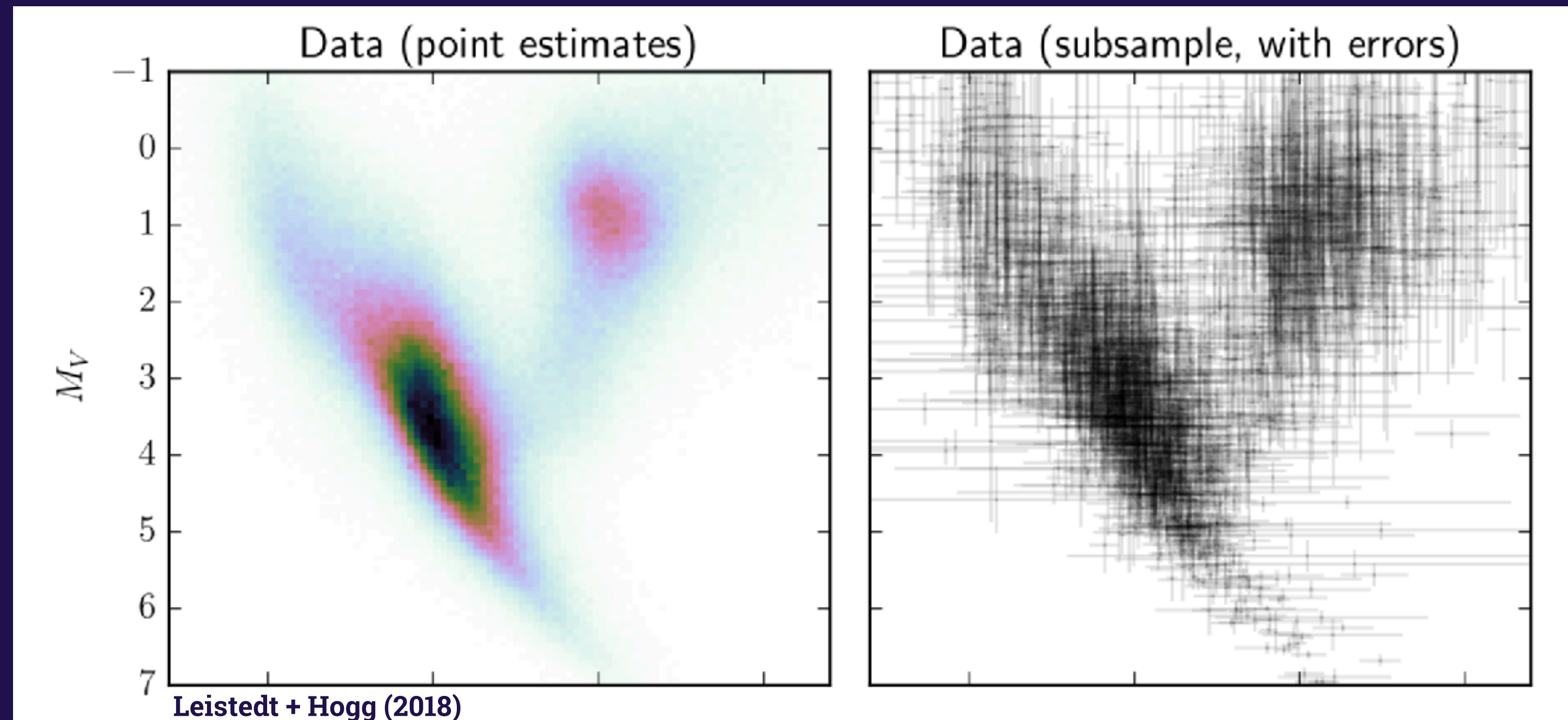
We have shown that we can write down arbitrarily complex probability distributions ...

$$p(\text{😎}) p(\text{☀️}, \text{🌴}, \text{🕒}, \text{🏠}, \text{🌡️} | \text{😎}) = \prod_{i=1}^7 p(\text{😎} | \text{☀️}, \text{🏠}) p(\text{☀️} | \text{🌴}, \text{🕒}) \\ \times p(\text{🏠} | \text{🌡️}) p(\text{🌡️}) p(\text{🕒})$$

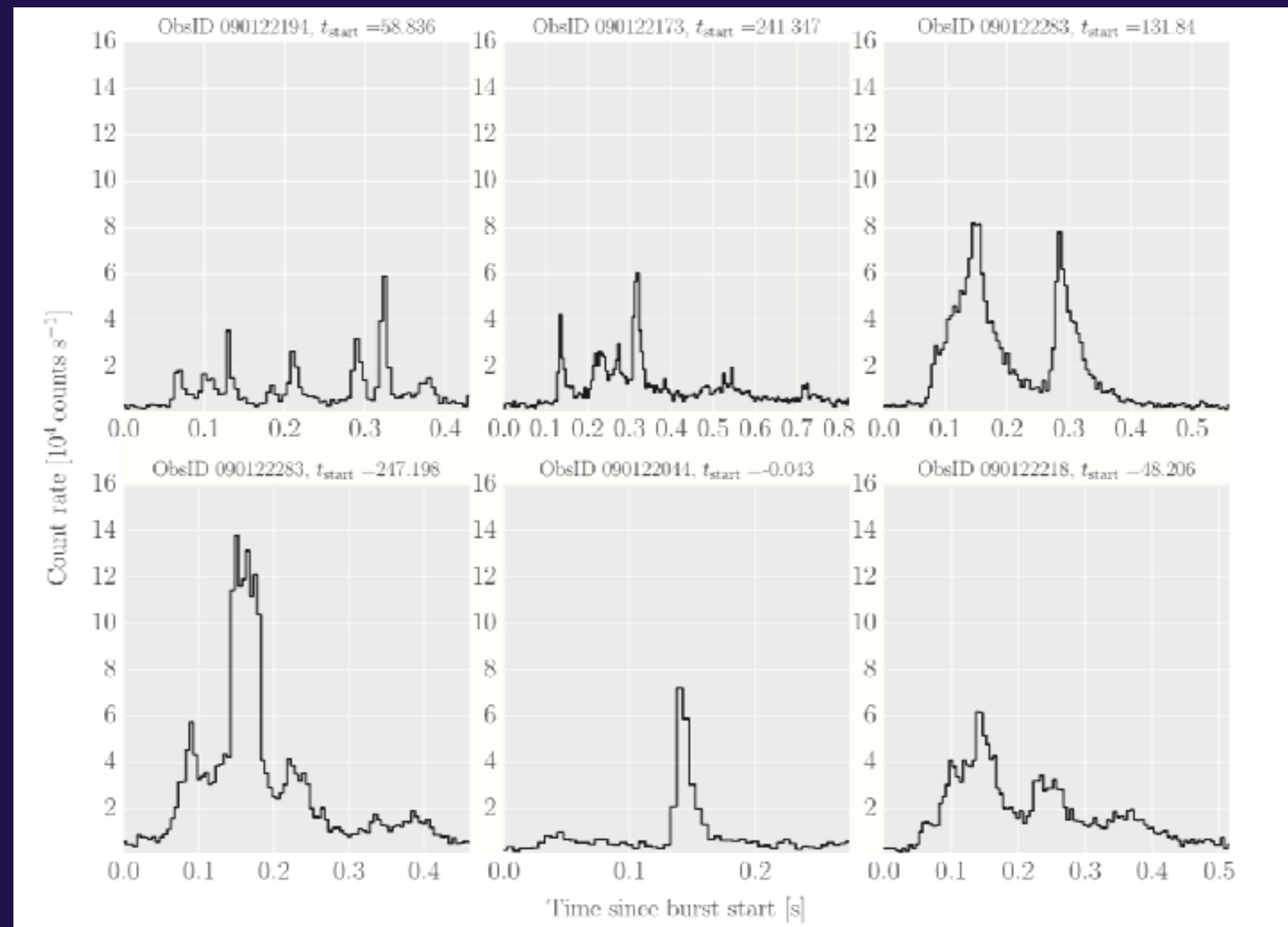
... now what?

# Might have many objects ...

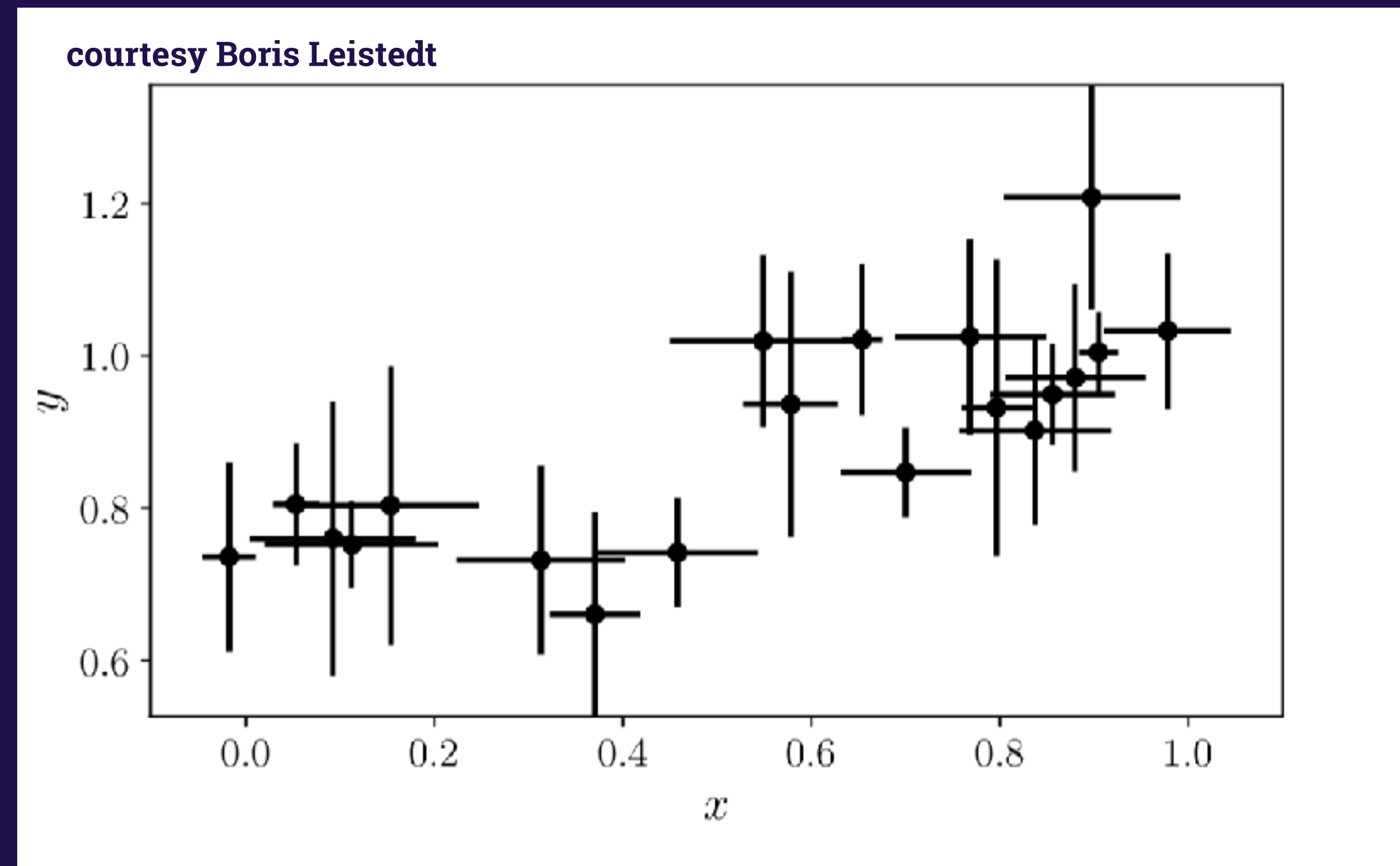
---



# ... or many observations per object

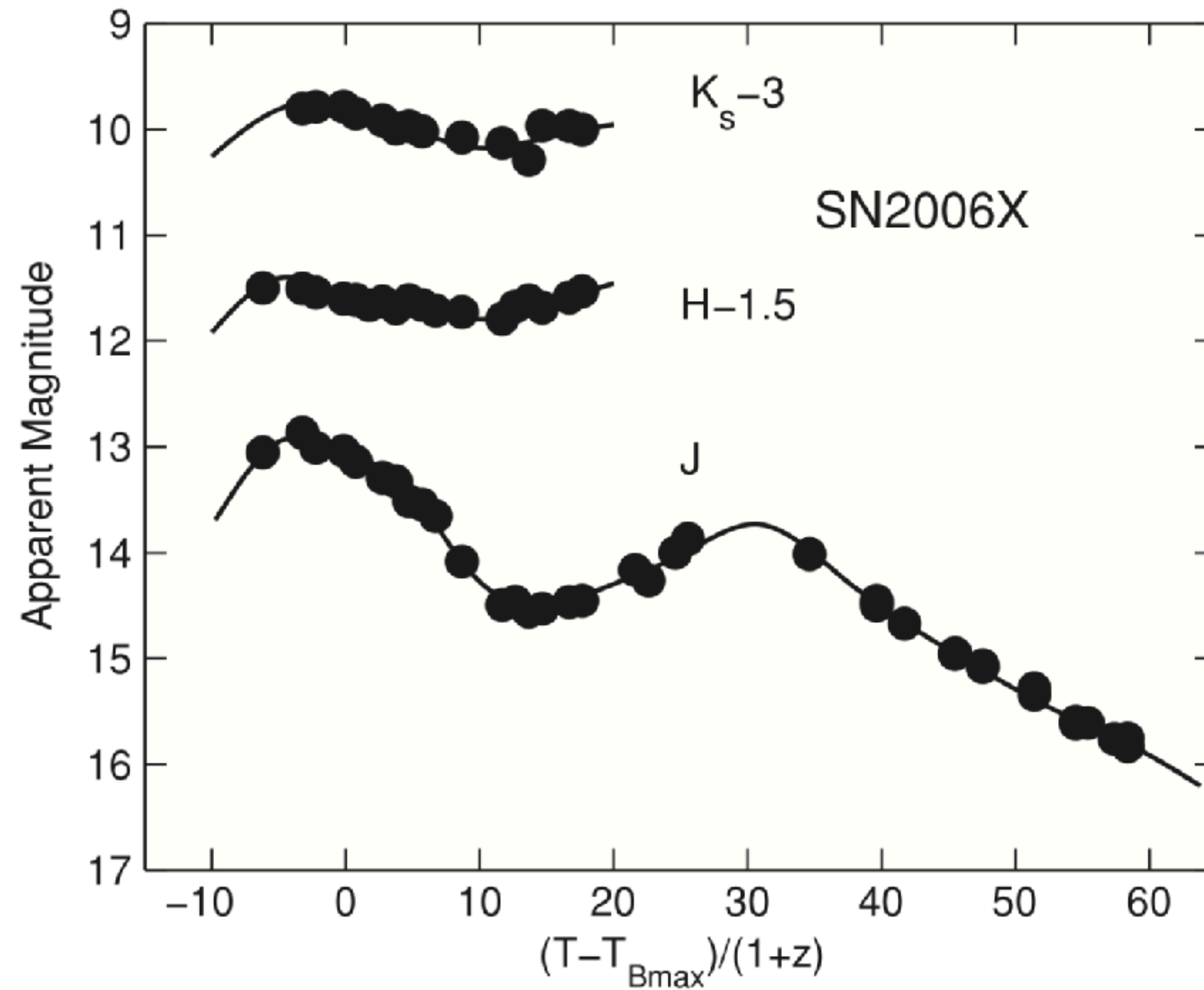


# ... or several types of uncertainties



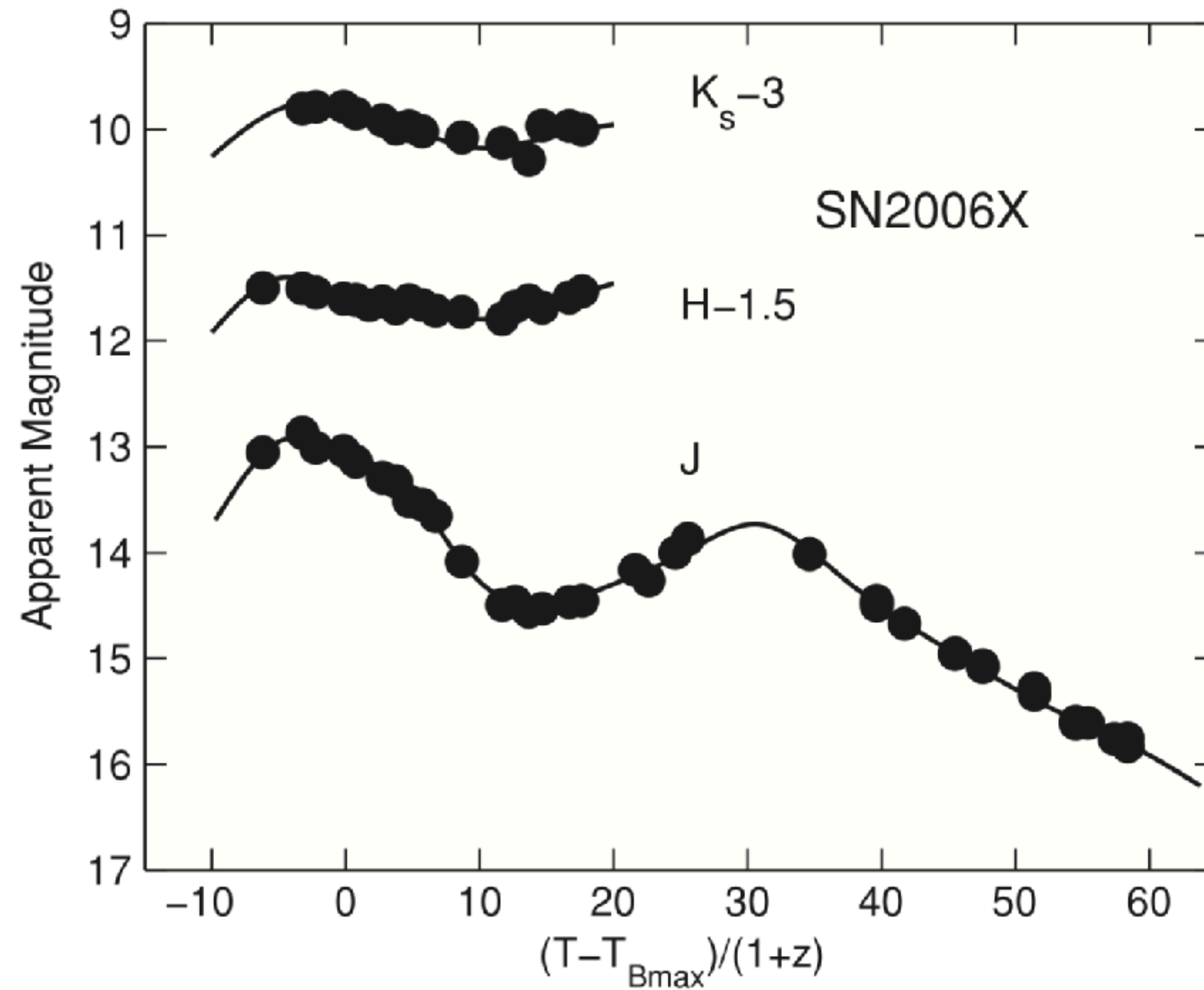


# A motivating example ...



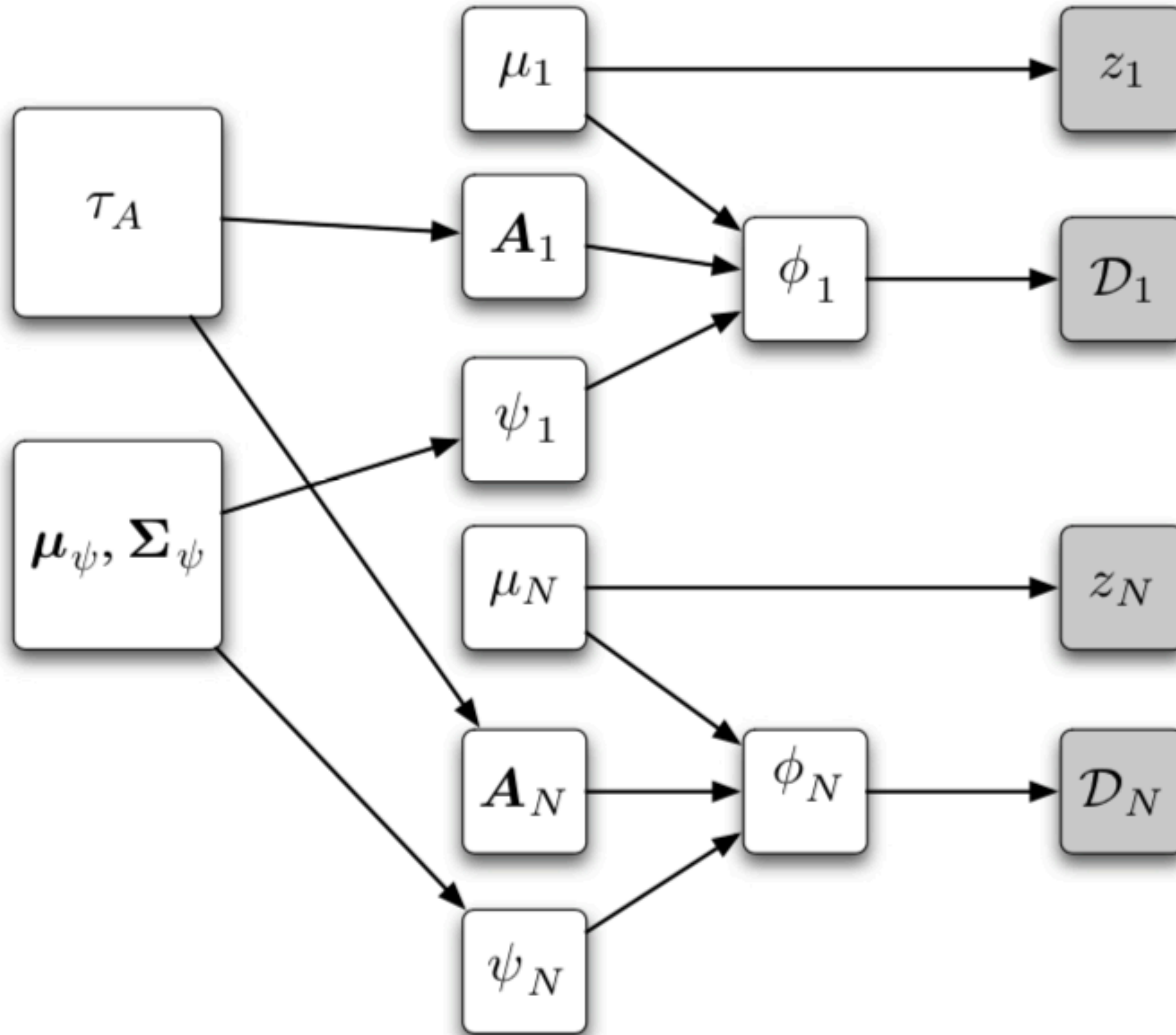


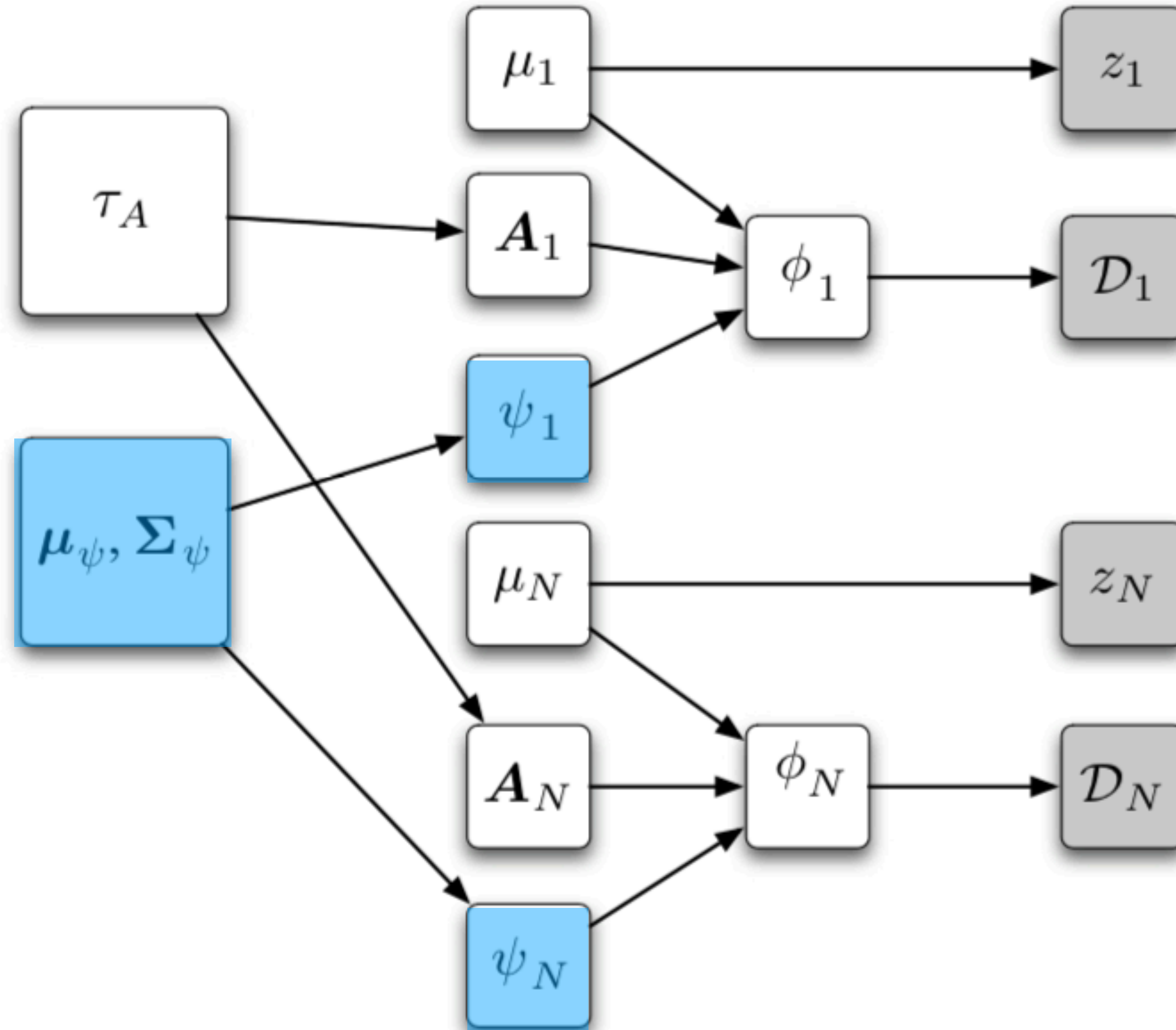
# SN Ia Light Curves

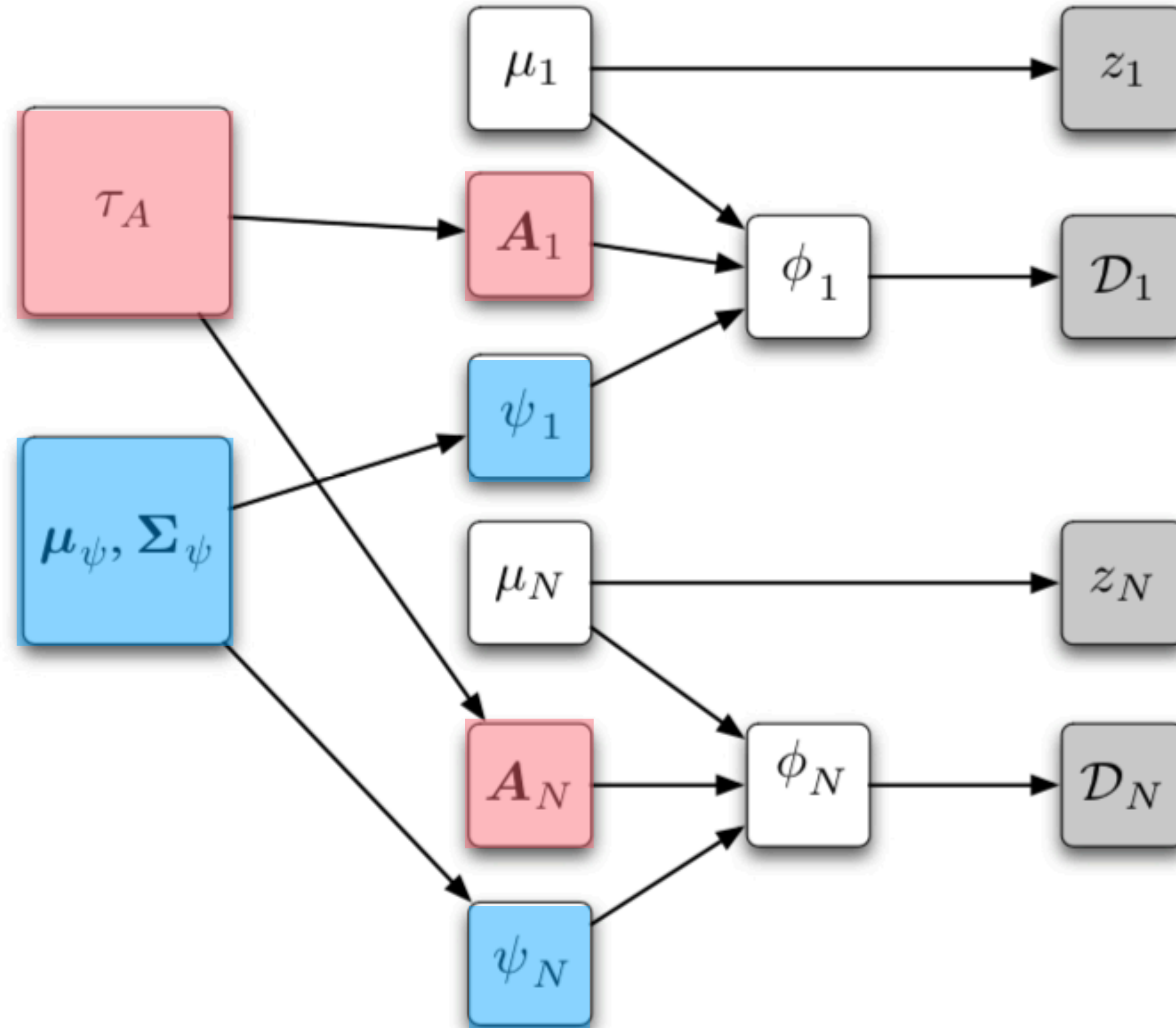




What observables and parameters do  
you have in SN Ia light curves?

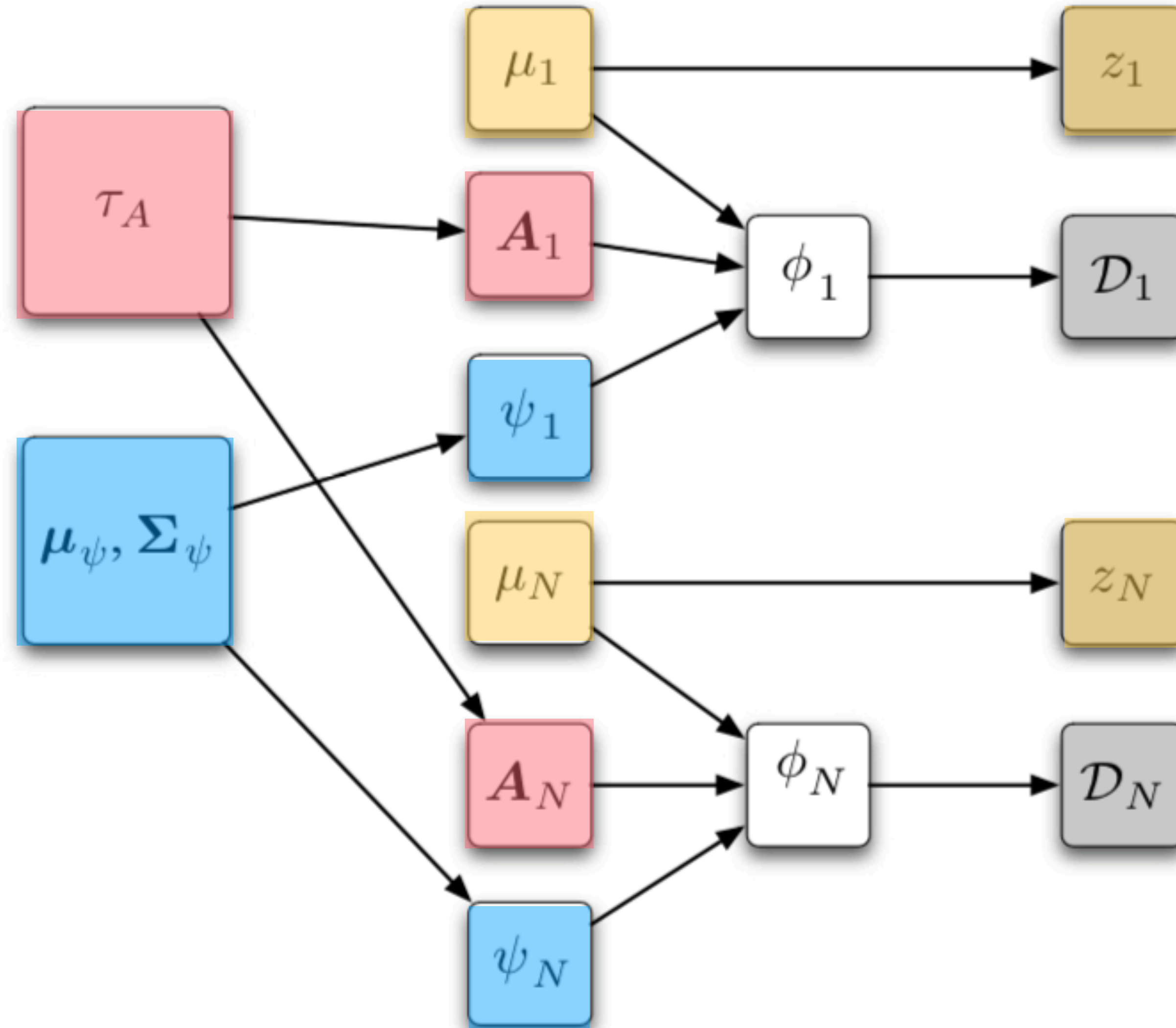






supernova physics

dust extinction/reddening

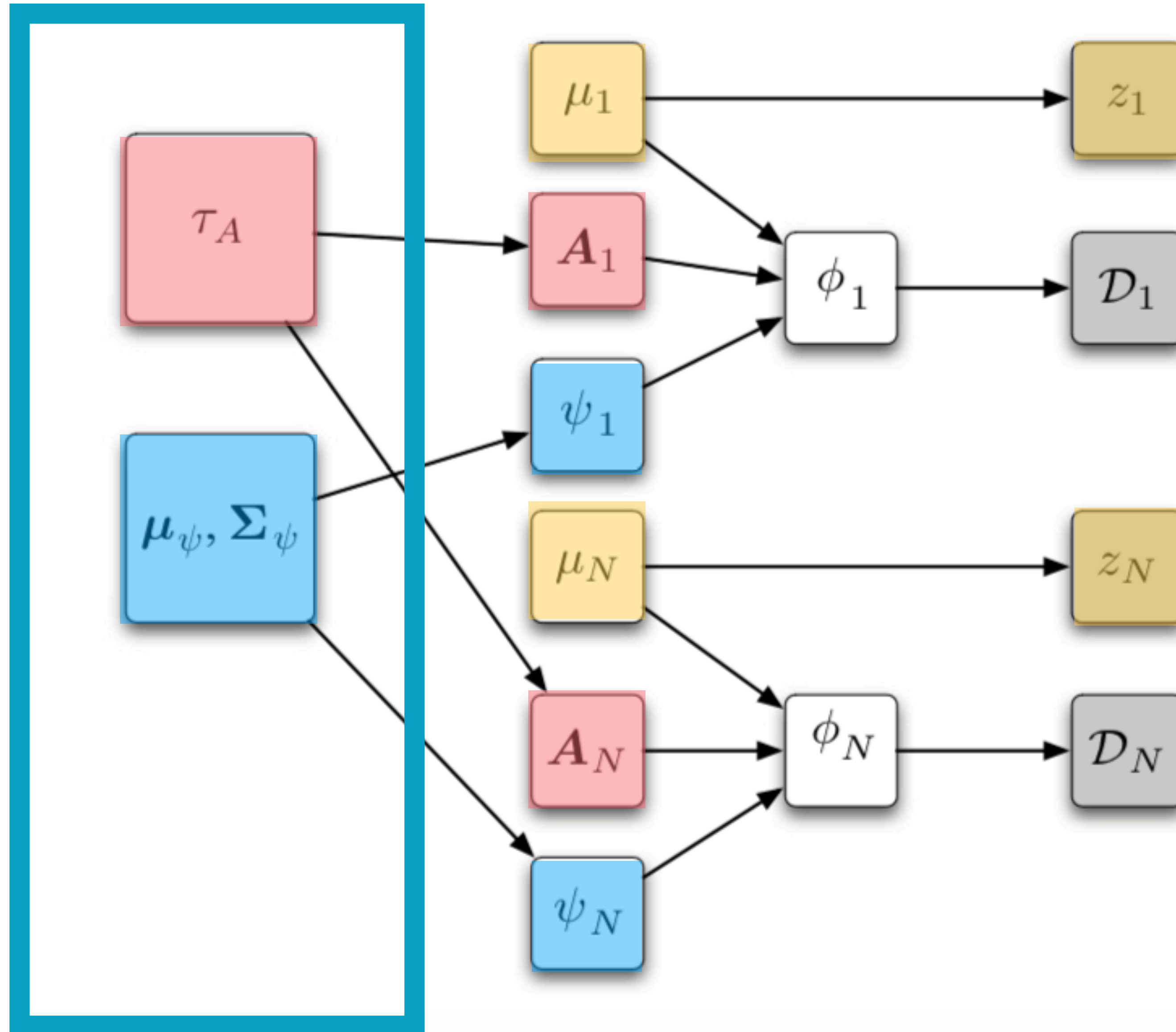


supernova physics

dust extinction/reddening

distance modulus

population-level  
(global) parameters



supernova physics

dust extinction/reddening

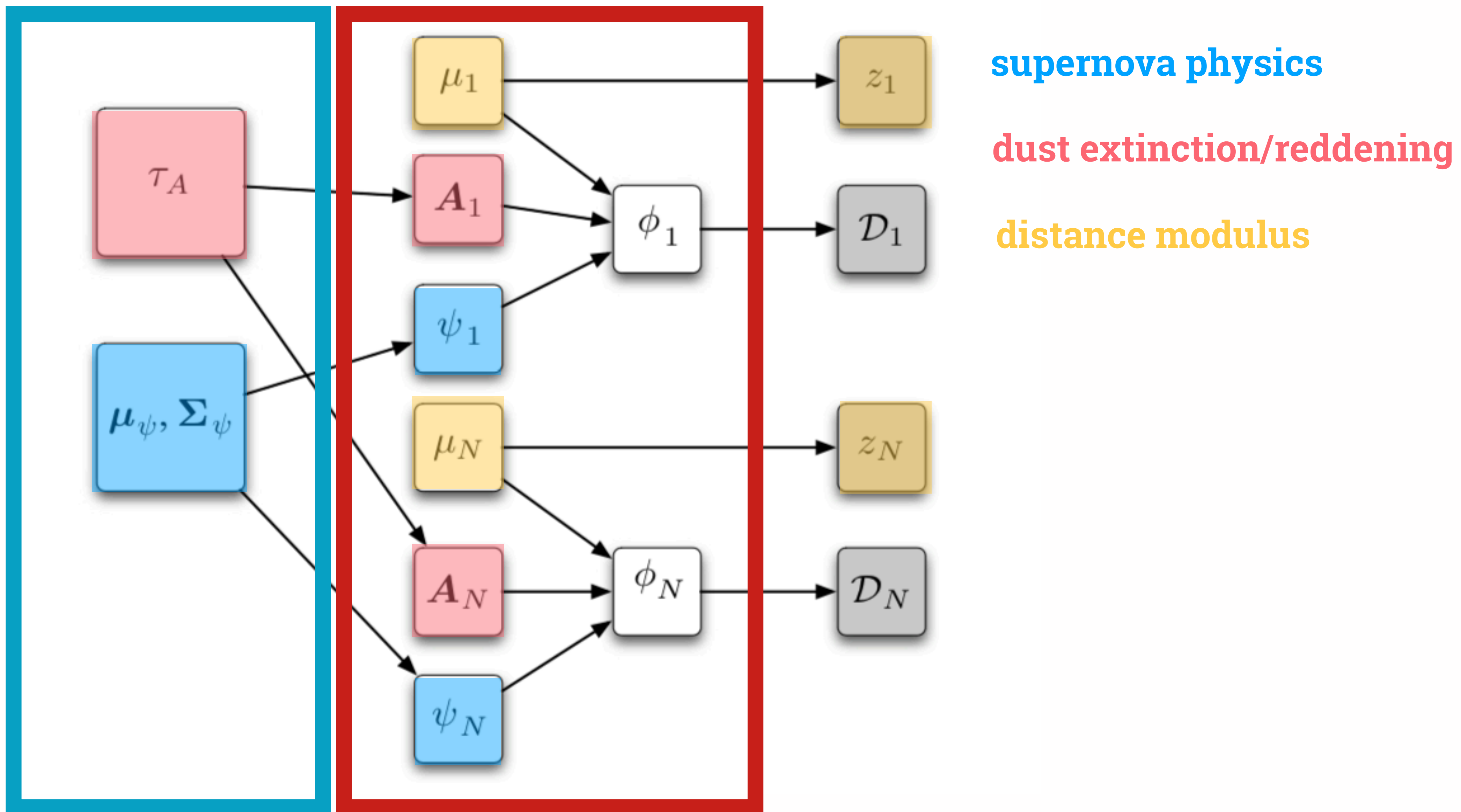
distance modulus





population-level  
(global) parameters

single-object (local)  
parameters



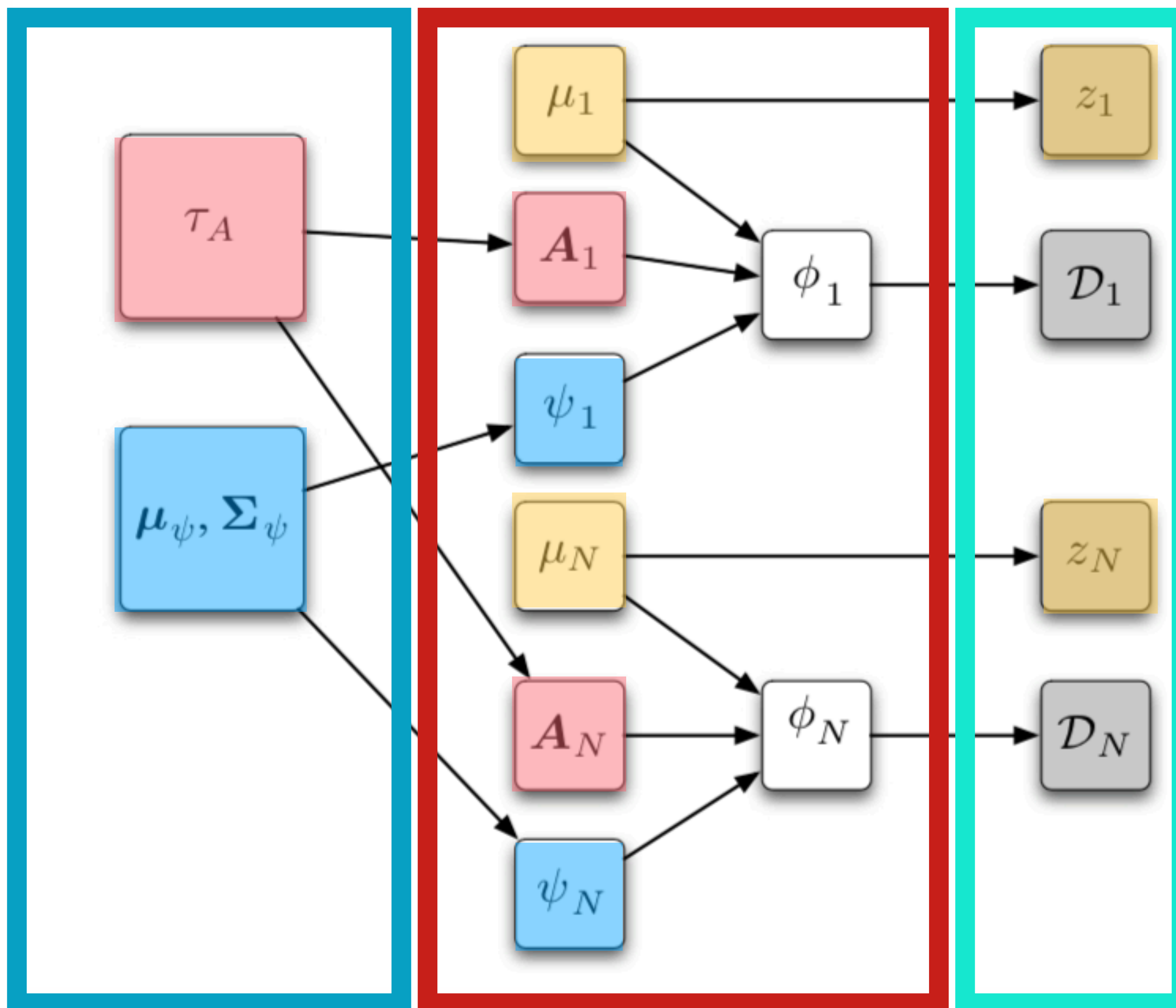




population-level  
(global) parameters

single-object (local)  
parameters

observations



supernova physics

dust extinction/reddening

distance modulus

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  $\alpha$  along with  $\theta$

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



# Could histogram individual parameters ...

---

... but how?



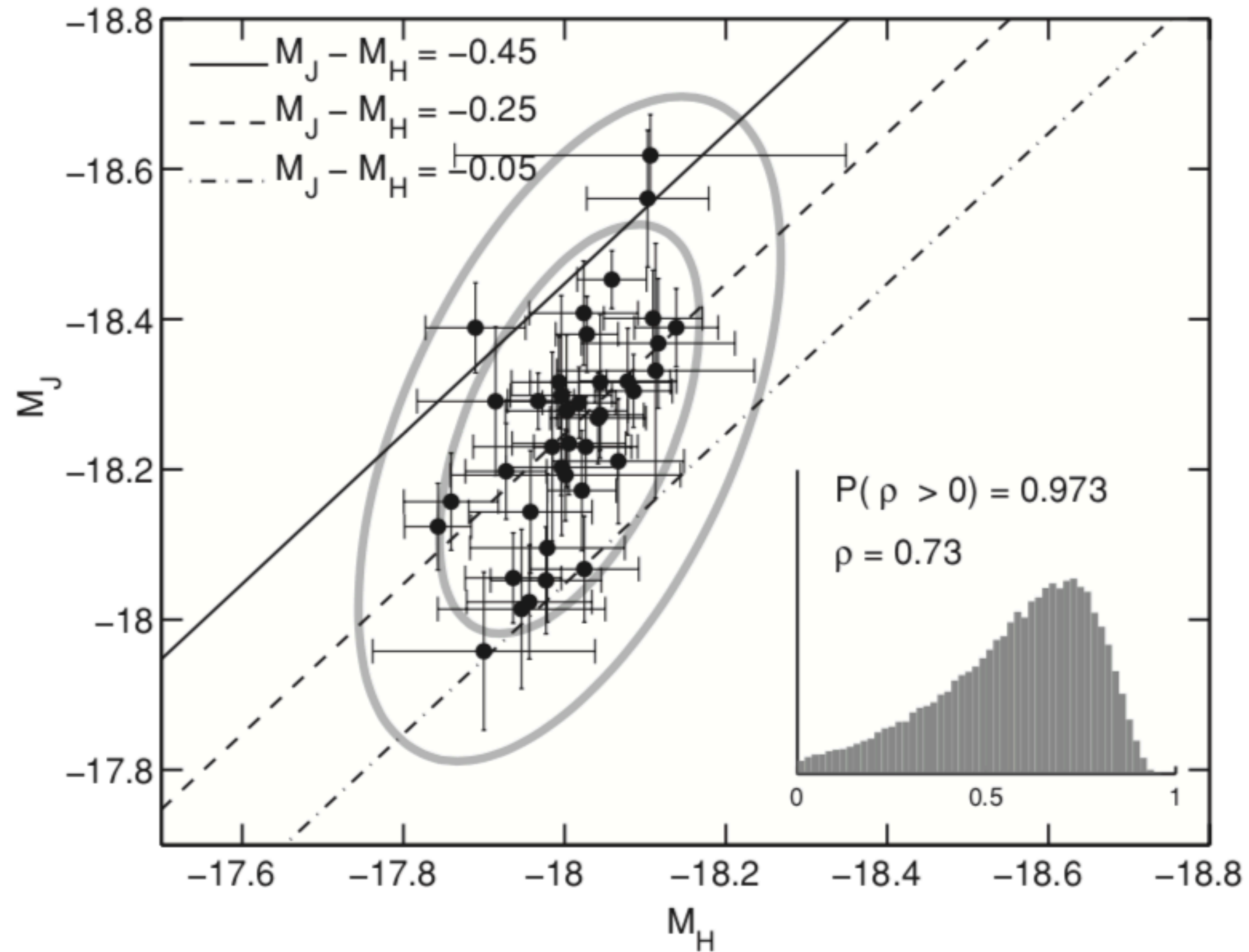
$$\begin{aligned} &P\left(\{\boldsymbol{\phi}_s, \mu_s, A_H^s, R_V^s\}; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \alpha_R \mid \mathcal{D}, \mathcal{Z}\right) \\ &\propto \left[ \prod_{s=1}^{N_{\text{SN}}} P(\boldsymbol{\phi}_s, \mu_s, A_H^s, R_V^s \mid \mathcal{D}_s, z_s; \boldsymbol{\mu}_\psi,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



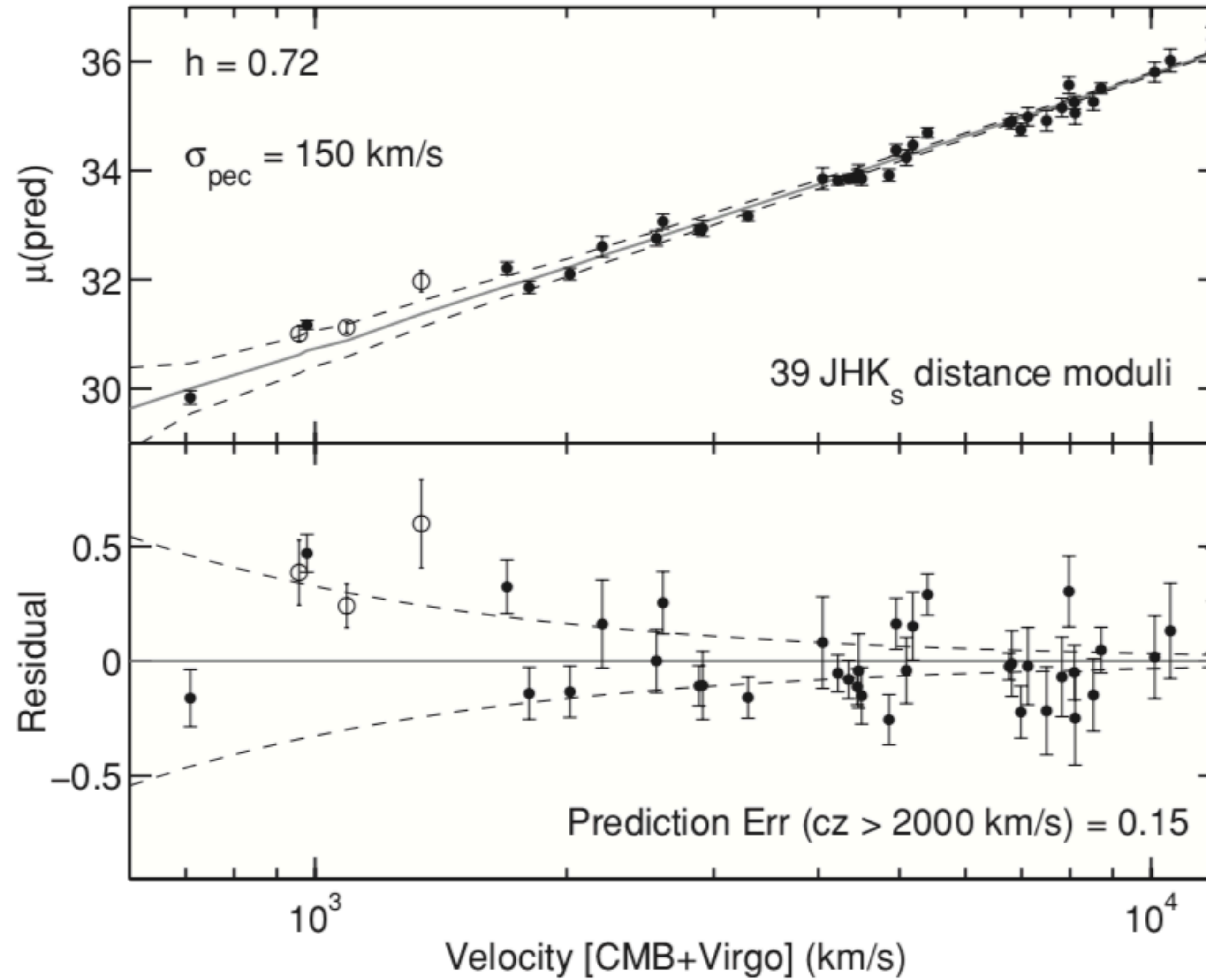
# Why?

---

- **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**



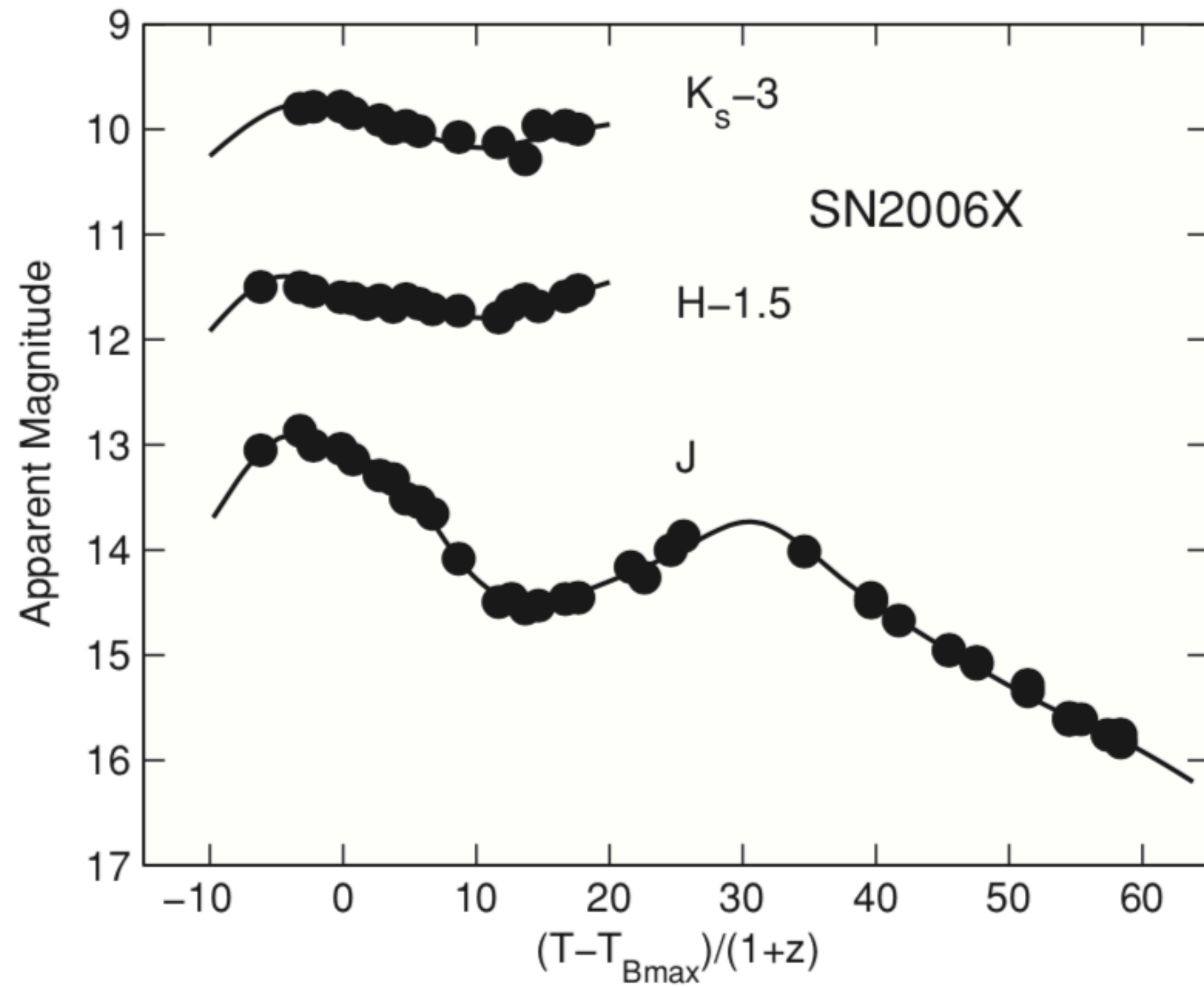
# Population Parameters



# Population Parameters



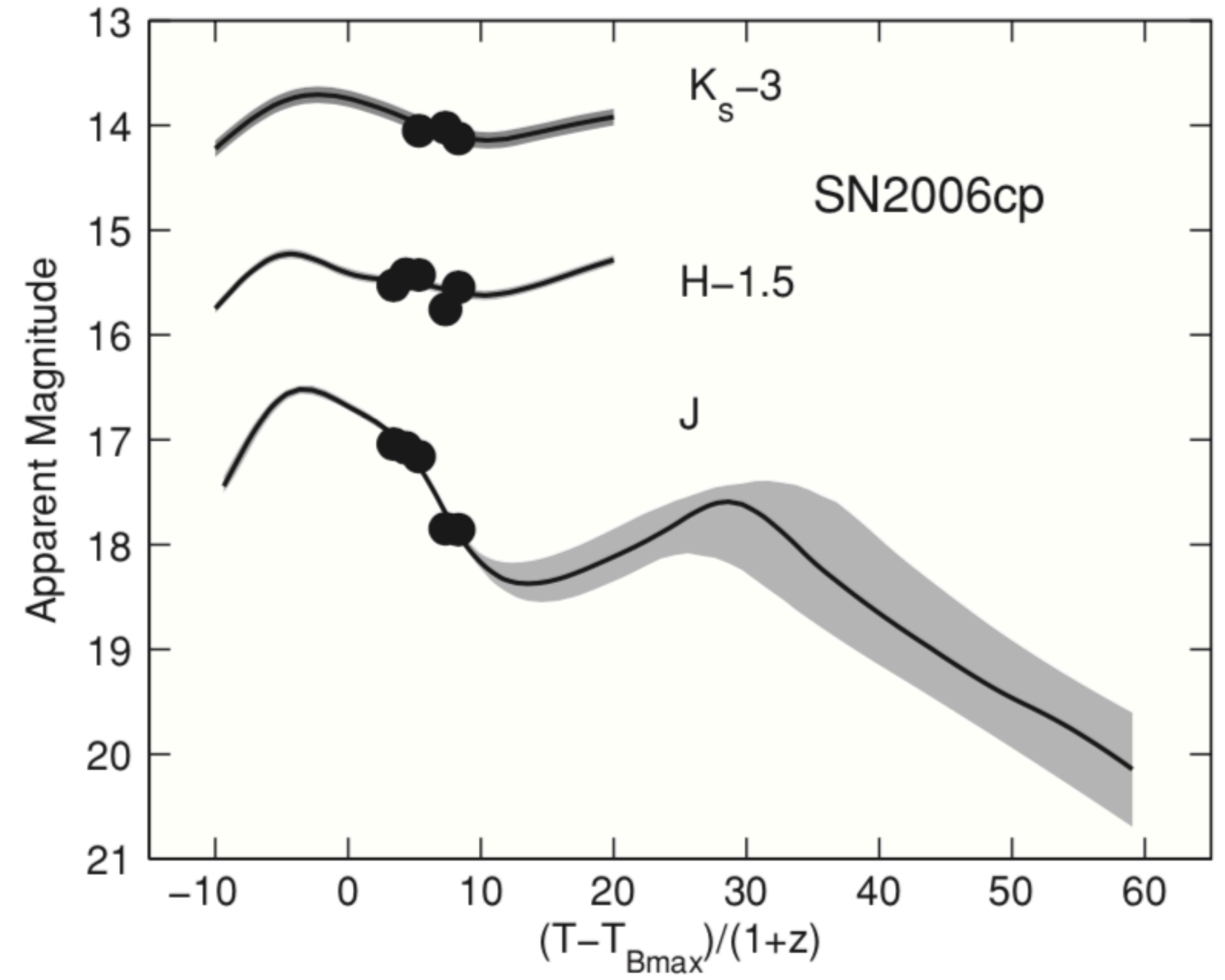
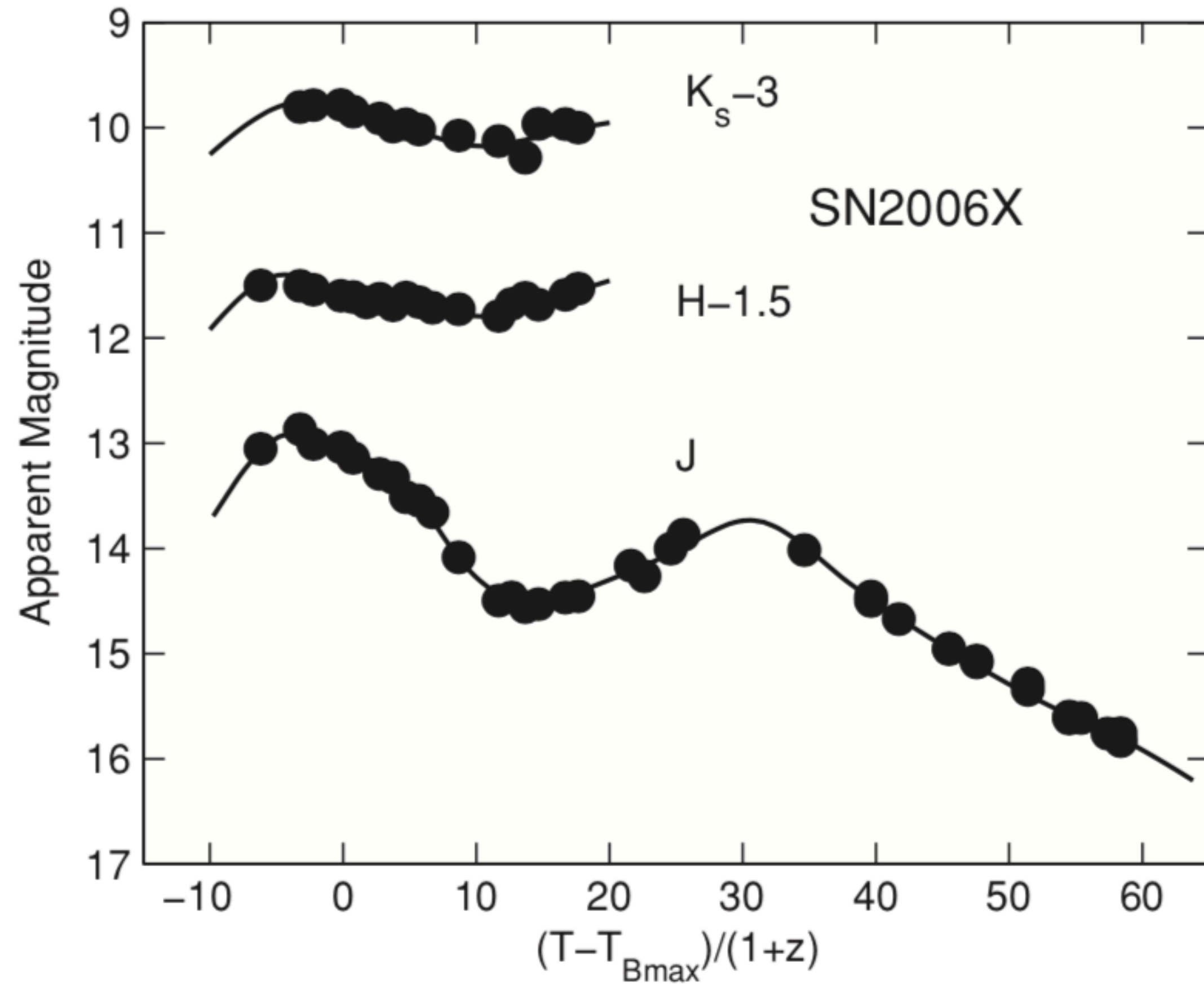
Mandel et al (2009)



# 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!

—

**[https://github.com/dhuppenkothen/cargese2018\\_tutorials](https://github.com/dhuppenkothen/cargese2018_tutorials)**

- **Option 1: git clone the repository, run notebook on your own laptop**
- **Option 2: go to website above, click on “mybinder” button in readme**
- **Option 3: go to <https://colab.research.google.com>, enter address above**
- **to save (options 2 and 3): go to File/Download as → notebook and download to your computer!**

**tutorial1/ChocolateProductivity.ipynb**