(Narratively) Generative Modeling

Michael Betancourt
Symplectomorphic, LLC
http://symplectomorphic.com

KBC Group ADAM Bootcamp
Gyöngyös, Hungary
May 20, 2026

We've seen that Bayesian inference is a mechanical way to quantify uncertainties *once we construct a full Bayesian model*.

$$\pi(y \mid \theta)$$

We've seen that Bayesian inference is a mechanical way to quantify uncertainties *once we construct a full Bayesian model.*

$$\pi(y \mid \theta) \, \pi(\theta)$$

We've seen that Bayesian inference is a mechanical way to quantify uncertainties *once we construct a full Bayesian model*.

$$\pi(y, \theta) = \pi(y \mid \theta) \, \pi(\theta)$$

We've seen that Bayesian inference is a mechanical way to quantify uncertainties *once we construct a full Bayesian model*.

$$\pi(\theta \mid \tilde{y}) \propto \pi(\tilde{y}, \theta)$$

We've seen that Bayesian inference is a mechanical way to quantify uncertainties *once we construct a full Bayesian model*.

$$\mathbb{E}_{\pi}[f] = \int \mathrm{d}\theta \, \pi(\theta \mid \tilde{y}) \, f(\theta)$$

We've seen that Bayesian inference is a mechanical way to quantify uncertainties *once we construct a full Bayesian model*.

$$\mathbb{E}_{\pi}[f] \approx \int \mathrm{d}\theta \, \pi(\theta \mid \tilde{y}) \, f(\theta)$$

But how do we construct a meaningful full Bayesian model, especially when the spaces are high-dimensional?

$$\pi(y_1, \ldots, y_n, \ldots, y_N, \theta_1, \ldots, \theta_i, \ldots, \theta_I)$$

Conditional decompositions of the full Bayesian model can reduce the modeling problem to more manageable pieces.

$$\pi(y_1, \ldots, y_N, \theta_1, \ldots, \theta_I) =$$

$$\prod_{n=1}^{N} \pi(y_n \mid y_1, \ldots, y_{n-1}, \theta_1, \ldots, \theta_I)$$

$$\cdot \prod_{i=1}^{I} \pi(\theta_i \mid \theta_1 \ldots \theta_{i-1})$$

Directed graphical models are a helpful way to communicate the structure of a conditional decomposition.

$$\pi(y_1, y_2, \theta_1, \theta_2, \theta_3) =$$

y_1

y_2

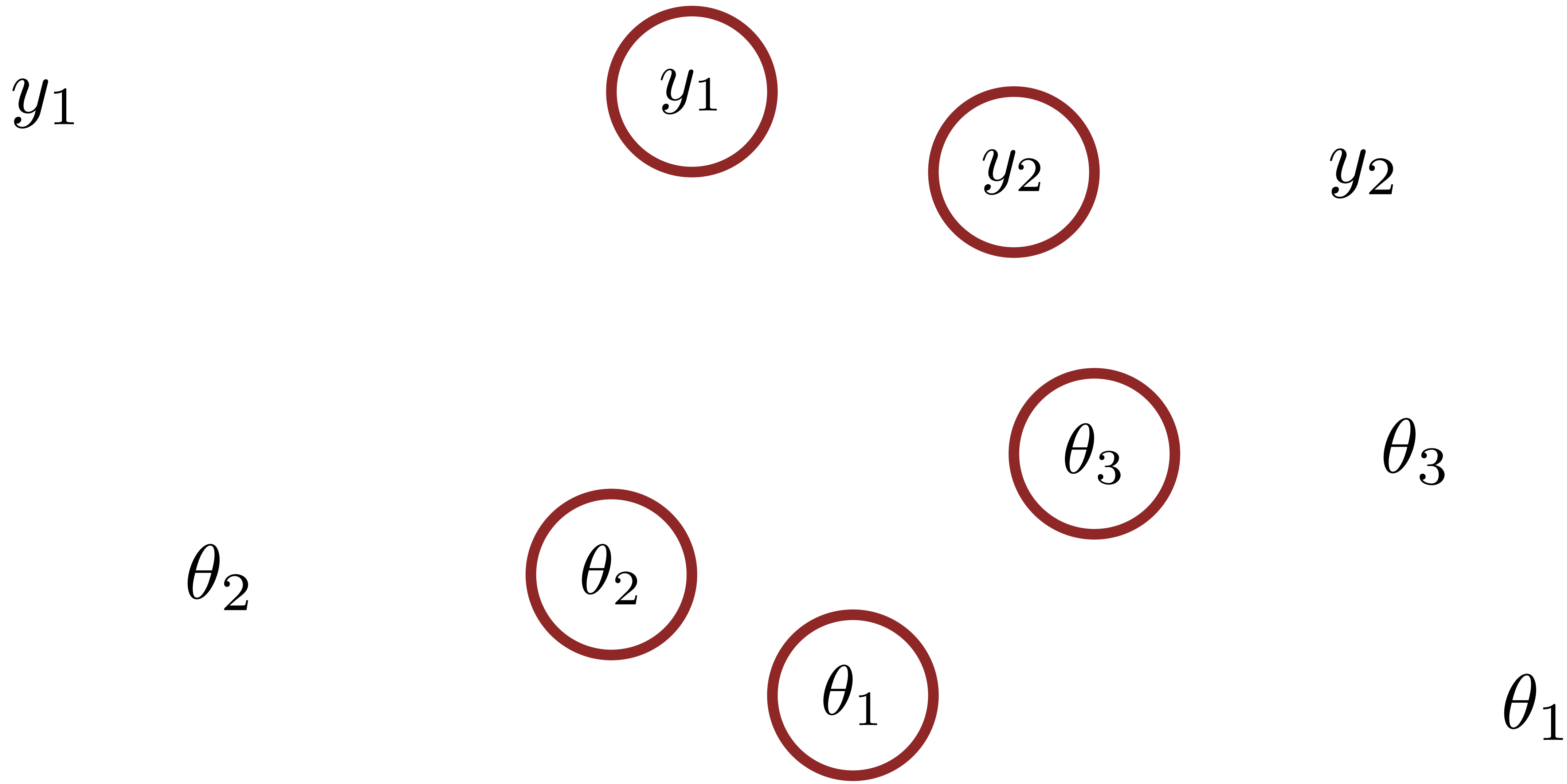
θ_3

θ_2

θ_1

Directed graphical models are a helpful way to communicate the structure of a conditional decomposition.

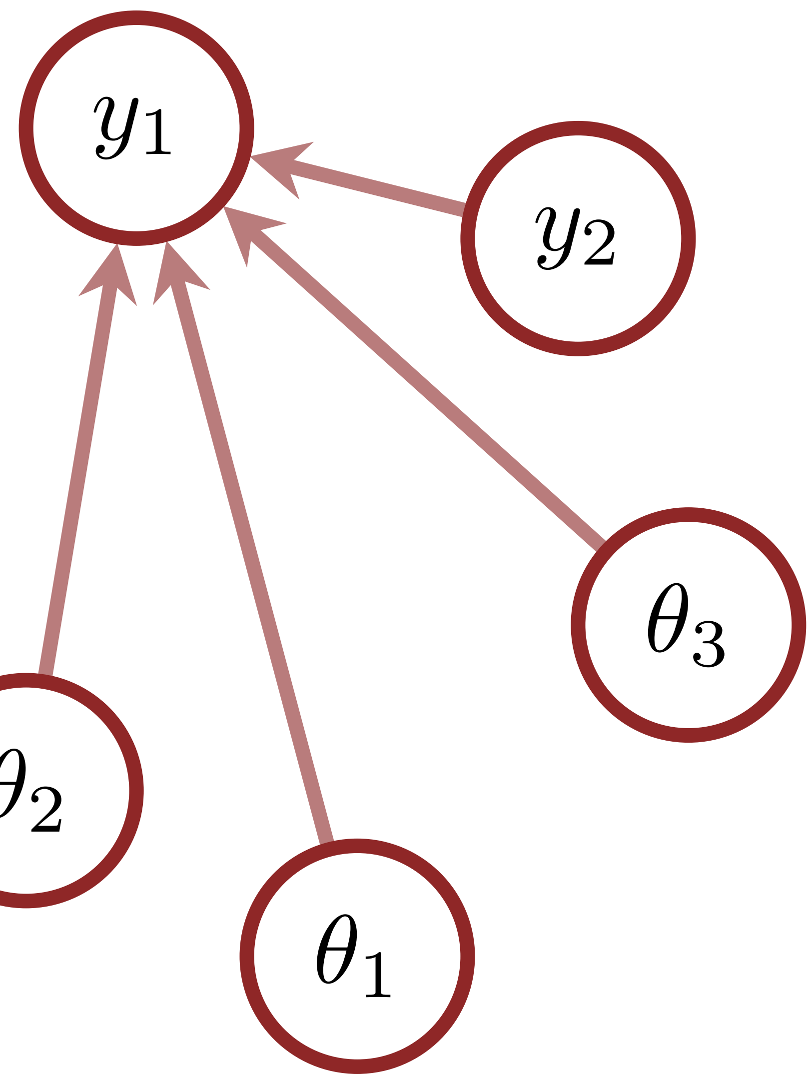
$$\pi(y_1, y_2, \theta_1, \theta_2, \theta_3) =$$



Directed graphical models are a helpful way to communicate the structure of a conditional decomposition.

$$\pi(y_1, y_2, \theta_1, \theta_2, \theta_3) =$$

$$\pi(y_1 \mid y_2, \theta_1, \theta_2, \theta_3)$$



y_2

θ_3

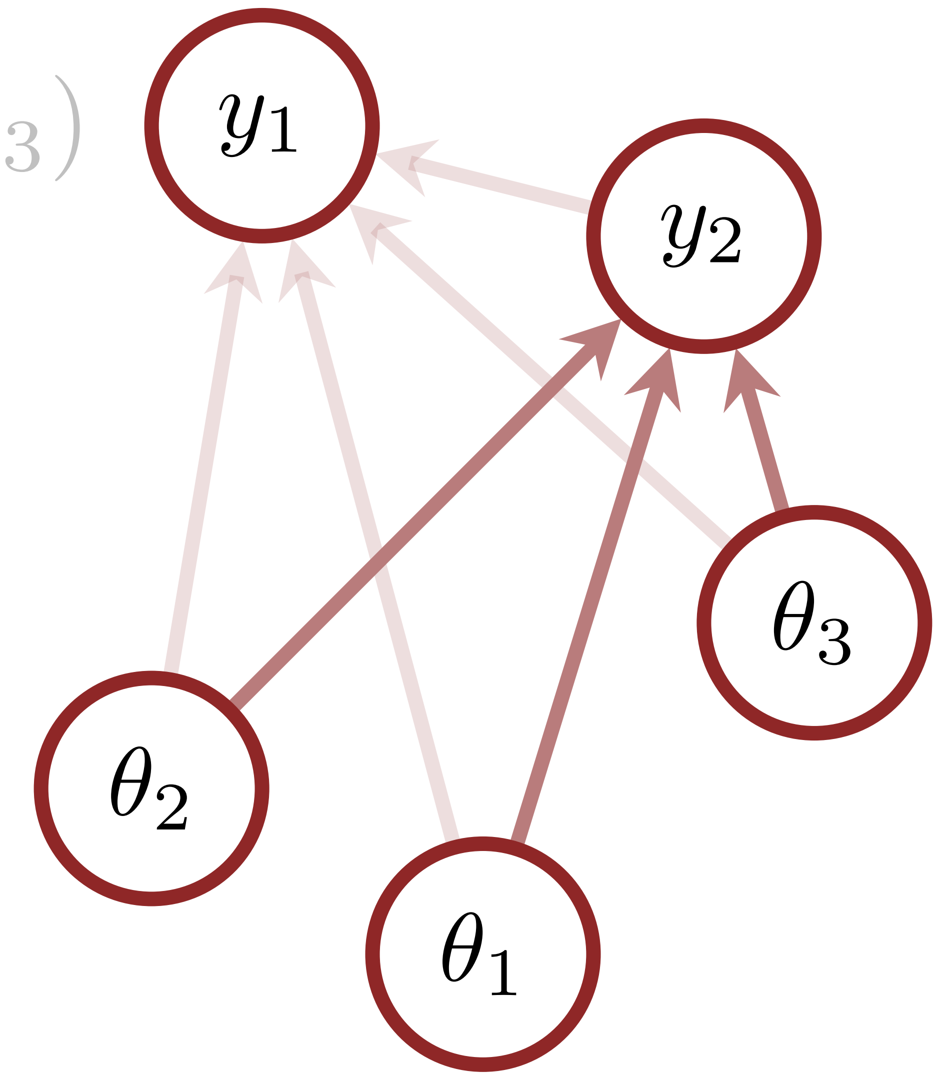
θ_2

θ_1

Directed graphical models are a helpful way to communicate the structure of a conditional decomposition.

$$\pi(y_1, y_2, \theta_1, \theta_2, \theta_3) =$$

$$\pi(y_1 \mid y_2, \theta_1, \theta_2, \theta_3)$$



$$\cdot \, \pi(y_2 \mid \theta_1, \theta_2, \theta_3)$$

$$\theta_3$$

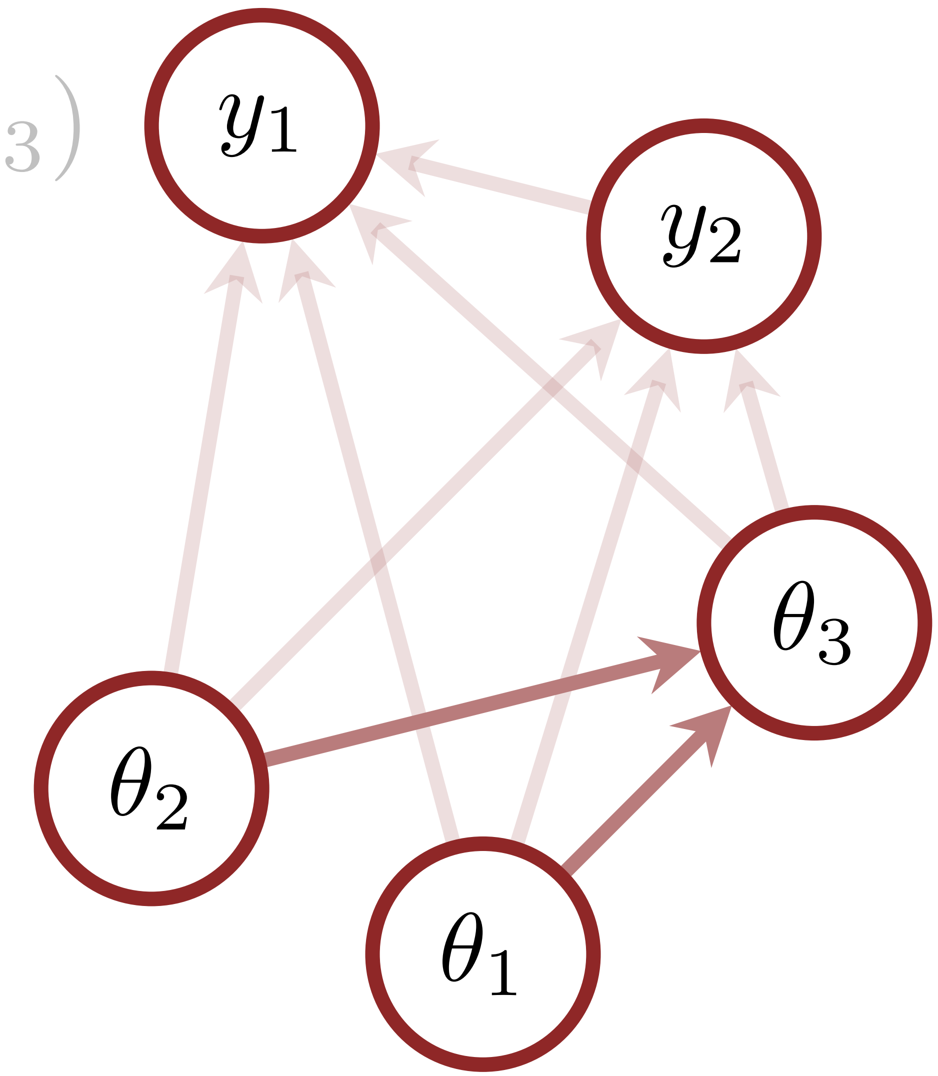
$$\theta_2$$

$$\theta_1$$

Directed graphical models are a helpful way to communicate the structure of a conditional decomposition.

$$\pi(y_1, y_2, \theta_1, \theta_2, \theta_3) =$$

$$\pi(y_1 \mid y_2, \theta_1, \theta_2, \theta_3)$$



$$\cdot \, \pi(y_2 \mid \theta_1, \theta_2, \theta_3)$$

$$\cdot \, \pi(\theta_3 \mid \theta_1, \theta_2)$$

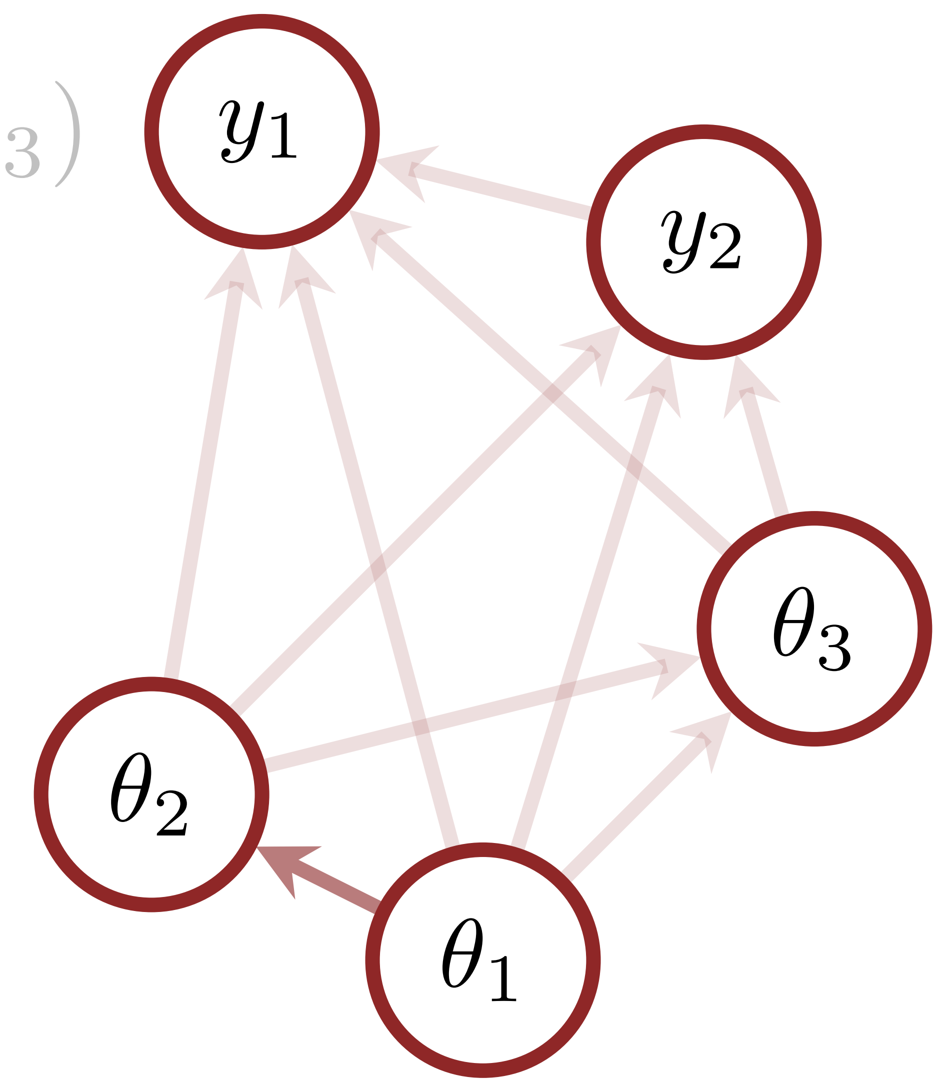
$$\theta_2$$

$$\theta_1$$

Directed graphical models are a helpful way to communicate the structure of a conditional decomposition.

$$\pi(y_1, y_2, \theta_1, \theta_2, \theta_3) =$$

$$\pi(y_1 \mid y_2, \theta_1, \theta_2, \theta_3)$$



$$\cdot \, \pi(y_2 \mid \theta_1, \theta_2, \theta_3)$$

$$\cdot \, \pi(\theta_3 \mid \theta_1, \theta_2)$$

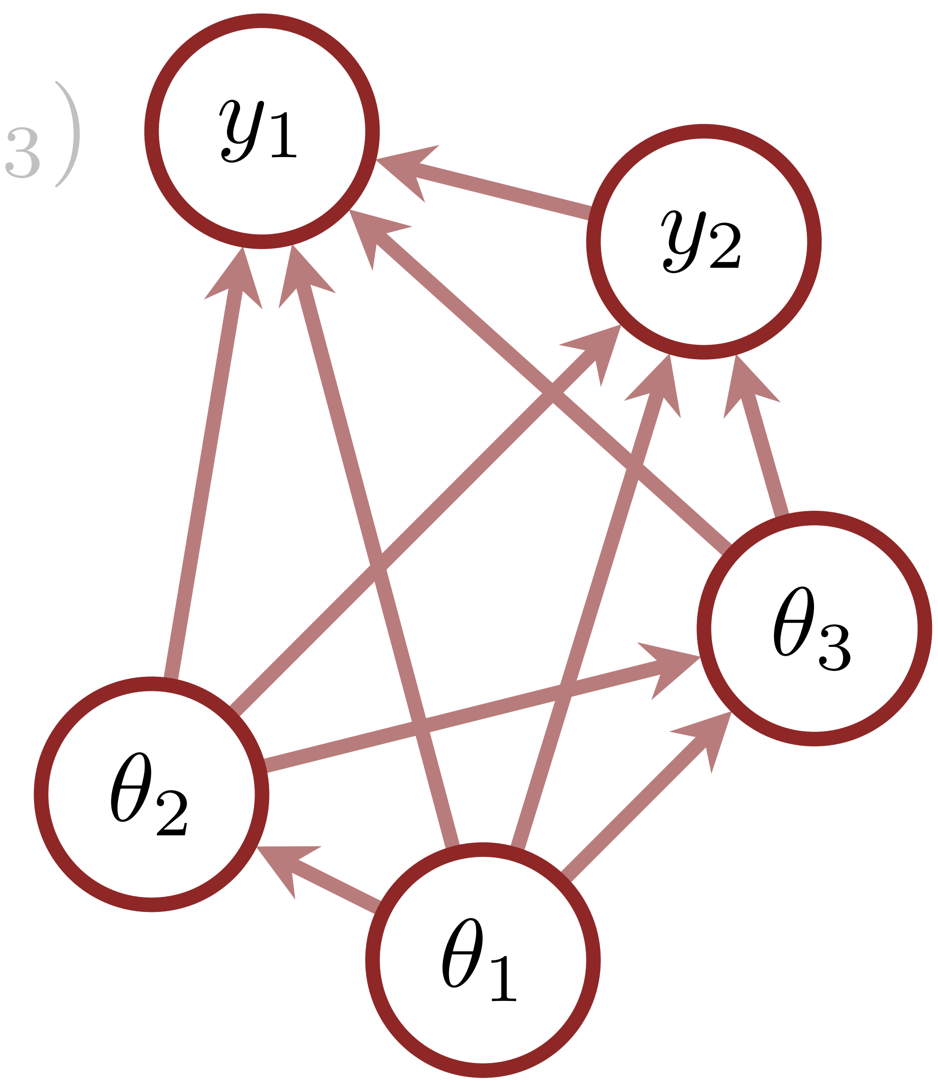
$$\cdot \, \pi(\theta_2 \mid \theta_1)$$

$$\theta_1$$

Directed graphical models are a helpful way to communicate the structure of a conditional decomposition.

$$\pi(y_1, y_2, \theta_1, \theta_2, \theta_3) =$$

$$\pi(y_1 \mid y_2, \theta_1, \theta_2, \theta_3)$$



$$\cdot\, \pi(y_2 \mid \theta_1, \theta_2, \theta_3)$$

$$\cdot\, \pi(\theta_3 \mid \theta_1, \theta_2)$$

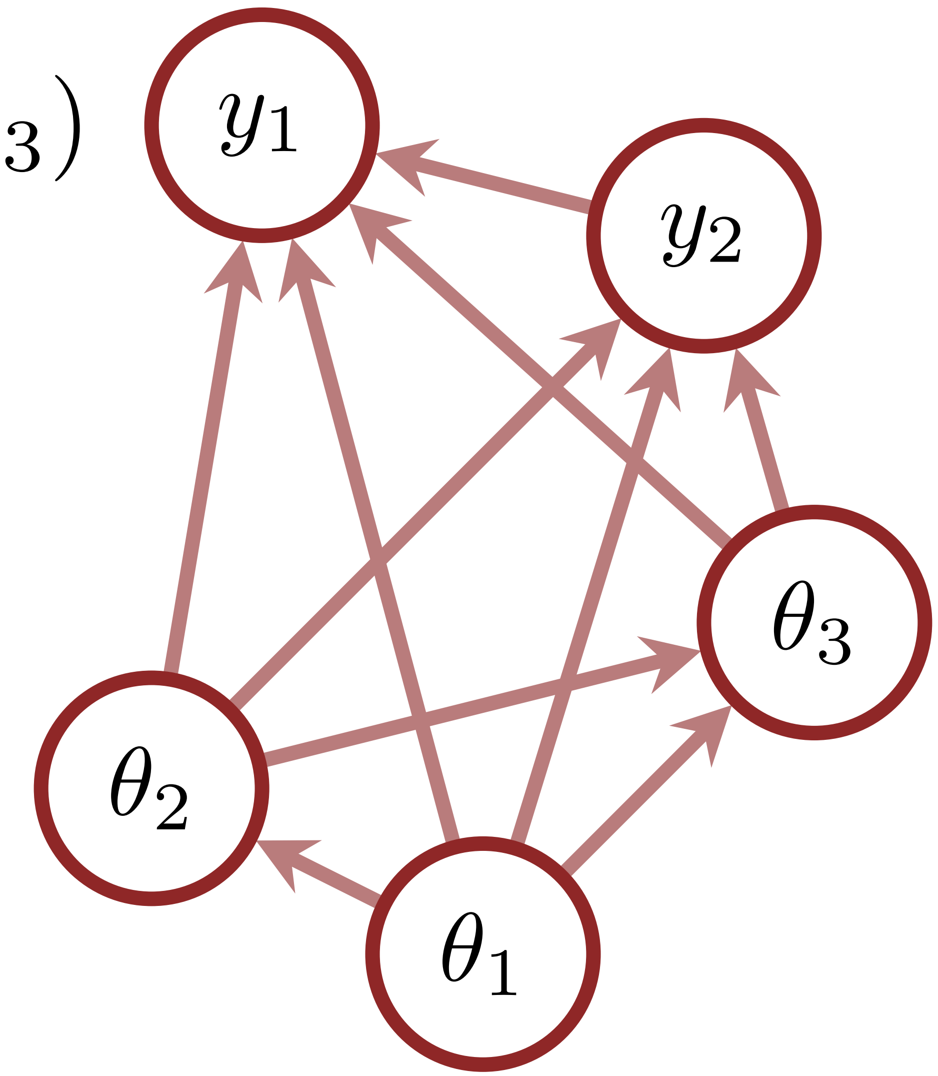
$$\cdot\, \pi(\theta_2 \mid \theta_1)$$

$$\cdot\, \pi(\theta_1)$$

Directed graphical models are a helpful way to communicate the structure of a conditional decomposition.

$$\pi(y_1, y_2, \theta_1, \theta_2, \theta_3) =$$

$$\pi(y_1 \mid y_2, \theta_1, \theta_2, \theta_3)$$



$$\cdot \, \pi(y_2 \mid \theta_1, \theta_2, \theta_3)$$

$$\cdot \, \pi(\theta_3 \mid \theta_1, \theta_2)$$

$$\cdot \, \pi(\theta_2 \mid \theta_1)$$

$$\cdot \, \pi(\theta_1)$$

Directed graphical models are a helpful way to communicate the structure of a conditional decomposition.

$$\pi(y_1, y_2, \theta_1, \theta_2, \theta_3) =$$

y_1

y_2

θ_3

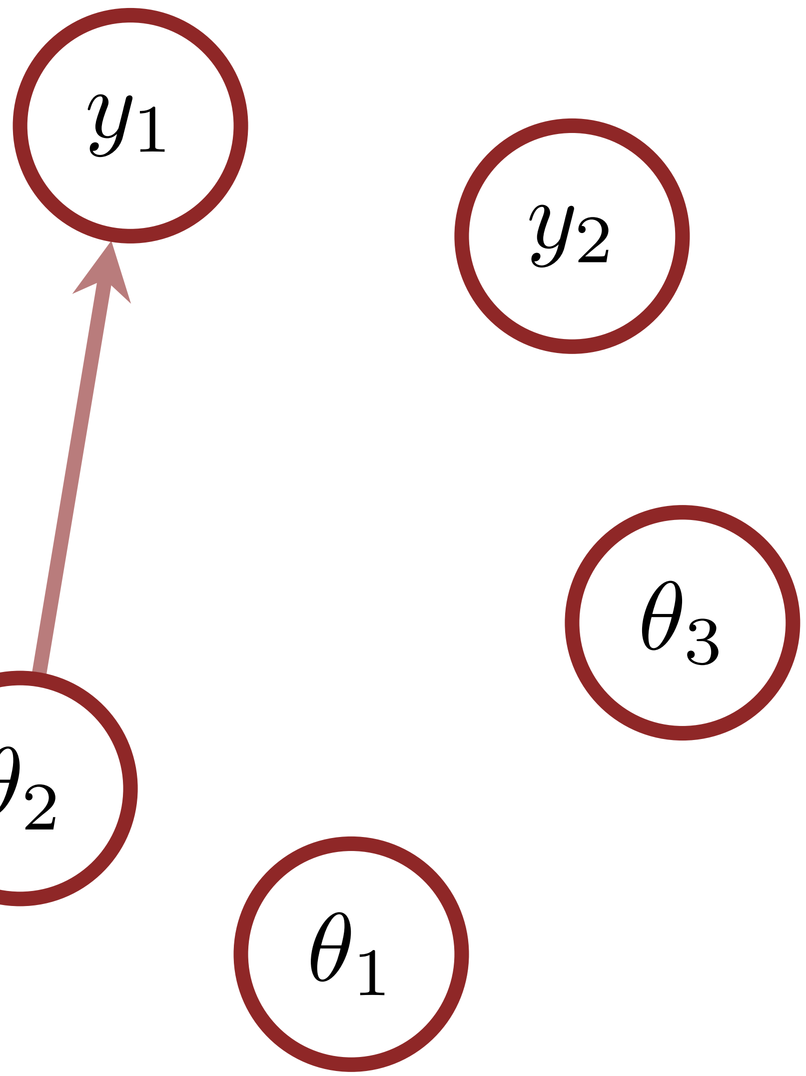
θ_2

θ_1

Directed graphical models are a helpful way to communicate the structure of a conditional decomposition.

$$\pi(y_1, y_2, \theta_1, \theta_2, \theta_3) =$$

$\pi(y_1 | \theta_2)$



y_2

θ_3

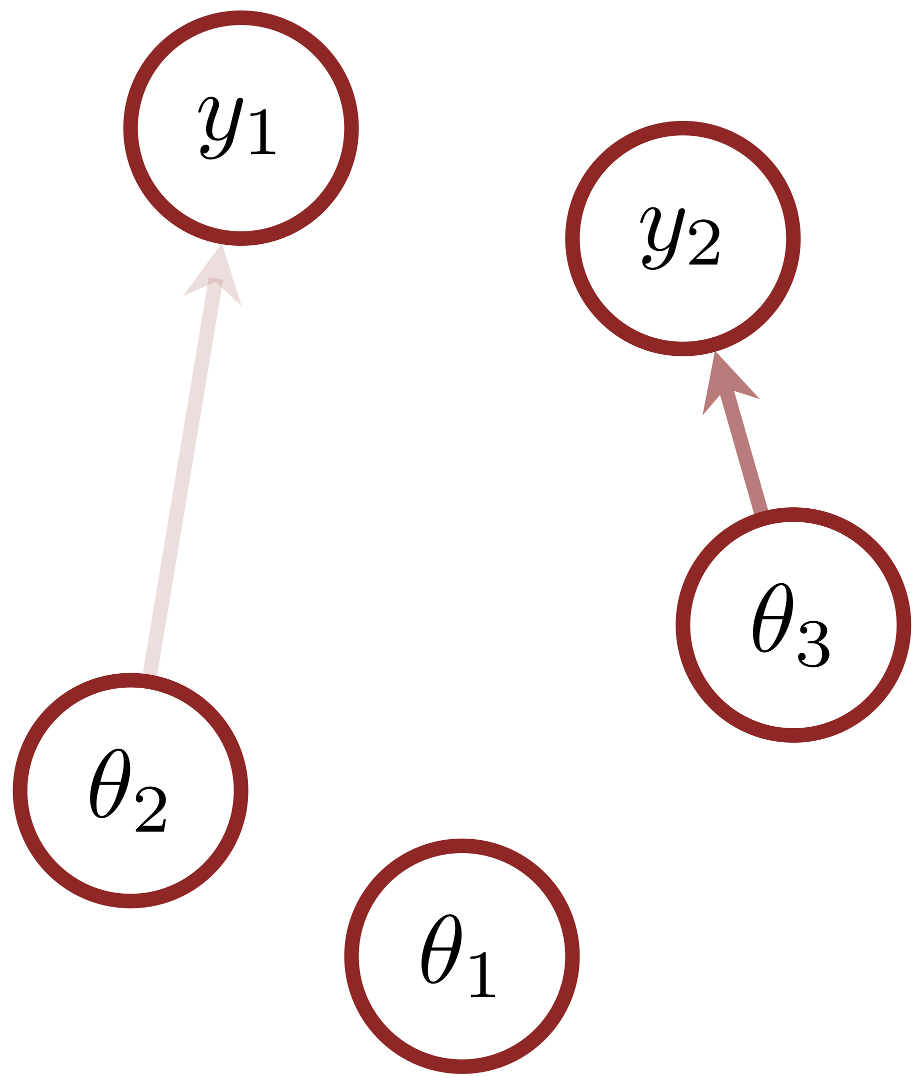
θ_2

θ_1

Directed graphical models are a helpful way to communicate the structure of a conditional decomposition.

$$\pi(y_1, y_2, \theta_1, \theta_2, \theta_3) =$$

$$\pi(y_1 | \theta_2)$$



$$\cdot \, \pi(y_2 | \theta_3)$$

$$\theta_3$$

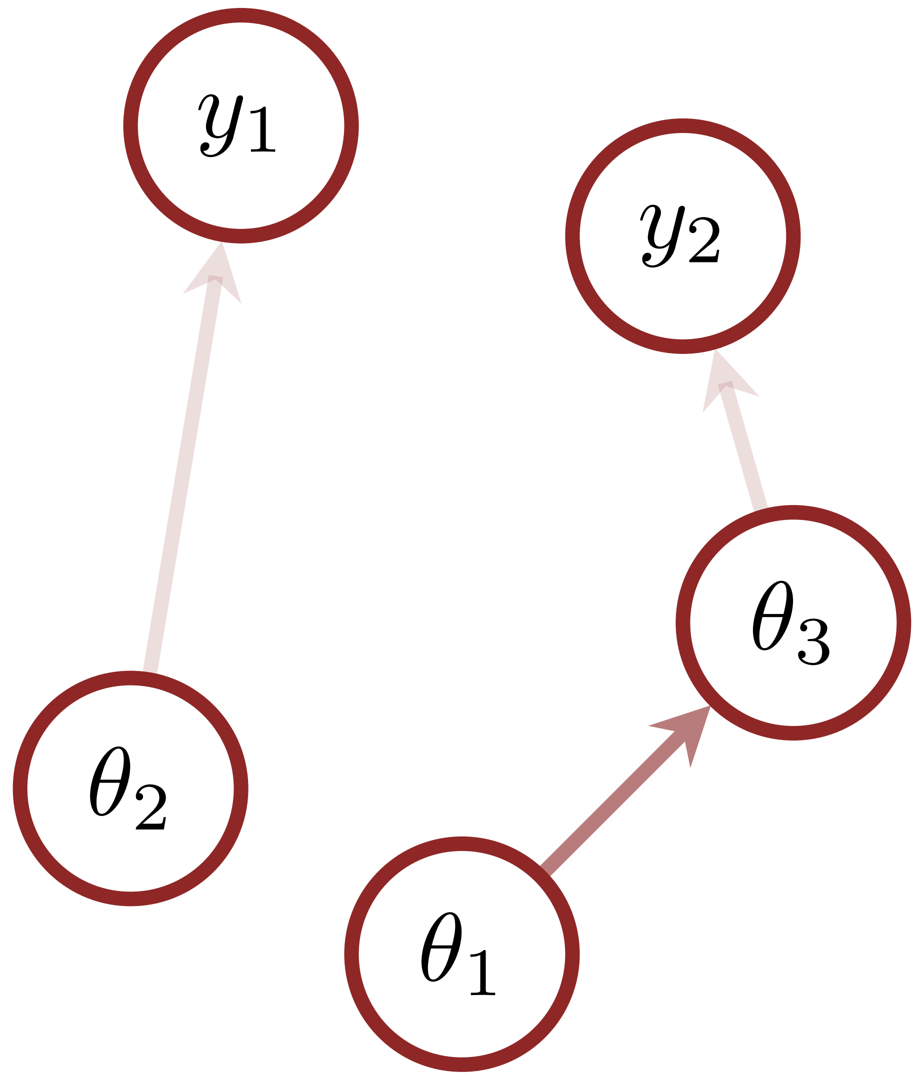
$$\theta_2$$

$$\theta_1$$

Directed graphical models are a helpful way to communicate the structure of a conditional decomposition.

$$\pi(y_1, y_2, \theta_1, \theta_2, \theta_3) =$$

$$\pi(y_1 | \theta_2)$$



$$\cdot \, \pi(y_2 | \theta_3)$$

$$\cdot \, \pi(\theta_3 | \theta_1)$$

$$\theta_2$$

$$\theta_1$$

Directed graphical models are a helpful way to communicate the structure of a conditional decomposition.

$$\pi(y_1, y_2, \theta_1, \theta_2, \theta_3) =$$

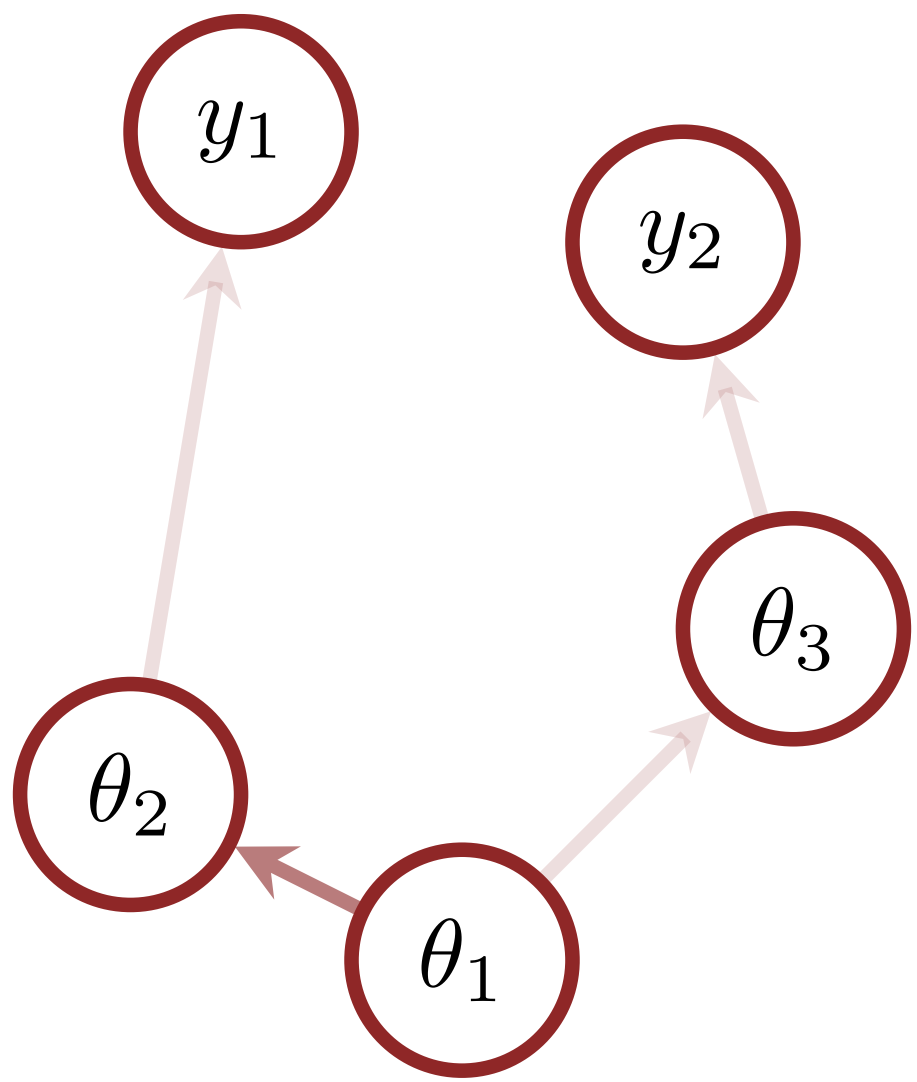
$$\pi(y_1 | \theta_2)$$

$$\cdot \, \pi(y_2 | \theta_3)$$

$$\cdot \, \pi(\theta_3 | \theta_1)$$

$$\cdot \, \pi(\theta_2 | \theta_1)$$

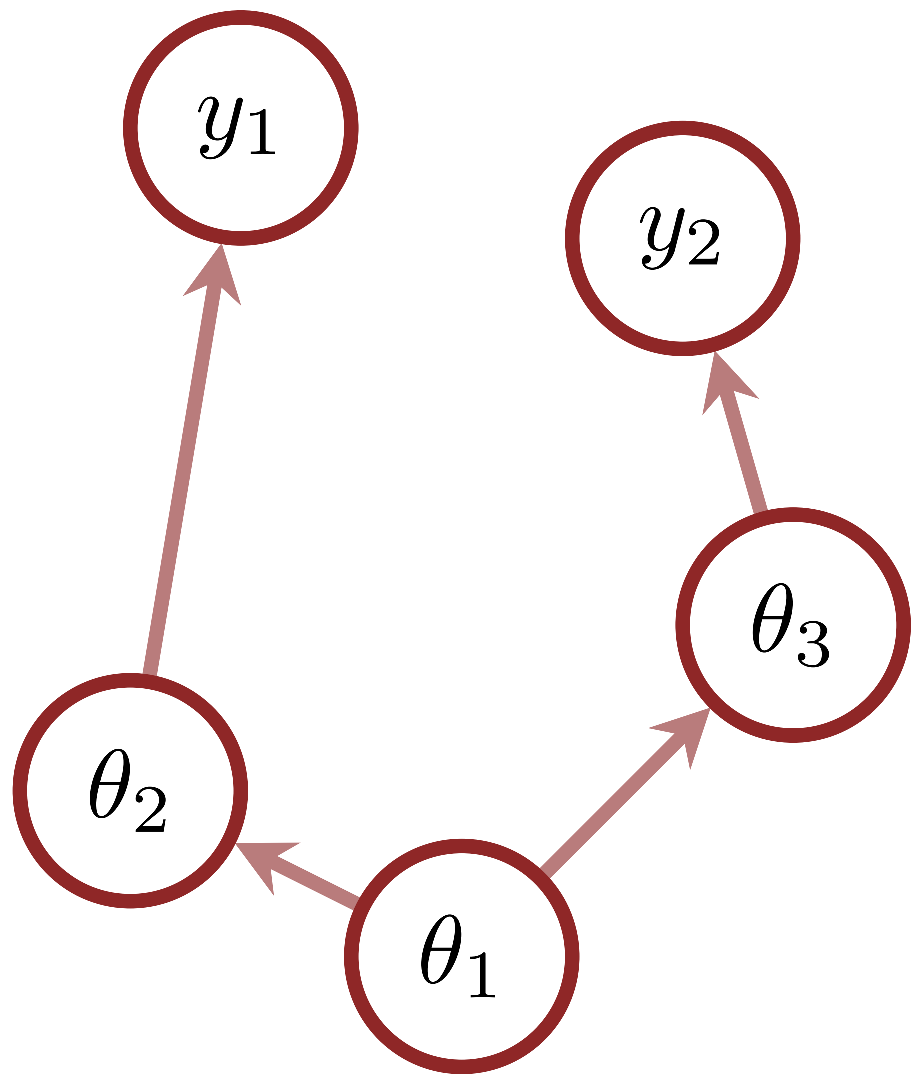
$$\theta_1$$



Directed graphical models are a helpful way to communicate the structure of a conditional decomposition.

$$\pi(y_1, y_2, \theta_1, \theta_2, \theta_3) =$$

$$\pi(y_1 | \theta_2)$$



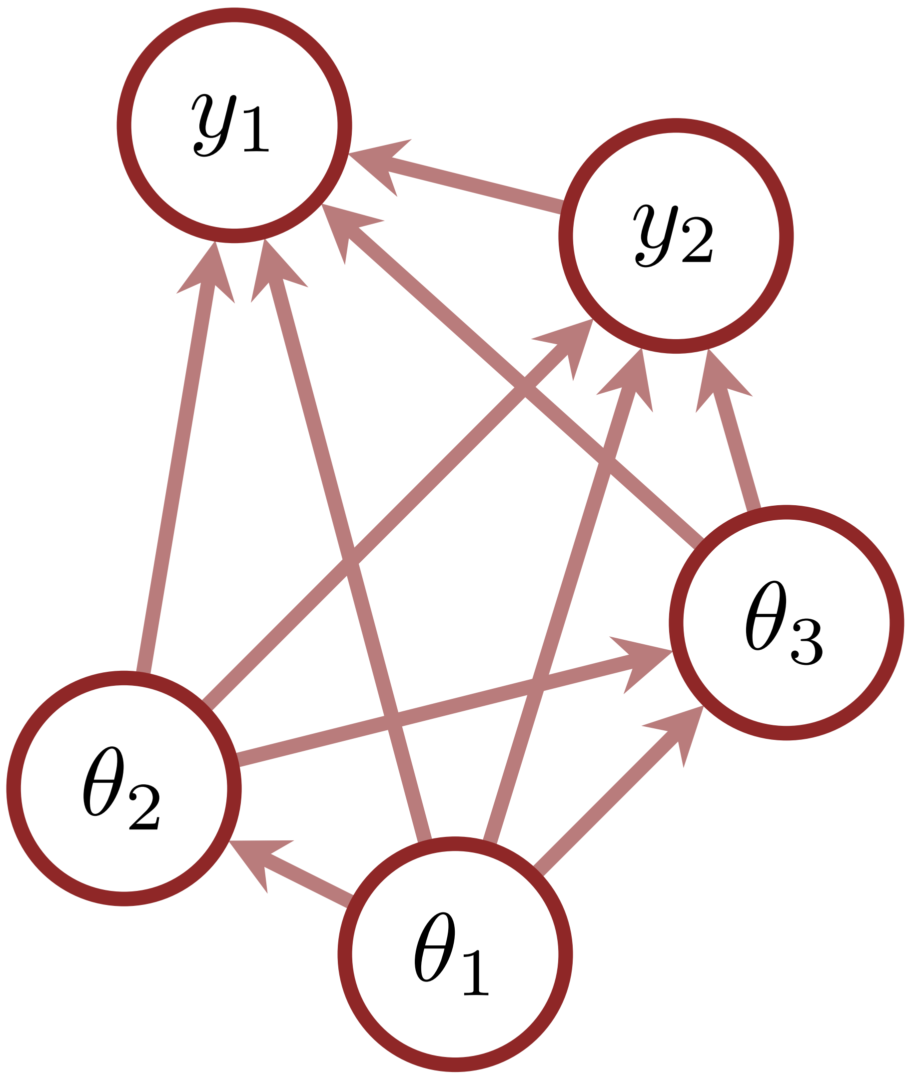
$$\cdot \, \pi(y_2 | \theta_3)$$

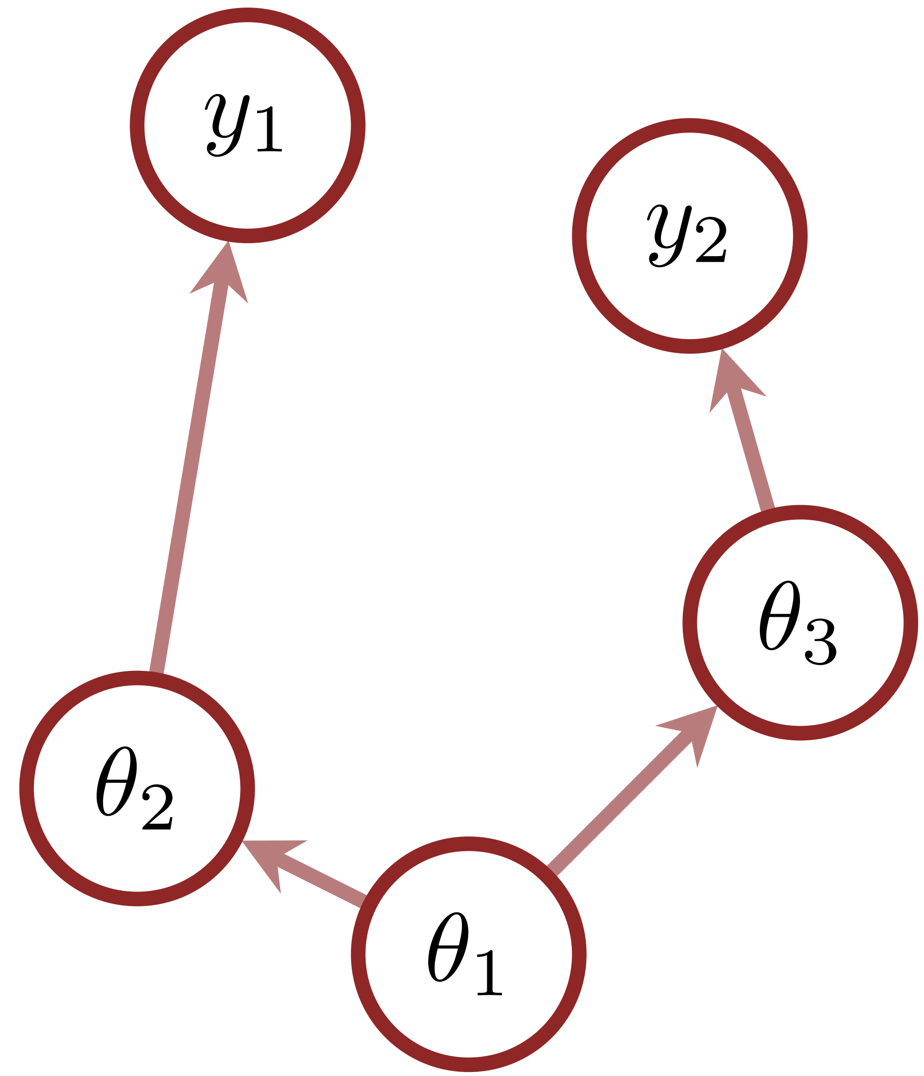
$$\cdot \, \pi(\theta_3 | \theta_1)$$

$$\cdot \, \pi(\theta_2 | \theta_1)$$

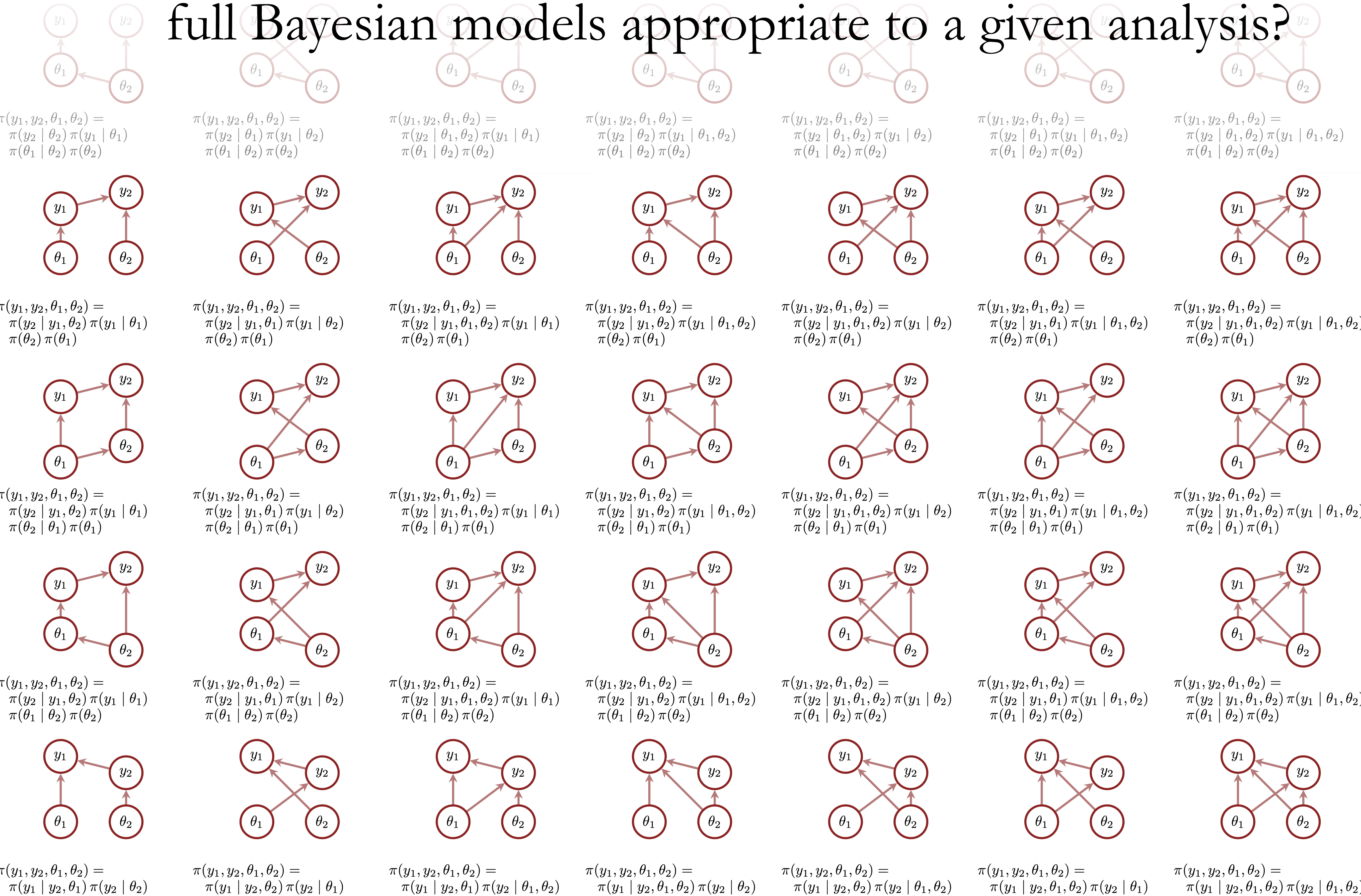
$$\cdot \, \pi(\theta_1)$$

Directed graphical models are a helpful way to communicate the structure of a conditional decomposition.

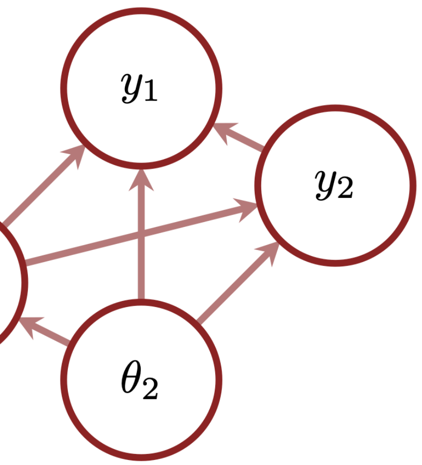


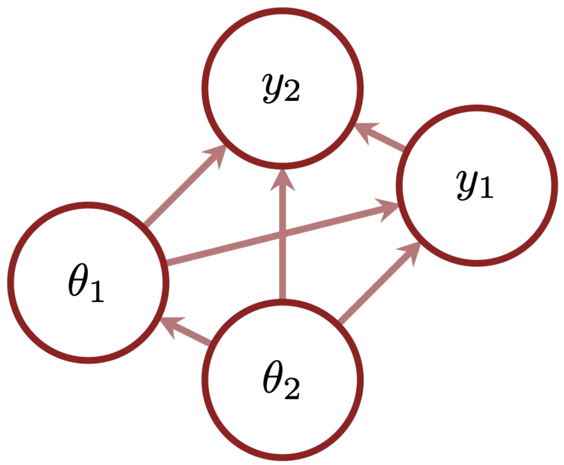


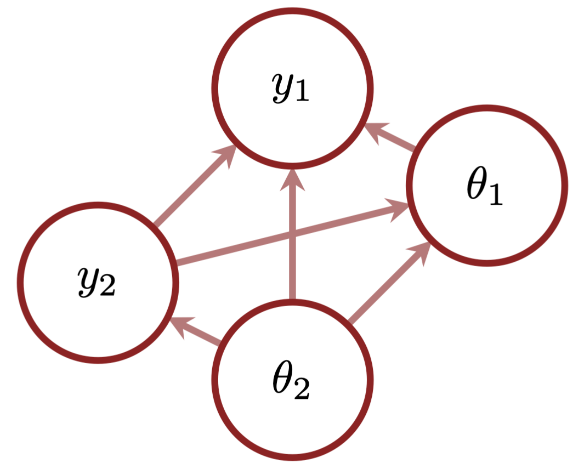
What conditional decompositions, however, result in full Bayesian models appropriate to a given analysis?

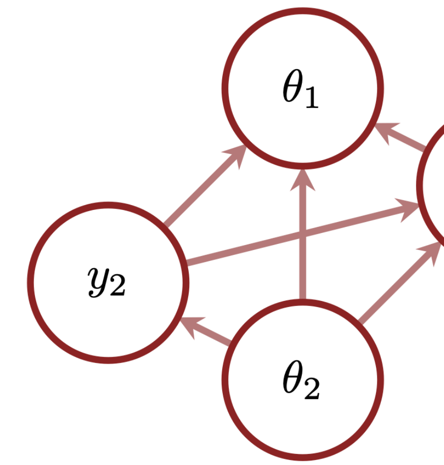


Procedurally Generative Modeling







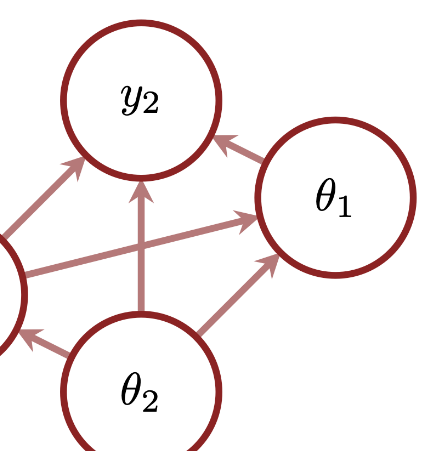


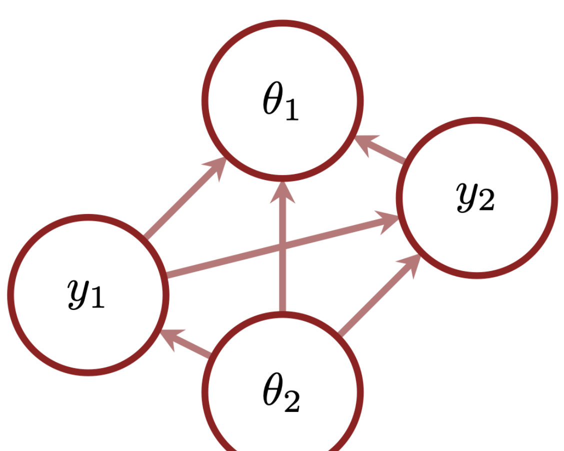
$$_1, \theta_2) = \pi(y_1 \mid y_2, \theta_1, \theta_2) \; \pi(y_2 \mid \theta_1, \theta_2) \; \pi(\theta_1 \mid \theta_2) \; \pi(\theta_2)$$

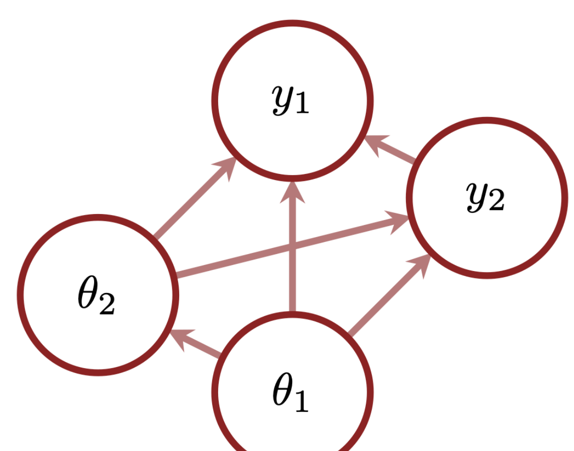
$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_2 \mid y_1, \theta_1, \theta_2) \; \pi(y_1 \mid \theta_1, \theta_2) \; \pi(\theta_1 \mid \theta_2) \; \pi(\theta_2)$$

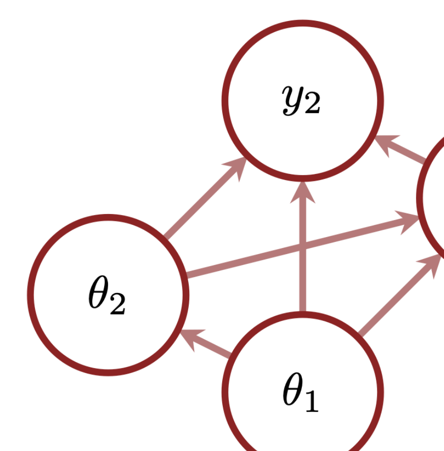
$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1 \mid \theta_1, y_2, \theta_2) \; \pi(\theta_1 \mid y_2, \theta_2) \; \pi(y_2 \mid \theta_2) \; \pi(\theta_2)$$

$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(\theta_1 \mid \; \pi(y_1 \mid \; \pi(y_2 \mid \; \pi(\theta_2)$$









Procedurally generative models are defined through an explicit, exact sampling mechanism.

$$\{\tilde{y}, \tilde{\theta}\} \sim \pi(y, \theta)$$

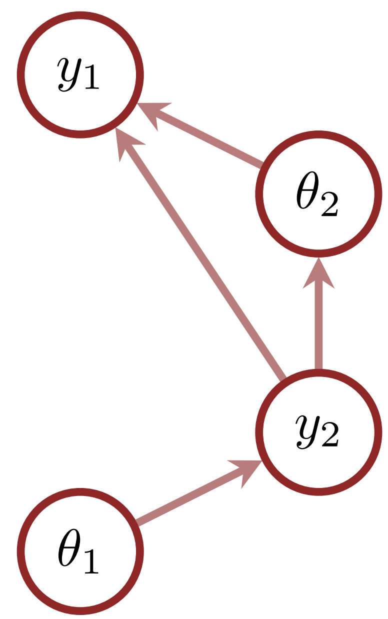
A procedurally generative model specification might push forward latent samples through a complicated function.

$$\pi(y, \theta) = \phi_{*} \pi(x)$$

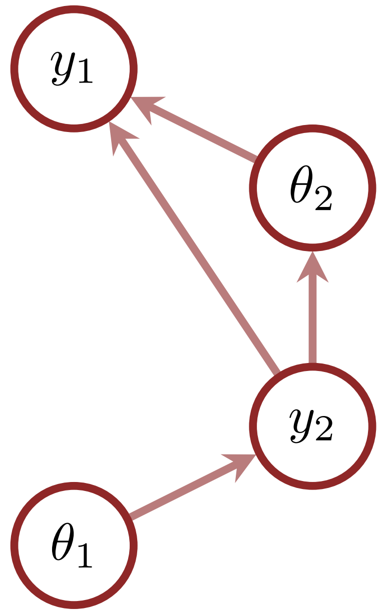
$$\tilde{x} \sim \pi(x)$$

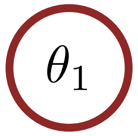
$$\left\{ \tilde{y}, \tilde{\theta} \right\} = \phi(\tilde{x})$$

Some conditional decompositions allow us generate joint samples with ancestral sampling.



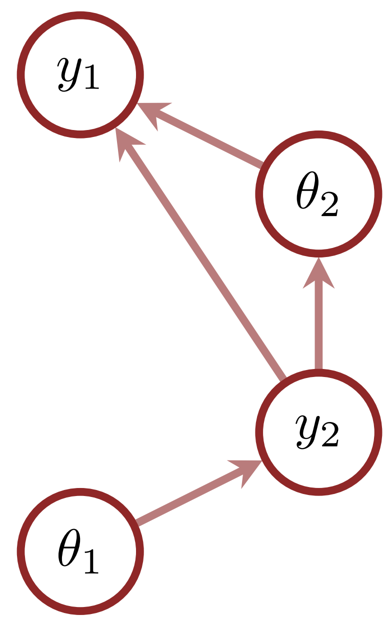
Some conditional decompositions allow us generate joint samples with ancestral sampling.

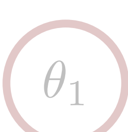


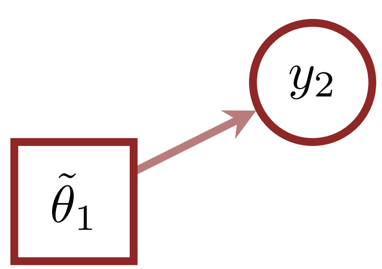


$$\tilde{\theta}_1 \sim \pi(\theta_1)$$

Some conditional decompositions allow us generate joint samples with ancestral sampling.



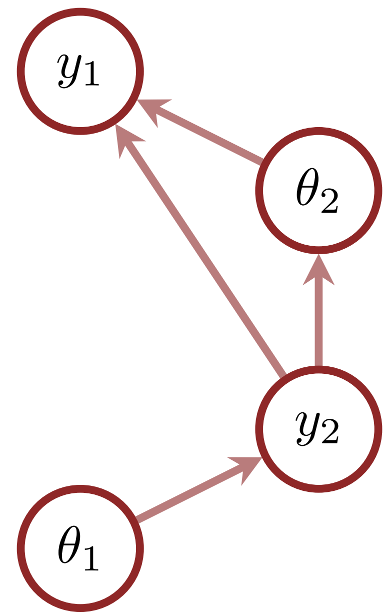


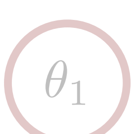


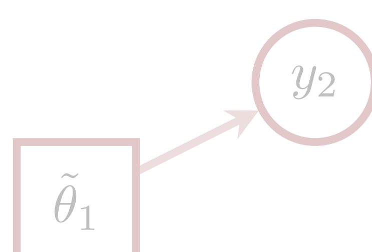
$\tilde{\theta}_1 \sim \pi(\theta_1)$

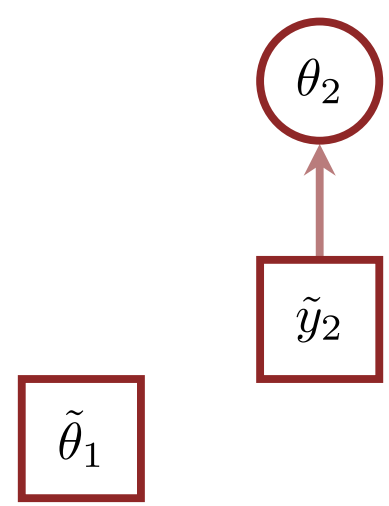
$\tilde{y}_2 \sim \pi(y_2 \mid \tilde{\theta}_1)$

Some conditional decompositions allow us generate joint samples with ancestral sampling.



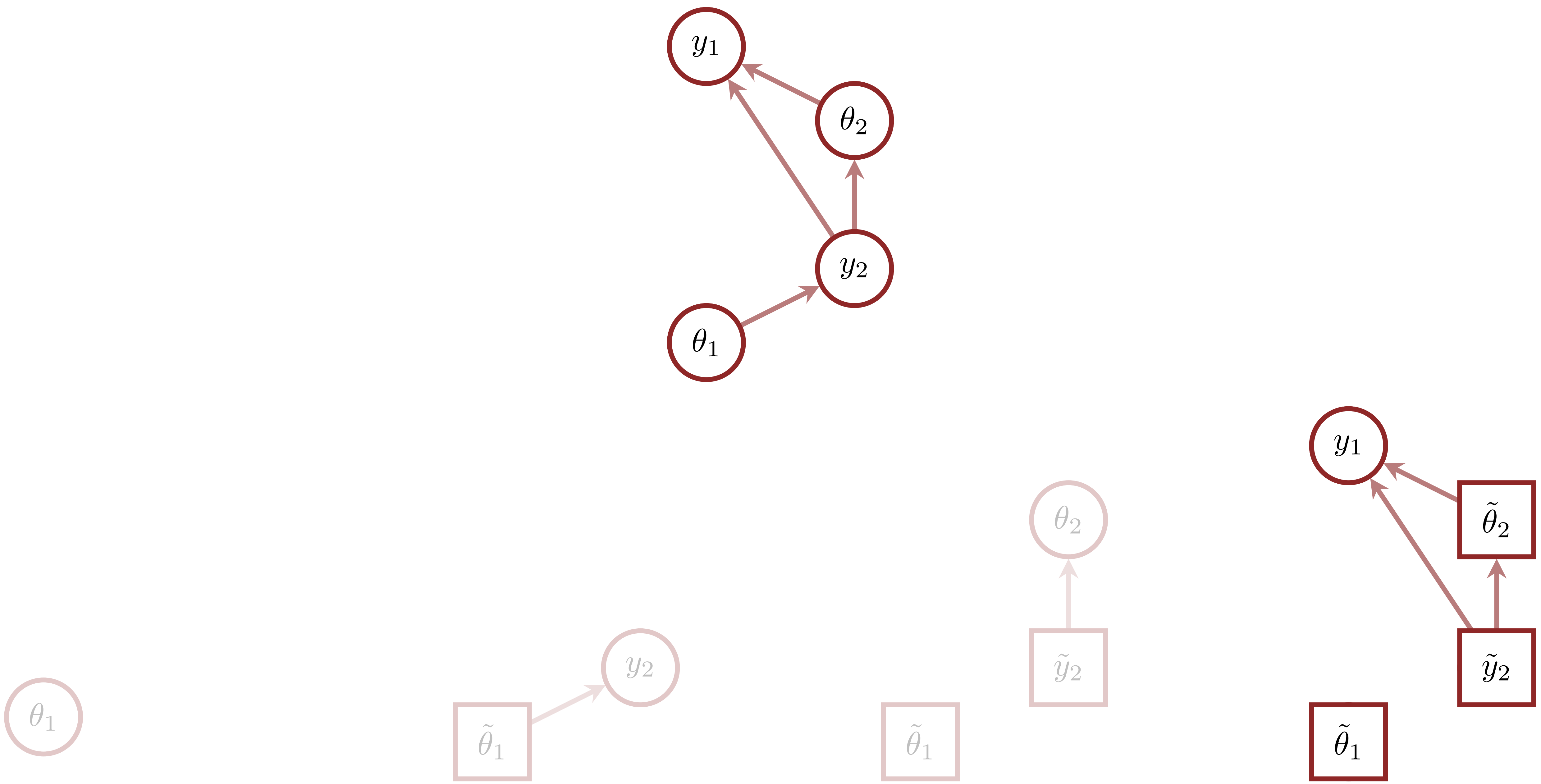






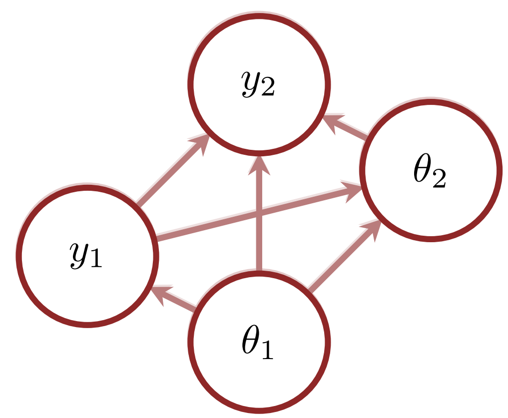
$\tilde{\theta}_1 \sim \pi(\theta_1)$ $\tilde{y}_2 \sim \pi(y_2 \mid \tilde{\theta}_1)$ $\tilde{\theta}_2 \sim \pi(\theta_2 \mid \tilde{y}_2, \tilde{\theta}_1)$

Some conditional decompositions allow us generate joint samples with ancestral sampling.



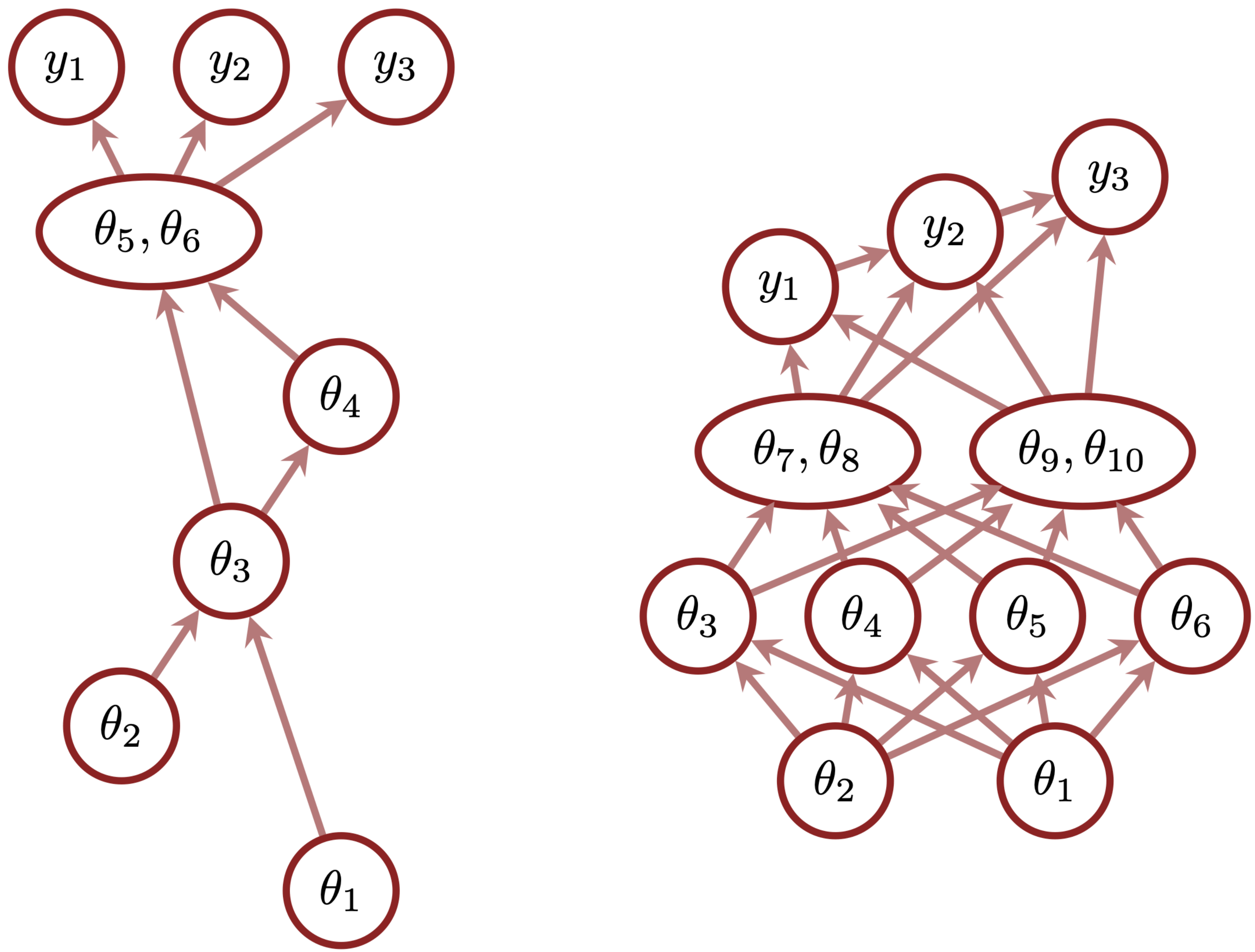
$\tilde{\theta}_1 \sim \pi(\theta_1)$ $\quad$ $\tilde{y}_2 \sim \pi(y_2 \mid \tilde{\theta}_1)$ $\quad$ $\tilde{\theta}_2 \sim \pi(\theta_2 \mid \tilde{y}_2, \tilde{\theta}_1)$ $\quad$ $\tilde{y}_1 \sim \pi(y_1 \mid \tilde{y}_2, \tilde{\theta}_1, \tilde{\theta}_2)$

There may be many different ways to procedurally generatively specify the *same* full Bayesian model.

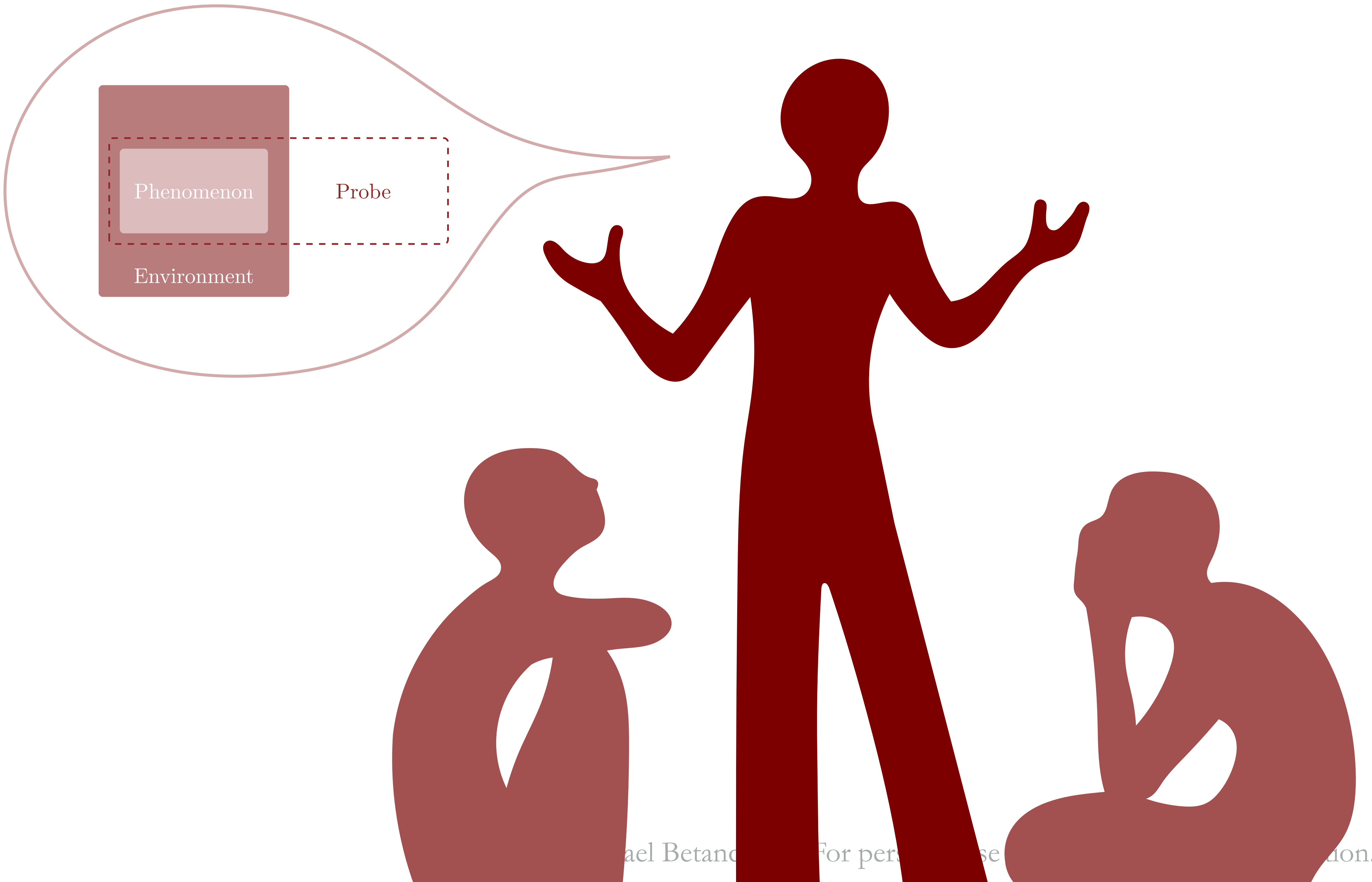


$$
\begin{aligned}
\pi(y_1, y_2, \theta_1, \theta_2) = \; & \pi(y_2 \mid \theta_2, y_1, \theta_1) \\
& \pi(\theta_2 \mid y_1, \theta_1) \\
& \pi(y_1 \mid \theta_1) \\
& \pi(\theta_1)
\end{aligned}
$$

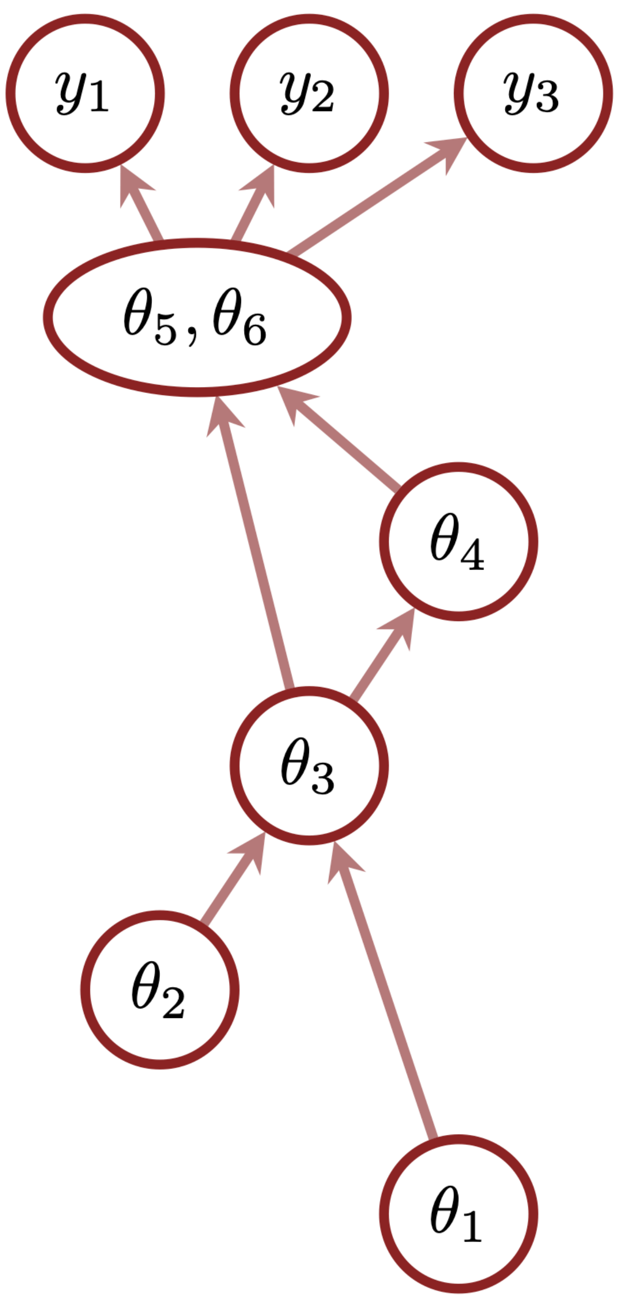
Narratively Generative Modeling



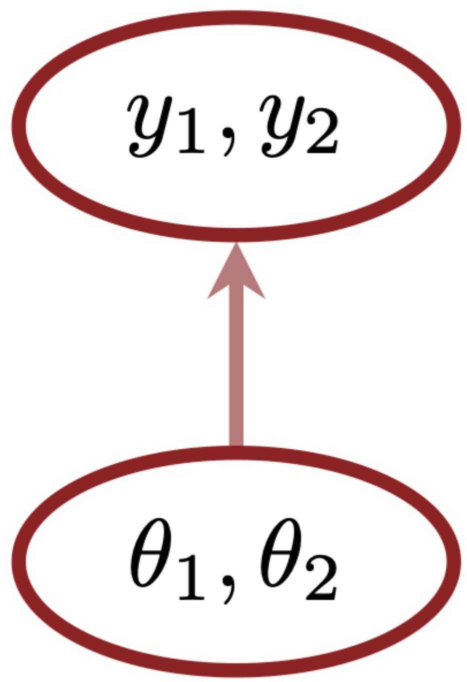
Narratively generative models aren't defined by their properties but rather by the mathematical stories they tell.



Each conditional decomposition of a model captures different stories of how observations are generated.

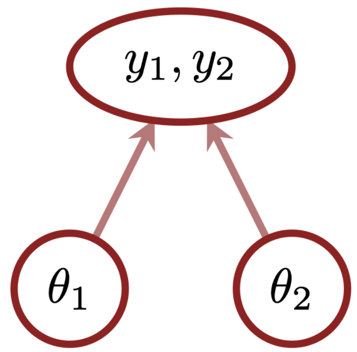


Each conditional decomposition of a model captures different stories of how observations are generated.

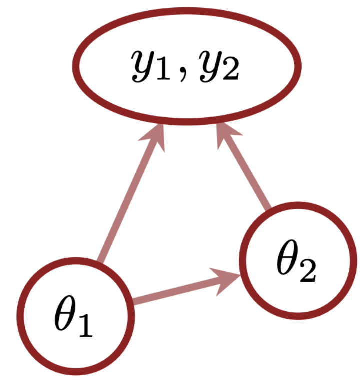


$$
\begin{aligned}
\pi(y_1, y_2, \theta_1, \theta_2) = & \\
& \pi(y_1, y_2 \mid \theta_1, \theta_2) \\
& \pi(\theta_1, \theta_2)
\end{aligned}
$$

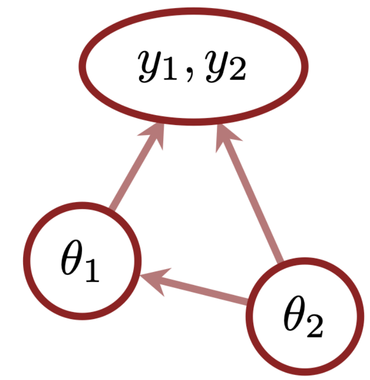
The more complete the conditional decomposition the more detailed the stories the model captures.



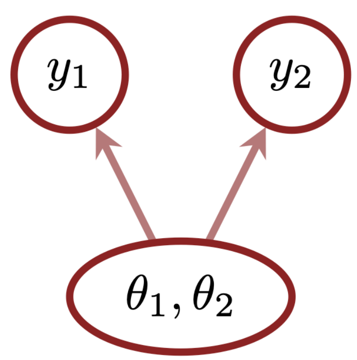
$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1, y_2 \mid \theta_1, \theta_2) \, \pi(\theta_1) \, \pi(\theta_2)$$



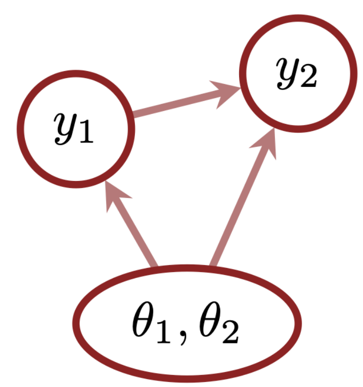
$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1, y_2 \mid \theta_1, \theta_2) \, \pi(\theta_2 \mid \theta_1) \, \pi(\theta_1)$$



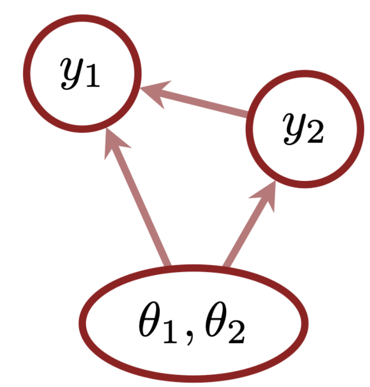
$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1, y_2 \mid \theta_1, \theta_2) \, \pi(\theta_1 \mid \theta_2) \, \pi(\theta_2)$$



$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1 \mid \theta_1, \theta_2) \, \pi(y_2 \mid \theta_1, \theta_2) \, \pi(\theta_1, \theta_2)$$

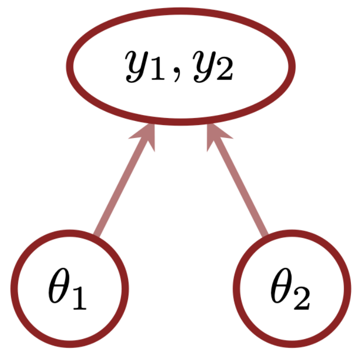


$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_2 \mid y_1, \theta_1, \theta_2) \, \pi(y_1 \mid \theta_1, \theta_2) \, \pi(\theta_1, \theta_2)$$

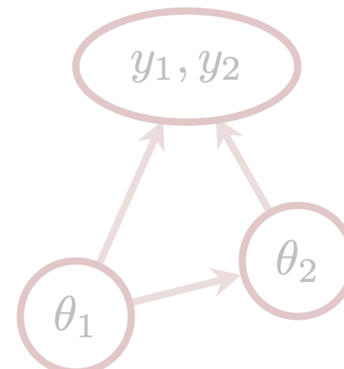


$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1 \mid y_2, \theta_1, \theta_2) \, \pi(y_2 \mid \theta_1, \theta_2) \, \pi(\theta_1, \theta_2)$$

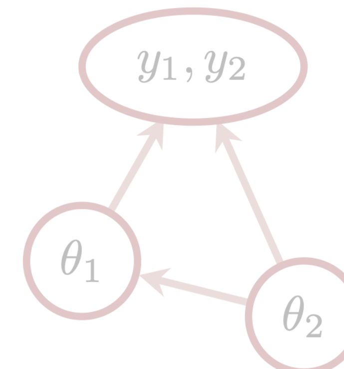
The more complete the conditional decomposition the more detailed the stories the model captures.



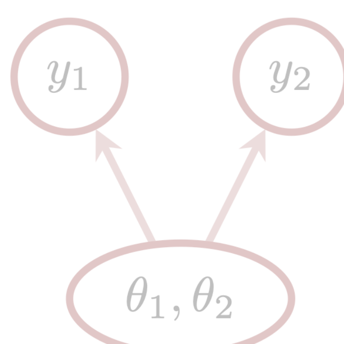
$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1, y_2 \mid \theta_1, \theta_2) \, \pi(\theta_1) \, \pi(\theta_2)$$



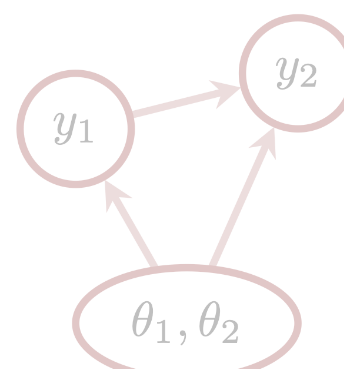
$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1, y_2 \mid \theta_1, \theta_2) \, \pi(\theta_2 \mid \theta_1) \, \pi(\theta_1)$$



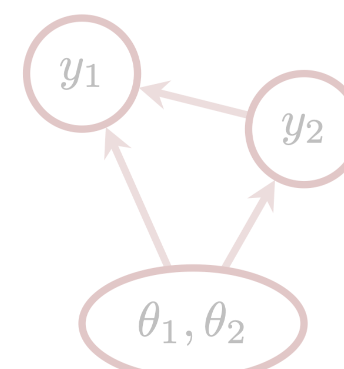
$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1, y_2 \mid \theta_1, \theta_2) \, \pi(\theta_1 \mid \theta_2) \, \pi(\theta_2)$$



$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1 \mid \theta_1, \theta_2) \, \pi(y_2 \mid \theta_1, \theta_2) \, \pi(\theta_1, \theta_2)$$

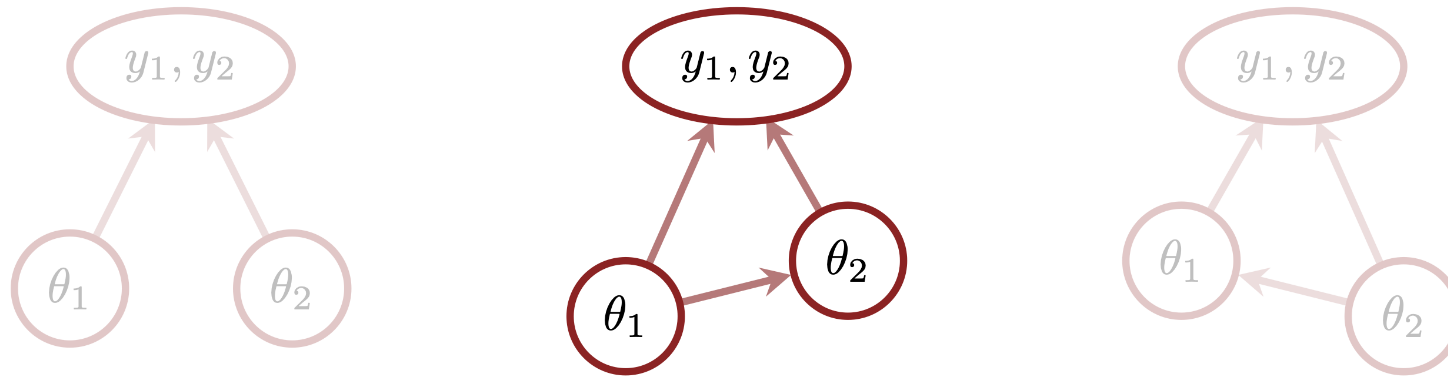


$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_2 \mid y_1, \theta_1, \theta_2) \, \pi(y_1 \mid \theta_1, \theta_2) \, \pi(\theta_1, \theta_2)$$



$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1 \mid y_2, \theta_1, \theta_2) \, \pi(y_2 \mid \theta_1, \theta_2) \, \pi(\theta_1, \theta_2)$$

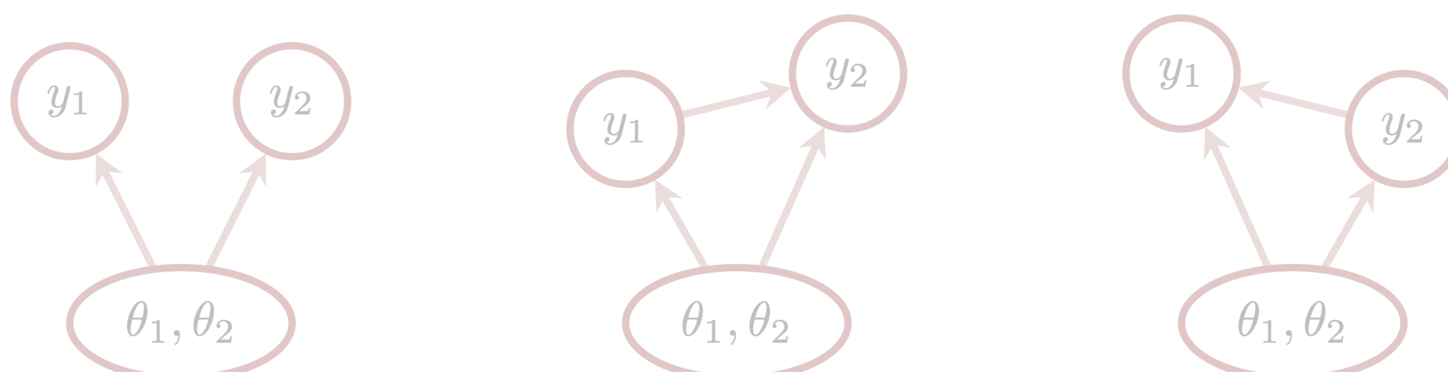
The more complete the conditional decomposition the more detailed the stories the model captures.



$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1, y_2 \mid \theta_1, \theta_2) \, \pi(\theta_1) \, \pi(\theta_2)$$

$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1, y_2 \mid \theta_1, \theta_2) \, \pi(\theta_2 \mid \theta_1) \, \pi(\theta_1)$$

$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1, y_2 \mid \theta_1, \theta_2) \, \pi(\theta_1 \mid \theta_2) \, \pi(\theta_2)$$

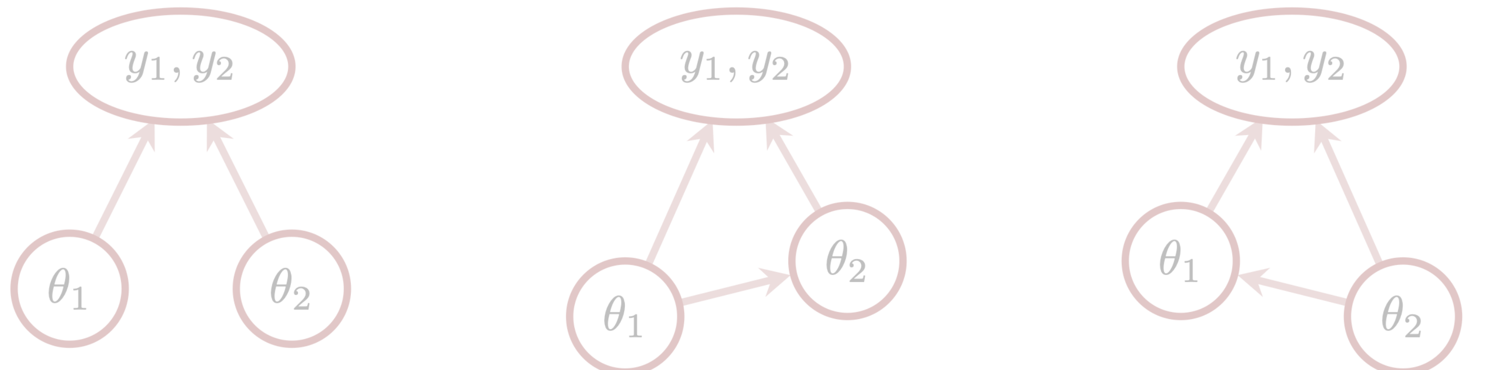


$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1 \mid \theta_1, \theta_2) \, \pi(y_2 \mid \theta_1, \theta_2) \, \pi(\theta_1, \theta_2)$$

$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_2 \mid y_1, \theta_1, \theta_2) \, \pi(y_1 \mid \theta_1, \theta_2) \, \pi(\theta_1, \theta_2)$$

$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1 \mid y_2, \theta_1, \theta_2) \, \pi(y_2 \mid \theta_1, \theta_2) \, \pi(\theta_1, \theta_2)$$

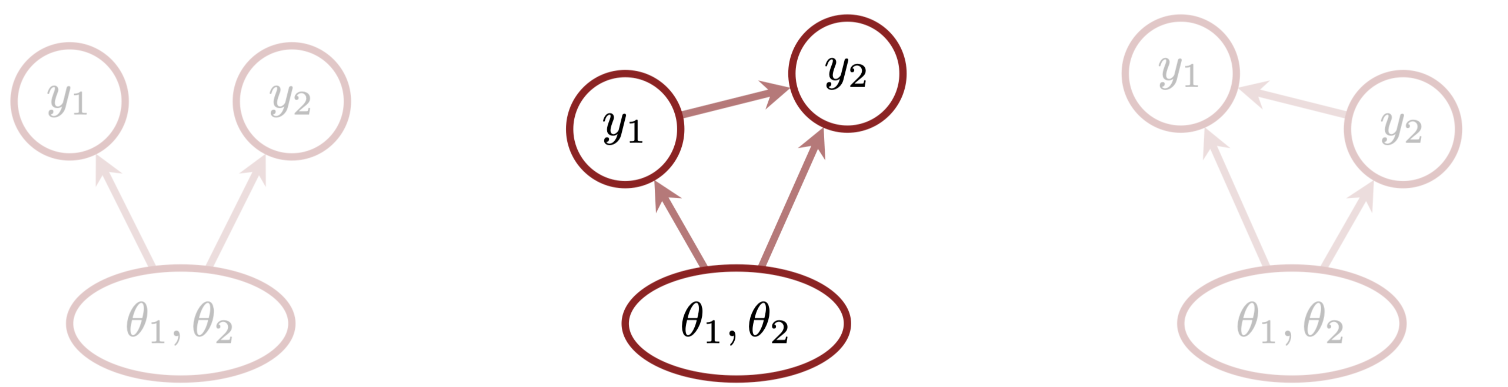
The more complete the conditional decomposition the more detailed the stories the model captures.



$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1, y_2 \mid \theta_1, \theta_2) \, \pi(\theta_1) \, \pi(\theta_2)$$

$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1, y_2 \mid \theta_1, \theta_2) \, \pi(\theta_2 \mid \theta_1) \, \pi(\theta_1)$$

$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1, y_2 \mid \theta_1, \theta_2) \, \pi(\theta_1 \mid \theta_2) \, \pi(\theta_2)$$

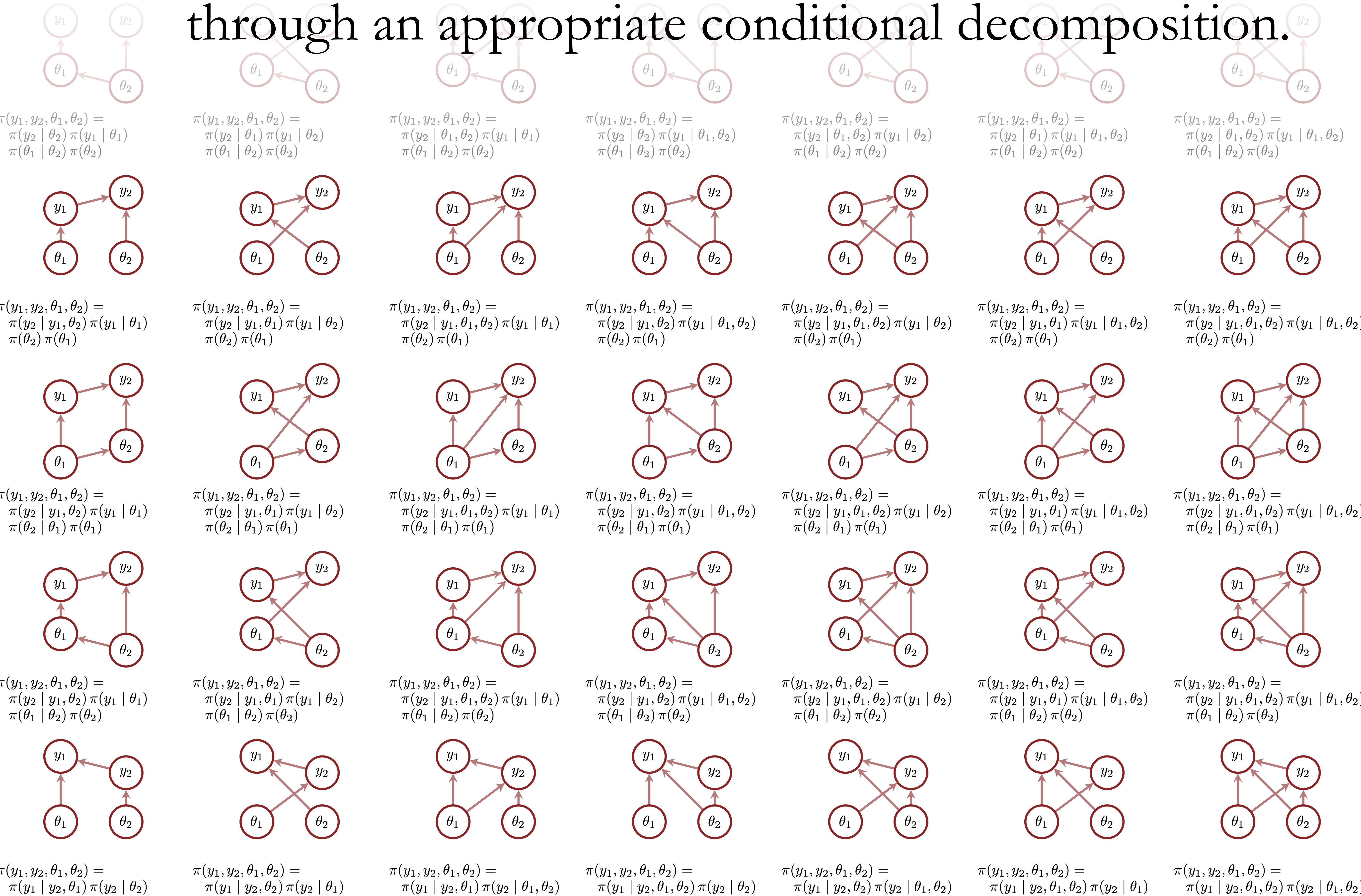


$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1 \mid \theta_1, \theta_2) \, \pi(y_2 \mid \theta_1, \theta_2) \, \pi(\theta_1, \theta_2)$$

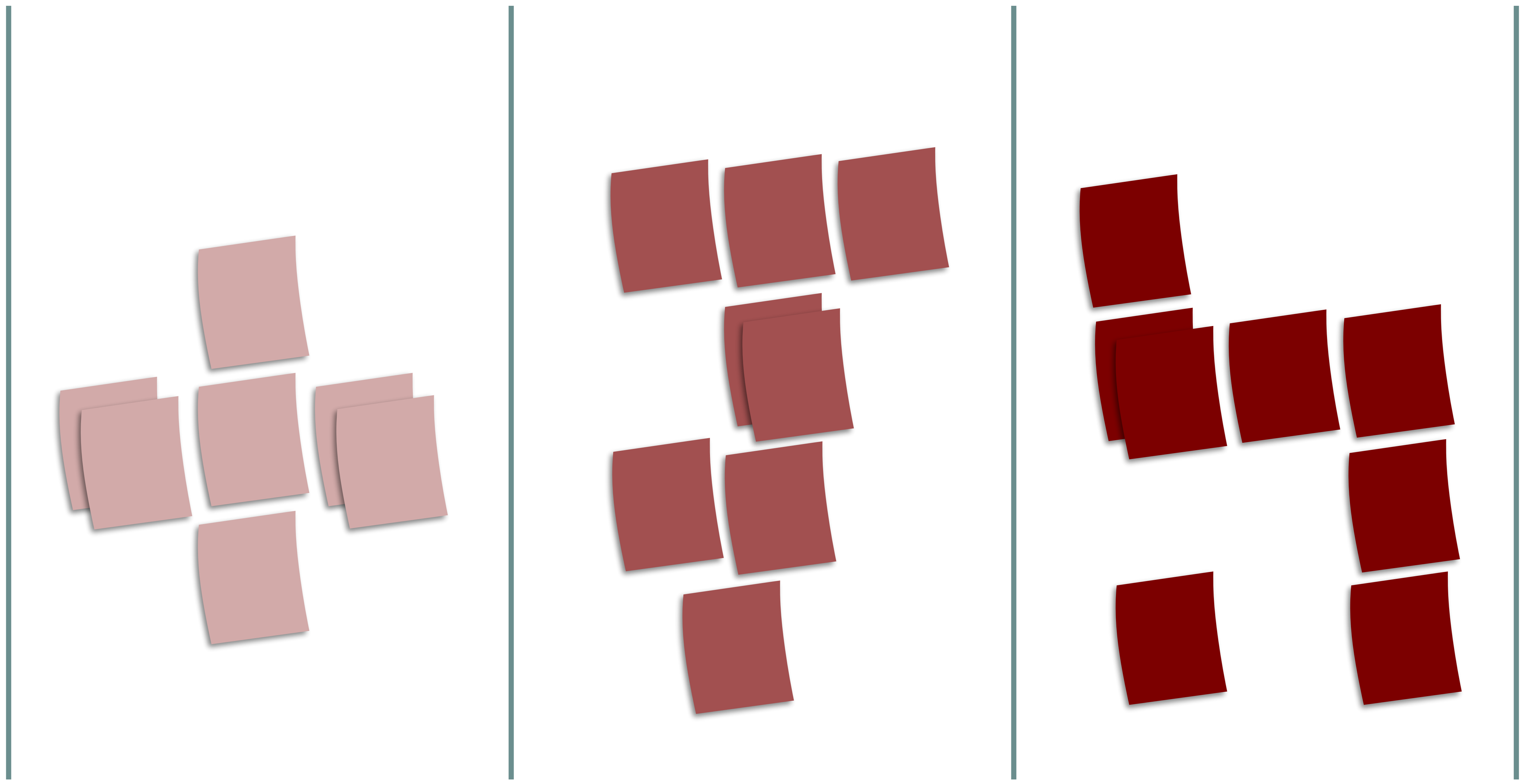
$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_2 \mid y_1, \theta_1, \theta_2) \, \pi(y_1 \mid \theta_1, \theta_2) \, \pi(\theta_1, \theta_2)$$

$$\pi(y_1, y_2, \theta_1, \theta_2) = \pi(y_1 \mid y_2, \theta_1, \theta_2) \, \pi(y_2 \mid \theta_1, \theta_2) \, \pi(\theta_1, \theta_2)$$

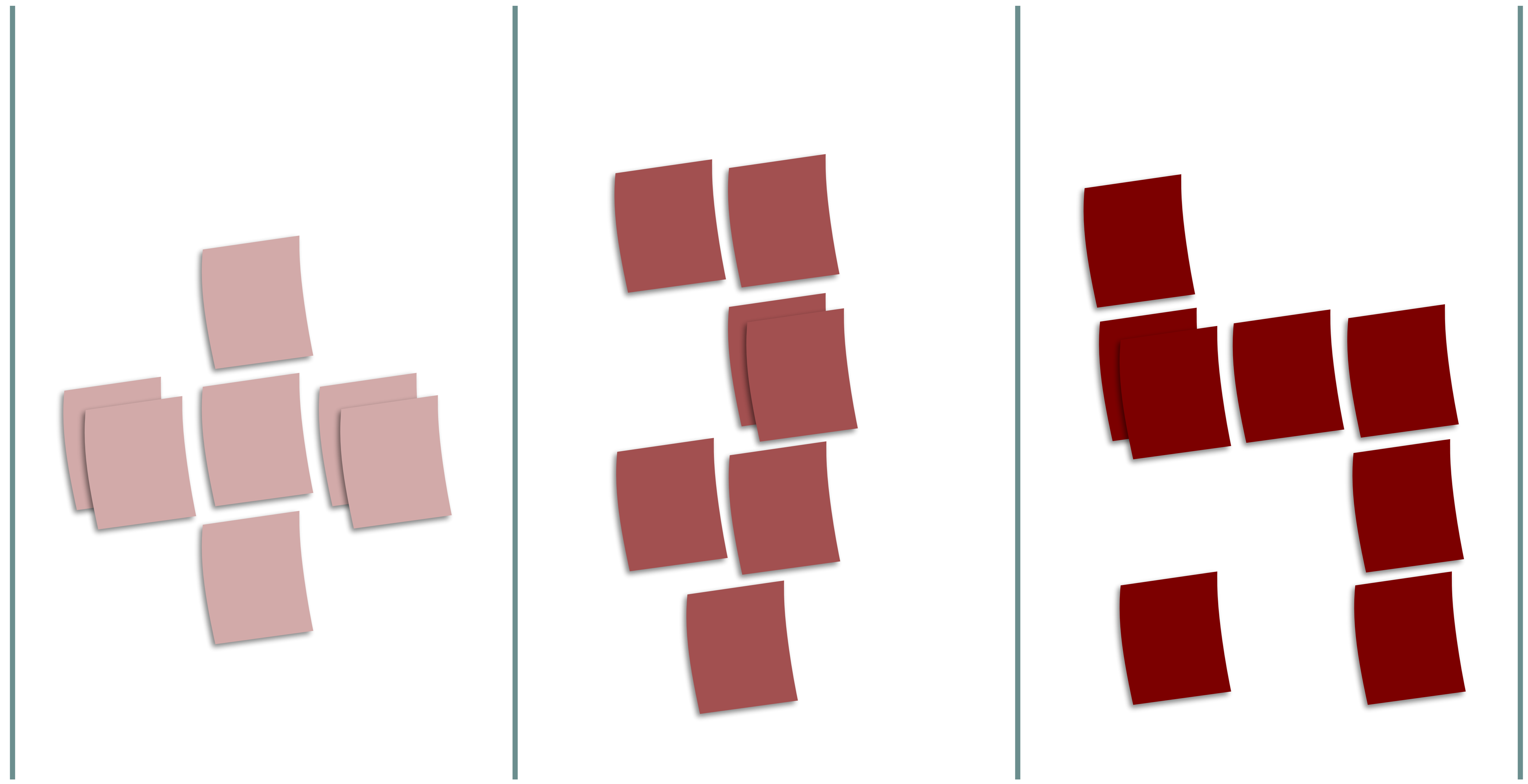
Domain expertise informs an appropriate Bayesian model through an appropriate conditional decomposition.



Advanced Narrative Techniques

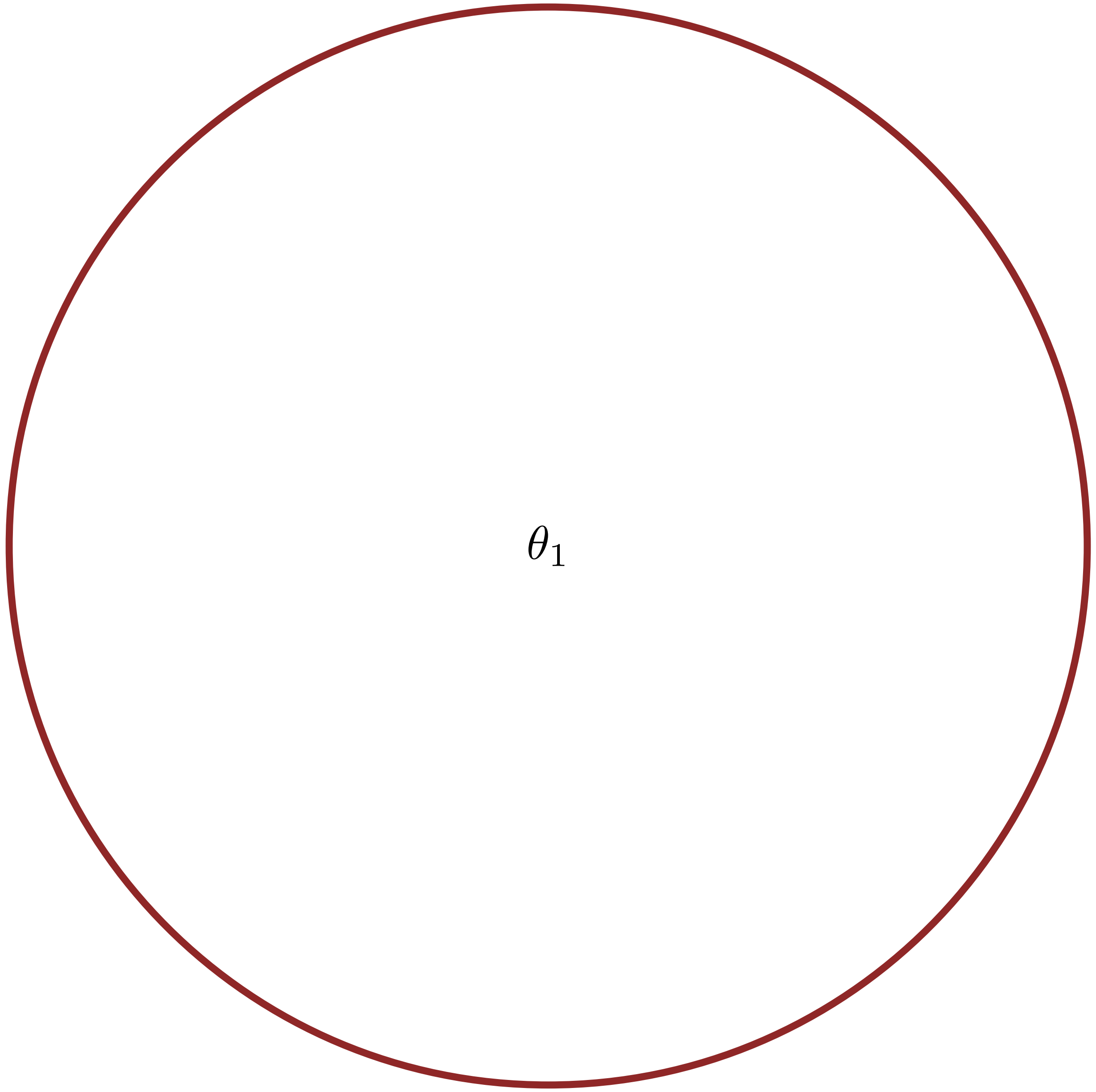


Advanced Narrative Techniques

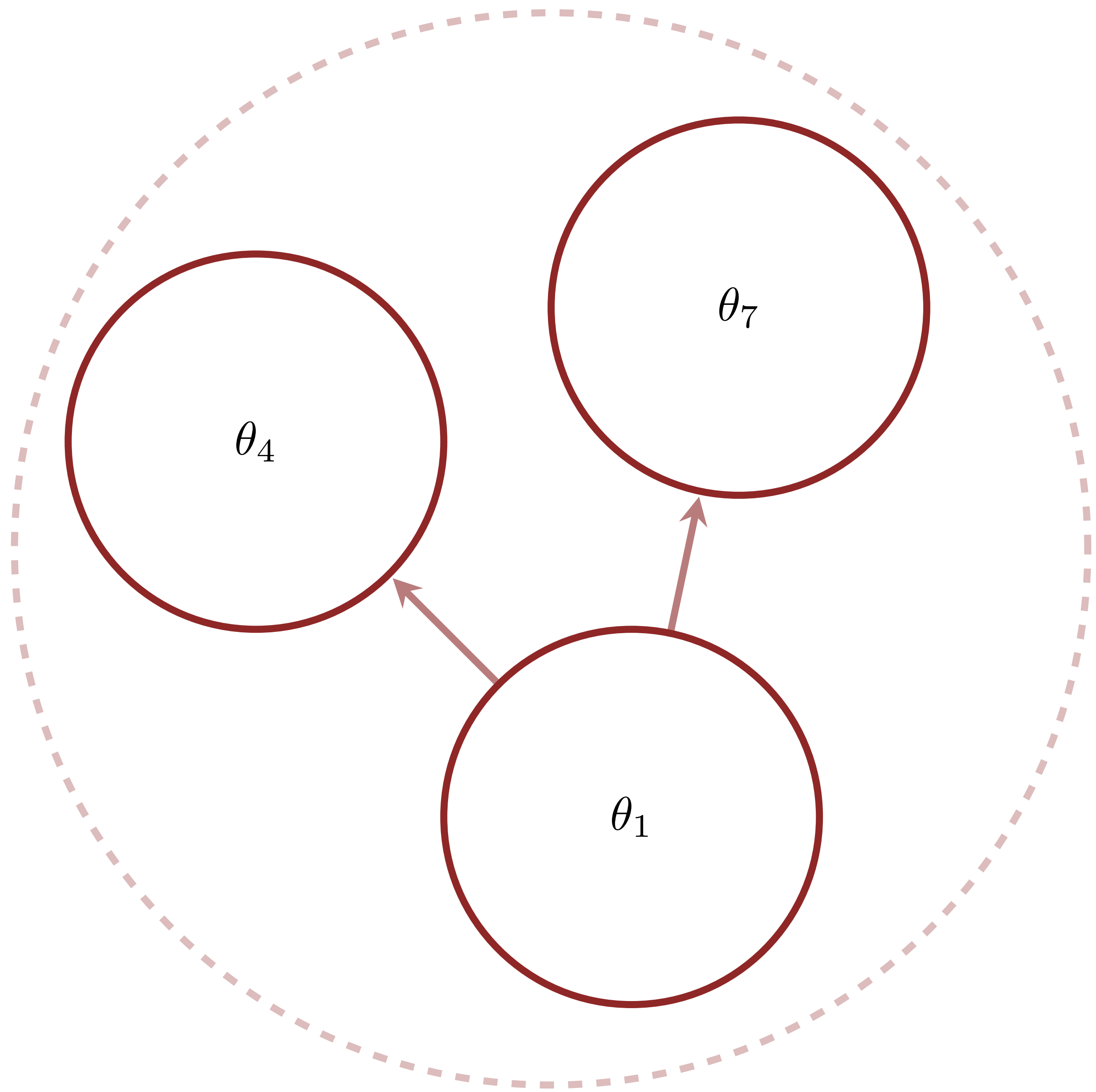


Serialization

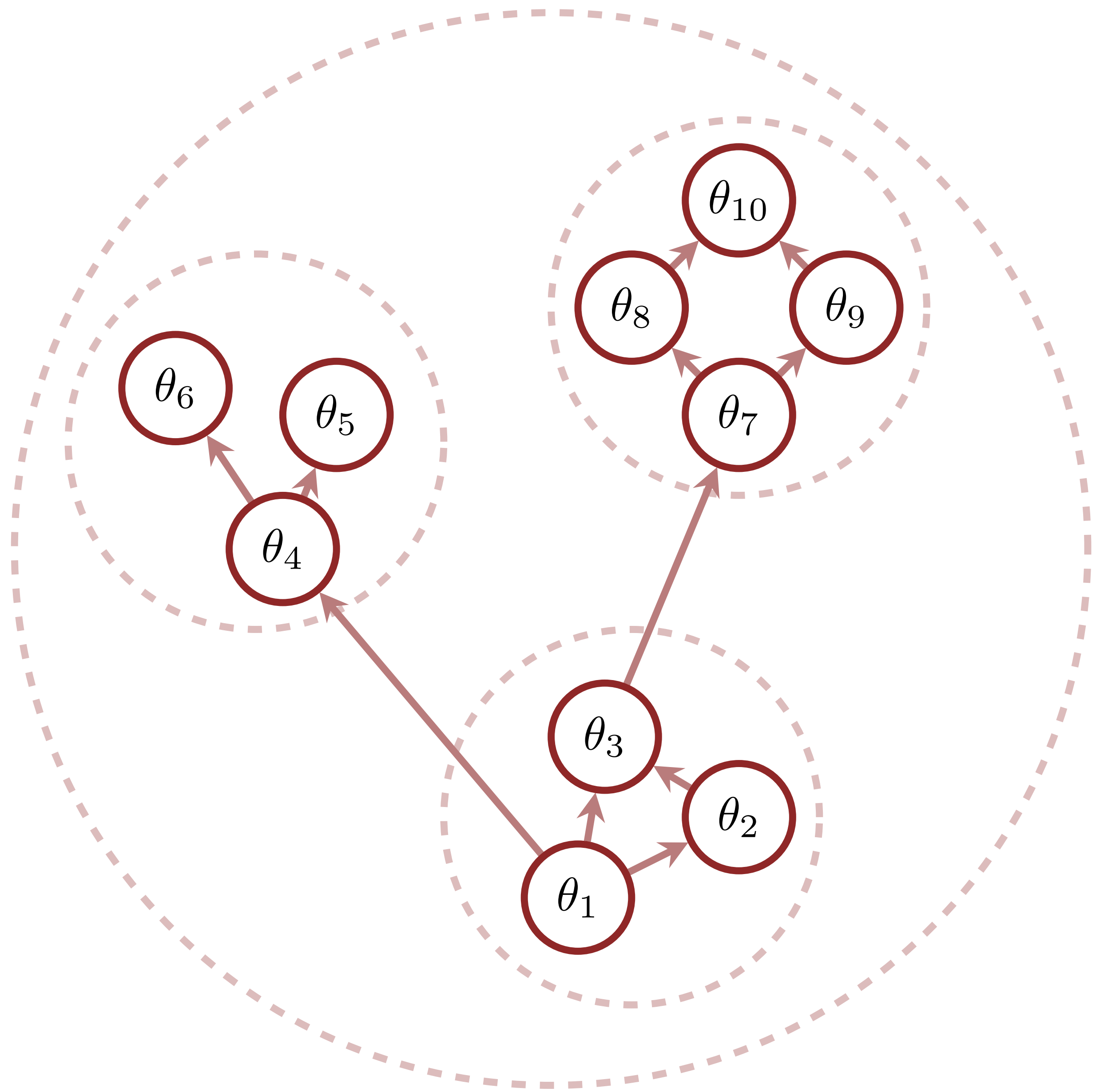
Stories are often comprised of constituent stories, which themselves might be comprised of smaller stories.



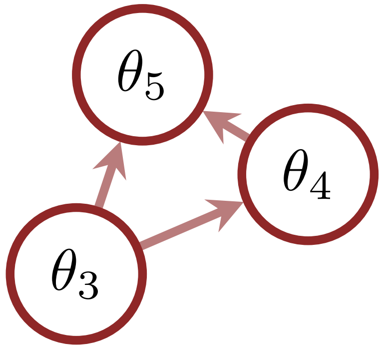
Stories are often comprised of constituent stories, which themselves might be comprised of smaller stories.



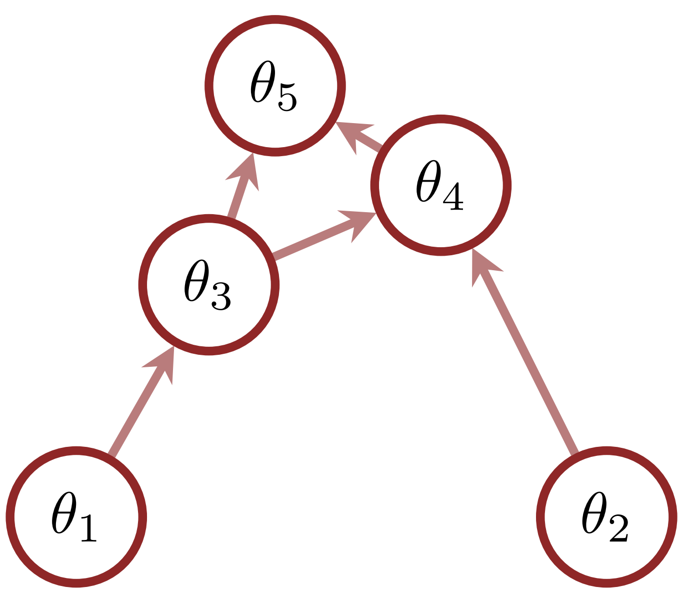
Stories are often comprised of constituent stories, which themselves might be comprised of smaller stories.



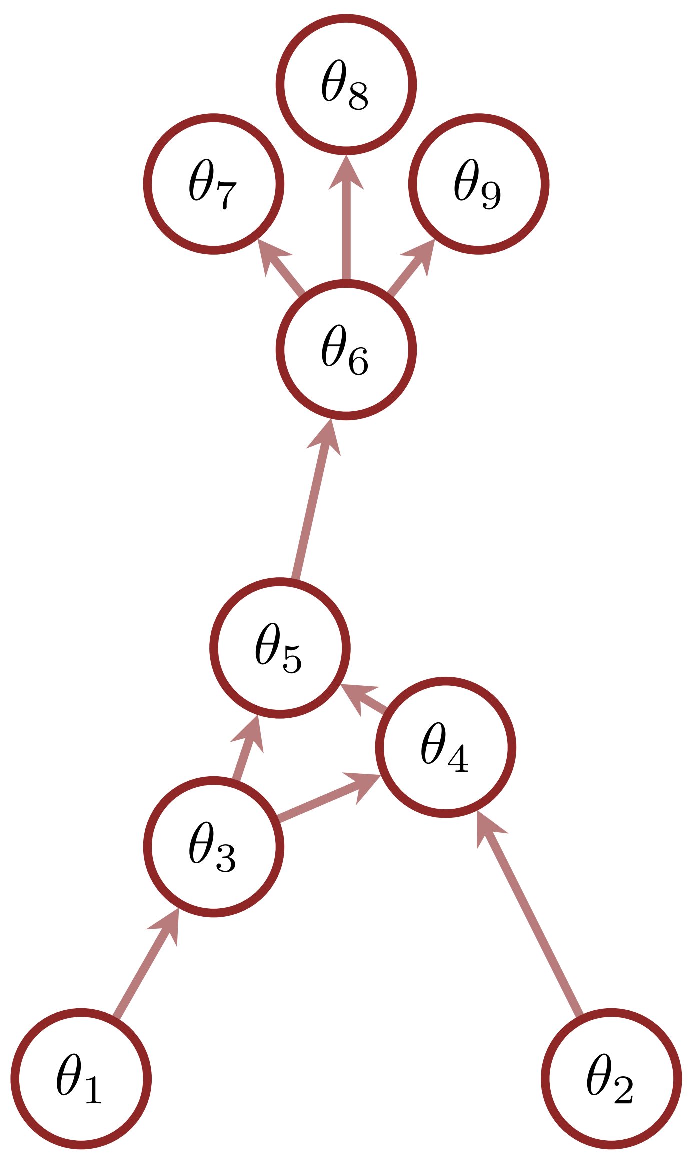
This *recursiveness* allows us to build up complex narratively generative models from simpler component models.



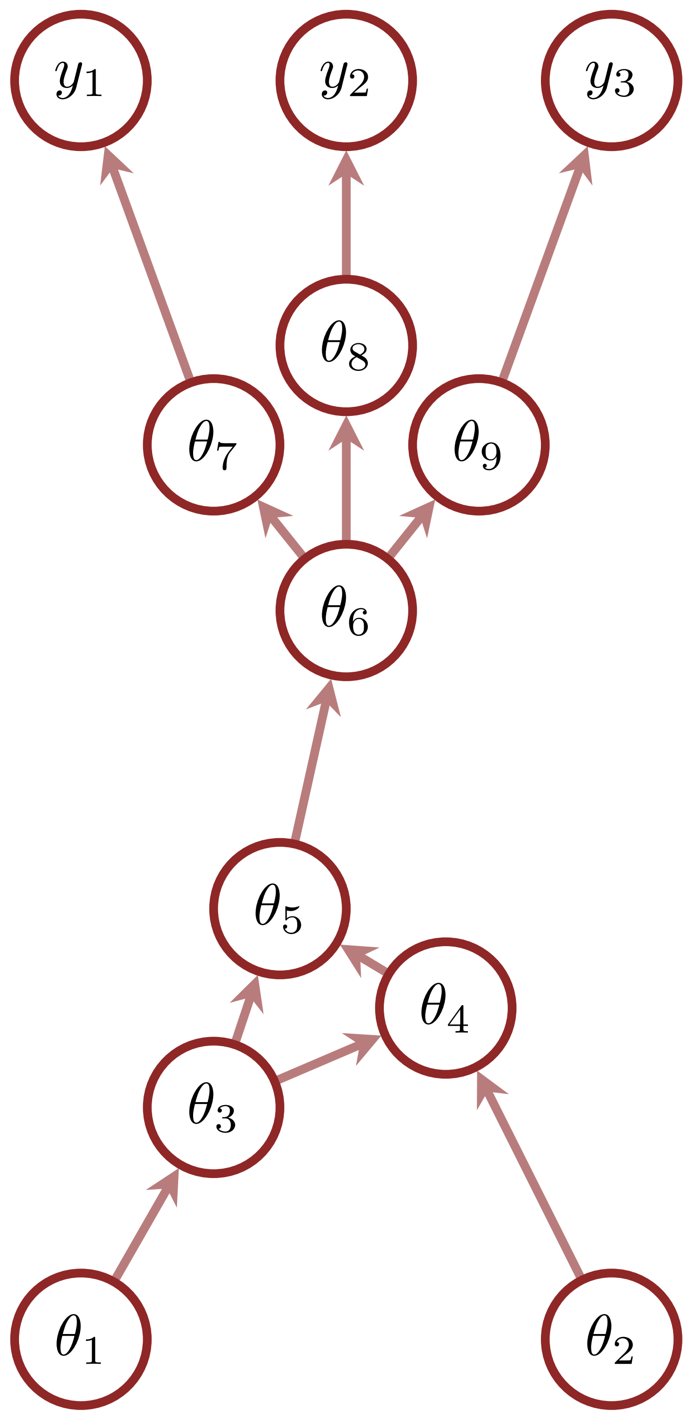
This *recursiveness* allows us to build up complex narratively generative models from simpler component models.



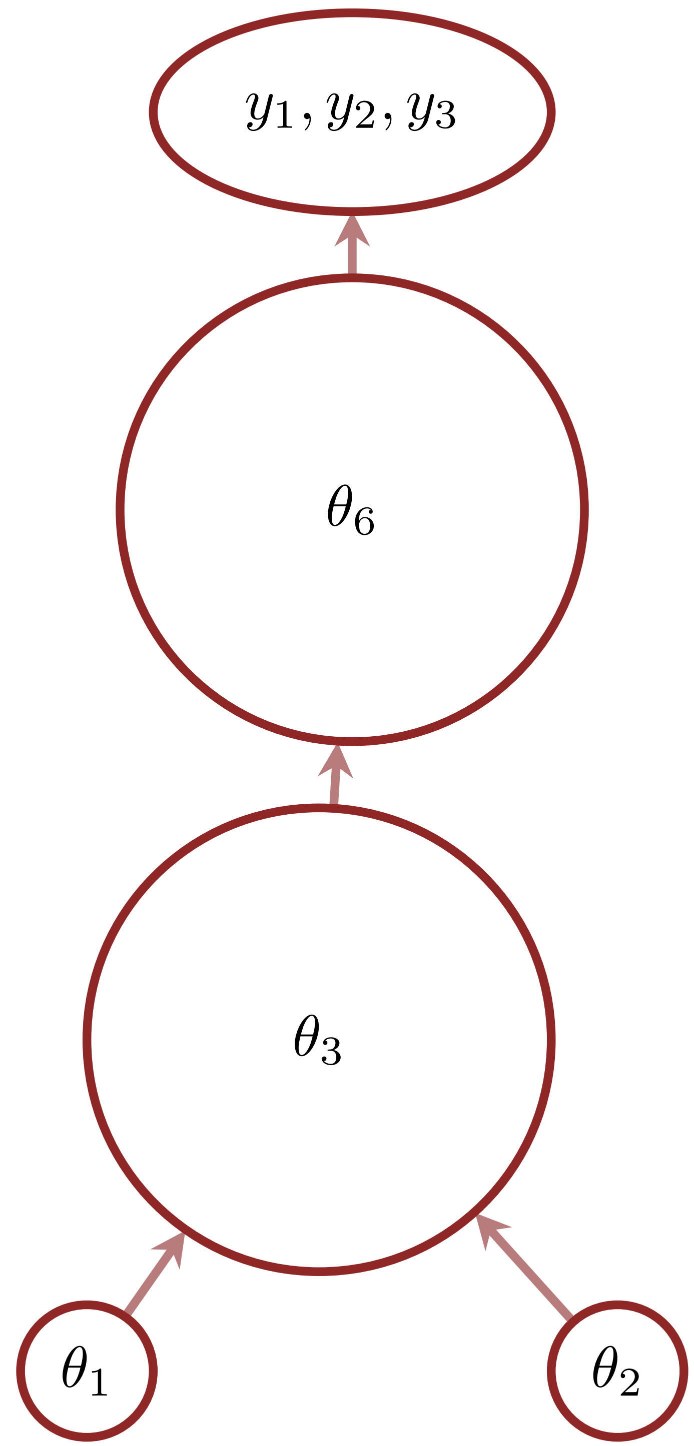
This *recursiveness* allows us to build up complex narratively generative models from simpler component models.



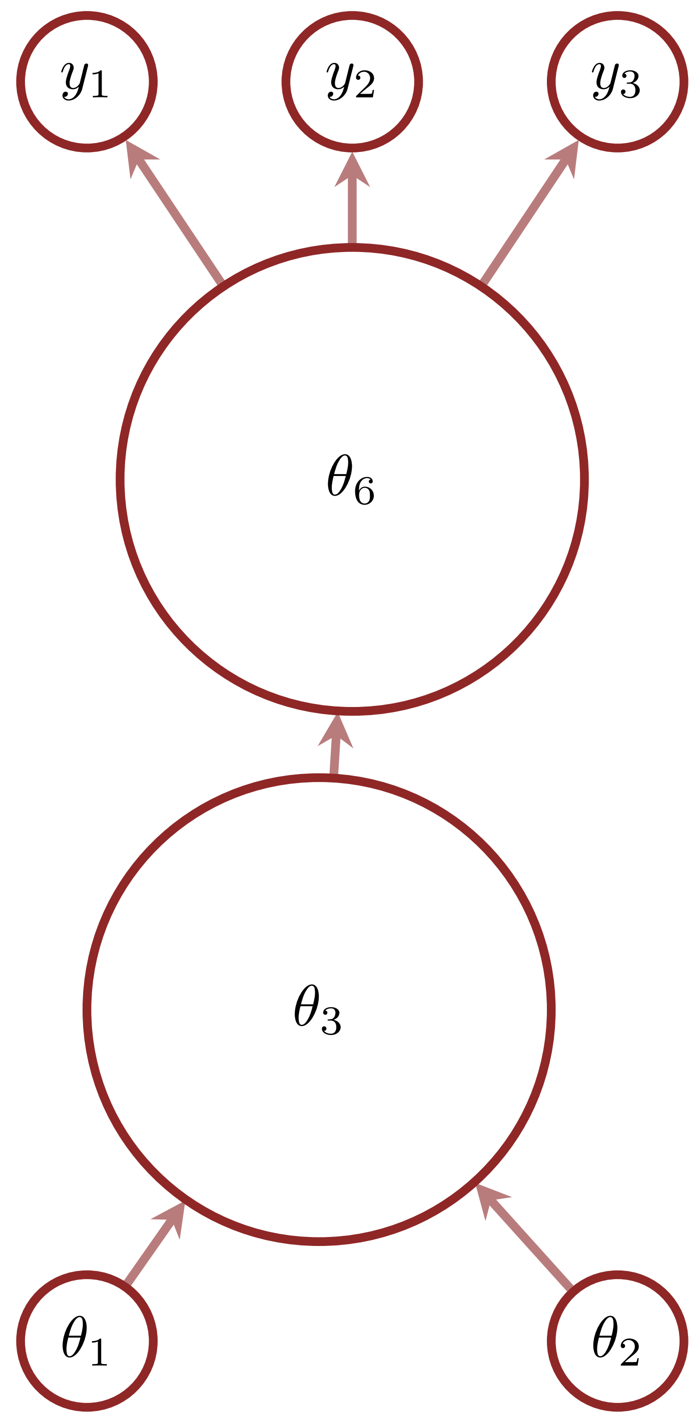
This *recursiveness* allows us to build up complex narratively generative models from simpler component models.



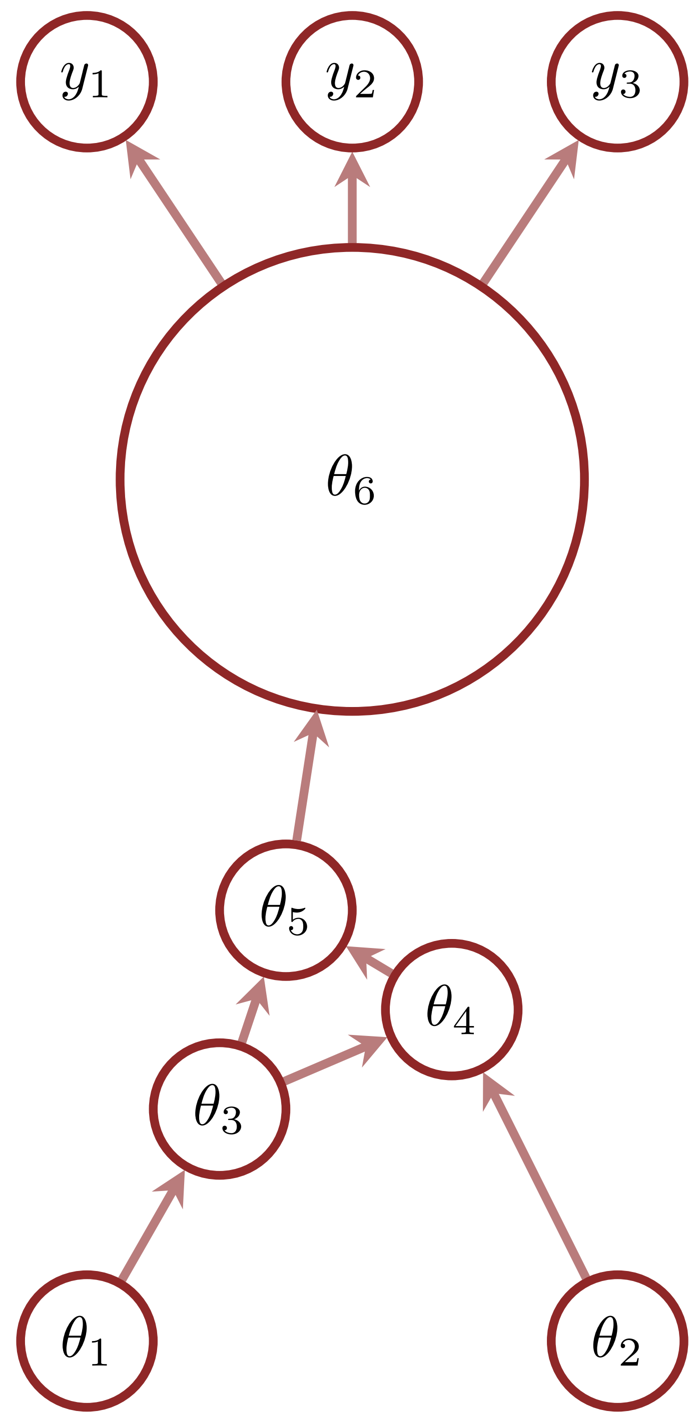
Alternatively we can develop a sophisticated narratively generative model by refining an initial coarse model.



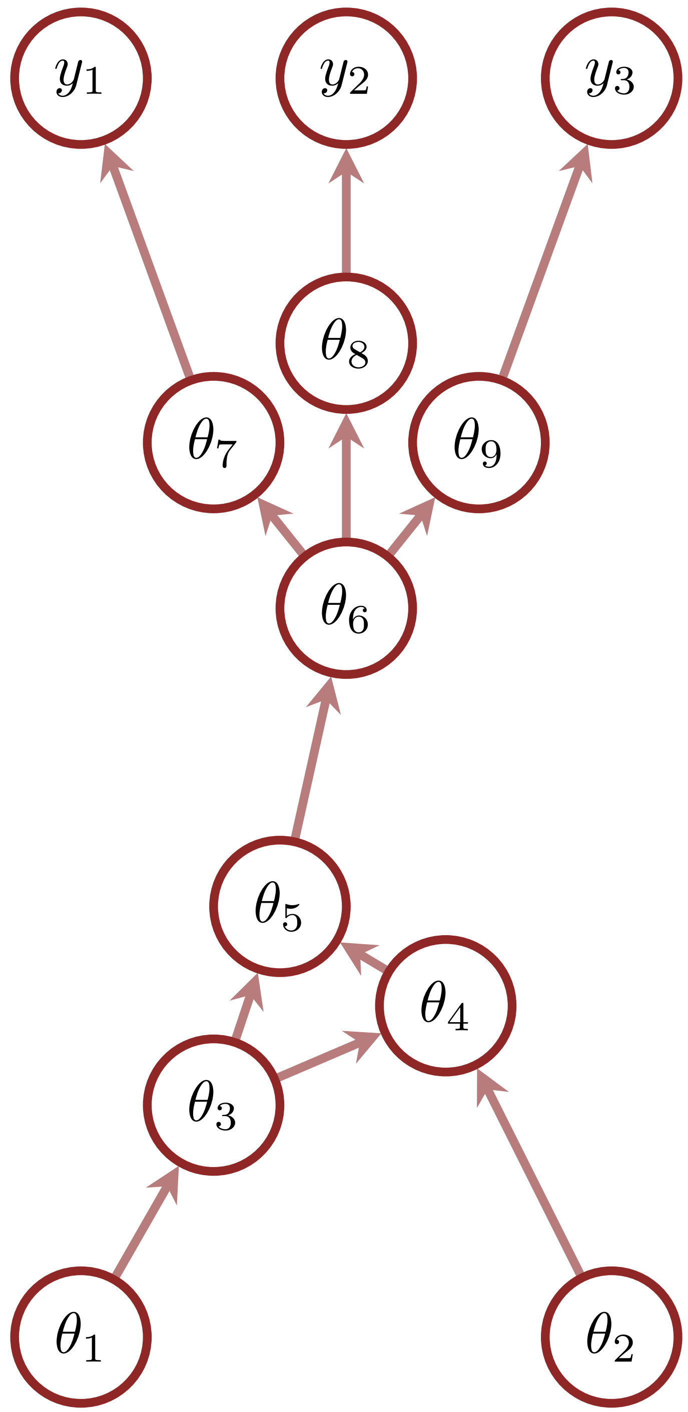
Alternatively we can develop a sophisticated narratively generative model by refining an initial coarse model.



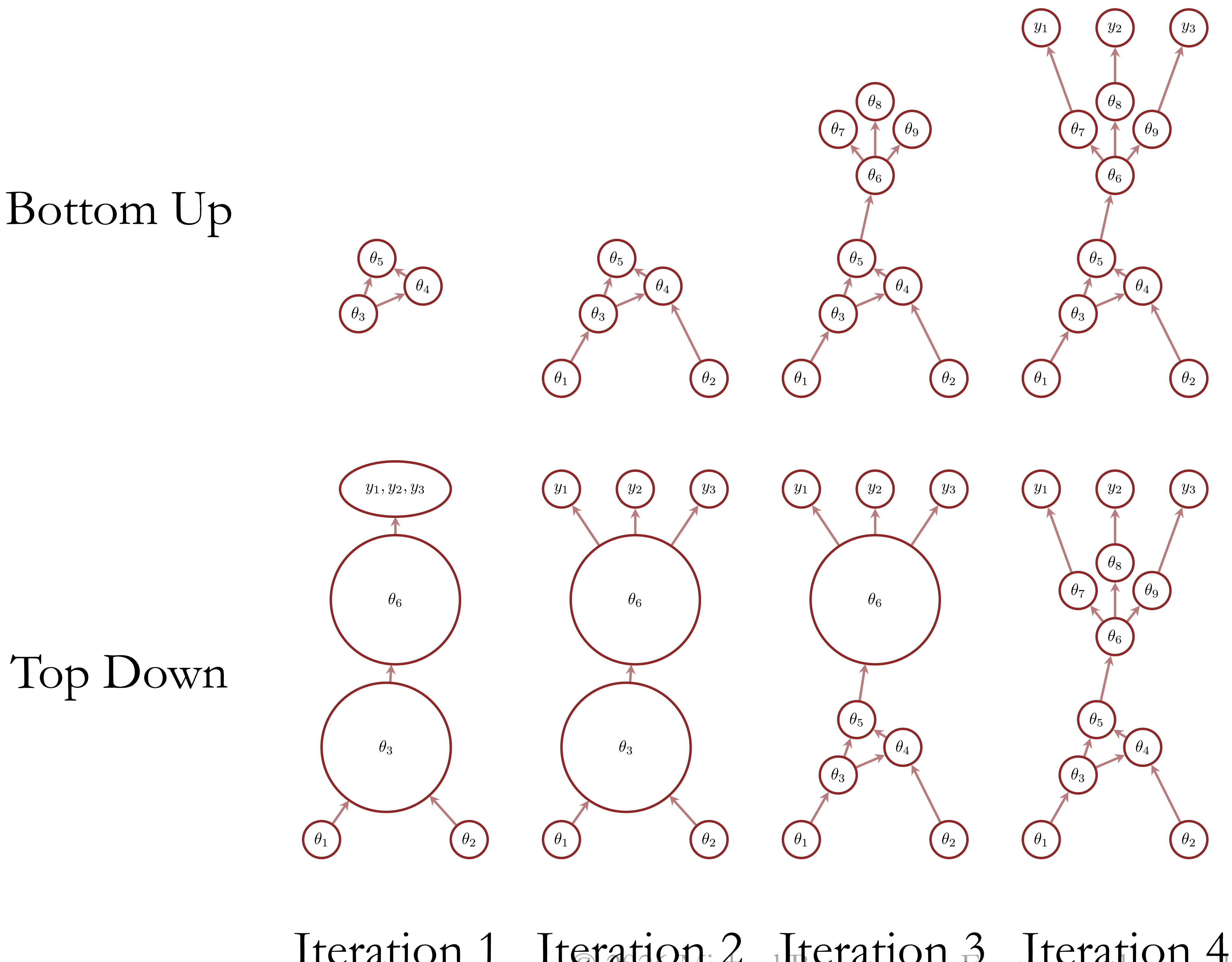
Alternatively we can develop a sophisticated narratively generative model by refining an initial coarse model.



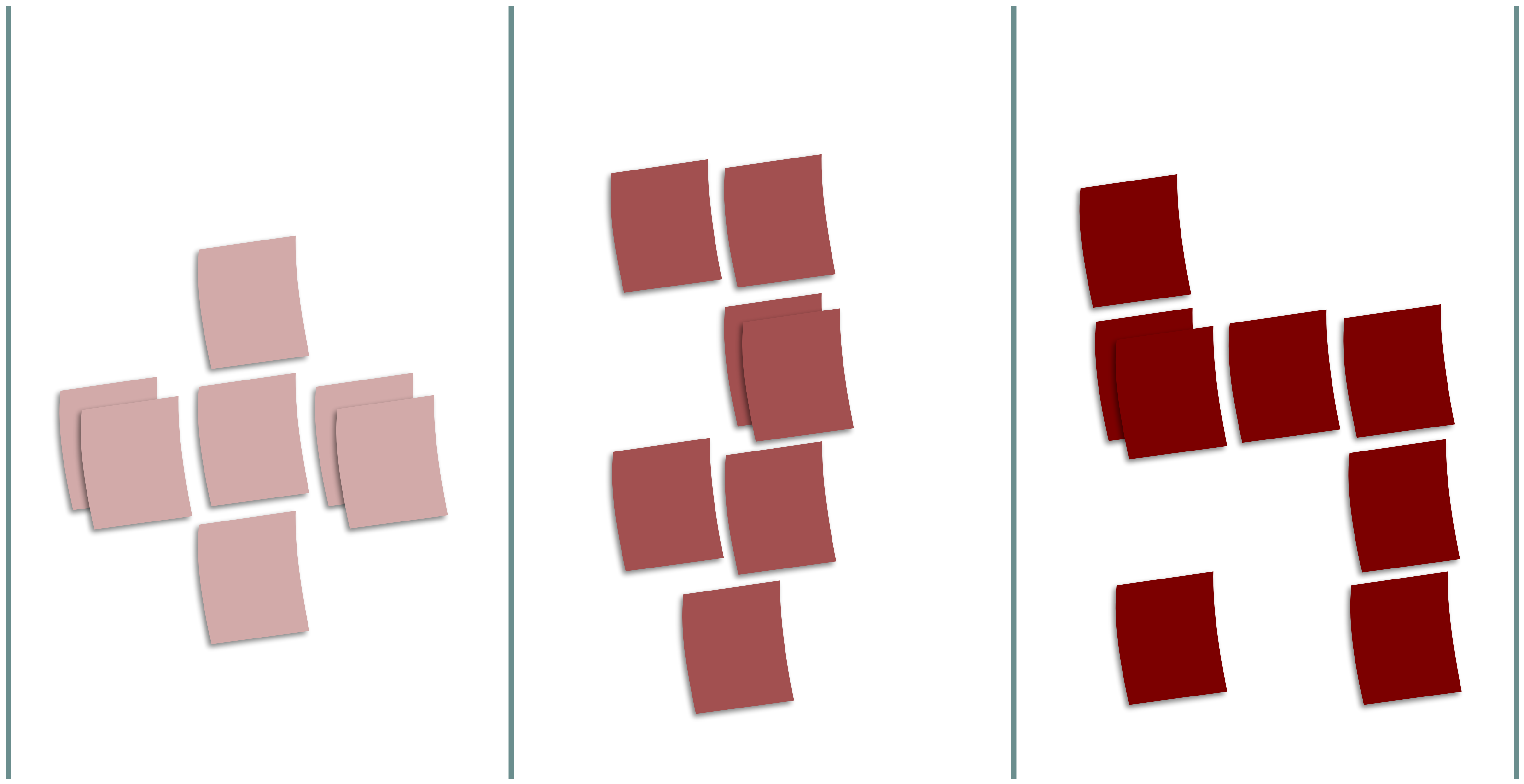
Alternatively we can develop a sophisticated narratively generative model by refining an initial coarse model.



In other words this compositional structure motivates building models from both the bottom up and top down.

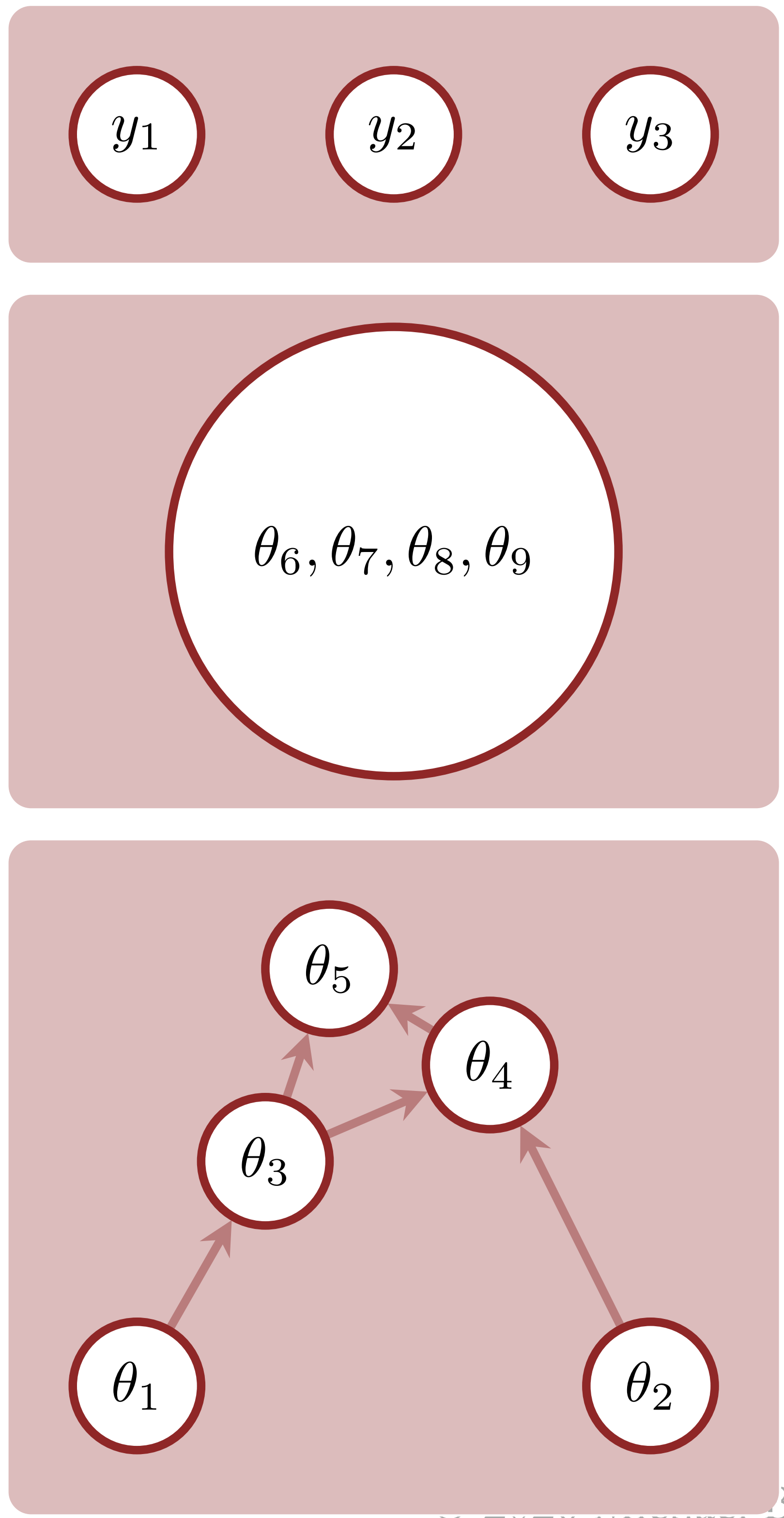


Advanced Narrative Techniques



Mixed Genres

We can build meaningfully narratively generative models from narratively non-generative components.

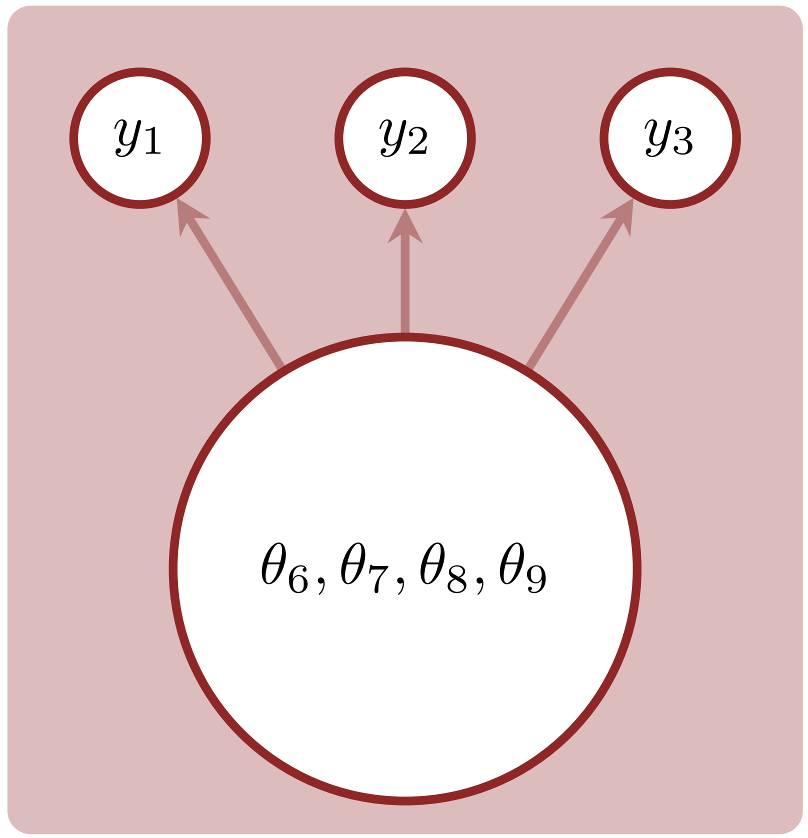


Narratively Generative Interpretation

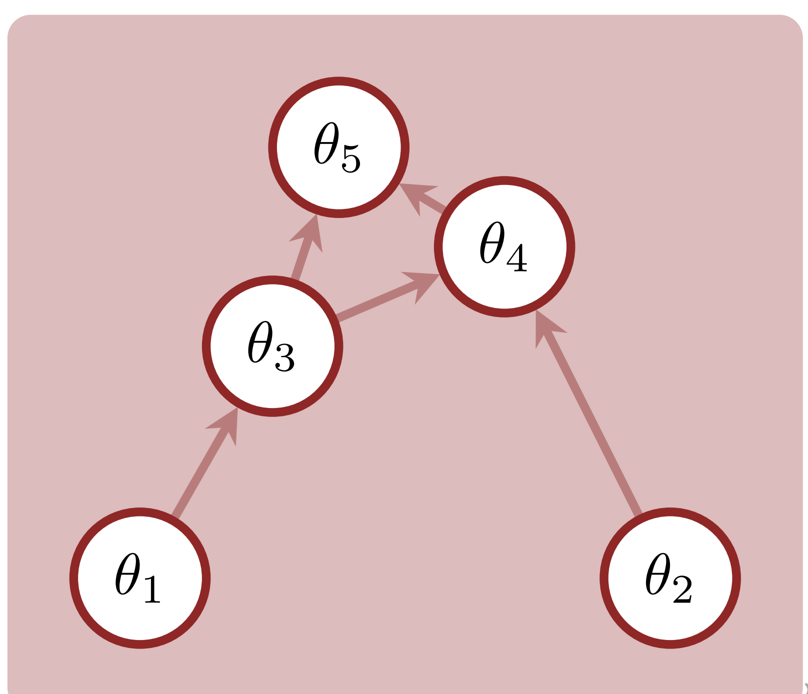
No Narratively Generative Interpretation

Narratively Generative Interpretation

We can build meaningfully narratively generative models from narratively non-generative components.

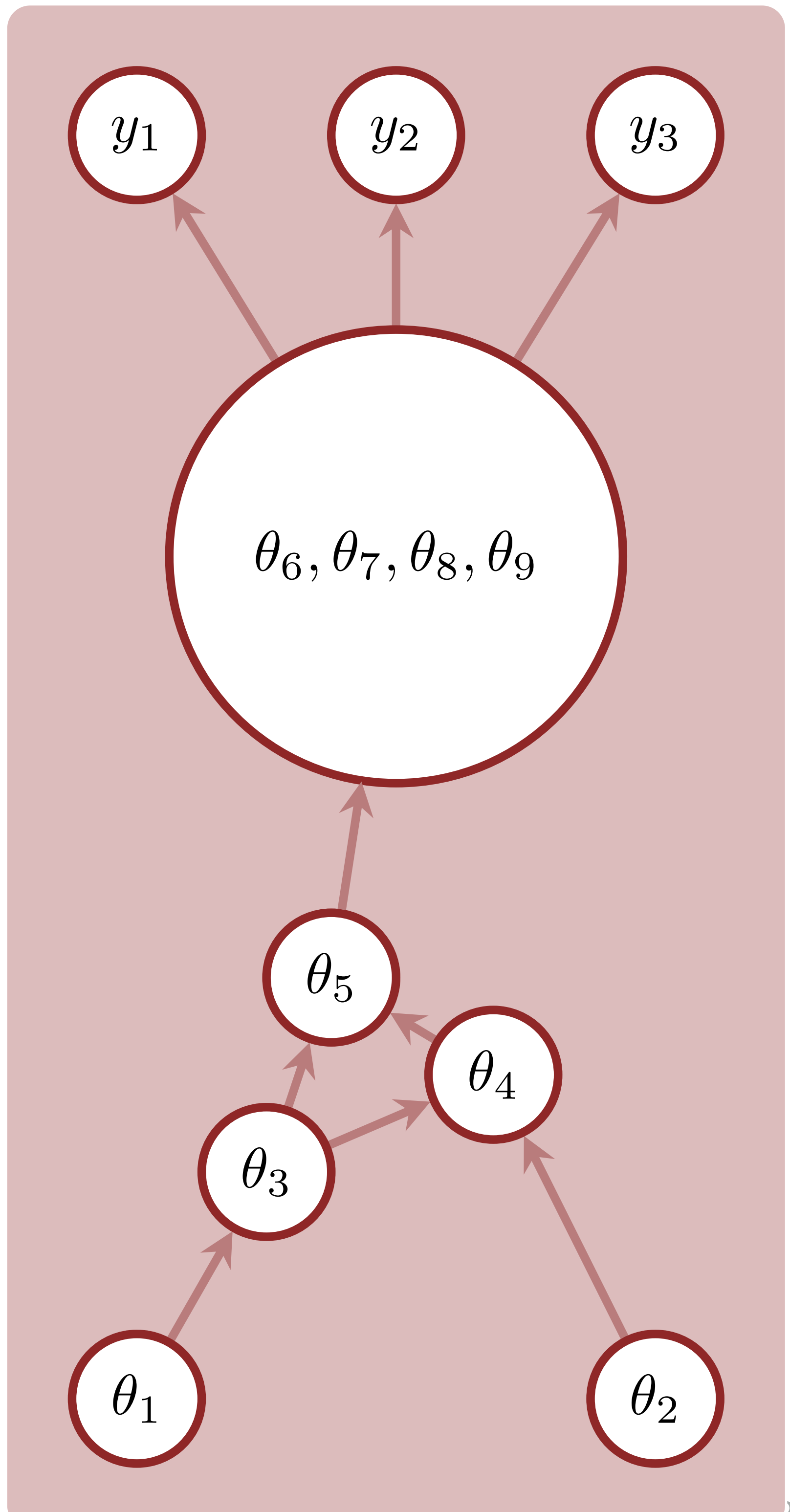


Narratively Generative Interpretation



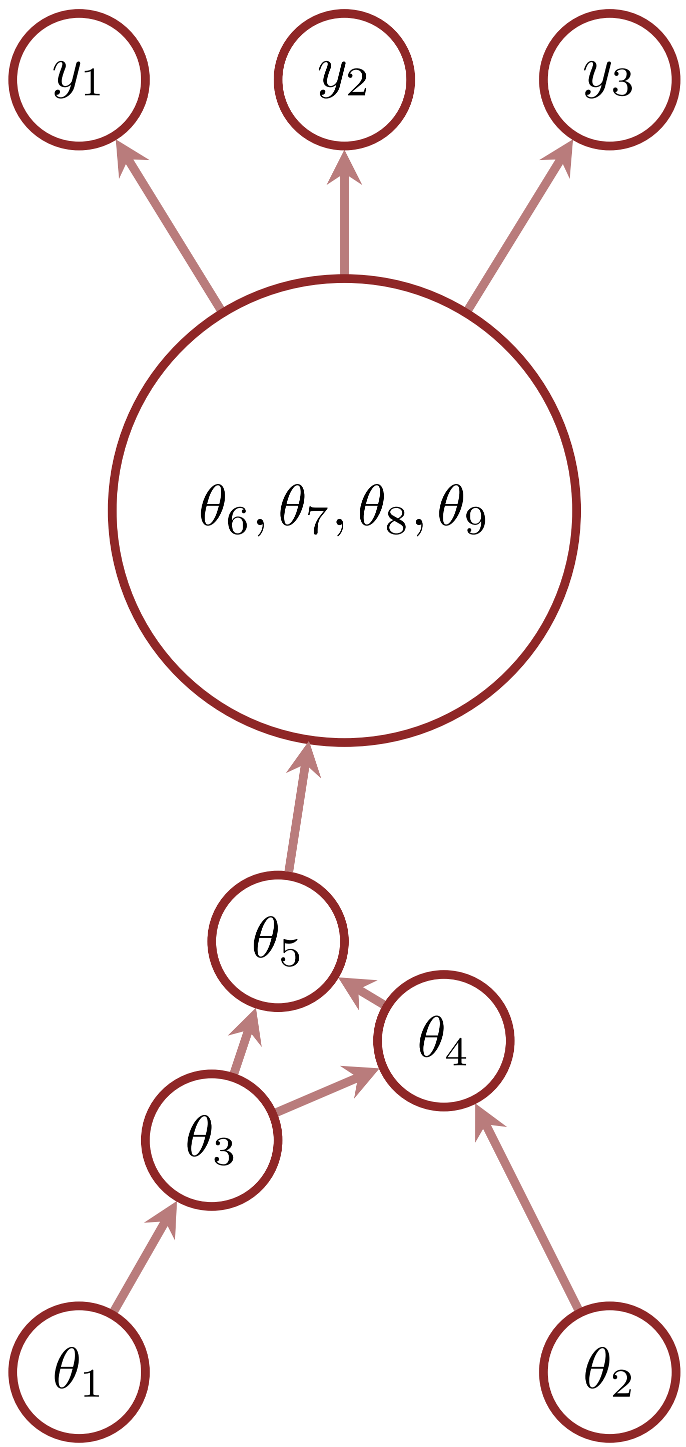
Narratively Generative Interpretation

We can build meaningfully narratively generative models from narratively non-generative components.



Narratively Generative Interpretation

In other words "narratively generative" is not a binary qualification of a full Bayesian model!

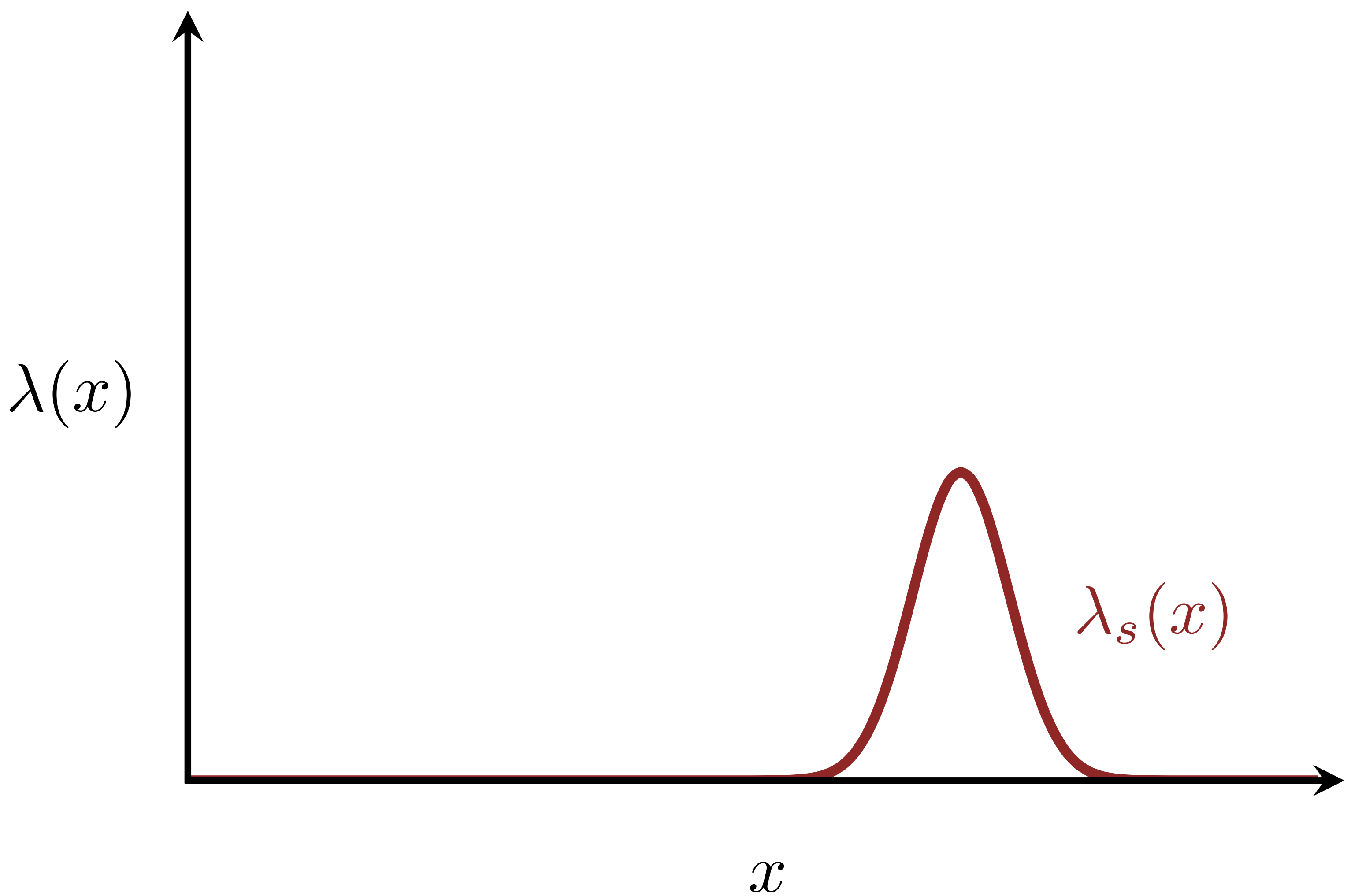


For example we might combine a theoretically-motivated signal model with a more heuristic background model.

$$\pi(y_n \mid \lambda_s(x_n) + \lambda_b(x_n))$$

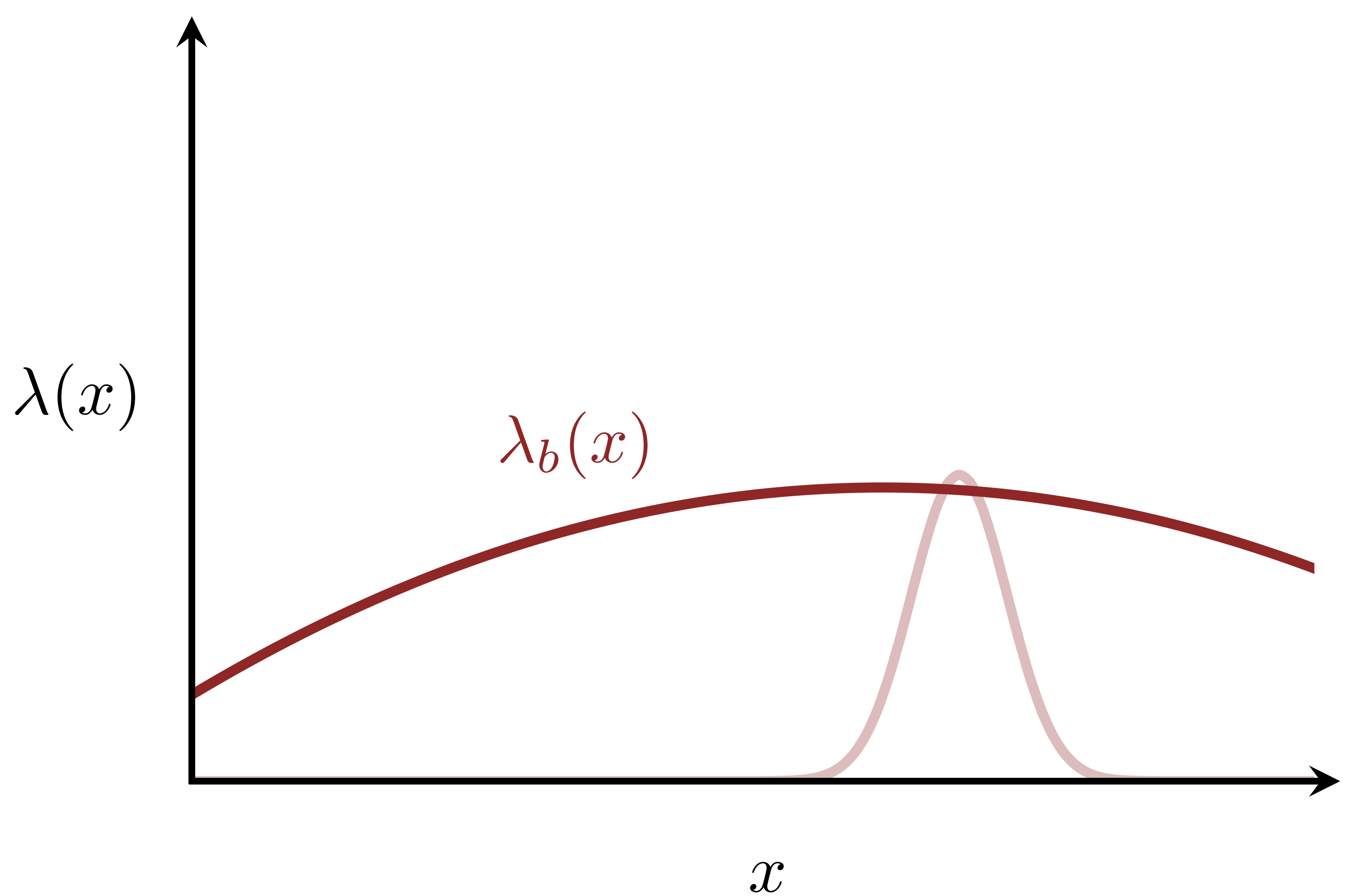
For example we might combine a theoretically-motivated signal model with a more heuristic background model.

$$\pi(y_n \mid \lambda_s(x_n) + \lambda_b(x_n))$$



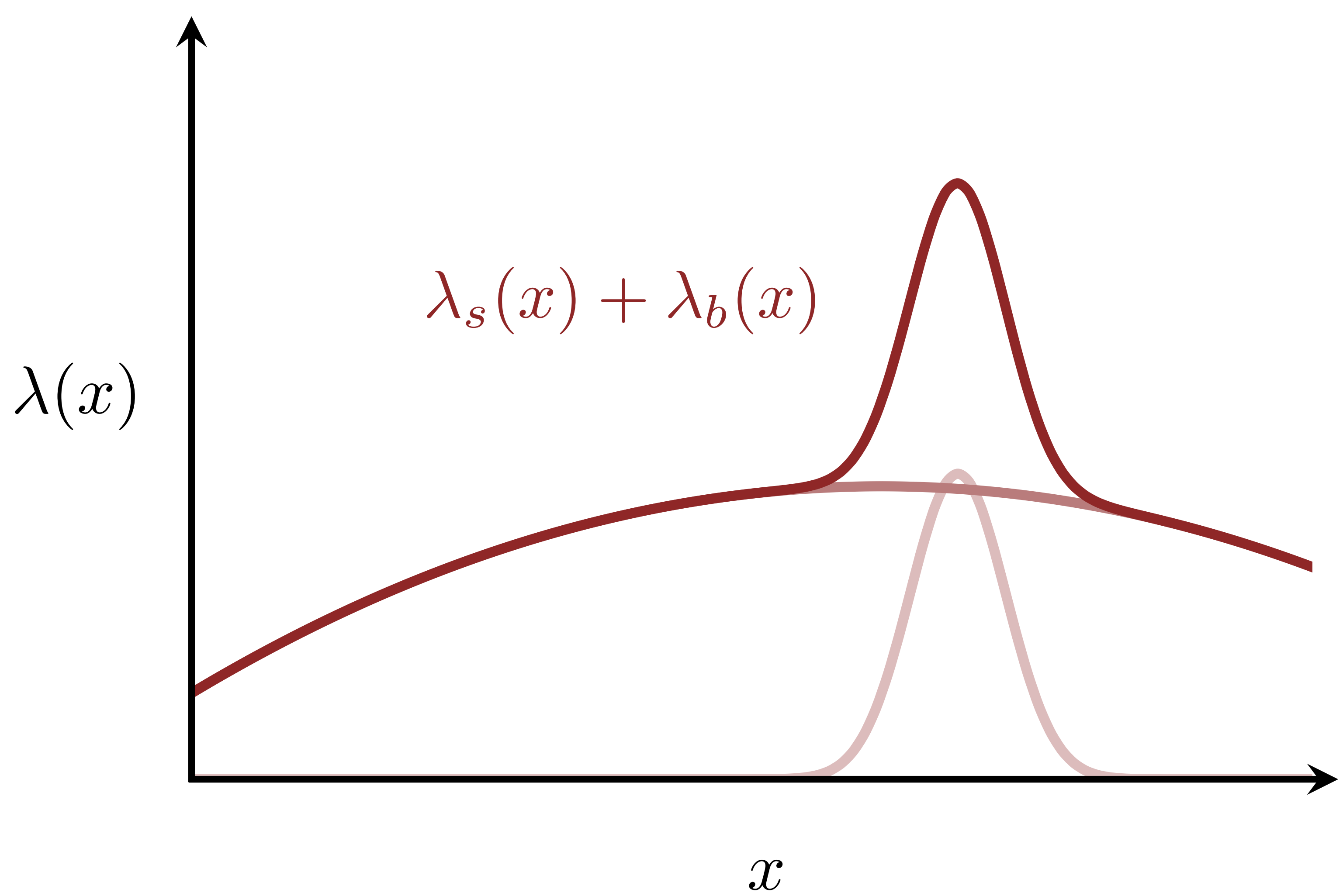
For example we might combine a theoretically-motivated signal model with a more heuristic background model.

$$\pi(y_n \mid \lambda_s(x_n) + \lambda_b(x_n))$$

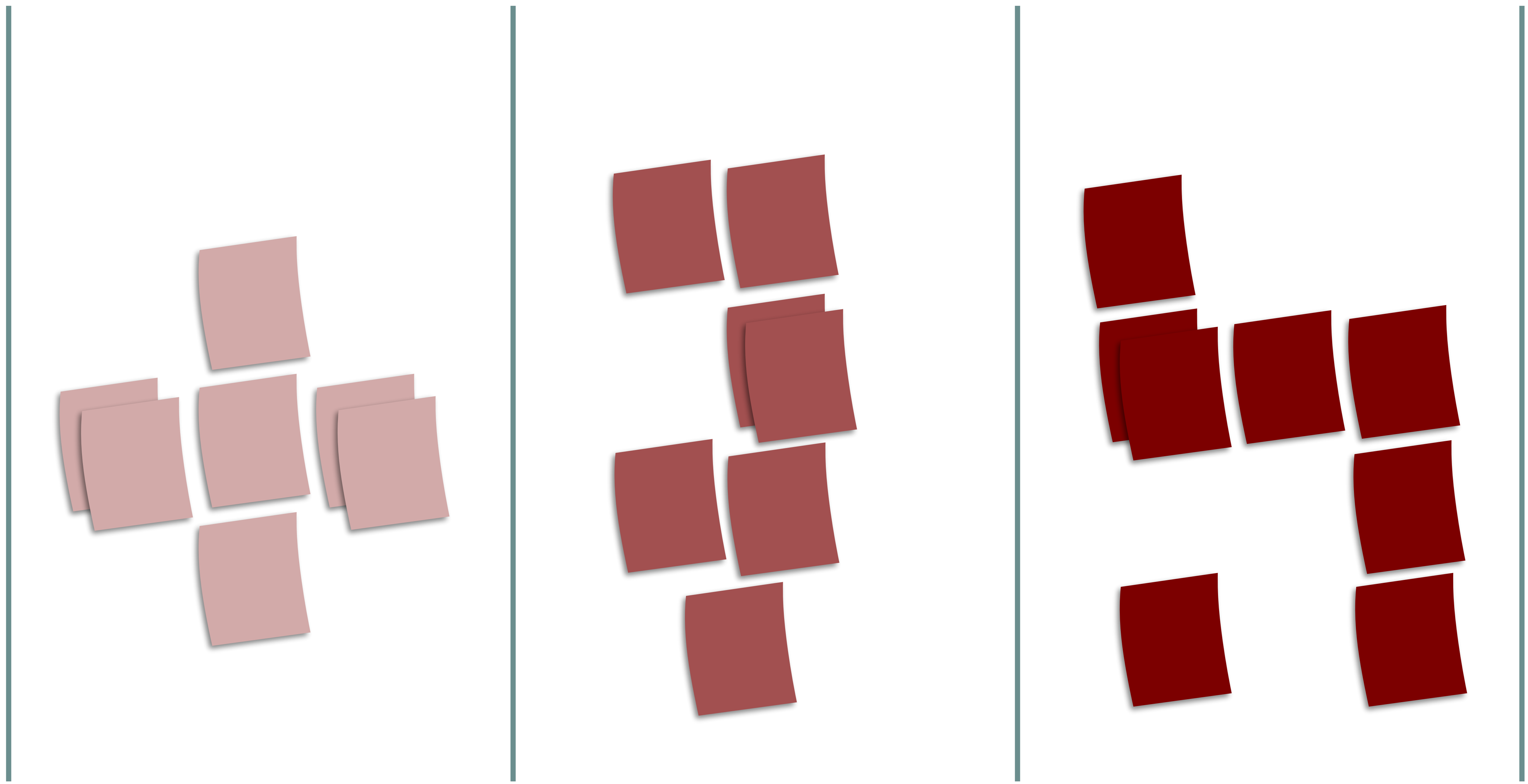


For example we might combine a theoretically-motivated signal model with a more heuristic background model.

$$\pi(y_n \mid \lambda_s(x_n) + \lambda_b(x_n))$$

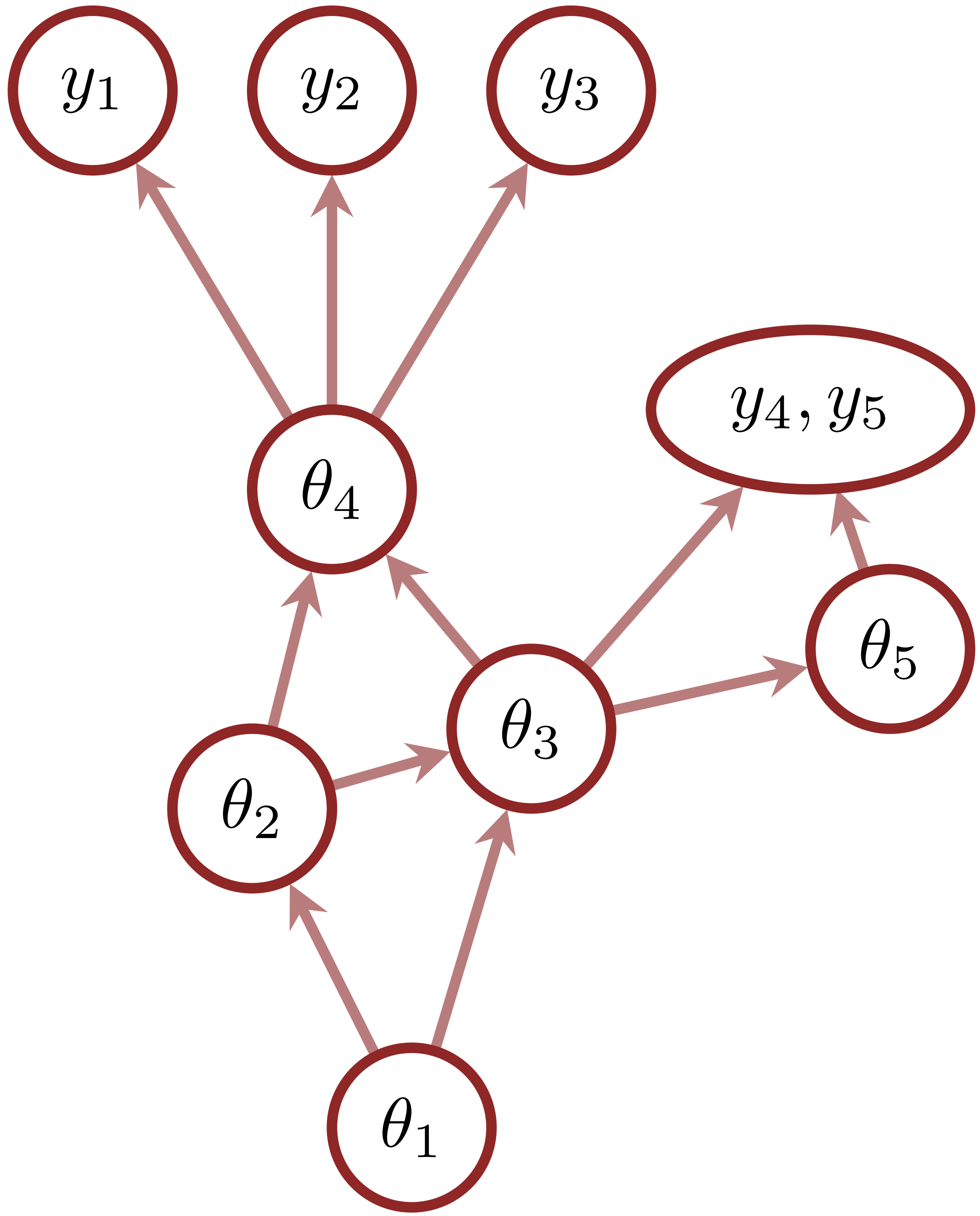


Advanced Narrative Techniques

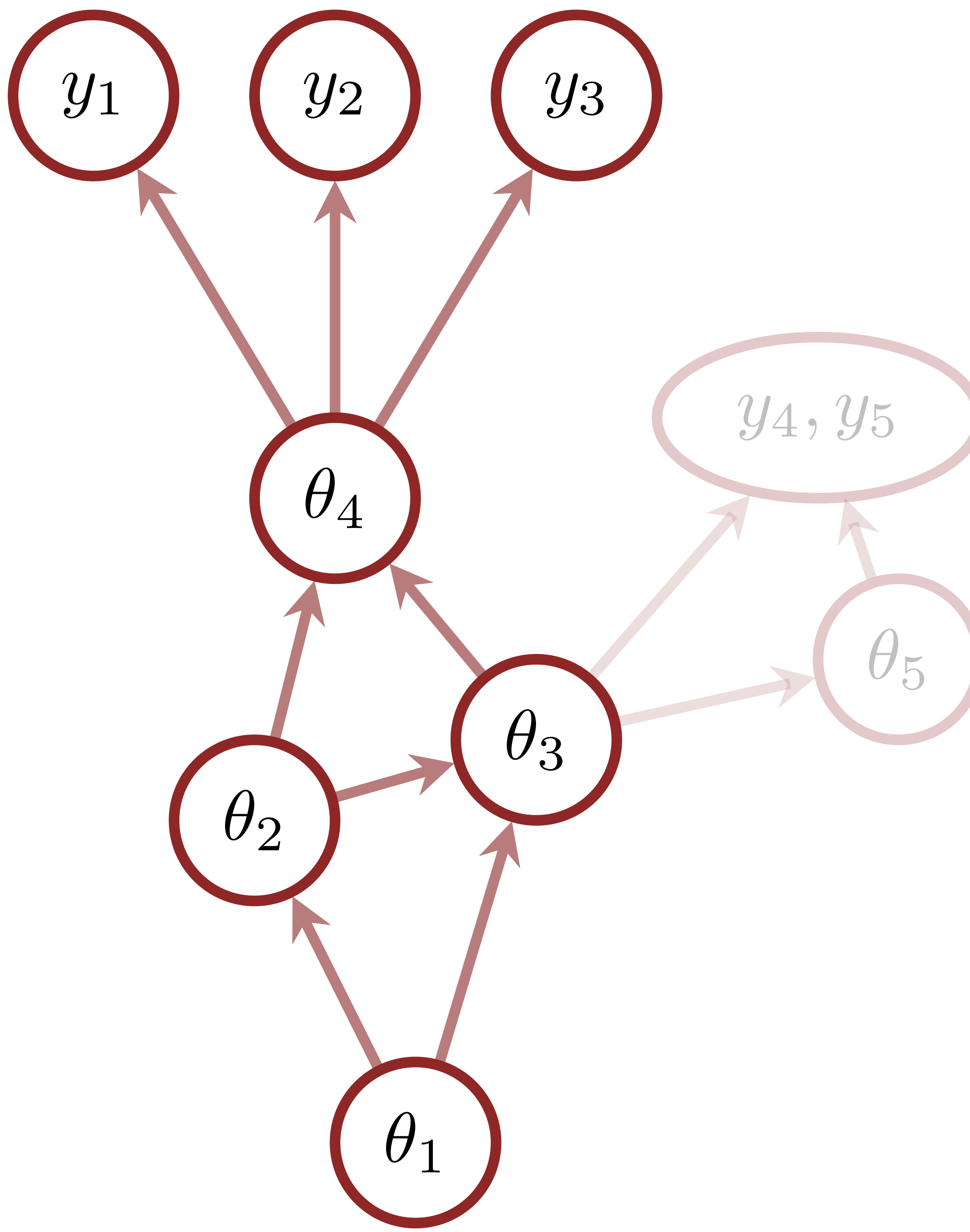


A Tangled Web

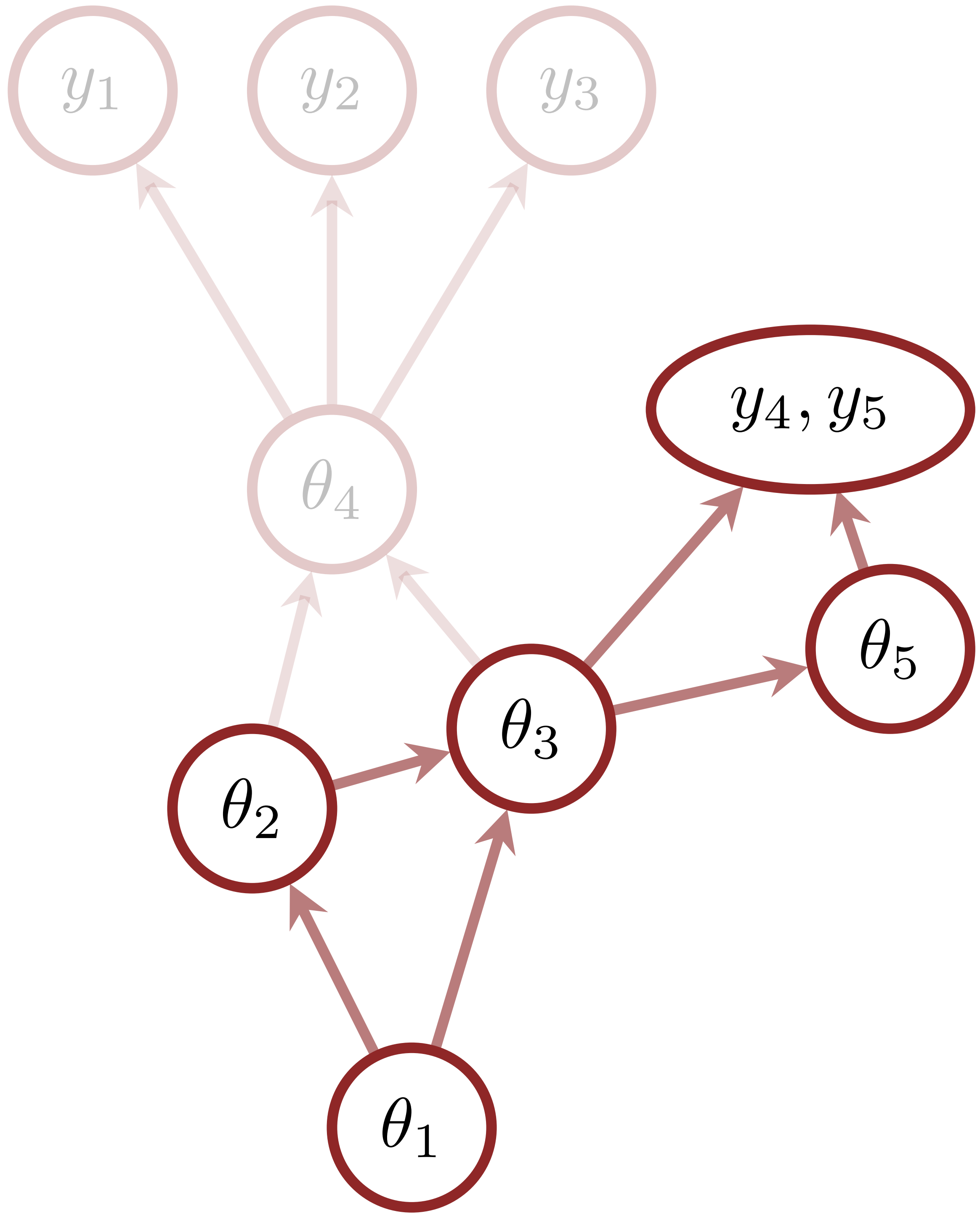
A story can tell multiple narratives, and a model can incorporate multiple data generating processes.



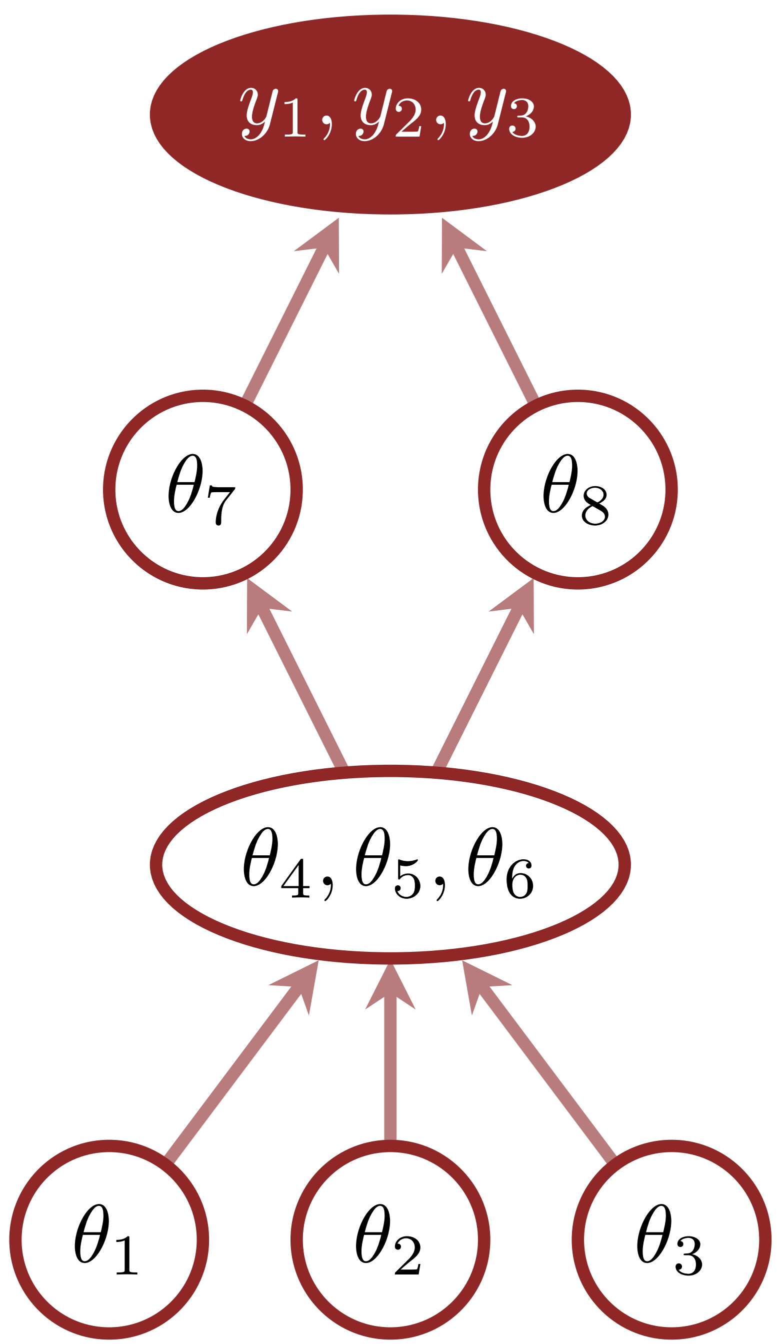
A story can tell multiple narratives, and a model can incorporate multiple data generating processes.



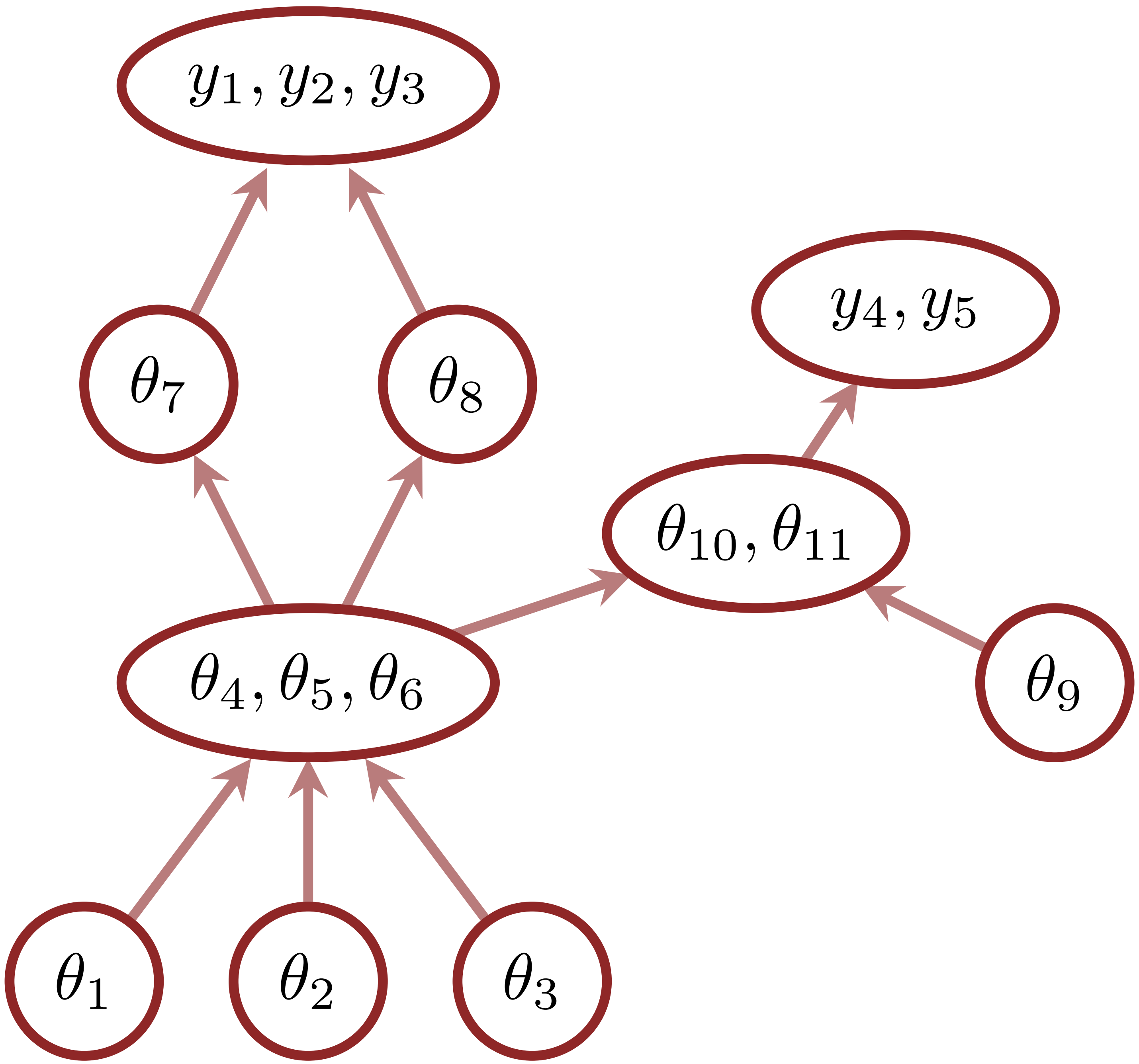
A story can tell multiple narratives, and a model can incorporate multiple data generating processes.



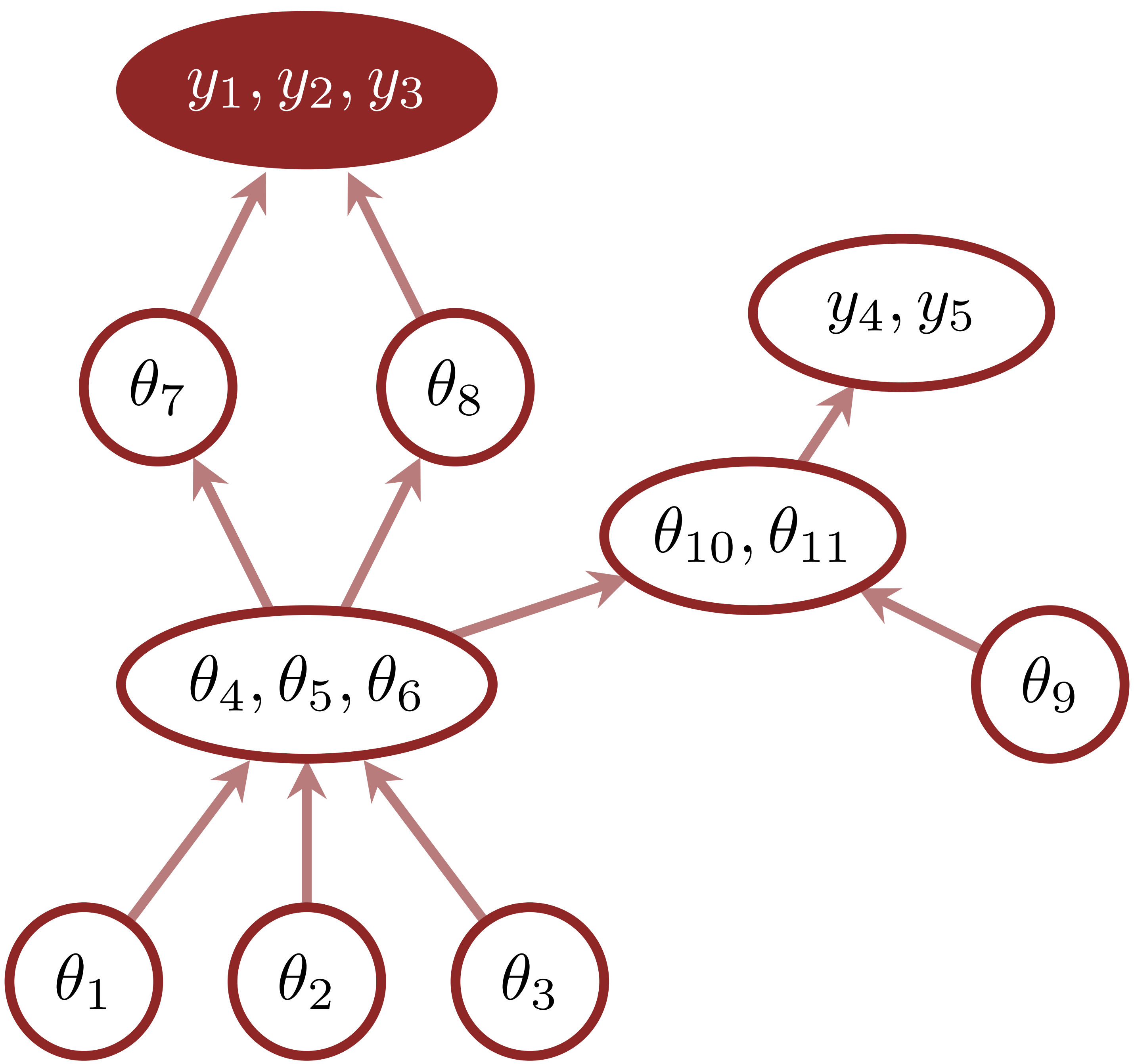
These joint models also facilitate predictions that generalize beyond the scope of an initial observation.



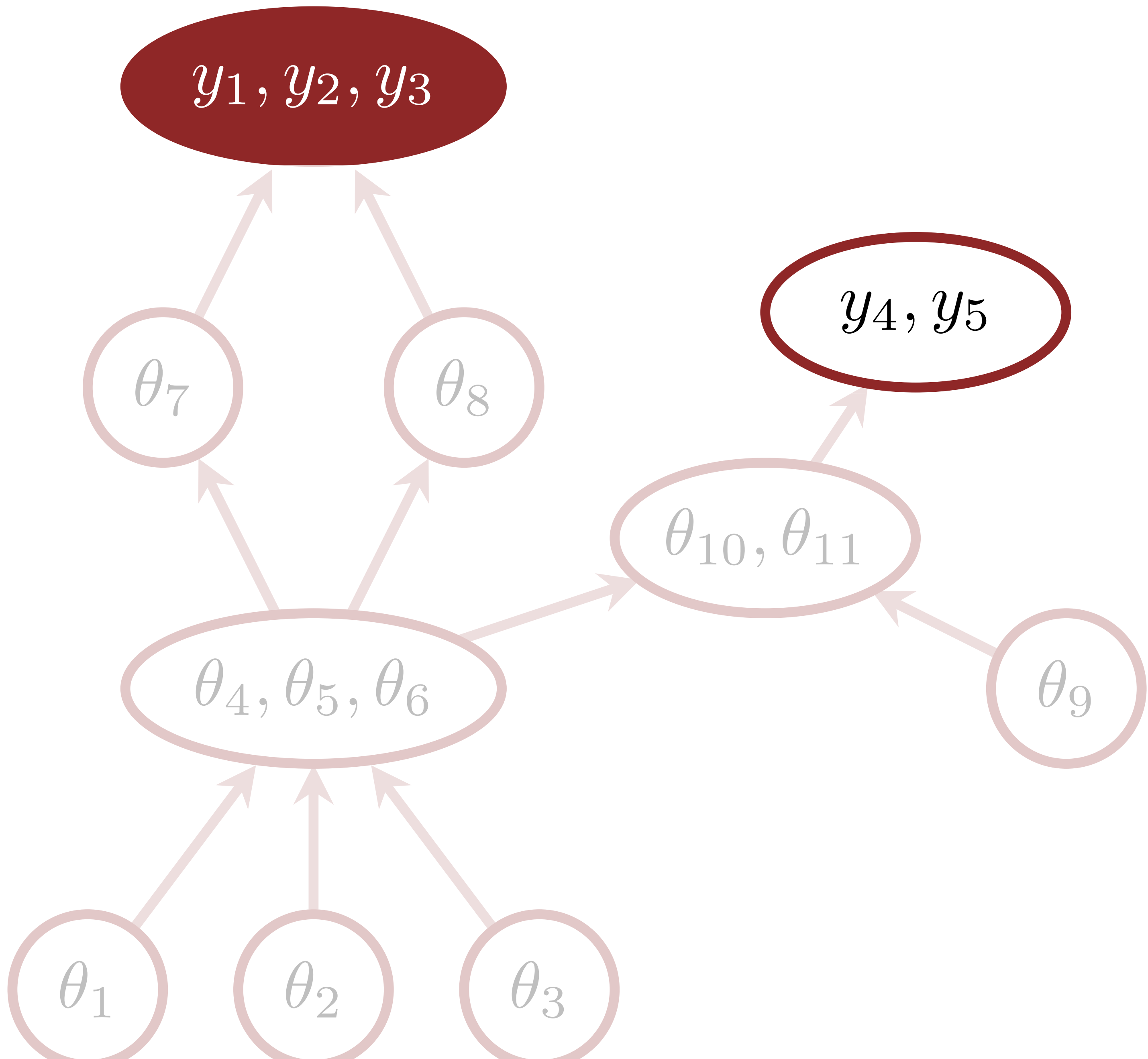
These joint models also facilitate predictions that generalize beyond the scope of an initial observation.



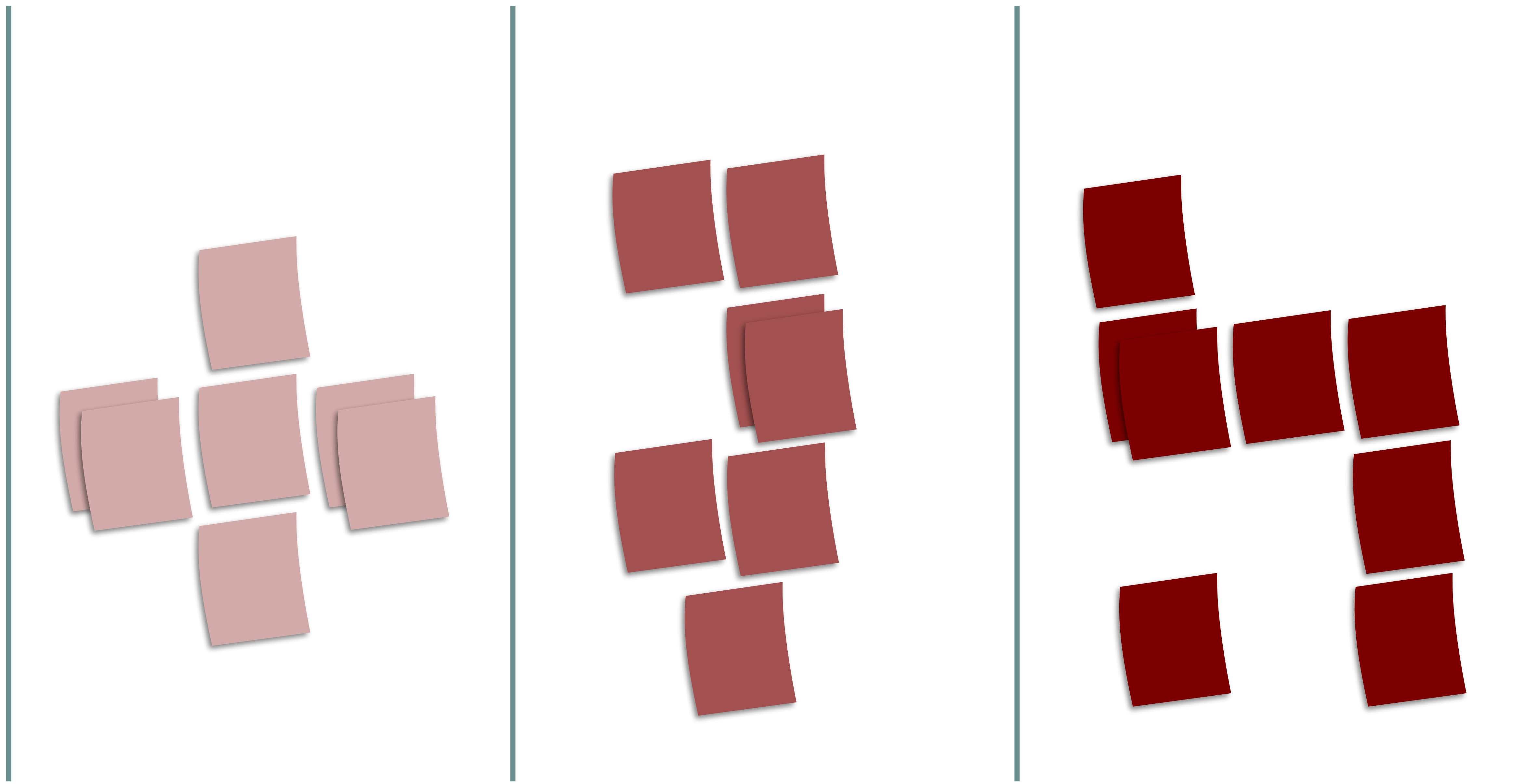
These joint models also facilitate predictions that generalize beyond the scope of an initial observation.



These joint models also facilitate predictions that generalize beyond the scope of an initial observation.

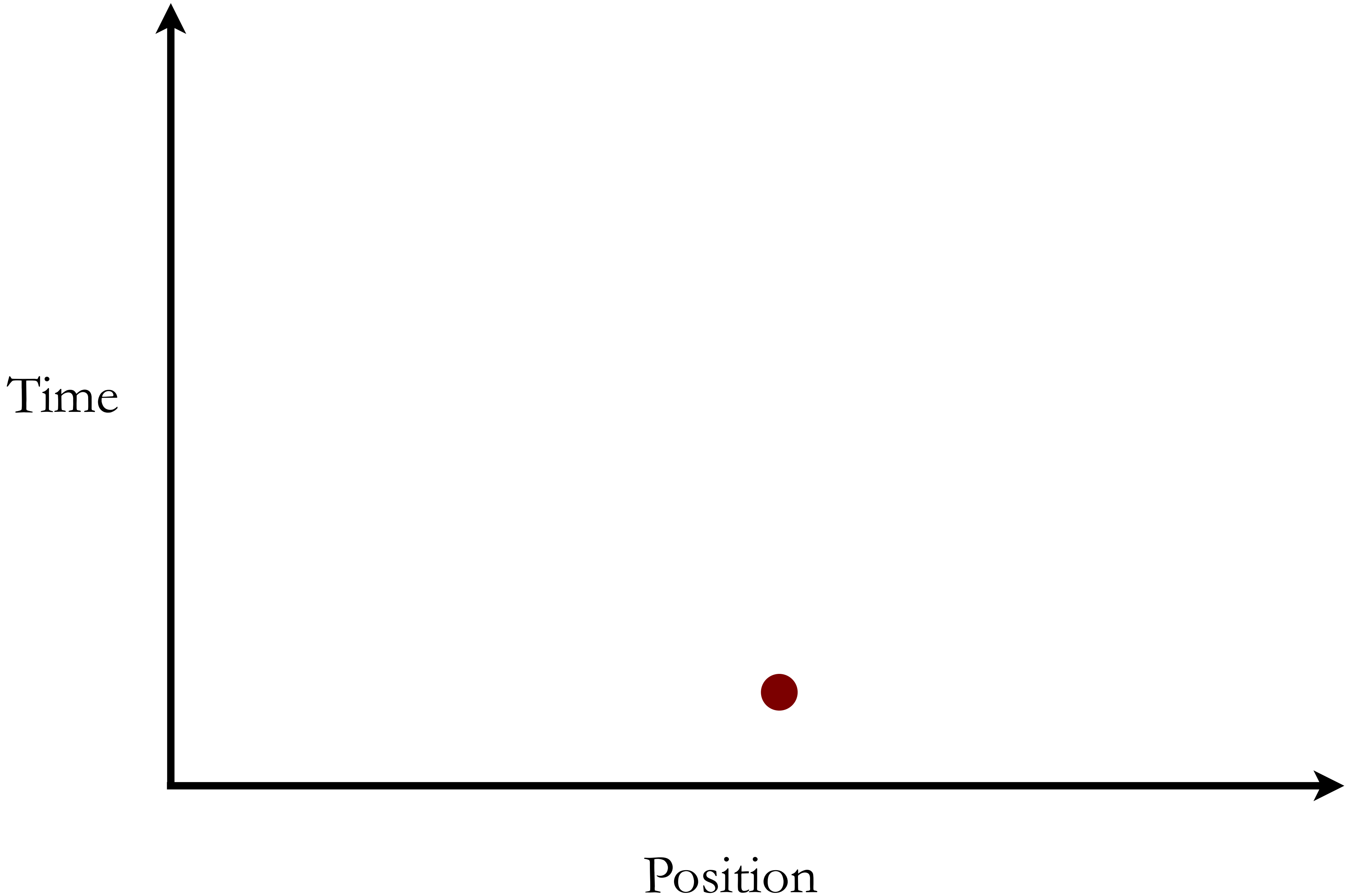


Advanced Narrative Techniques

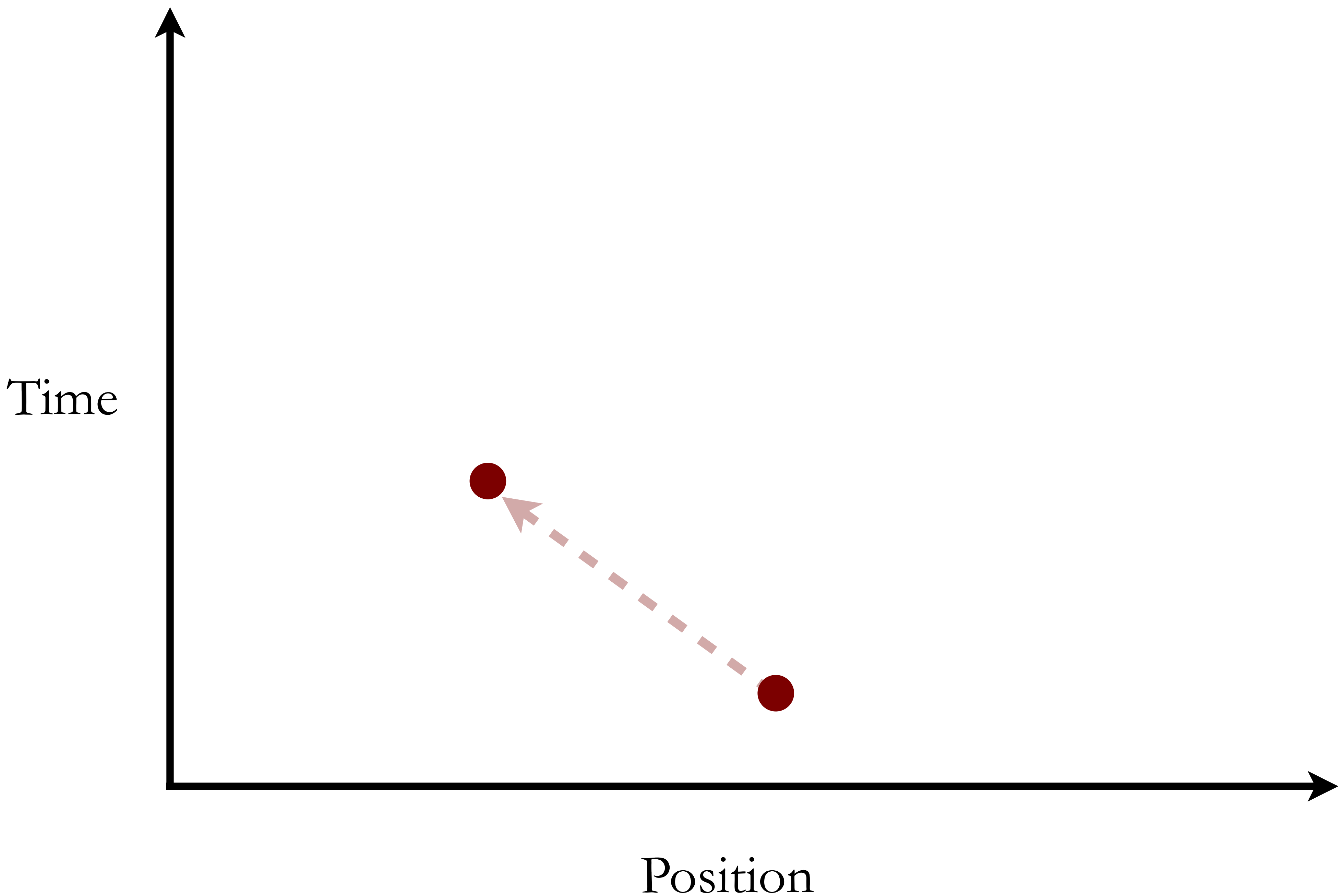


“Causal” “Inference”

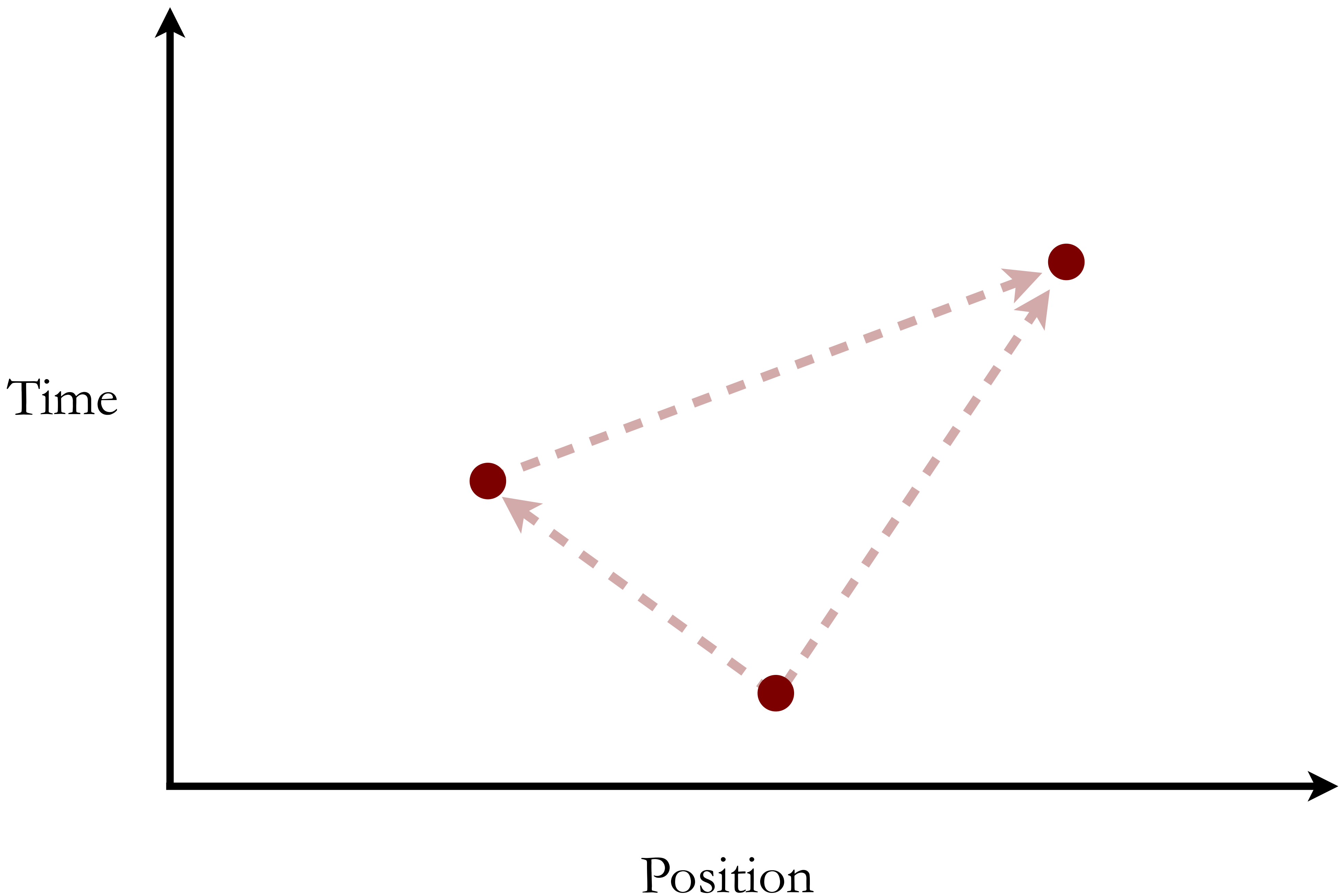
Many data generating stories follow a cascade of propagating causes and effects.



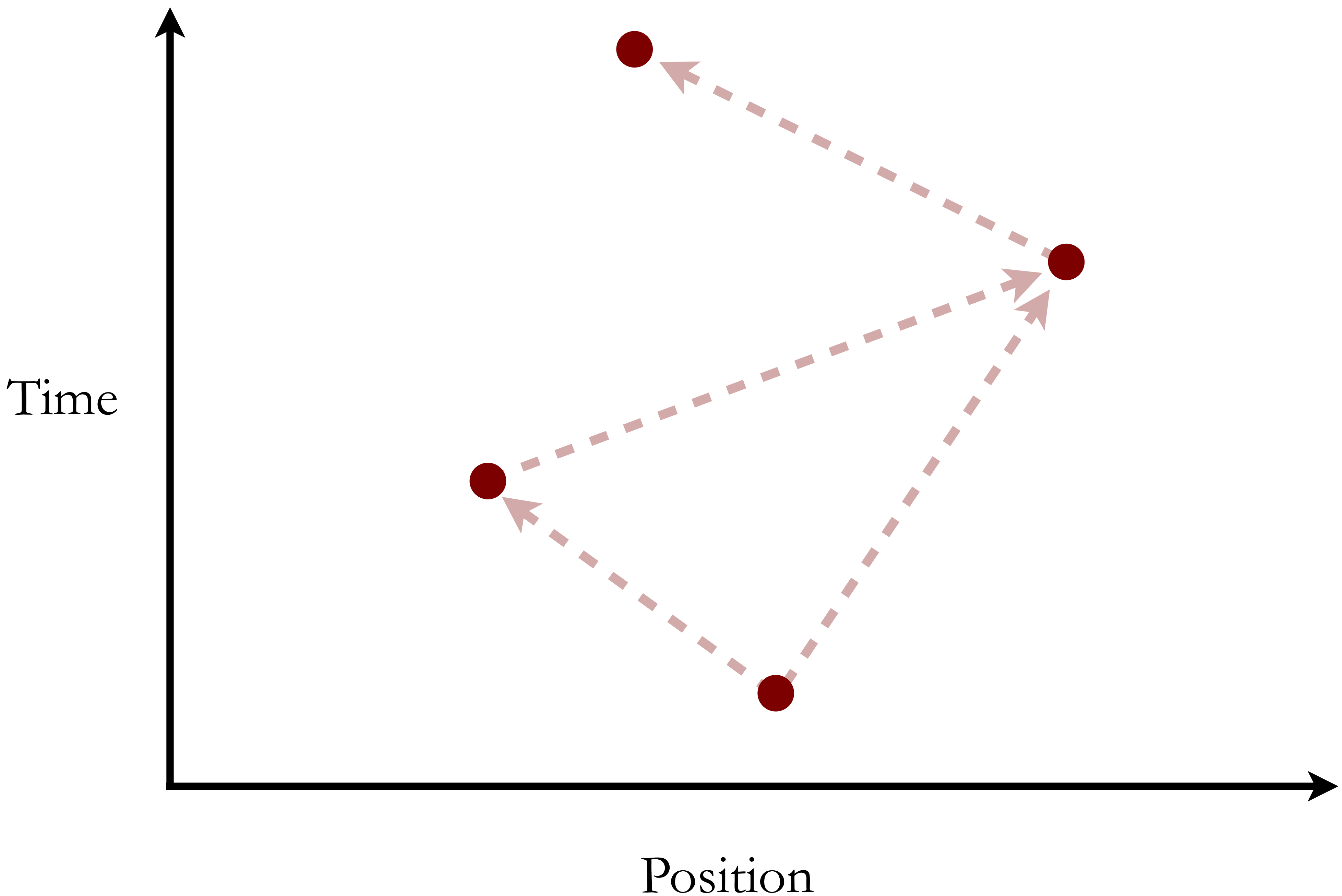
Many data generating stories follow a cascade of propagating causes and effects.



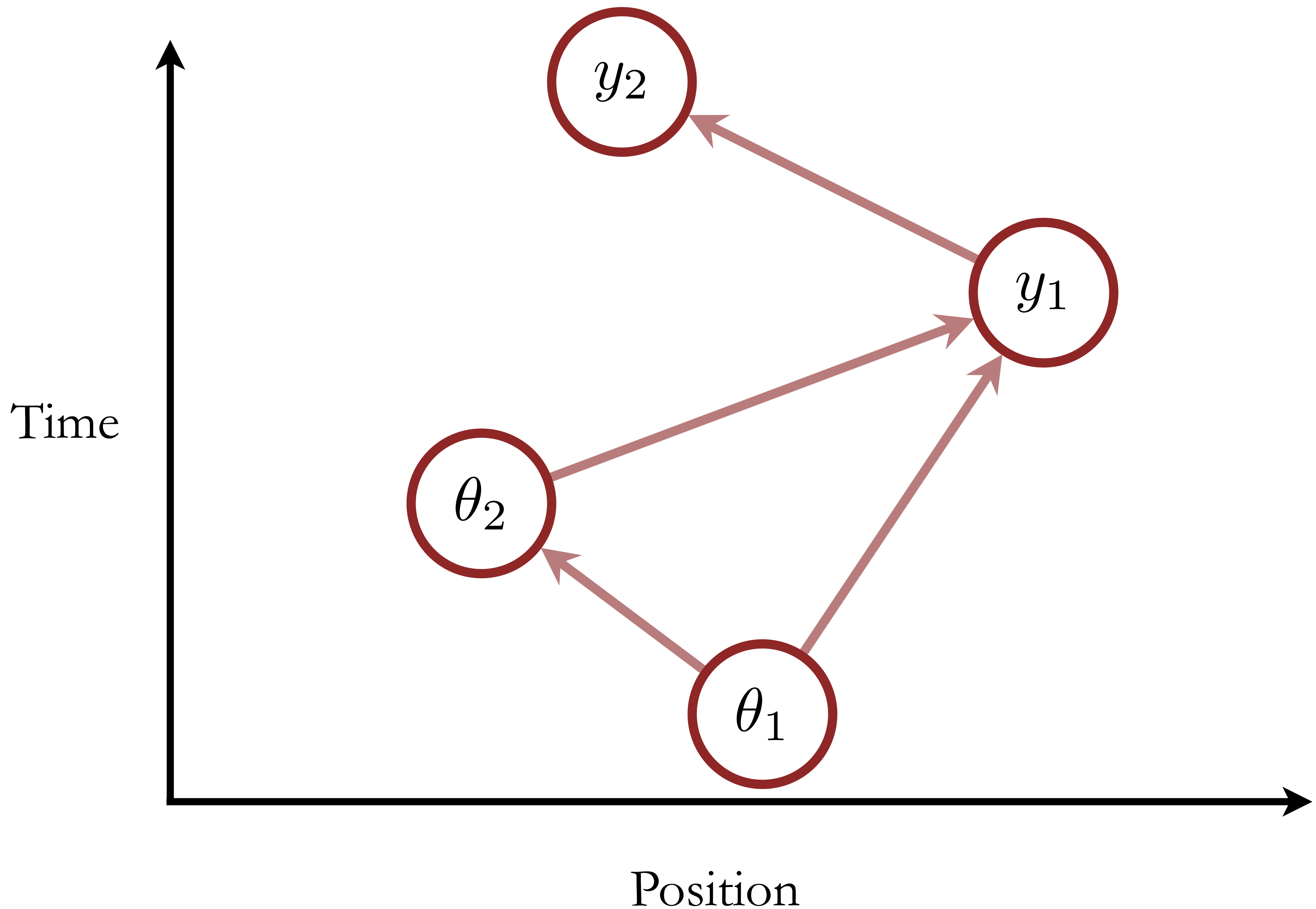
Many data generating stories follow a cascade of propagating causes and effects.



Many data generating stories follow a cascade of propagating causes and effects.



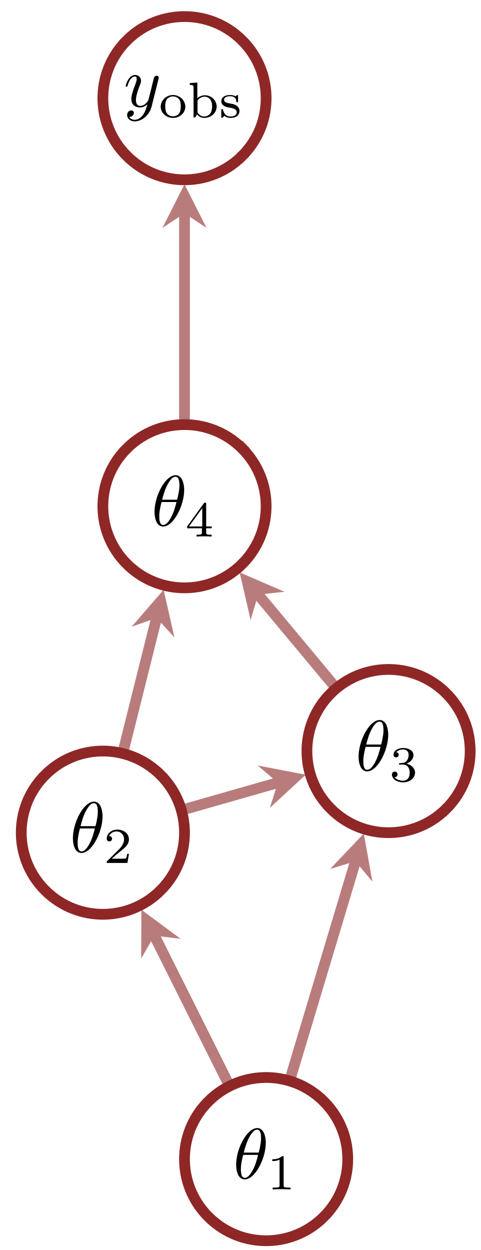
These *causal* influences often motivate particular narratively generative model structures.



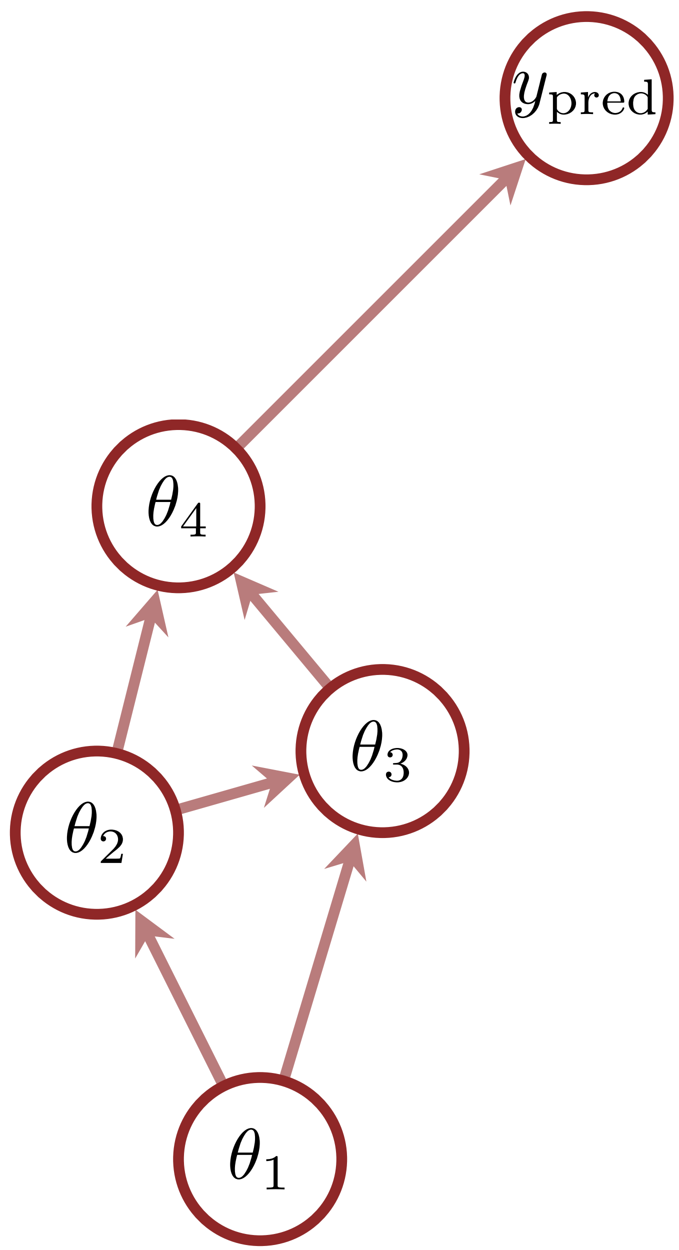
“Causal” often refers to inferential or predictive methods that allow for *any* change in the data generating process.

Making Predictions In A Static World	**Making Predictions In A Dynamic World**
Interpolation	Extrapolation
In-Sample	Out-of-Sample
...	Generalization
	Transfer Learning
	Model/Concept Drift
	Hypothetical
	Counterfactual
	...

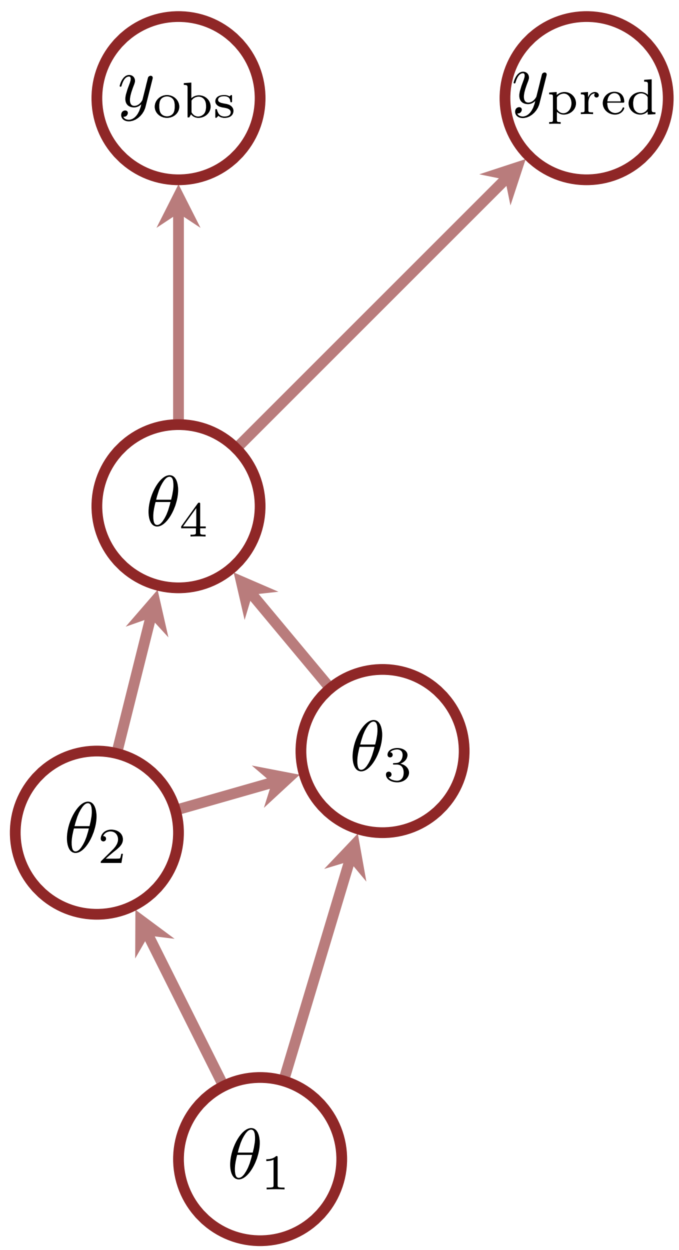
Both static and dynamic processes are straightforward to incorporate into narratively generative models.



Both static and dynamic processes are straightforward to incorporate into narratively generative models.

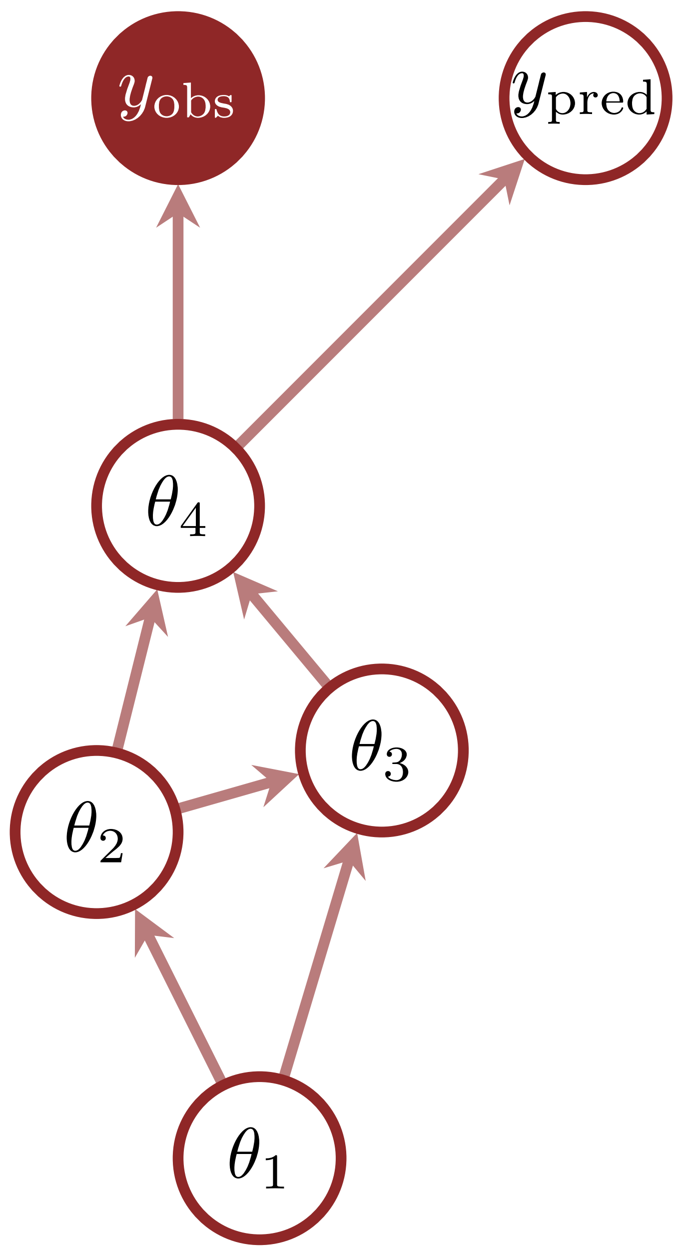


Both static and dynamic processes are straightforward to incorporate into narratively generative models.



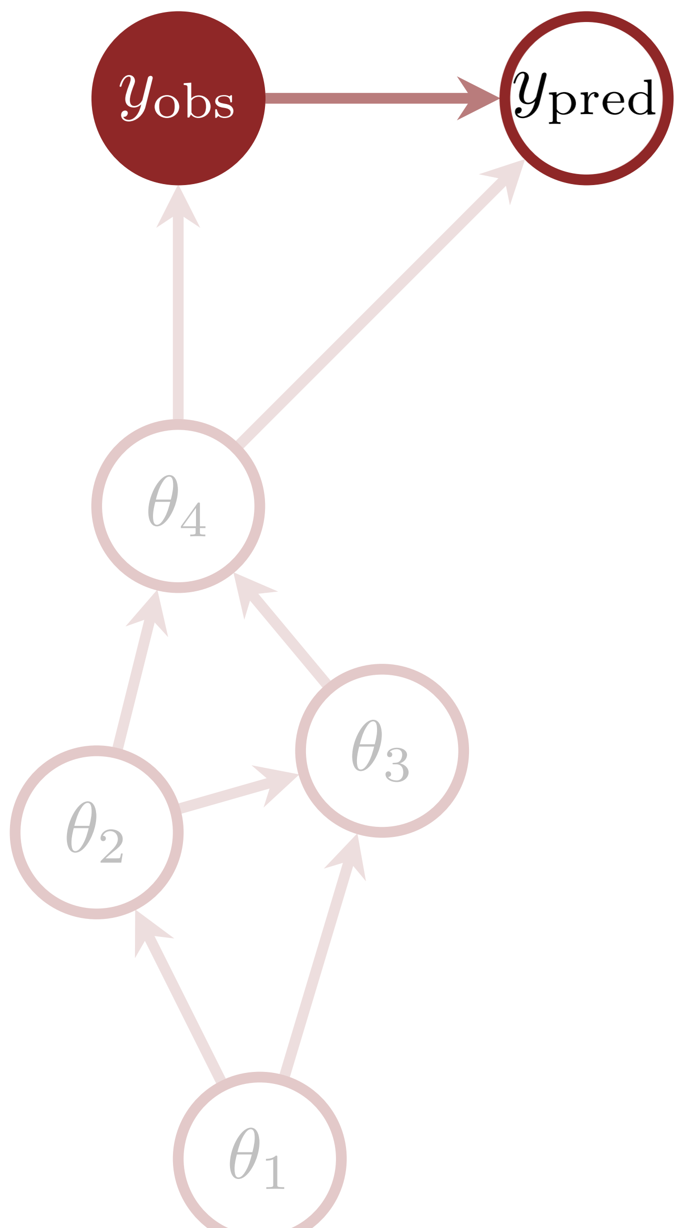
$$\pi(y_{\text{obs}}, y_{\text{pred}}, \theta_1, \theta_2, \theta_3, \theta_4)$$

Both static and dynamic processes are straightforward to incorporate into narratively generative models.



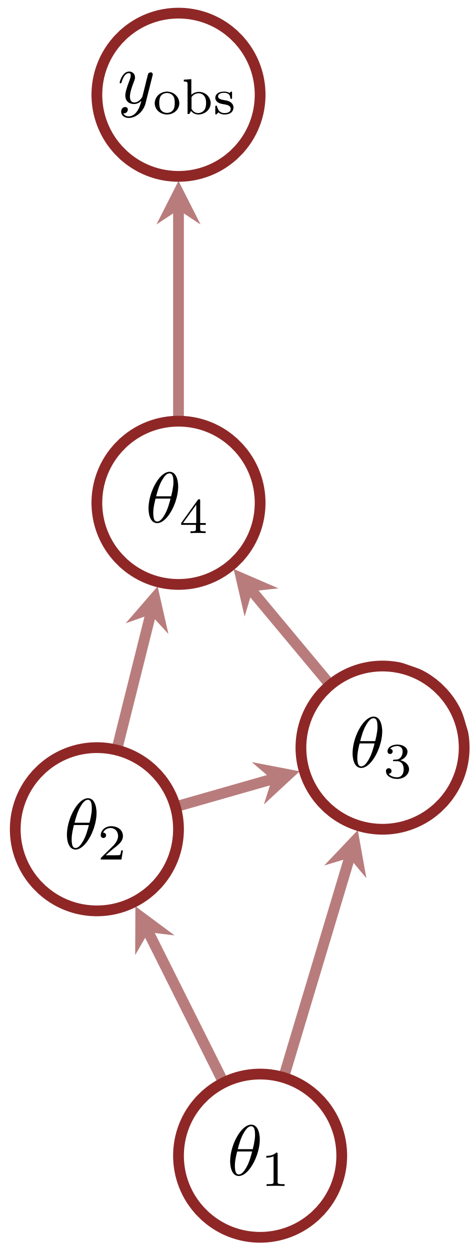
$$\pi(y_{\text{pred}}, \theta_1, \theta_2, \theta_3, \theta_4 \mid \tilde{y}_{\text{obs}}) \propto \pi(\tilde{y}_{\text{obs}}, y_{\text{pred}}, \theta_1, \theta_2, \theta_3, \theta_4)$$

Both static and dynamic processes are straightforward to incorporate into narratively generative models.

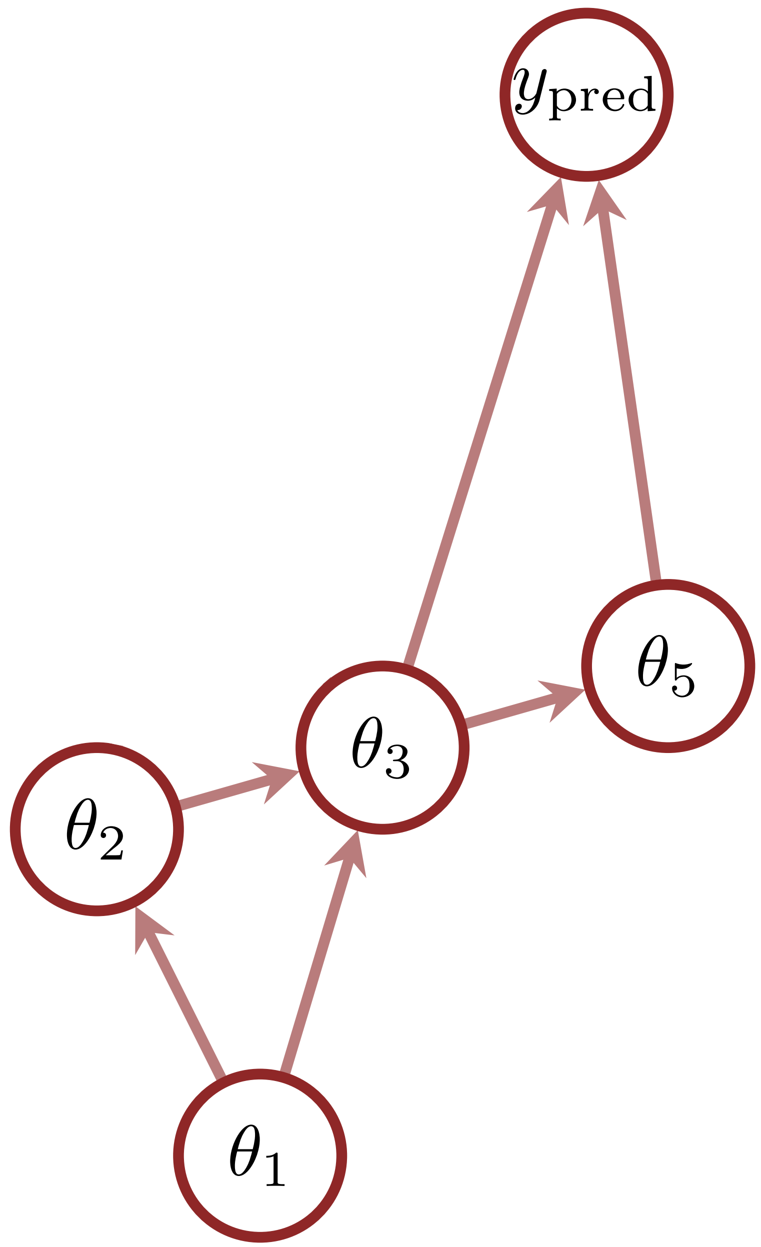


$$\pi(y_{\text{pred}} \mid \tilde{y}_{\text{obs}}) = \int \mathrm{d}\theta_1 \, \mathrm{d}\theta_2 \, \mathrm{d}\theta_3 \, \mathrm{d}\theta_4 \, \pi(y_{\text{pred}}, \theta_1, \theta_2, \theta_3, \theta_4 \mid \tilde{y}_{\text{obs}})$$

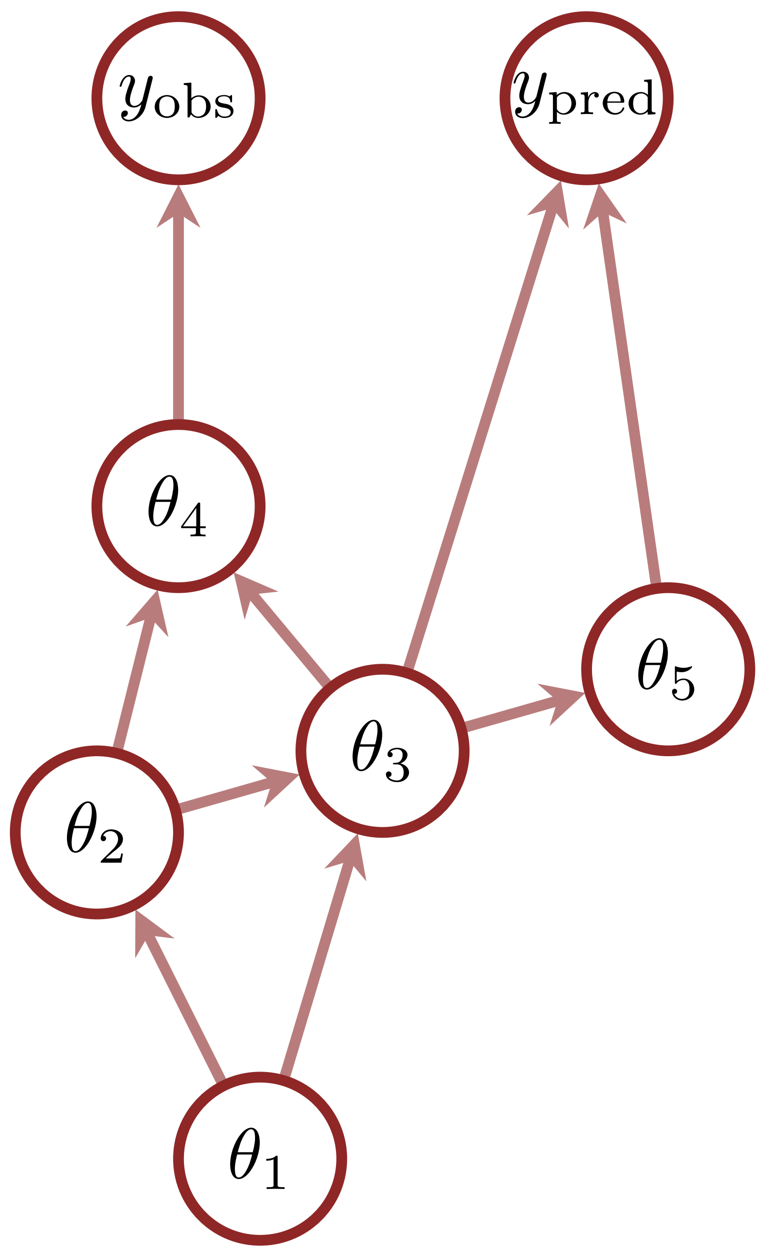
Both static and dynamic processes are straightforward to incorporate into narratively generative models.



Both static and dynamic processes are straightforward to incorporate into narratively generative models.

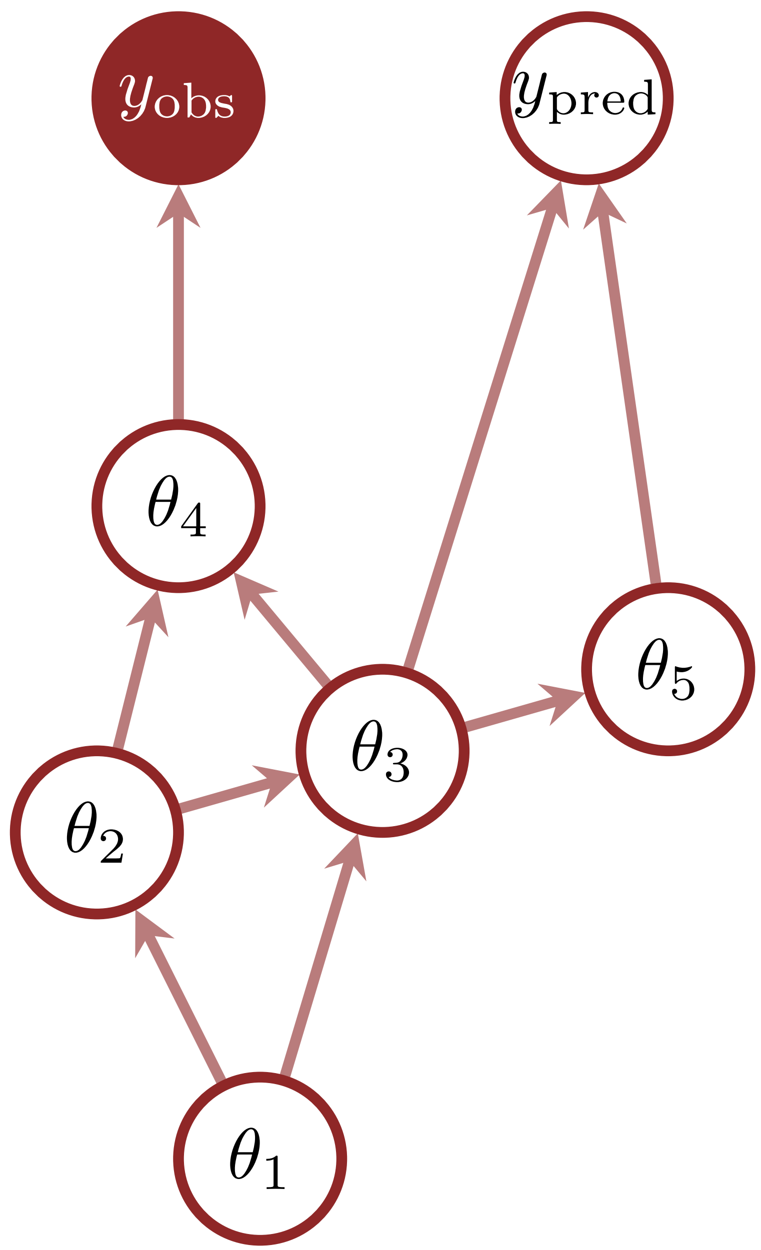


Both static and dynamic processes are straightforward to incorporate into narratively generative models.



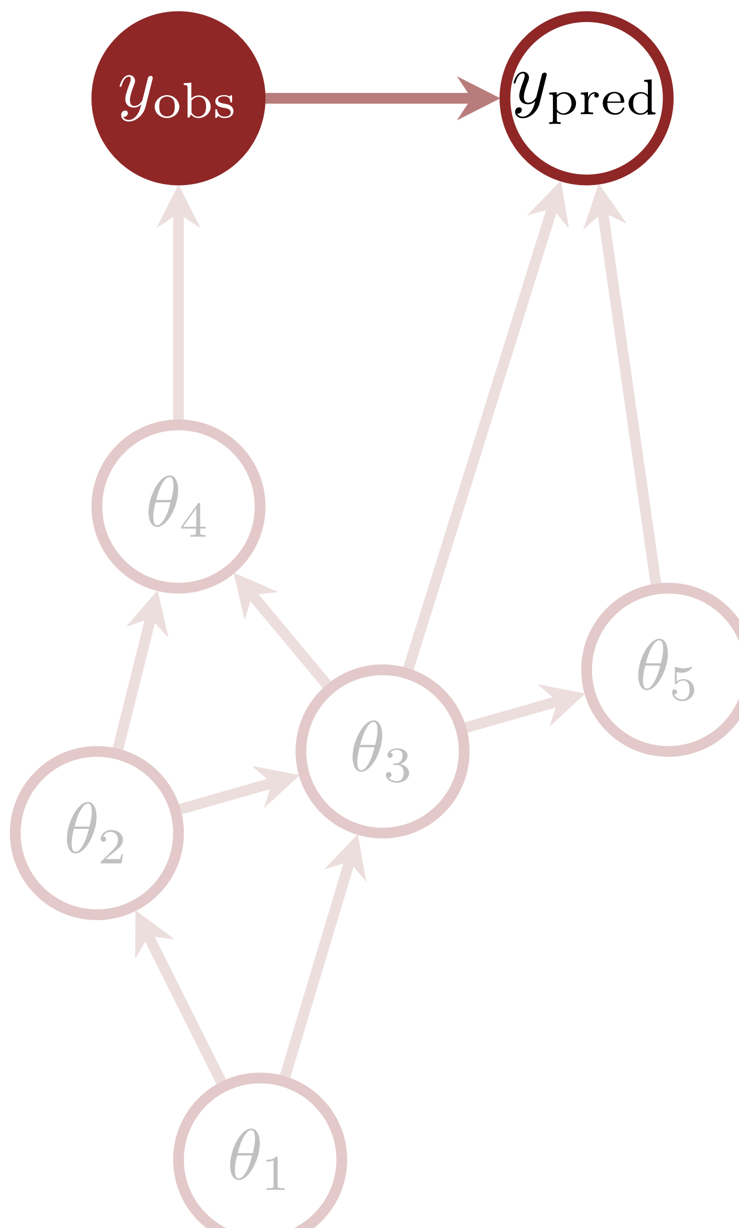
$$\pi(y_{\text{obs}}, y_{\text{pred}}, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$$

Both static and dynamic processes are straightforward to incorporate into narratively generative models.



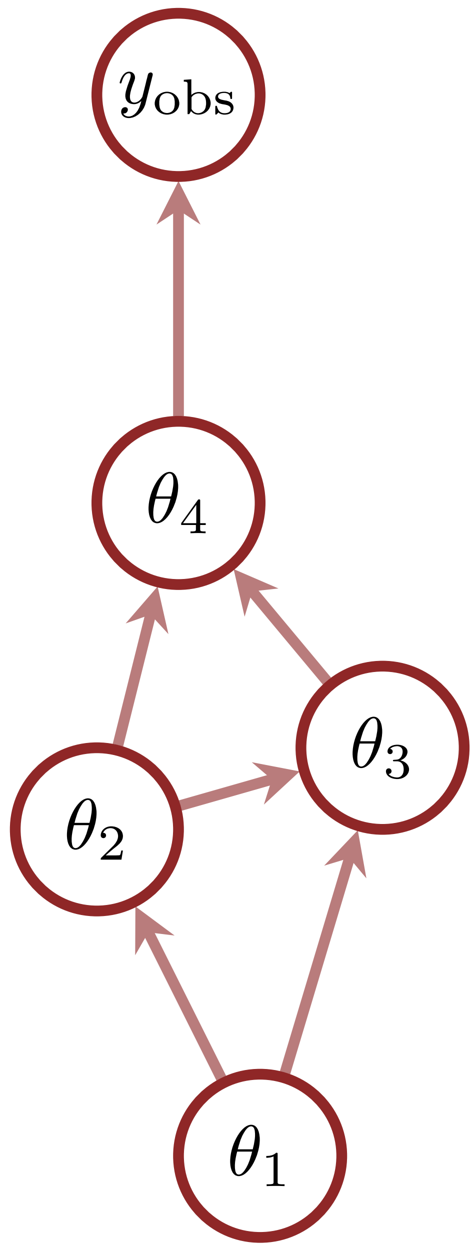
$$\pi(y_{\text{pred}}, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5 \mid \tilde{y}_{\text{obs}}) \propto \pi(\tilde{y}_{\text{obs}}, y_{\text{pred}}, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$$

Both static and dynamic processes are straightforward to incorporate into narratively generative models.



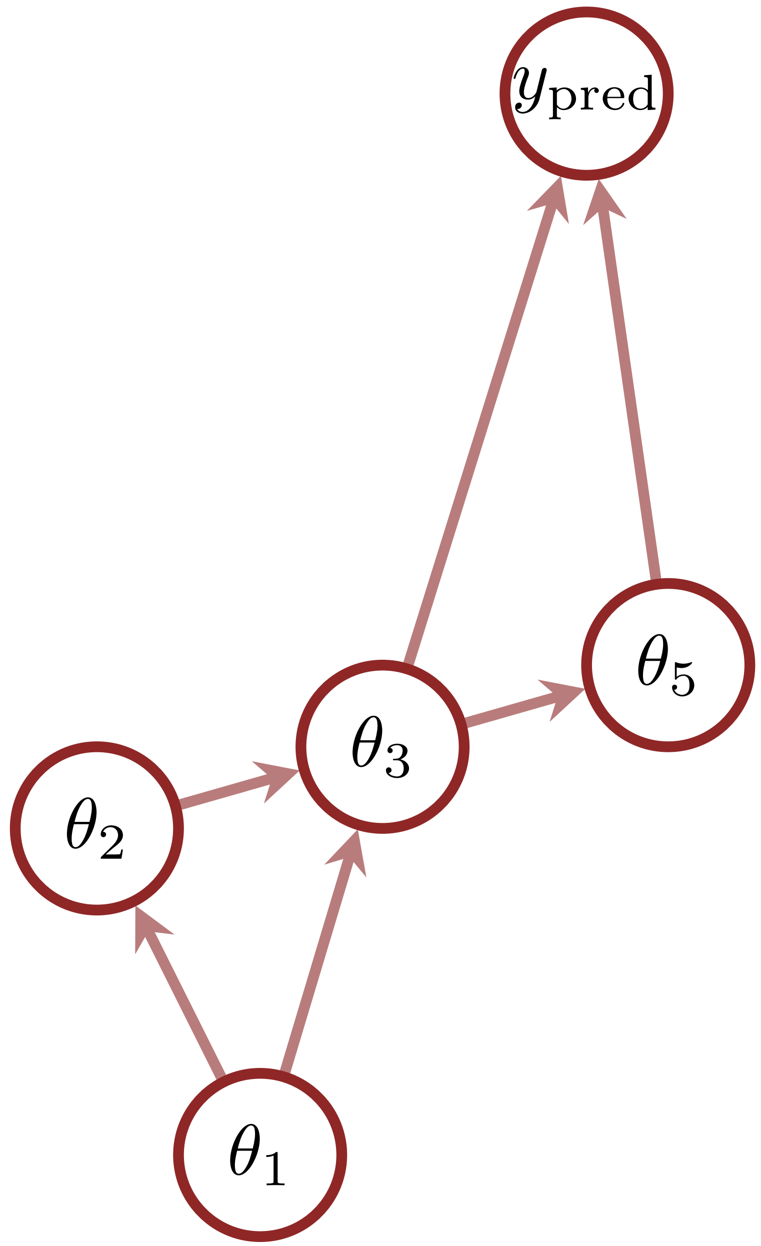
$$\pi(y_{\text{pred}} \mid \tilde{y}_{\text{obs}}) = \int \mathrm{d}\theta_1 \, \mathrm{d}\theta_2 \, \mathrm{d}\theta_3 \, \mathrm{d}\theta_4 \, \mathrm{d}\theta_5 \, \pi(y_{\text{pred}}, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5 \mid \tilde{y}_{\text{obs}})$$

Interventions are modeled the same as any change to the data generating process.



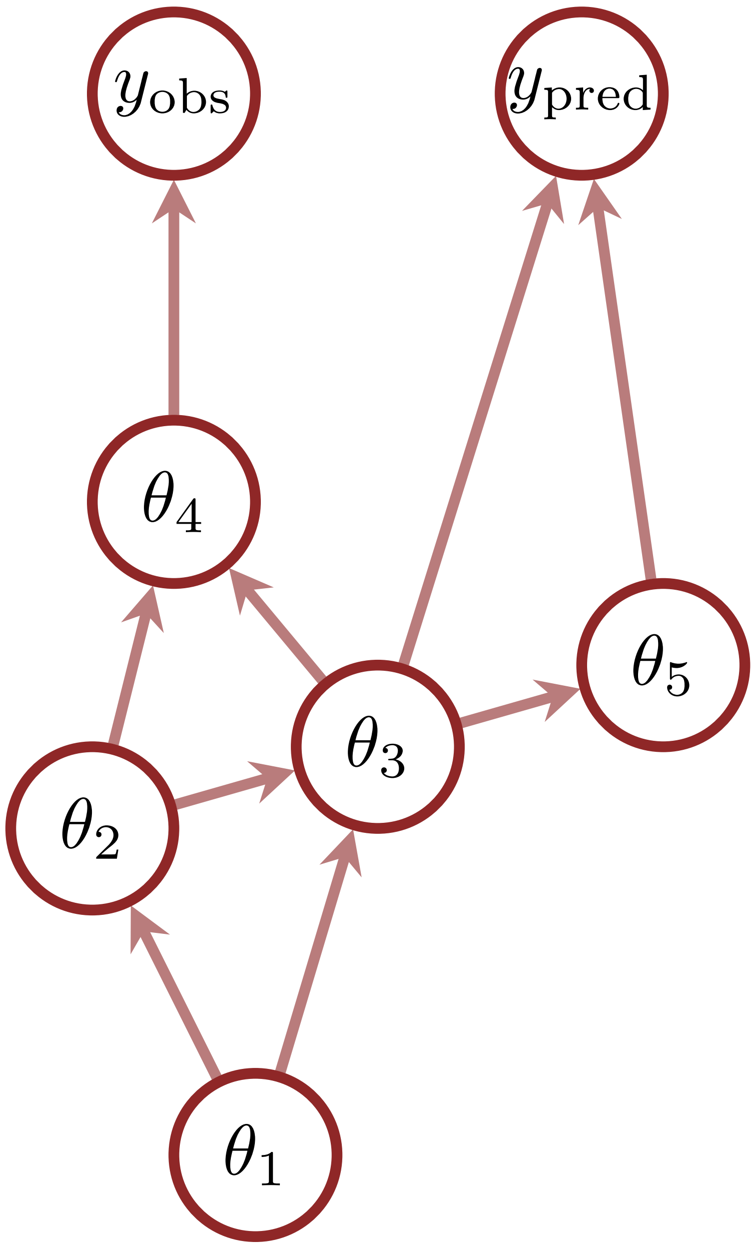
Before Intervention

Interventions are modeled the same as any change to the data generating process.



After Intervention

Interventions are modeled the same as any change to the data generating process.



Joint Model

https://betanalpha.github.io/writing/

Part II: Modeling and Inference

- Generative Modeling

Case Studies

https://betanalpha.github.io/writing/

Part II: Modeling and Inference

- Generative Modeling

Case Studies

https://betanalpha.github.io/speaking/

Generative Modeling Lecture and Exercise Review (YouTube)

Plurinational Bayes Webinar (July 2024)

https://betanalpha.github.io/writing/

Part II: Modeling and Inference

- Generative Modeling

Case Studies

https://betanalpha.github.io/speaking/

Generative Modeling Lecture and Exercise Review (YouTube)

Plurinational Bayes Webinar (July 2024)

Today's Exercise!

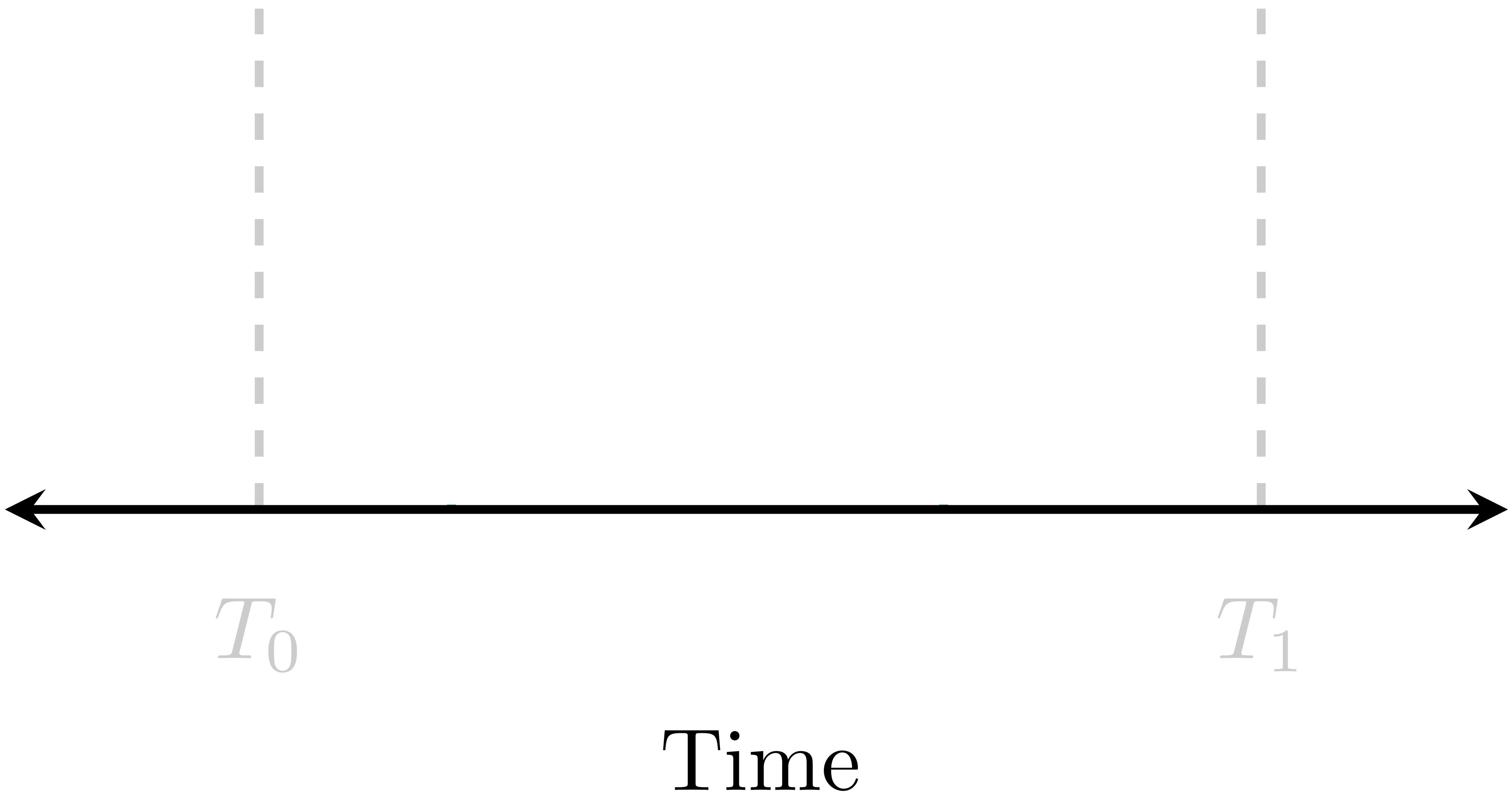
Socials and Mailing List

- https://betanalpha.github.io/contact/

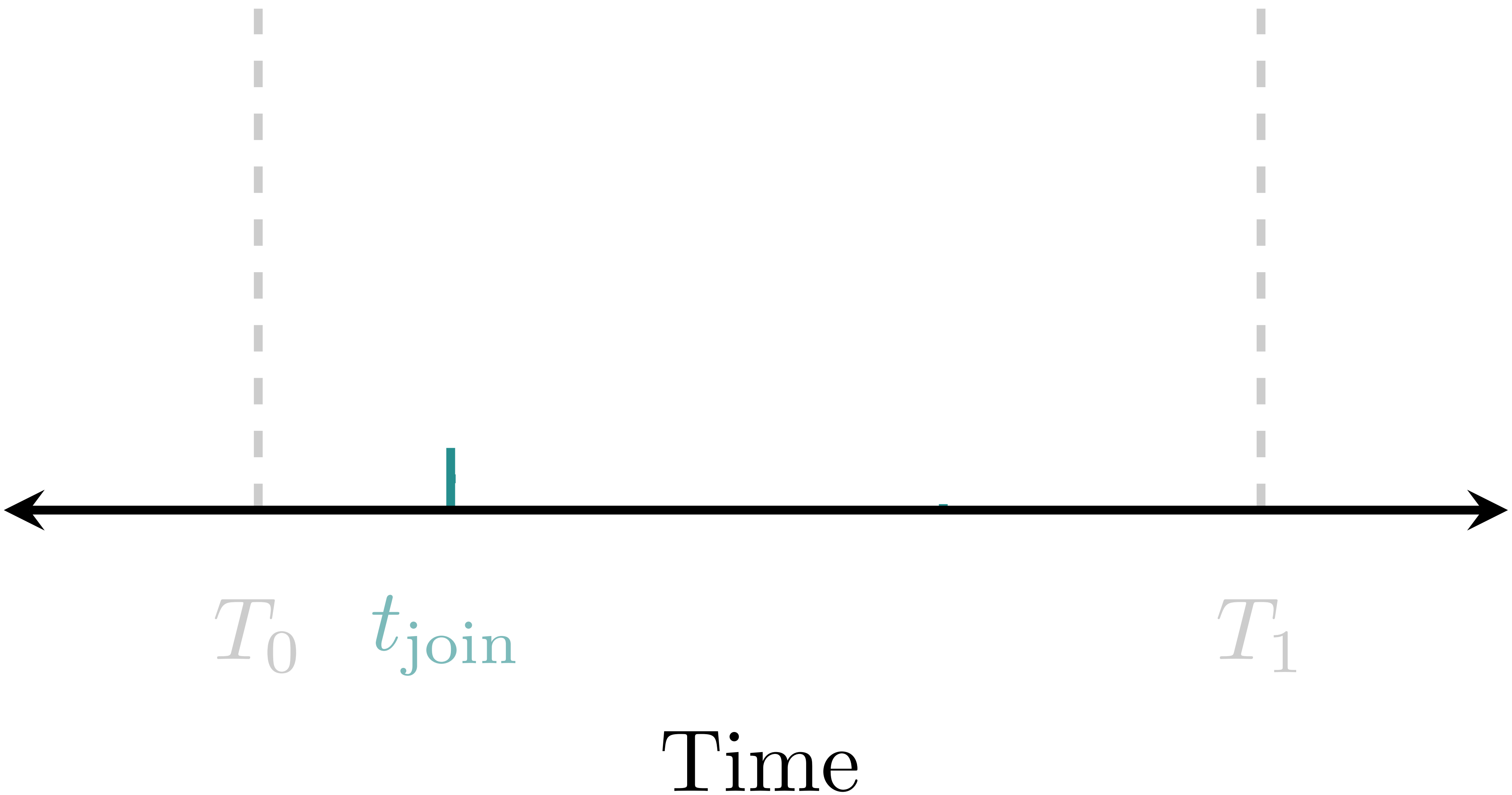
Patreon

- https://www.patreon.com/betanalpha

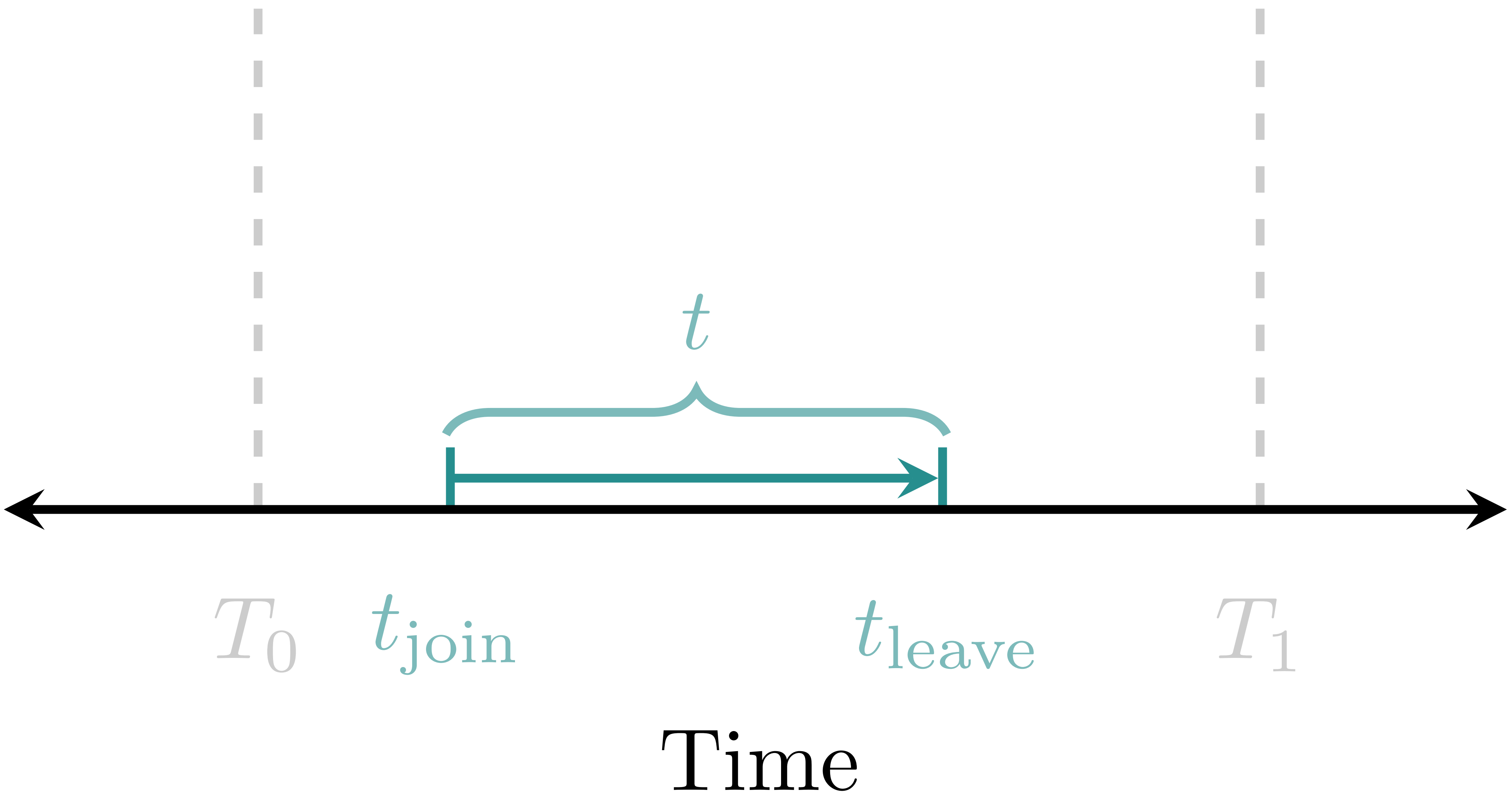
For the exercise, we'll be looking at customer lifetime data that was collected across a finite observation period.



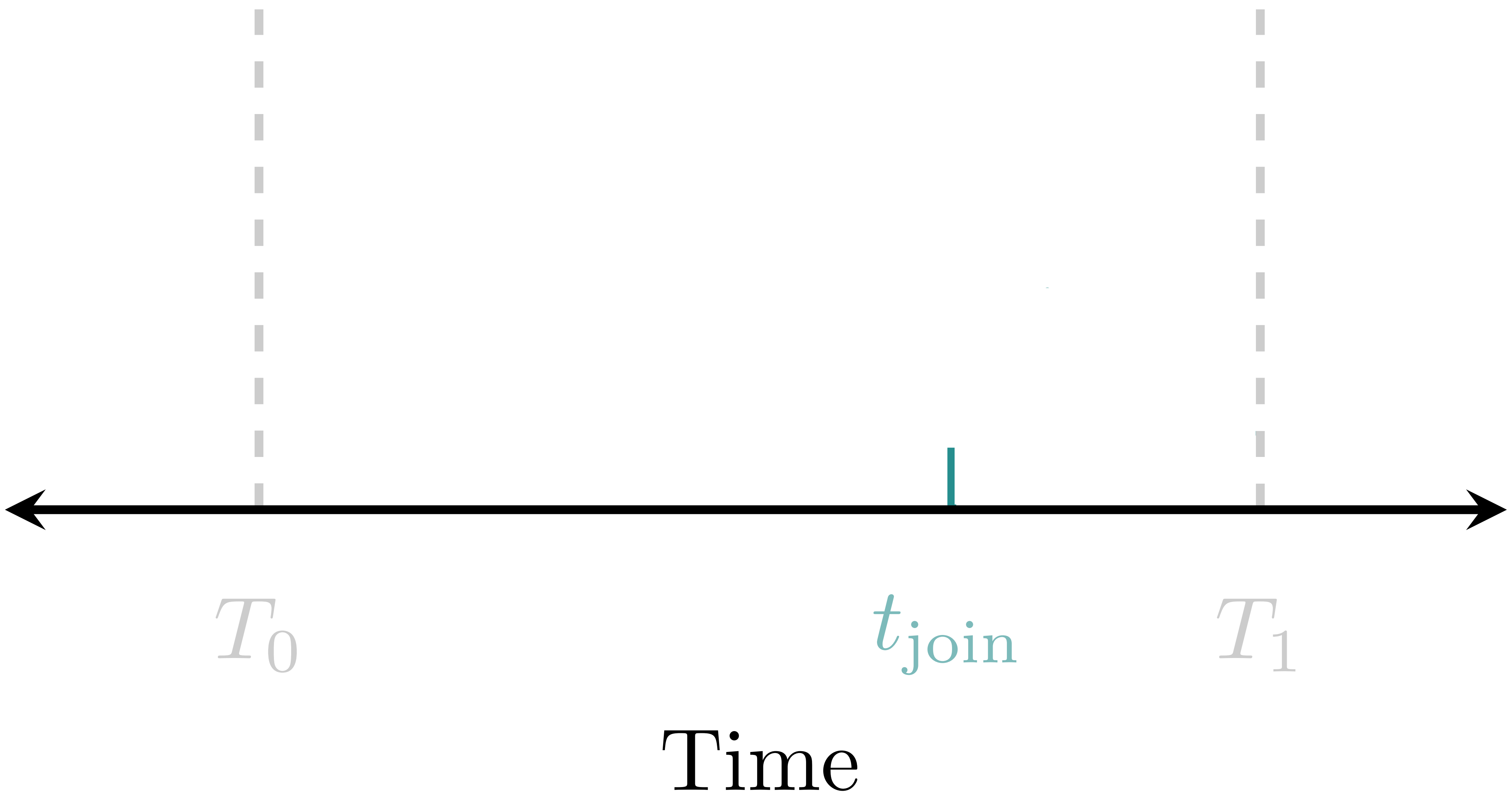
For the exercise, we'll be looking at customer lifetime data that was collected across a finite observation period.



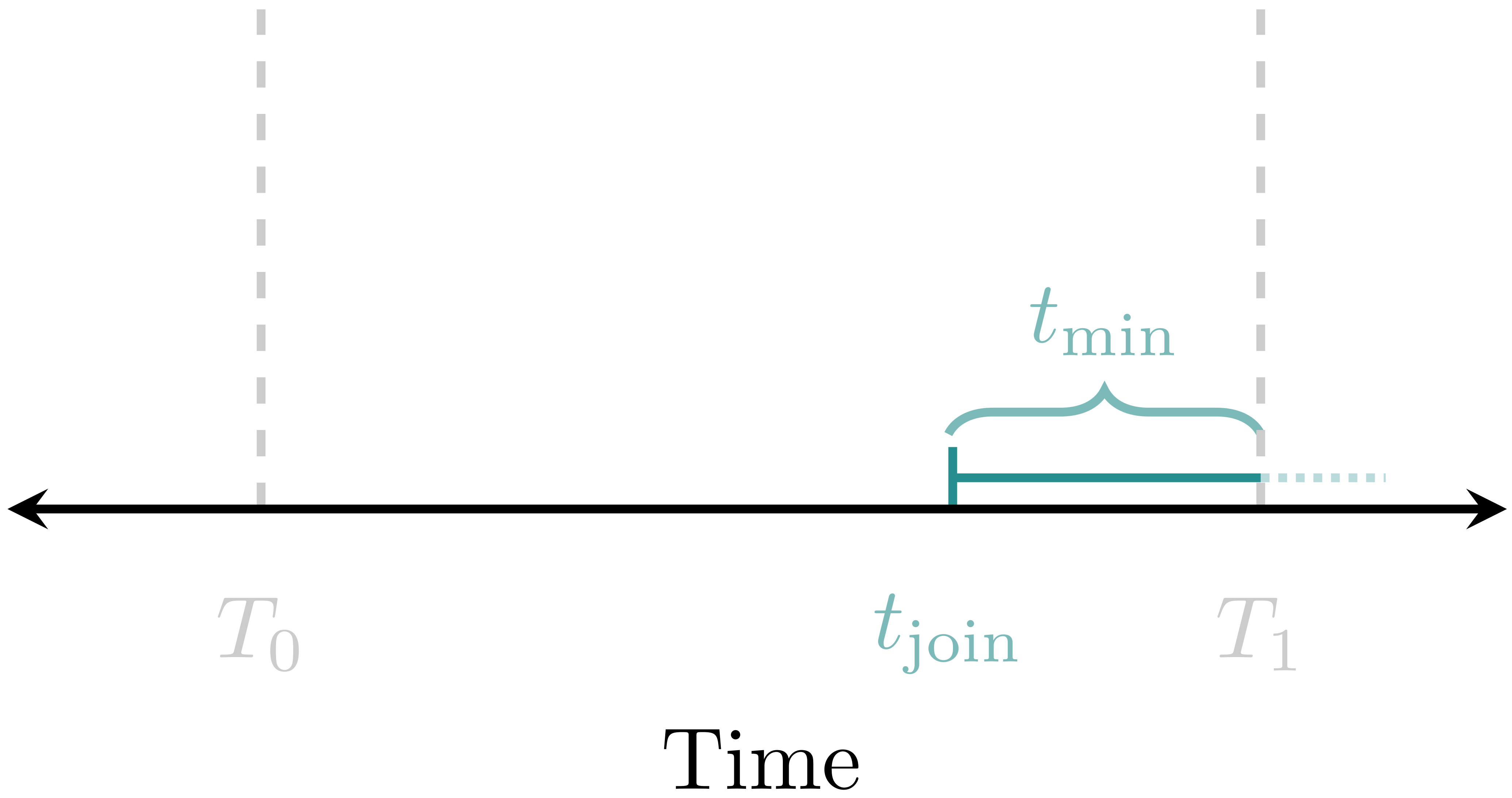
For the exercise, we'll be looking at customer lifetime data that was collected across a finite observation period.



For the exercise, we'll be looking at customer lifetime data that was collected across a finite observation period.



For the exercise, we'll be looking at customer lifetime data that was collected across a finite observation period.



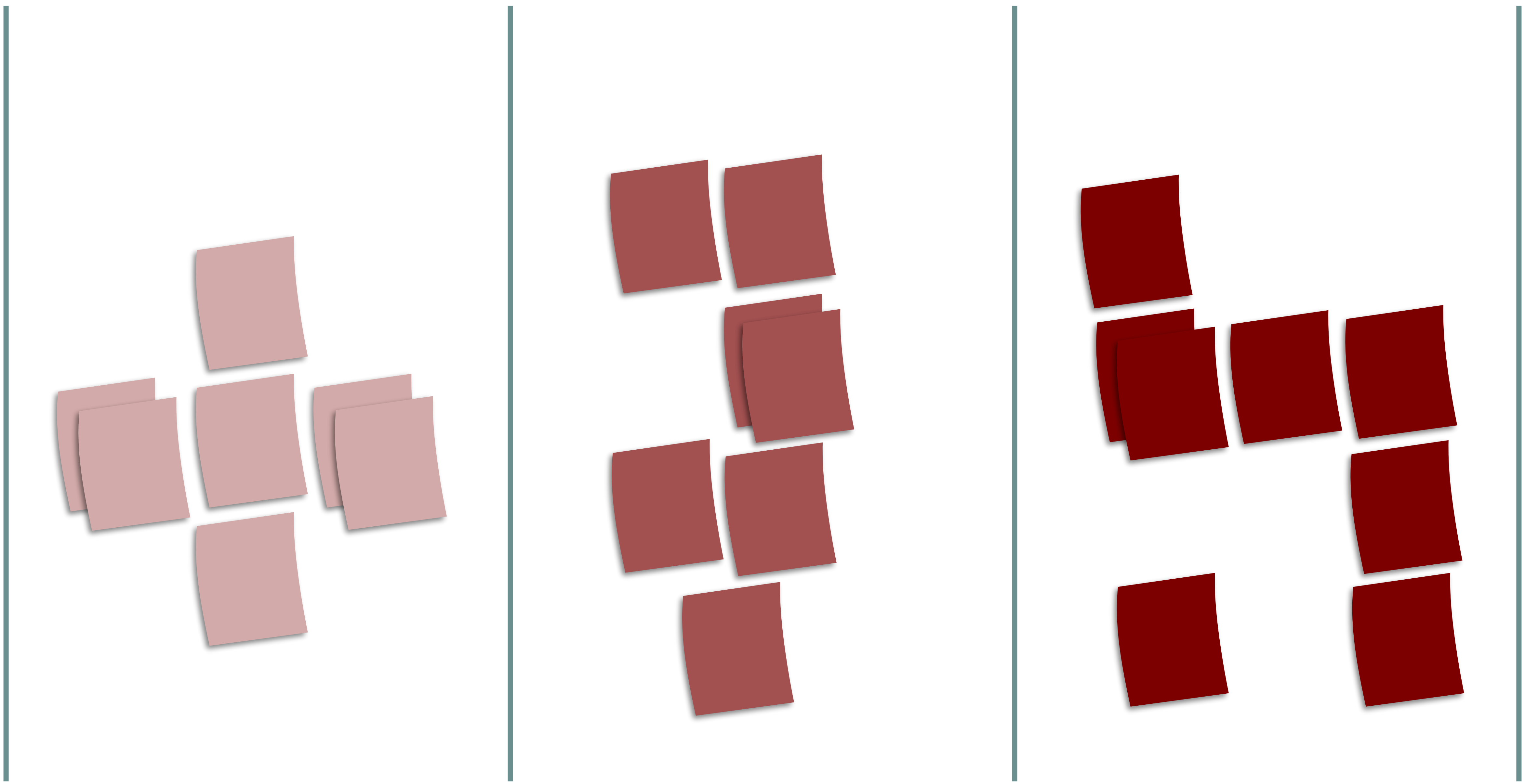
Exercises Material

part2/data/historical_data.json
part2/data/customer_segmentation.json

part2/stan_programs/model1_prior.stan
part2_stan_programs/model1.stan

part2/analysis.py

Advanced Narrative Techniques



Clear and Concise Language

Well-chosen parameterizations allow us to tell data generating stories elegantly and compactly.

$$\pi(y_1, \ldots, y_n, \ldots, y_N, \theta)$$

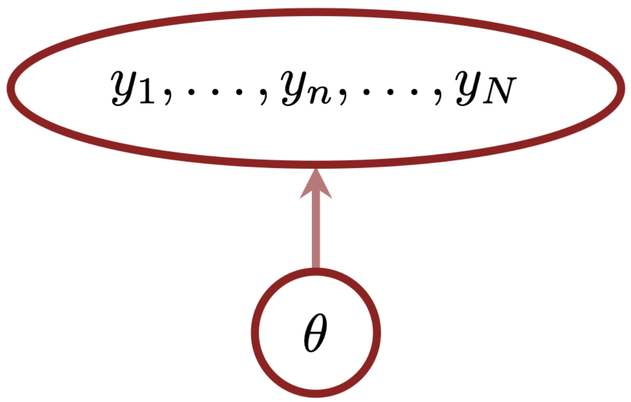
Well-chosen parameterizations allow us to tell data generating stories elegantly and compactly.

$$\pi(y_1, \ldots, y_n, \ldots, y_N, \theta) = \pi(y_1, \ldots, y_n, \ldots, y_N \mid \theta) \, \pi(\theta)$$

Well-chosen parameterizations allow us to tell data generating stories elegantly and compactly.

$$
\begin{aligned}
\pi(y_1, \ldots, y_n, \ldots, y_N, \theta) &= \pi(y_1, \ldots, y_n, \ldots, y_N \mid \theta) \, \pi(\theta) \\
&= \left[\prod_{n=1}^{N} \pi(y_n \mid \theta) \right] \pi(\theta)
\end{aligned}
$$

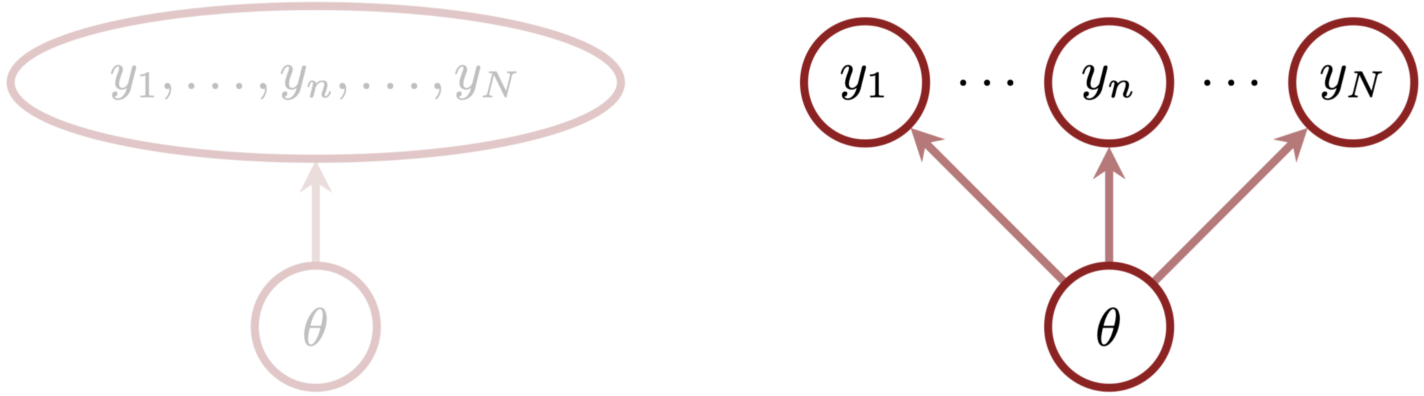
Well-chosen parameterizations allow us to tell data generating stories elegantly and compactly.



$$\pi(y_1, \ldots, y_n, \ldots, y_N, \theta) = \pi(y_1, \ldots, y_n, \ldots, y_N \mid \theta) \, \pi(\theta)$$

$$= \left[\prod_{n=1}^{N} \pi(y_n \mid \theta) \right] \pi(\theta)$$

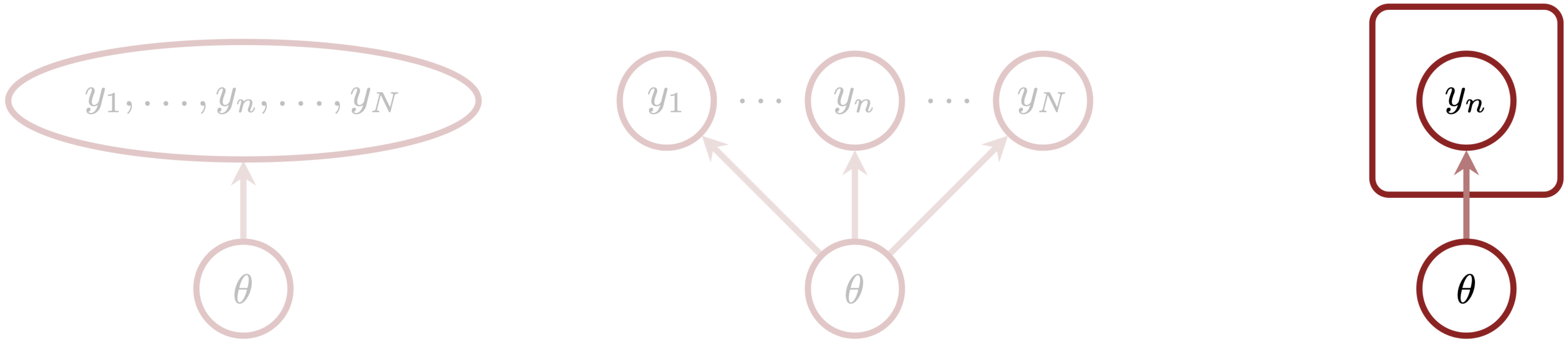
Well-chosen parameterizations allow us to tell data generating stories elegantly and compactly.



$$\pi(y_1, \ldots, y_n, \ldots, y_N, \theta) = \pi(y_1, \ldots, y_n, \ldots, y_N \mid \theta) \, \pi(\theta)$$

$$= \left[\prod_{n=1}^{N} \pi(y_n \mid \theta) \right] \pi(\theta)$$

Well-chosen parameterizations allow us to tell data generating stories elegantly and compactly.



$$\pi(y_1, \ldots, y_n, \ldots, y_N, \theta) = \pi(y_1, \ldots, y_n, \ldots, y_N \mid \theta) \, \pi(\theta)$$

$$= \left[\prod_{n=1}^{N} \pi(y_n \mid \theta) \right] \pi(\theta)$$

More incongruous parameterizations require much more complexity to tell the same stories.

$$
\begin{aligned}
\Delta_1 &= y_1 \\
\Delta_2 &= y_2 - y_1 \\
&\ldots \\
\Delta_n &= y_n - y_{n-1} \\
&\ldots \\
\Delta_N &= y_N - y_{N-1}
\end{aligned}
$$

More incongruous parameterizations require much more complexity to tell the same stories.

$$
\begin{aligned}
\Delta_1 &= y_1 \\
\Delta_2 &= y_2 - y_1 \\
&\ldots \\
\Delta_n &= y_n - y_{n-1} \\
&\ldots \\
\Delta_N &= y_N - y_{N-1}
\end{aligned}
$$

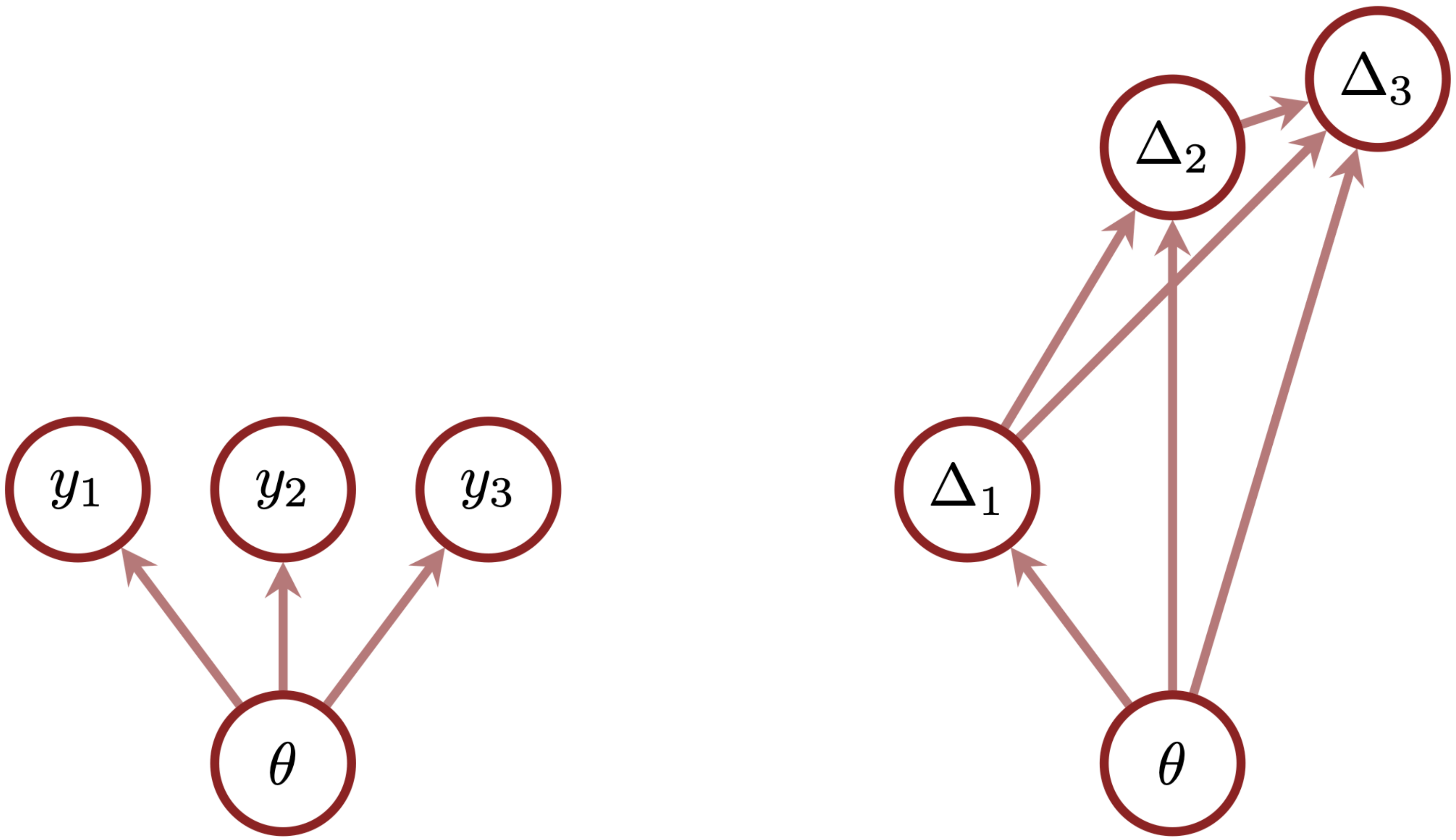
$$
\pi(\Delta_1, \ldots, \Delta_n, \ldots, \Delta_N, \theta) = \left[\prod_{n=2}^{N} \pi(\Delta_n \mid \Delta_1, \ldots, \Delta_{n-1}, \theta) \right] \pi(\Delta_1 \mid \theta) \, \pi(\theta)
$$

More incongruous parameterizations require much more complexity to tell the same stories.

$$\Delta_1 = y_1$$
$$\Delta_2 = y_2 - y_1$$
$$\ldots$$
$$\Delta_n = y_n - y_{n=1}$$
$$\ldots$$
$$\Delta_N = y_N - y_{N-1}$$

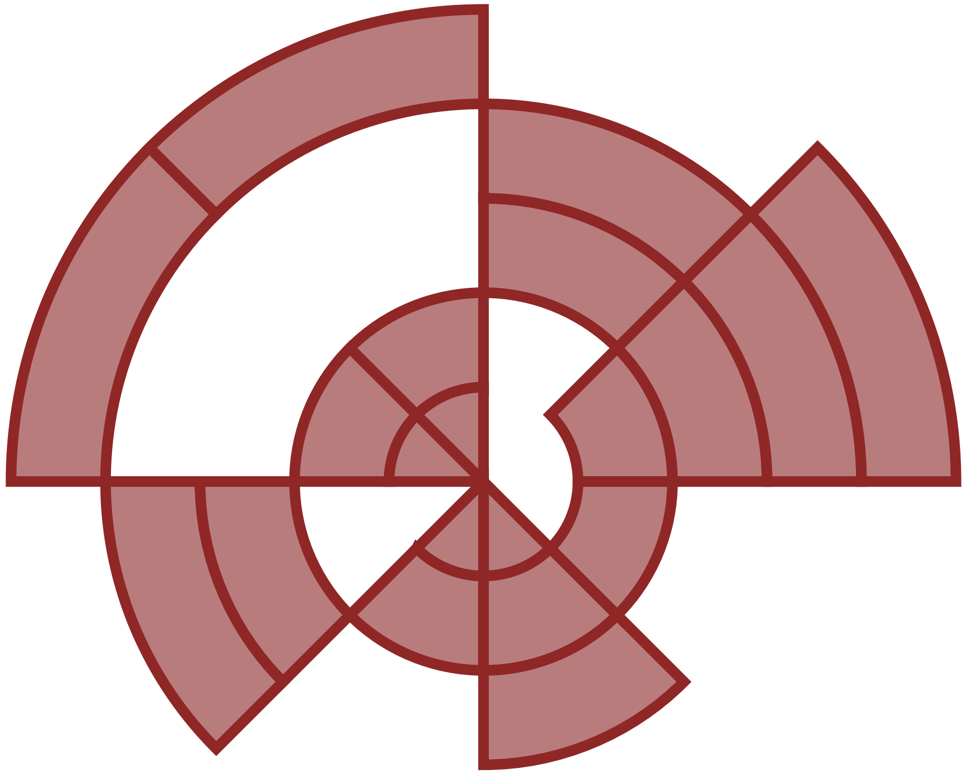
$$\pi(\Delta_1, \ldots, \Delta_n, \ldots, \Delta_N, \theta) = \left[\prod_{n=2}^{N} \pi(\Delta_n \mid \Delta_1, \ldots, \Delta_{n-1}, \theta) \right] \pi(\Delta_1 \mid \theta) \, \pi(\theta)$$

More incongruous parameterizations require much more complexity to tell the same stories.

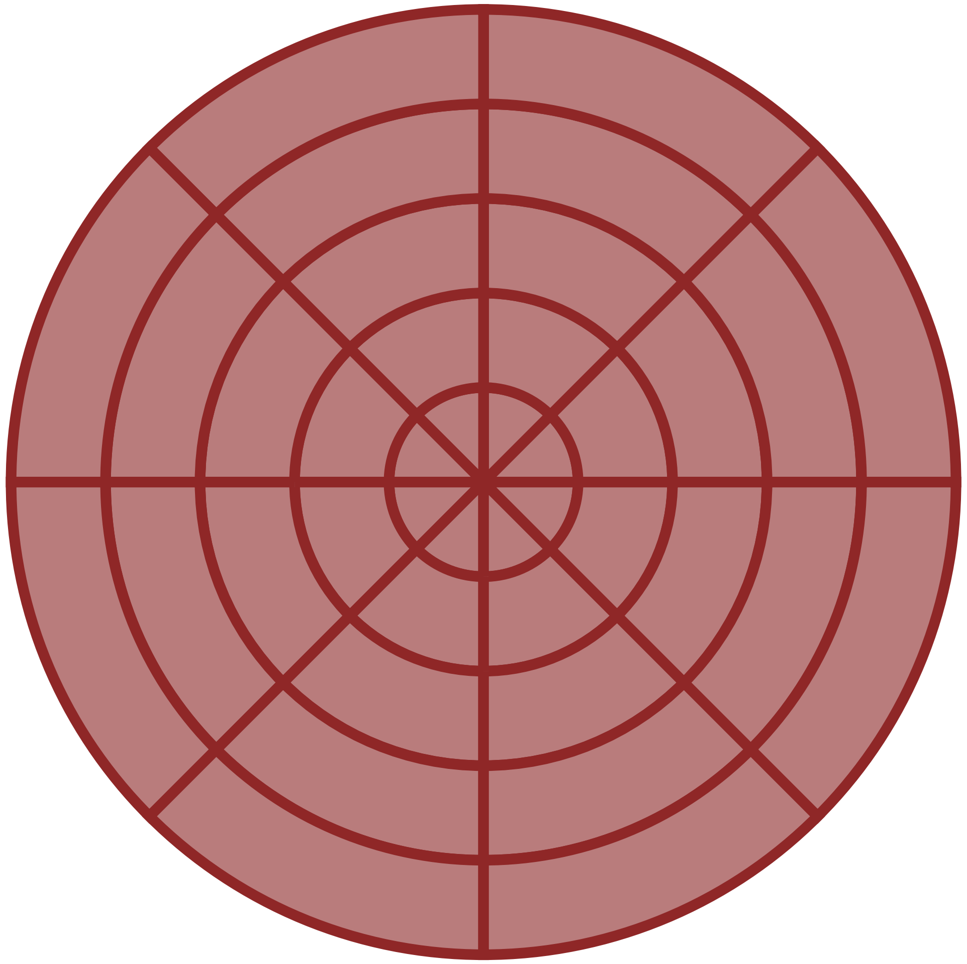


Similarly we might develop a full Bayesian model one constituent model one at a time.

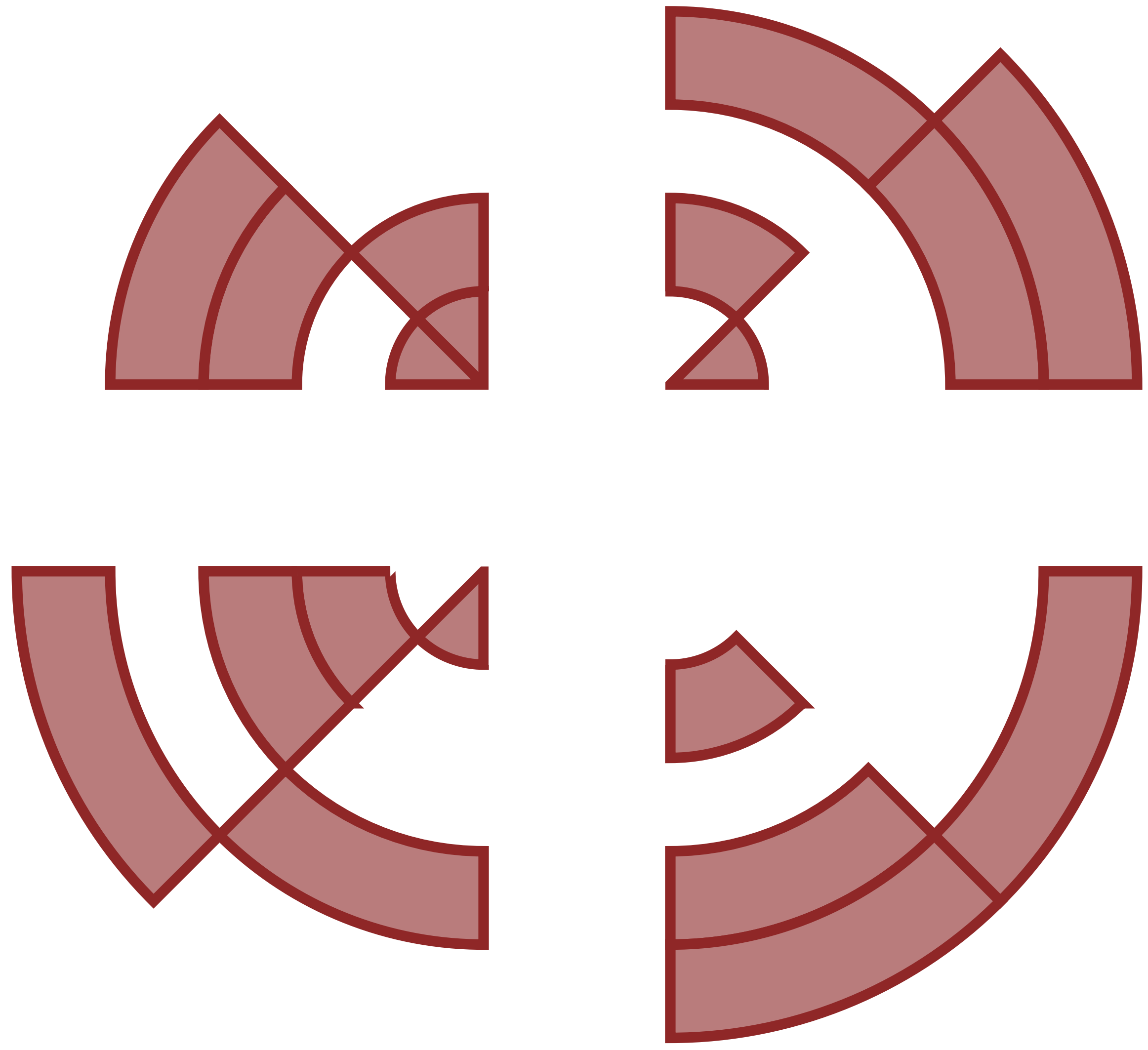
Similarly we might develop a full Bayesian model one constituent model one at a time.



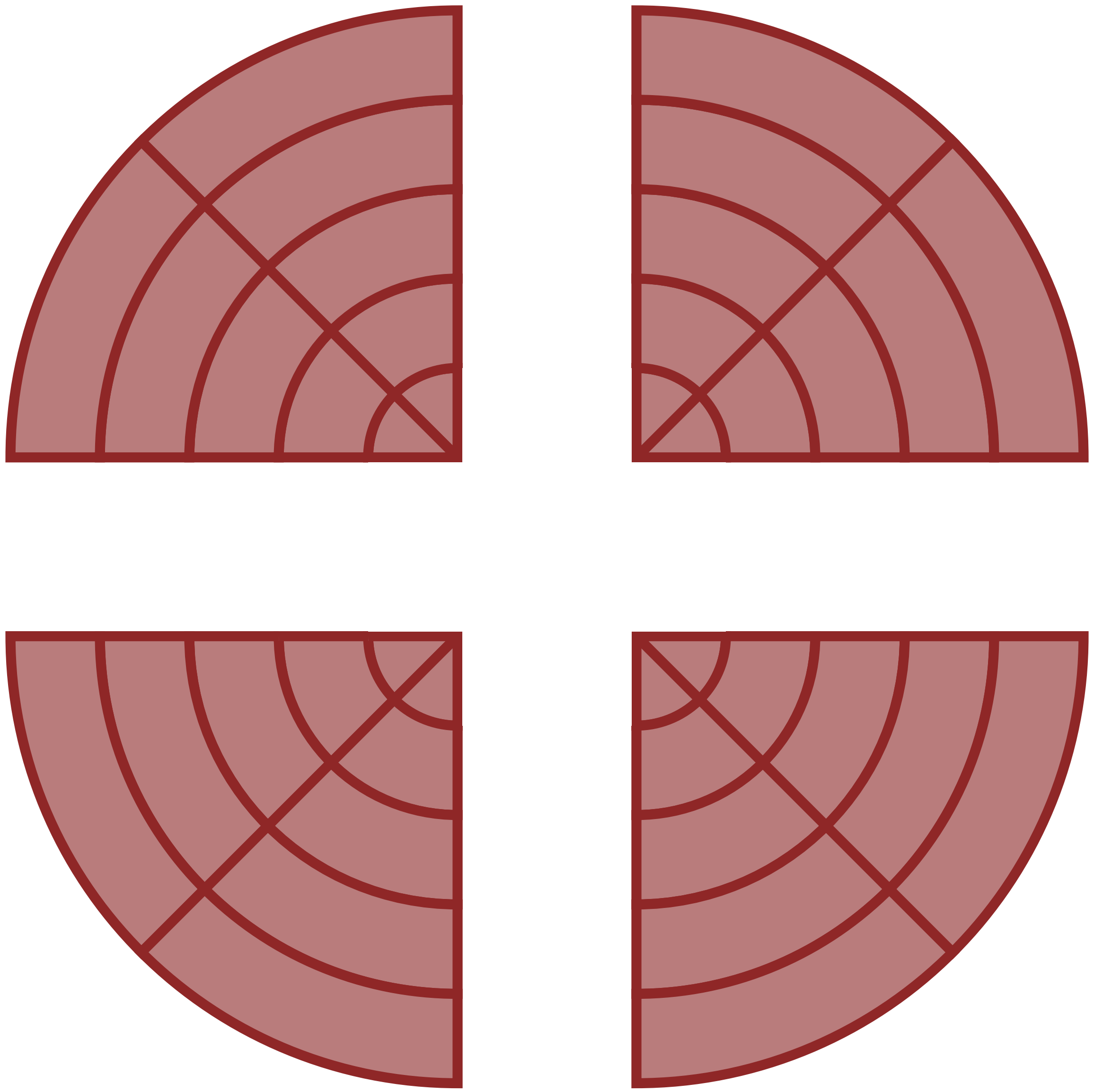
Similarly we might develop a full Bayesian model one constituent model one at a time.



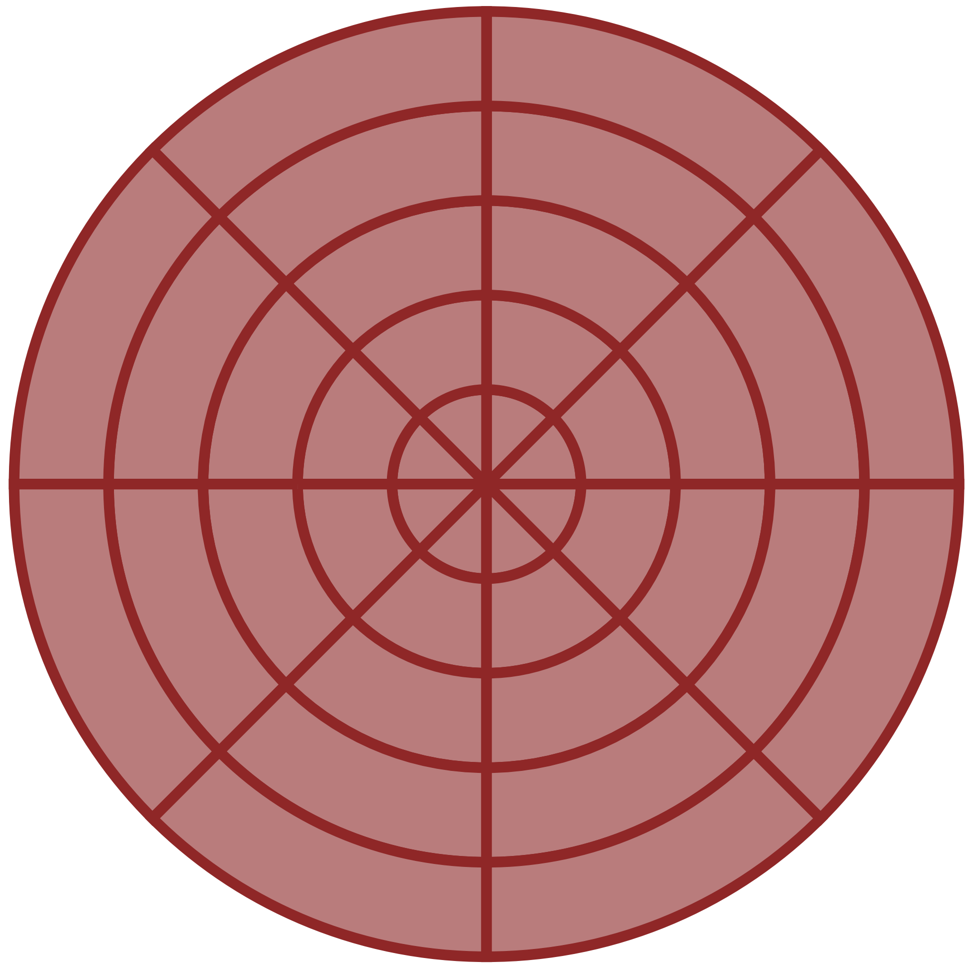
Or we might integrate the constituents models into intermediate models that we then piece together.



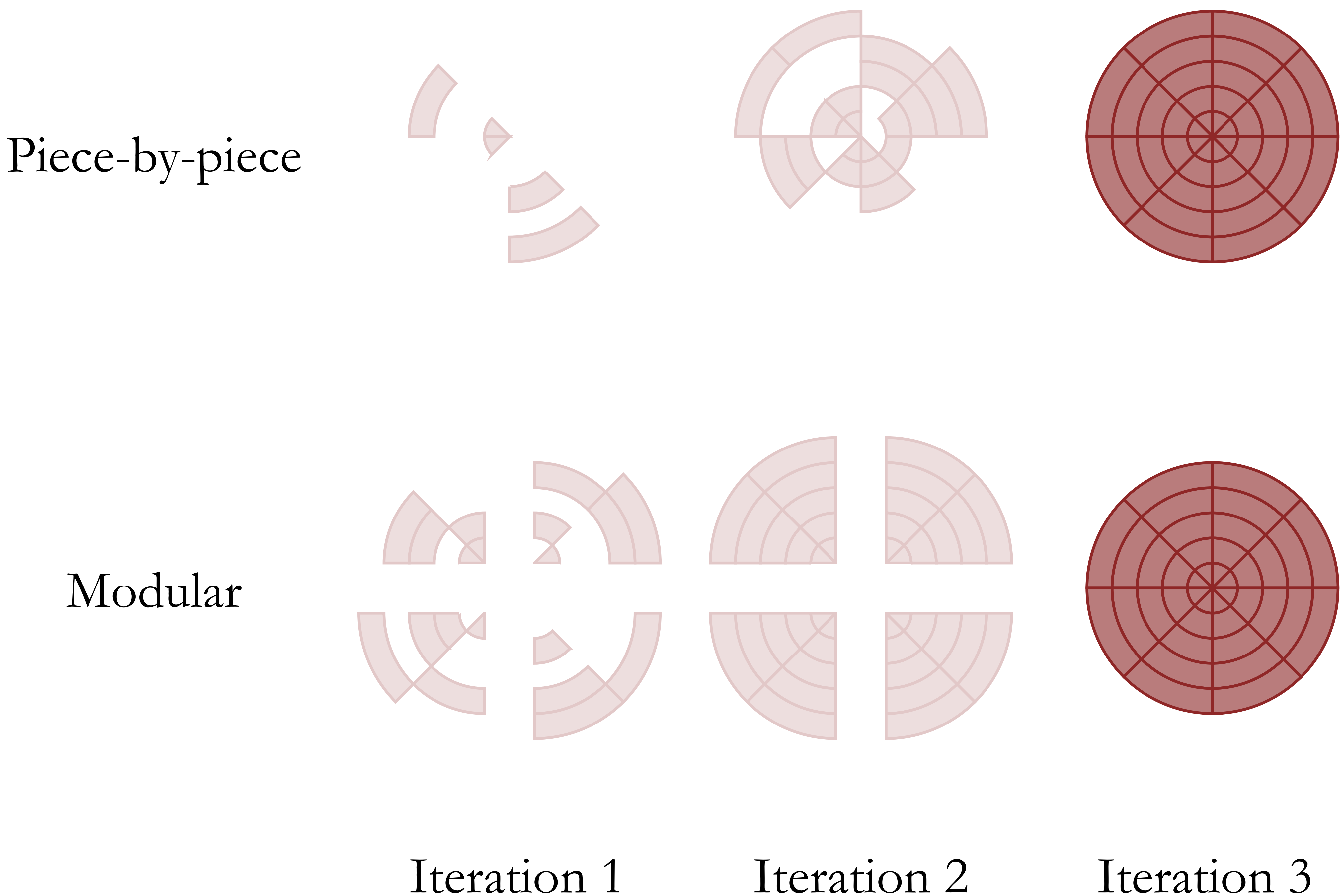
Or we might integrate the constituents models into intermediate models that we then piece together.



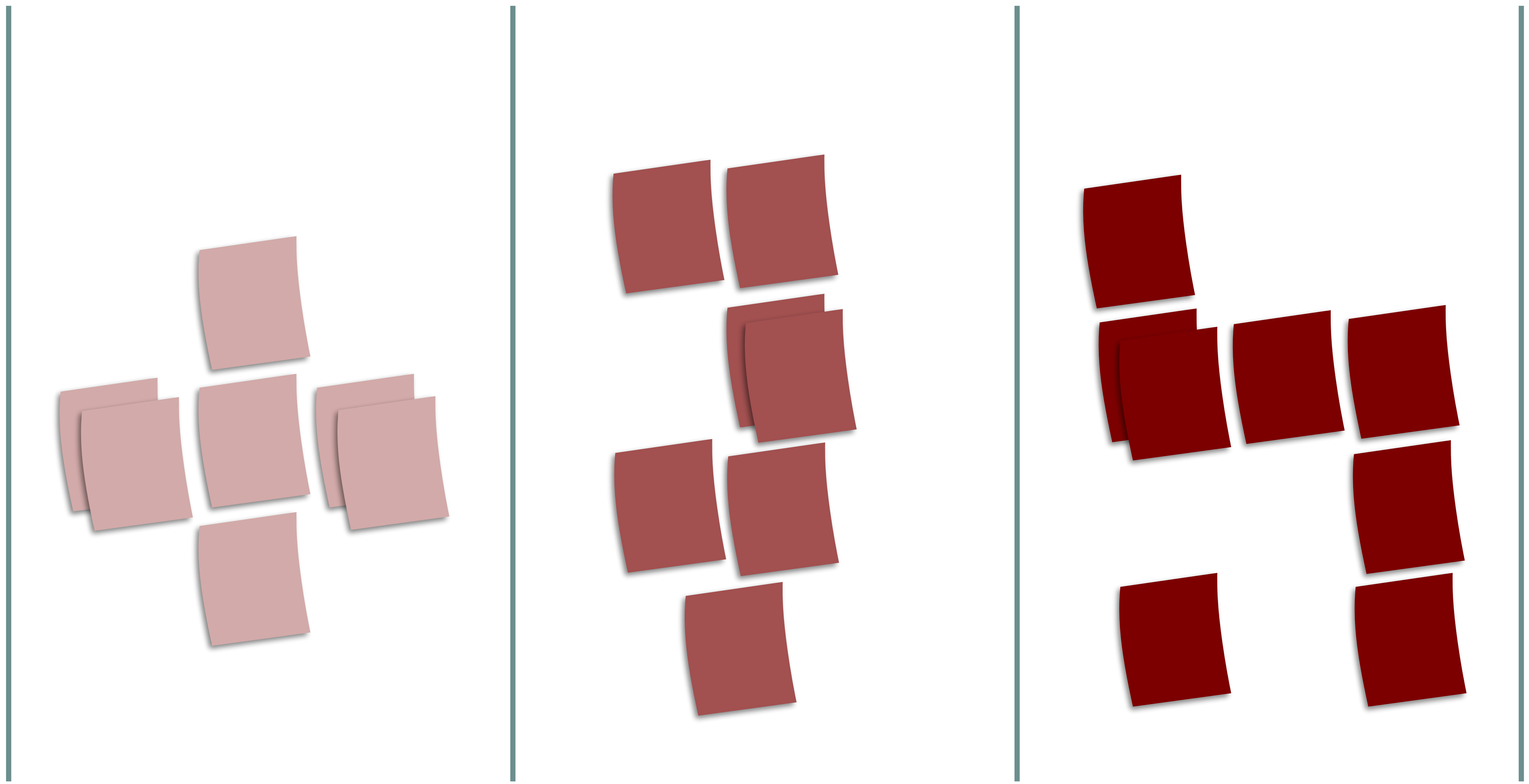
Or we might integrate the constituents models into intermediate models that we then piece together.



While both approaches can lead to the same final model the progression can be easier one way over the other.

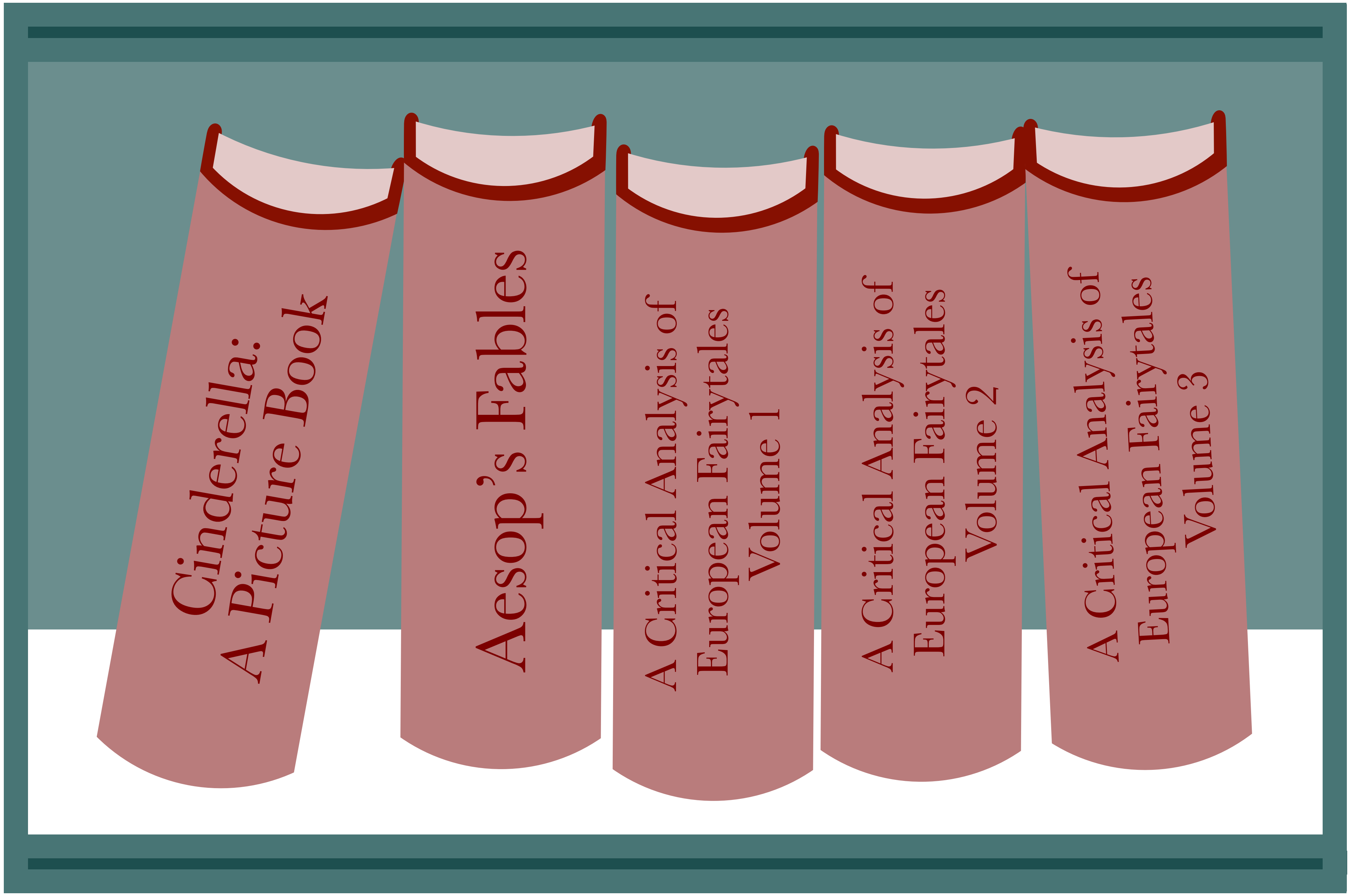


Advanced Narrative Techniques

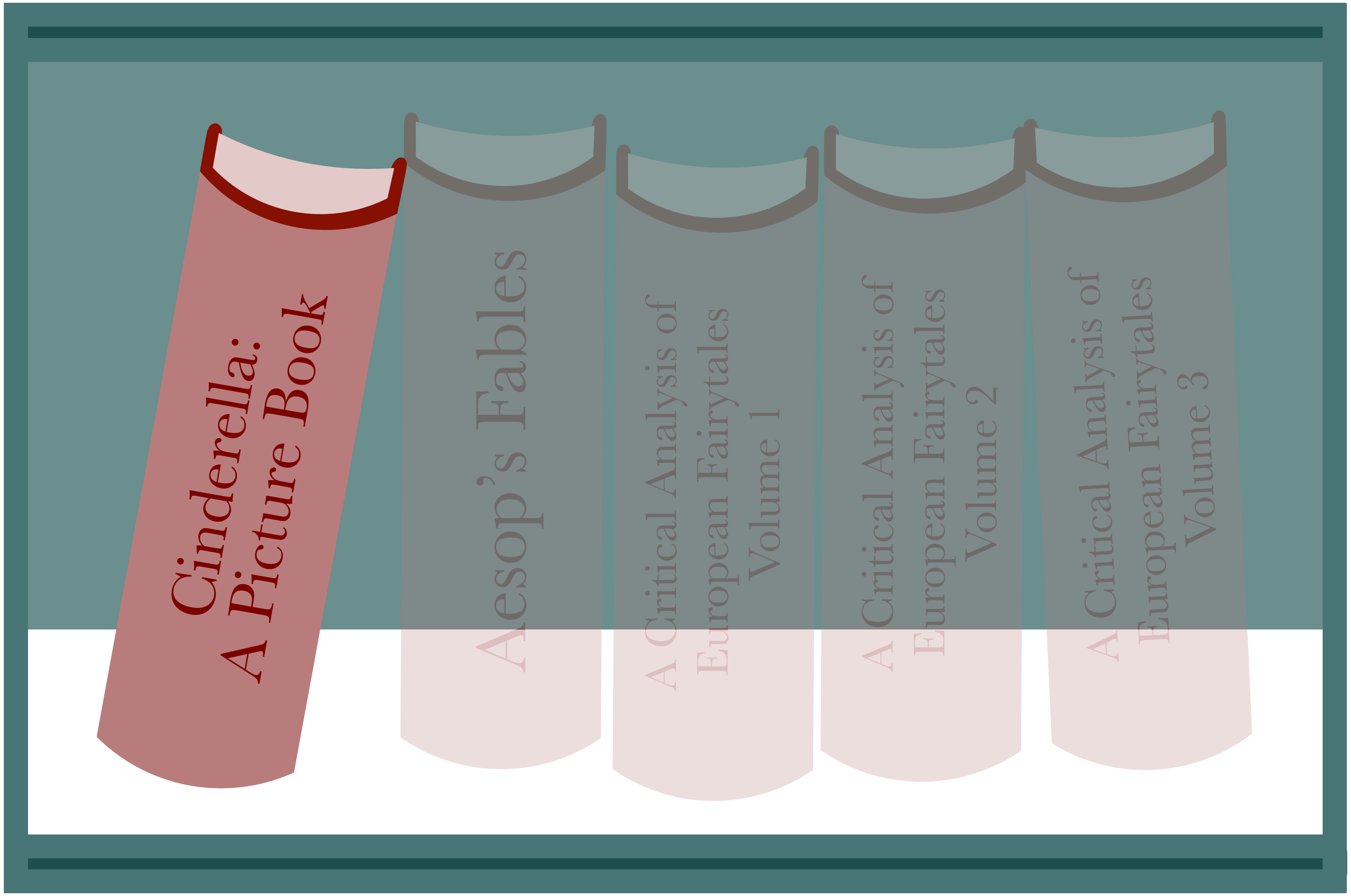


Writing To The Audience

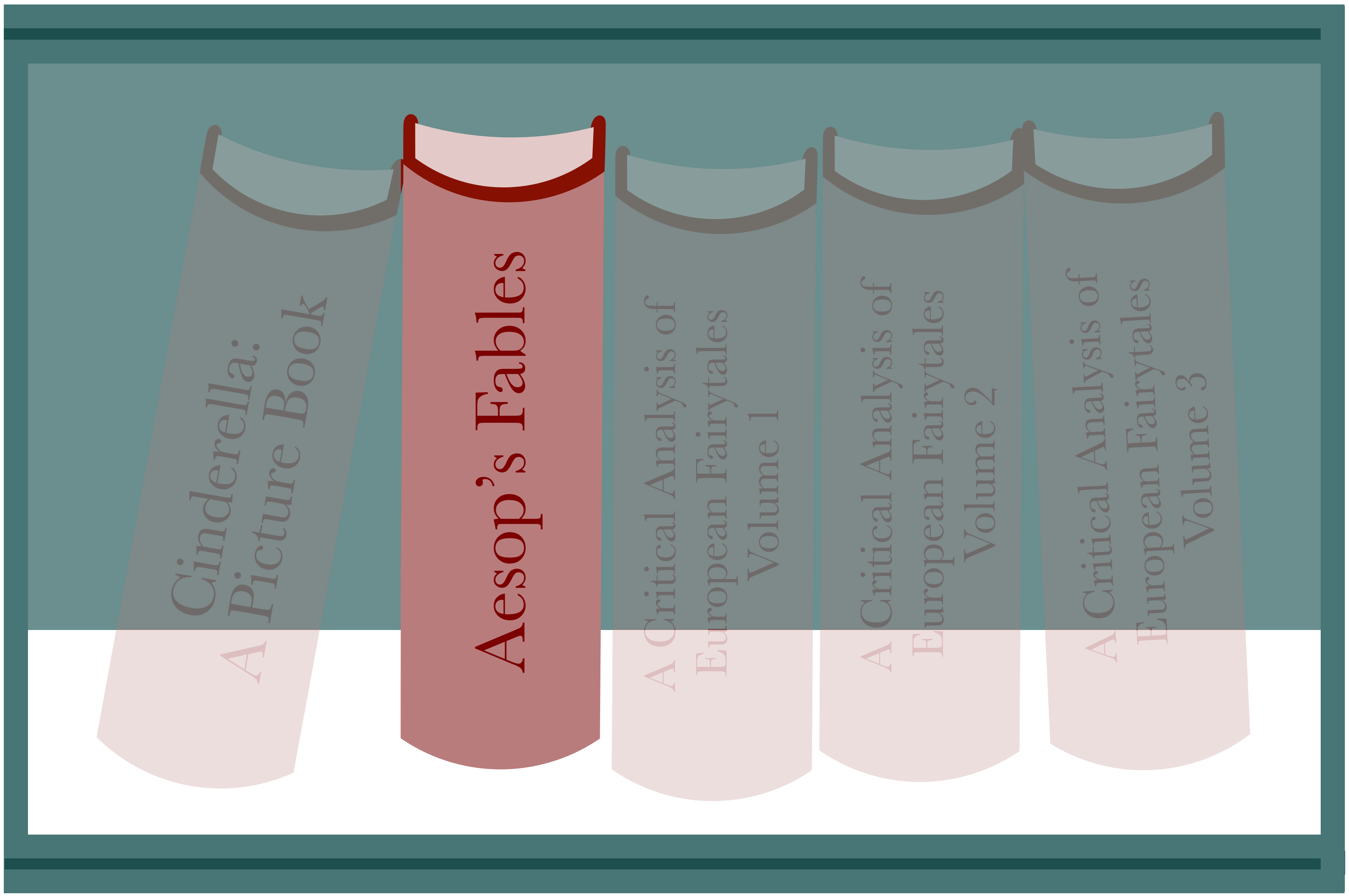
Stories can be told at different levels of sophistication, from simple summaries to richly detailed expositions.



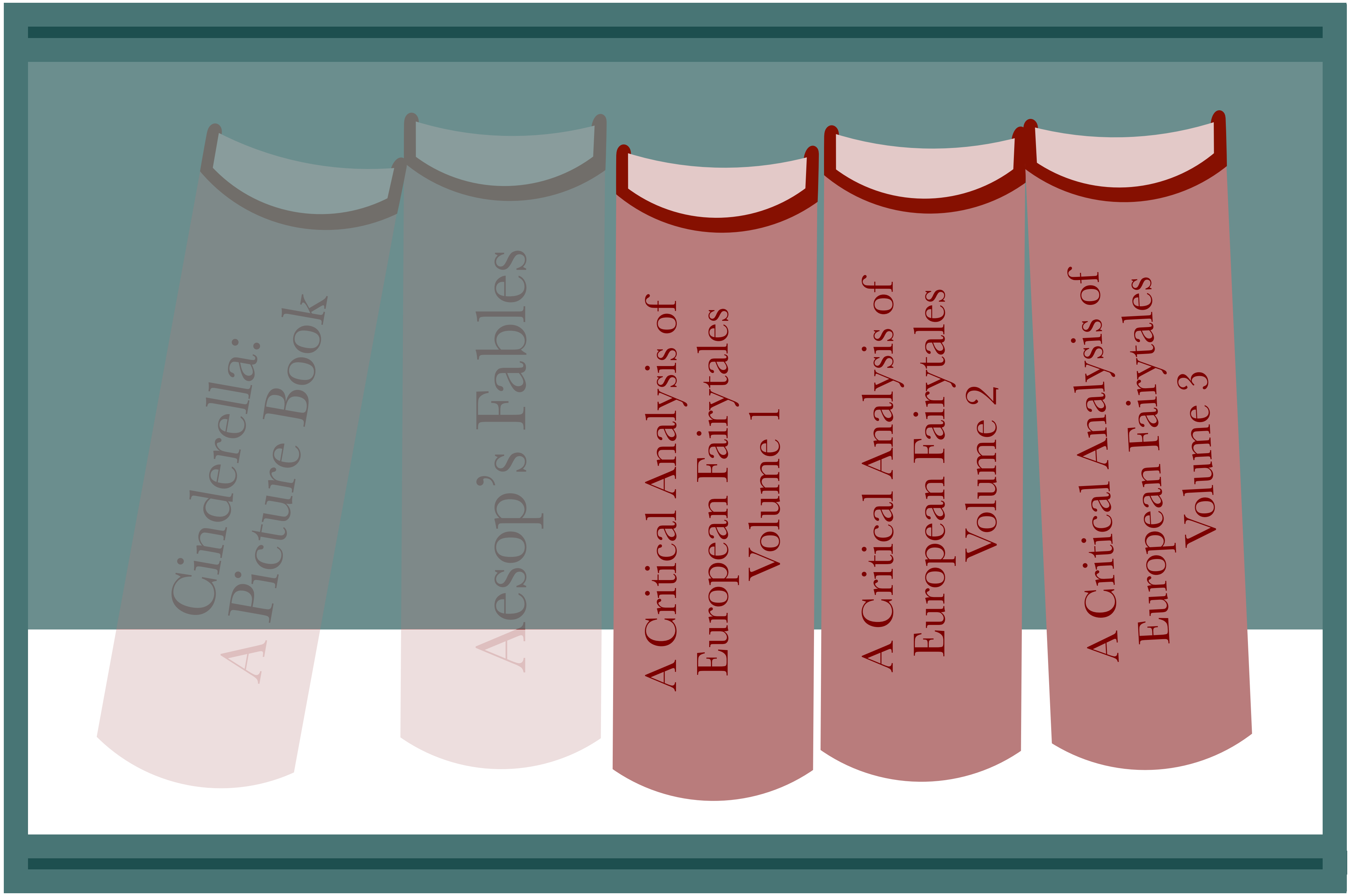
Stories can be told at different levels of sophistication, from simple summaries to richly detailed expositions.



Stories can be told at different levels of sophistication, from simple summaries to richly detailed expositions.



Stories can be told at different levels of sophistication, from simple summaries to richly detailed expositions.

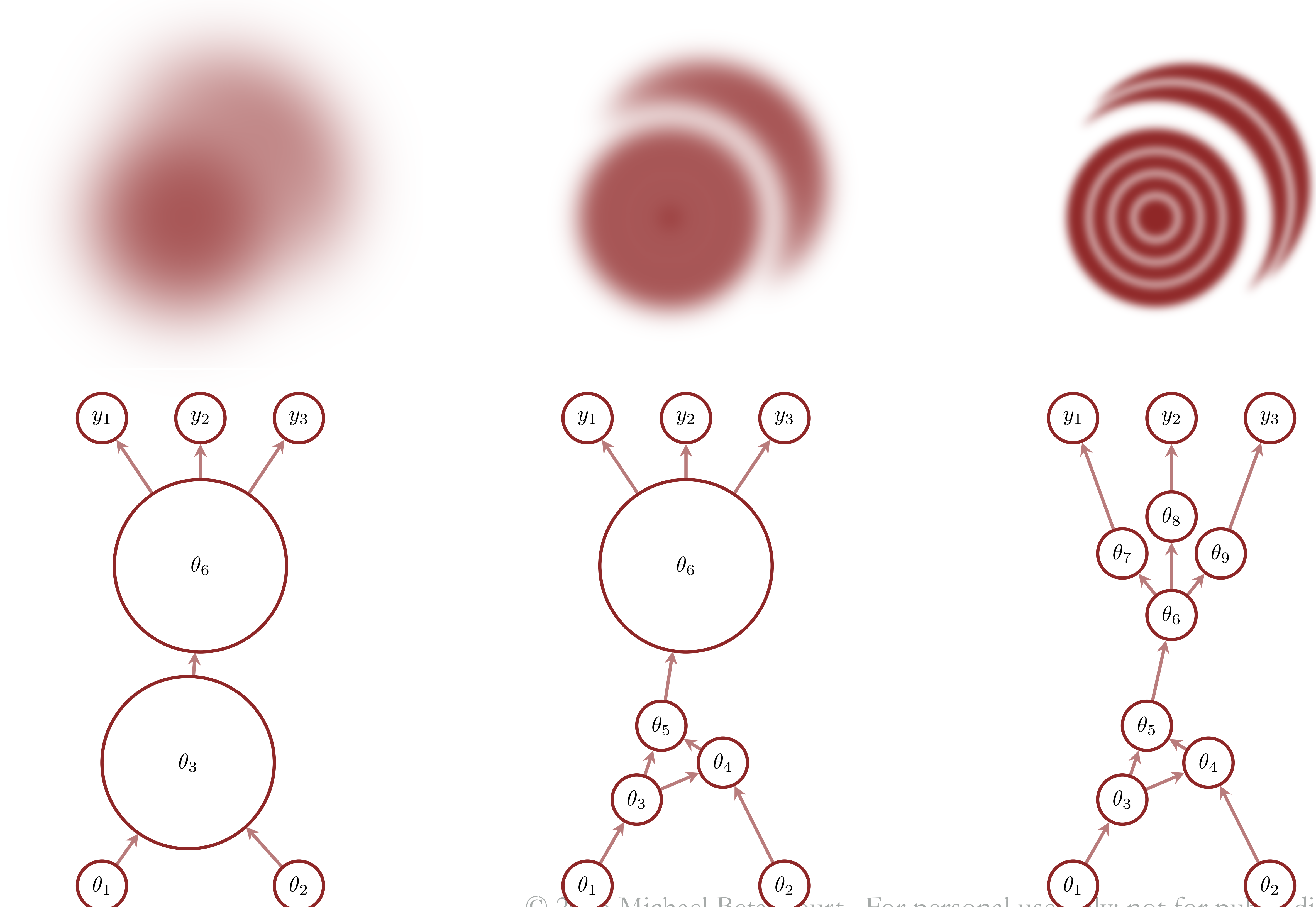


The better we can resolve a system the more sophisticated of a narratively generative model we need to analyze it.

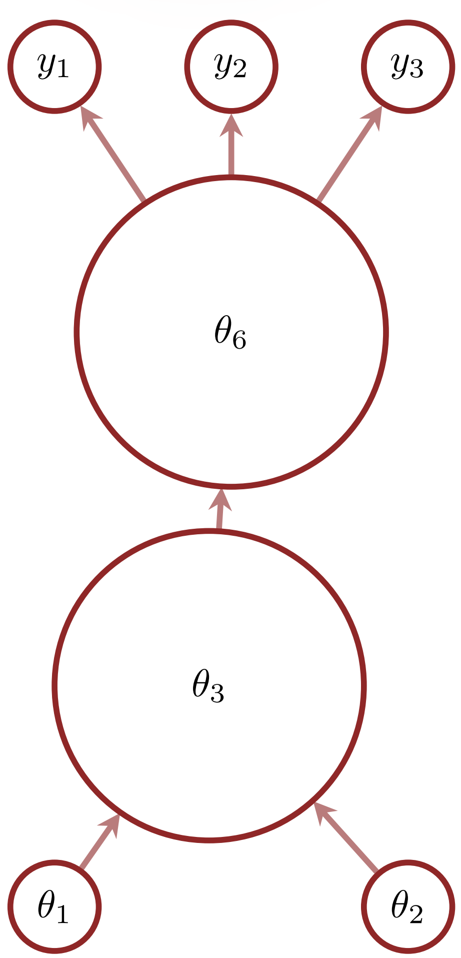
The better we can resolve a system the more sophisticated of a narratively generative model we need to analyze it.

The better we can resolve a system the more sophisticated of a narratively generative model we need to analyze it.

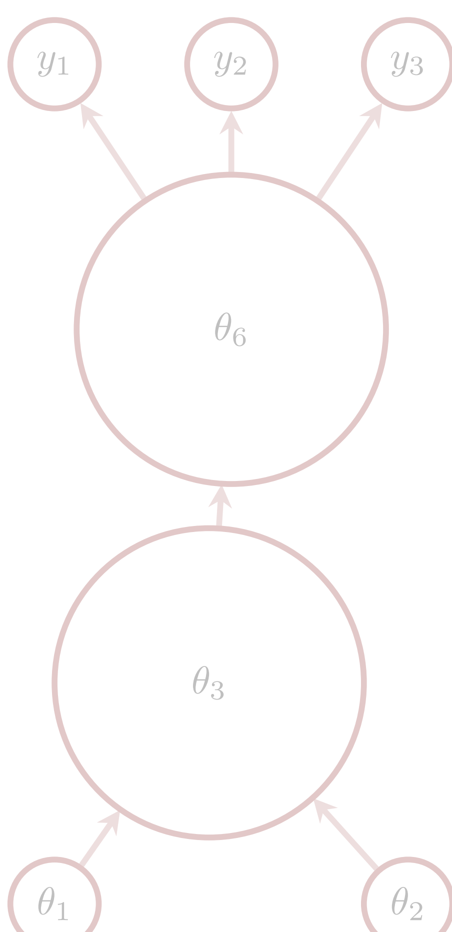
An ideal model captures the *emergent* features of the system that we can resolve in a given observation.

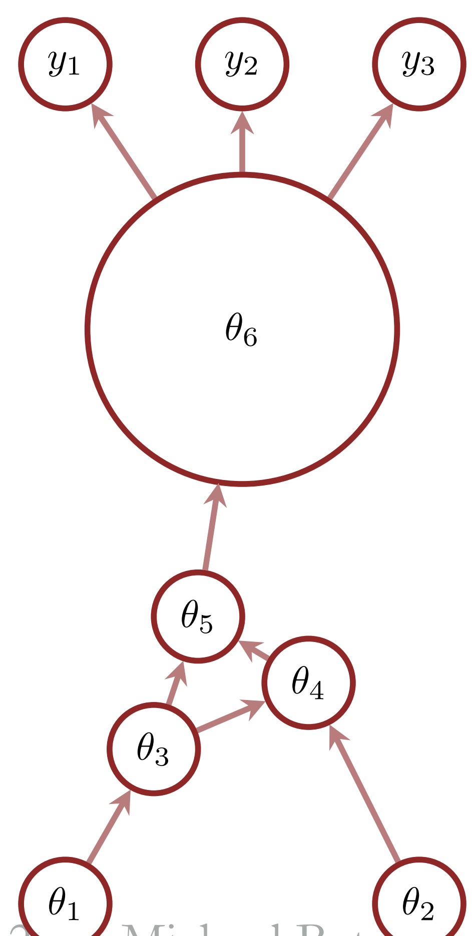


Top-down model building is particularly well-suited to the development of increasingly sophisticated models.

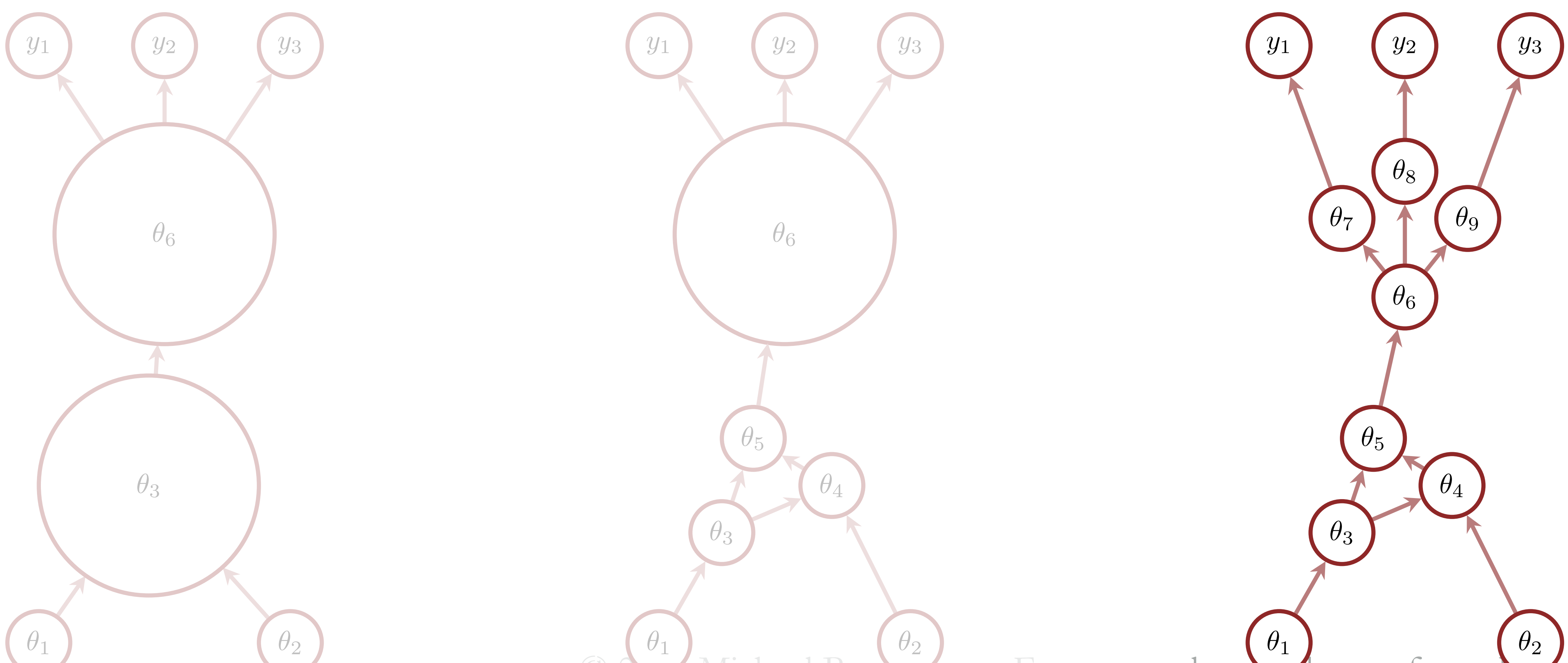


Top-down model building is particularly well-suited to the development of increasingly sophisticated models.

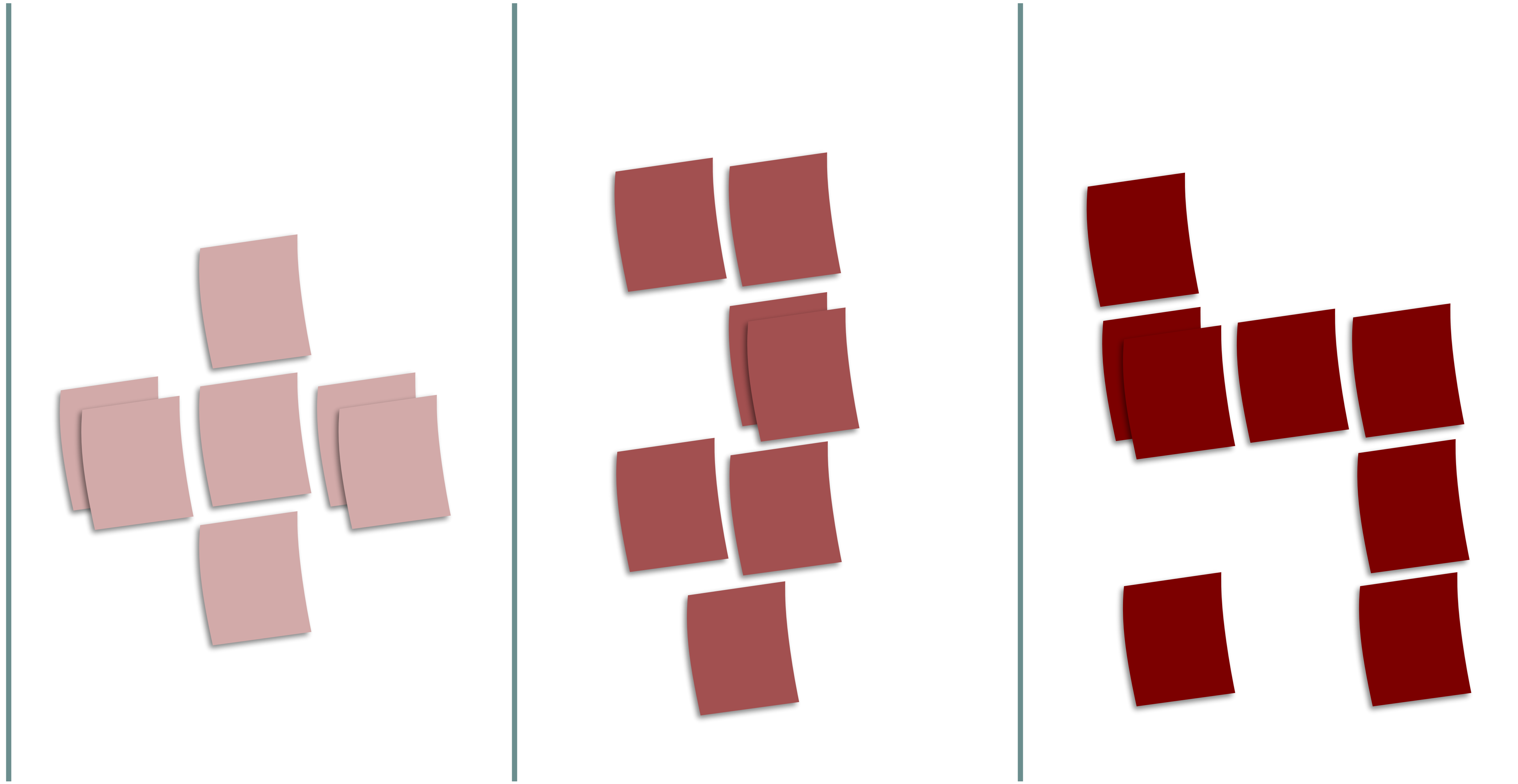




Top-down model building is particularly well-suited to the development of increasingly sophisticated models.

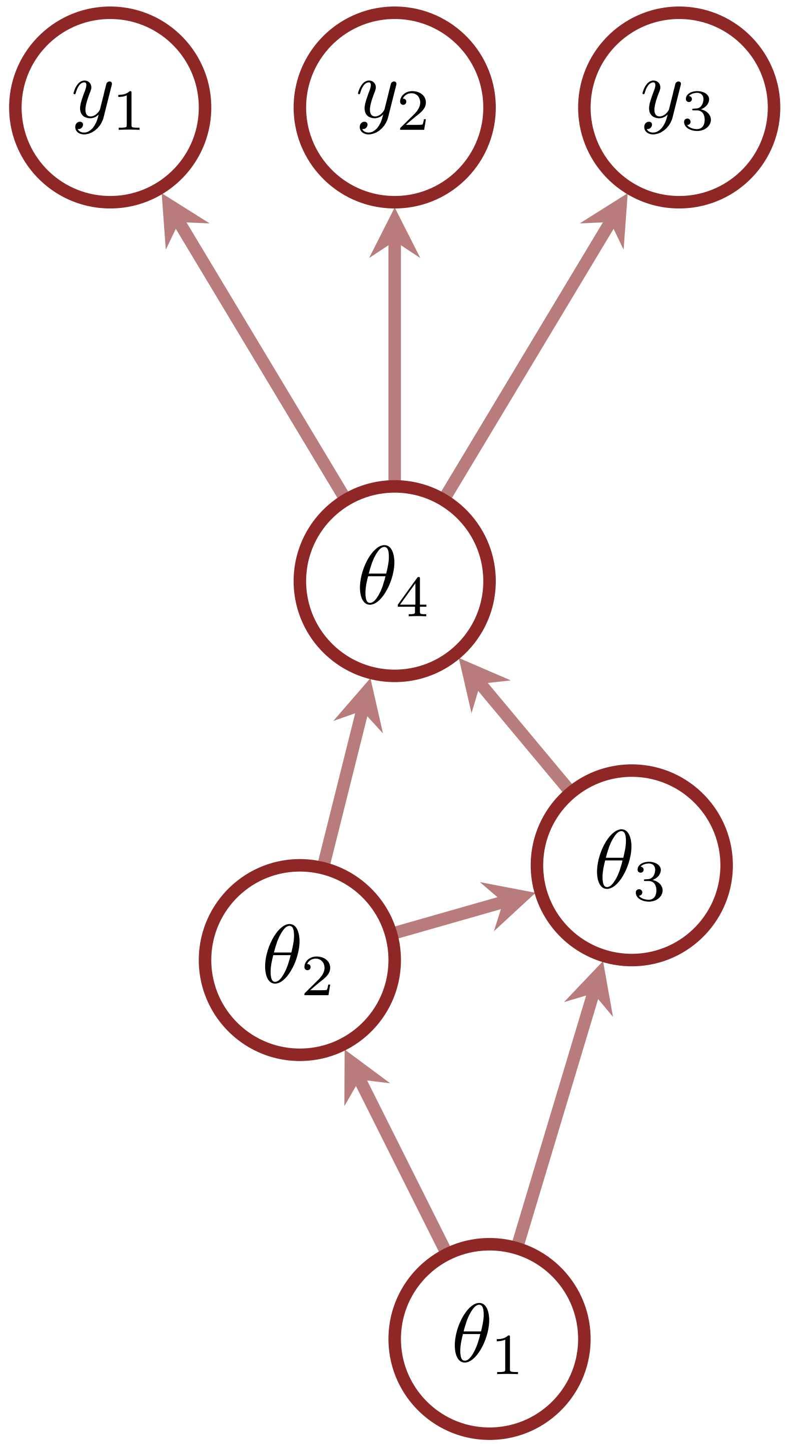


Advanced Narrative Techniques

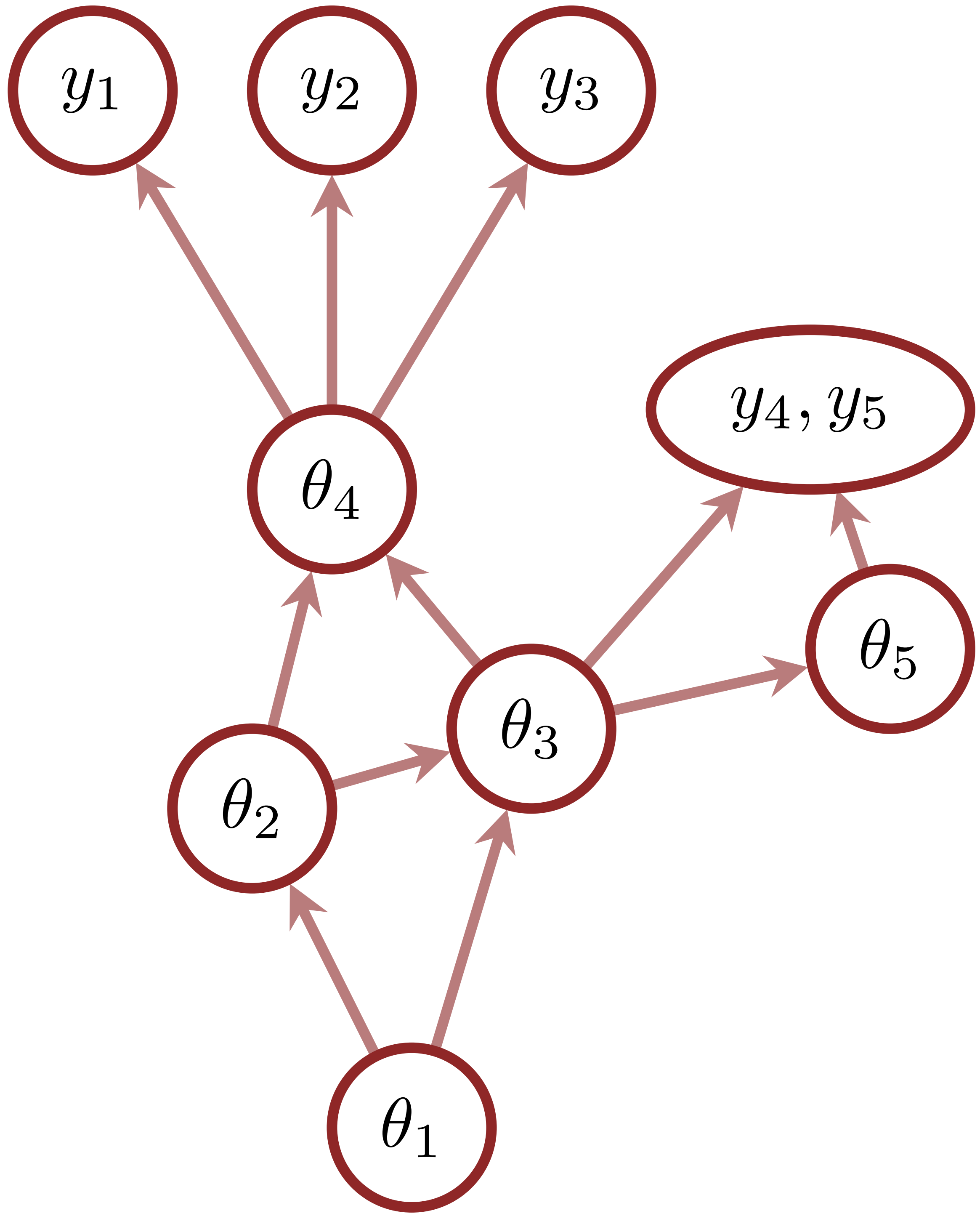


A Tangled Web

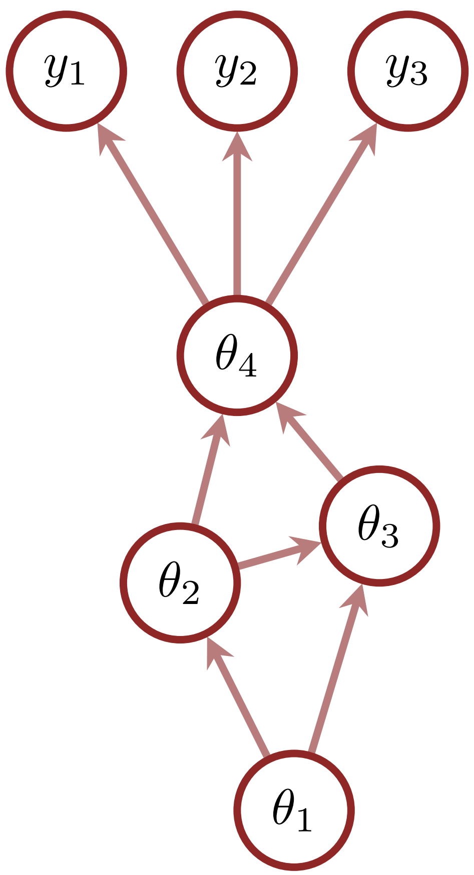
These amalgamated narratively generative models can be developed by *branching* one process off of another.

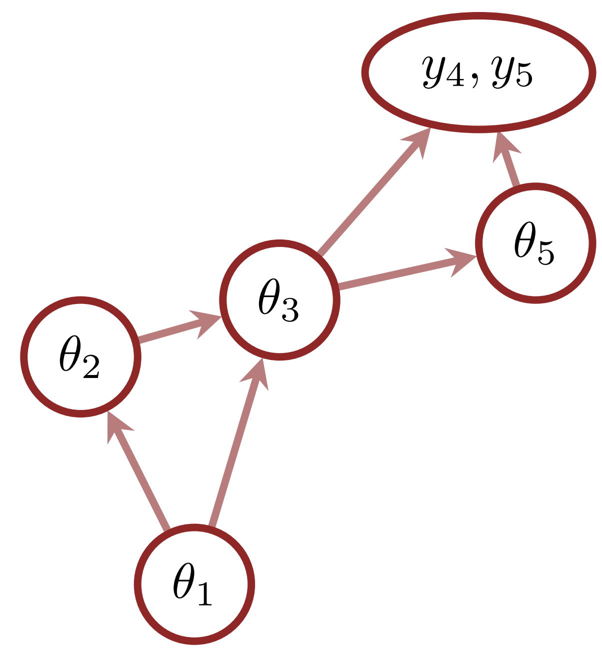


These amalgamated narratively generative models can be developed by *branching* one process off of another.

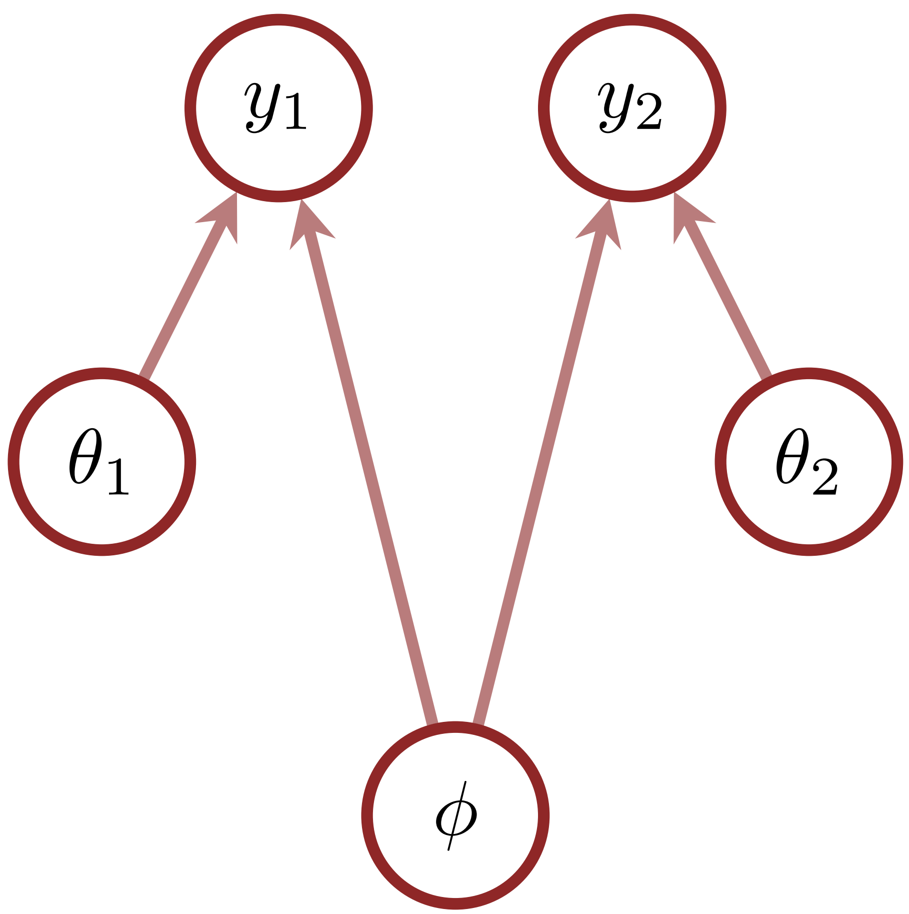


Alternatively we can build full models for each and them *merge* them together by connecting related components.



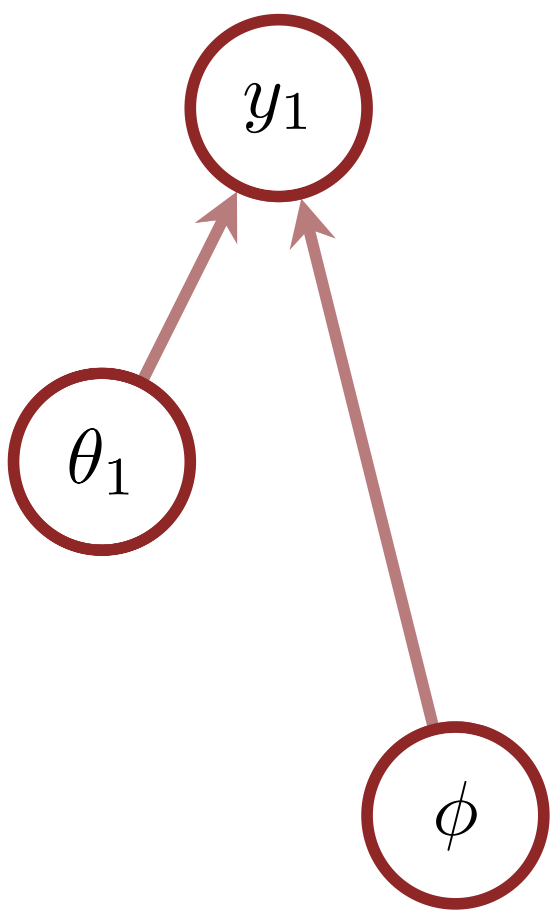


These joints models allow us to fuse inferences from different observations together in a consistent way.



$$
\begin{aligned}
\pi(y_1, y_2, \theta_1, \theta_2, \phi) = & \ \pi(y_1 \mid \theta_1, \phi) \\
& \ \pi(y_2 \mid \theta_2, \phi) \\
& \ \pi(\theta_1) \, \pi(\theta_2) \, \pi(\phi)
\end{aligned}
$$

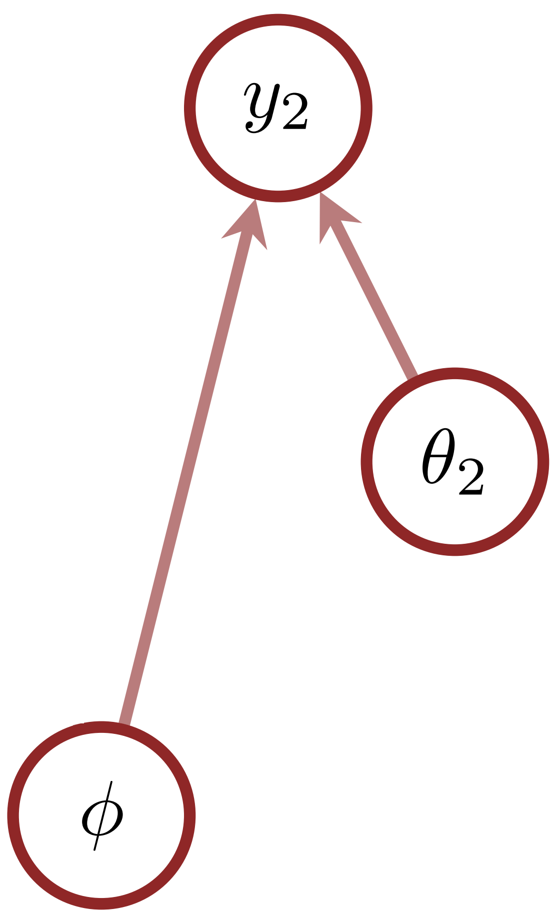
These joints models allow us to fuse inferences from different observations together in a consistent way.



$$\pi(y_1, \quad \theta_1, \quad \phi) = \pi(y_1 \mid \theta_1, \phi)$$

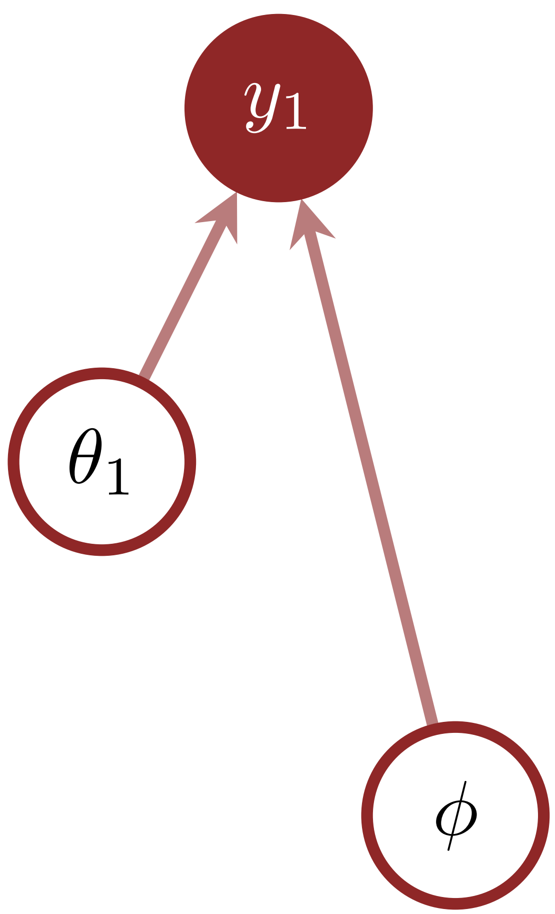
$$\pi(\theta_1) \qquad \pi(\phi)$$

These joints models allow us to fuse inferences from different observations together in a consistent way.



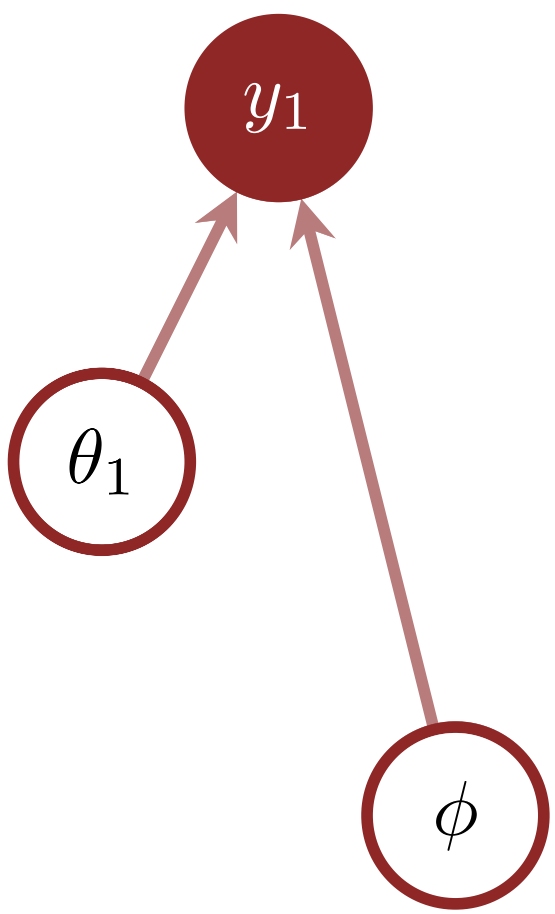
$$\pi(y_2, \theta_2, \phi) = \pi(y_2 \mid \theta_2, \phi) \, \pi(\theta_2) \, \pi(\phi)$$

Often this fusion is achieved through *sequential inferences*; one observation informs an initial posterior distribution.



$$\pi(\theta_1, \phi \mid \tilde{y}_1) \propto \pi(y_1 \mid \theta_1, \phi) \, \pi(\theta_1) \, \pi(\phi)$$

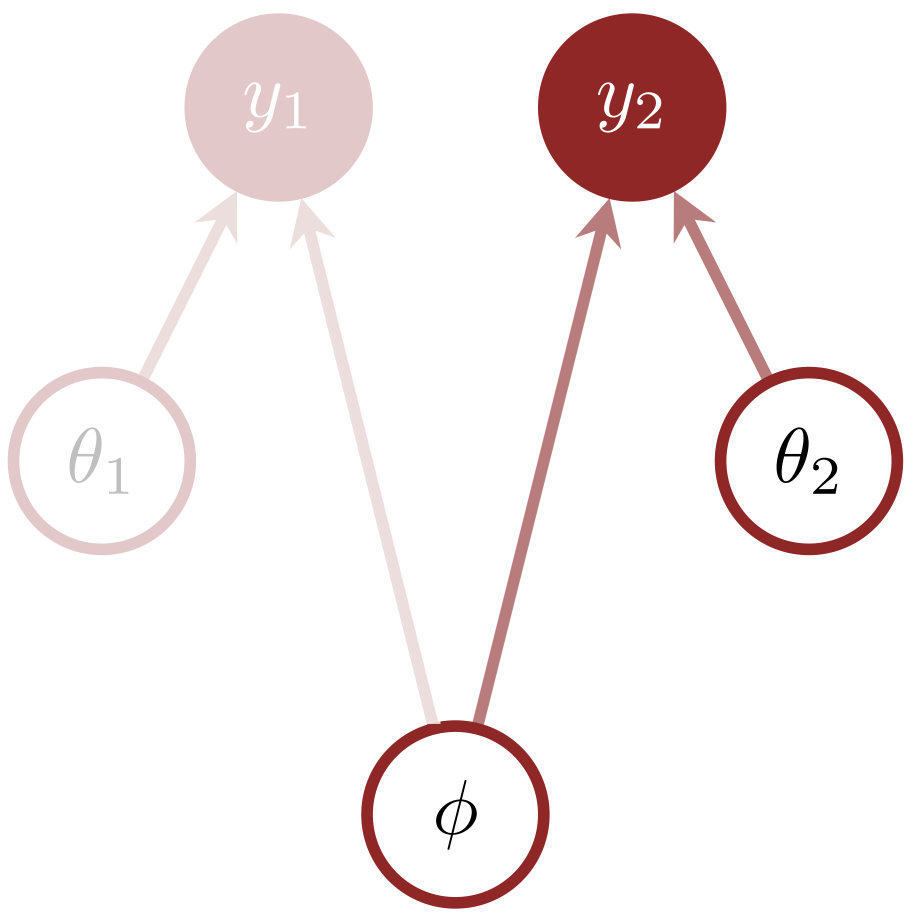
That posterior distribution is then marginalized into a prior model for the shared parameter.



$$\pi(\theta_1, \phi \mid \tilde{y}_1) \propto \pi(y_1 \mid \theta_1, \phi) \, \pi(\theta_1) \, \pi(\phi)$$

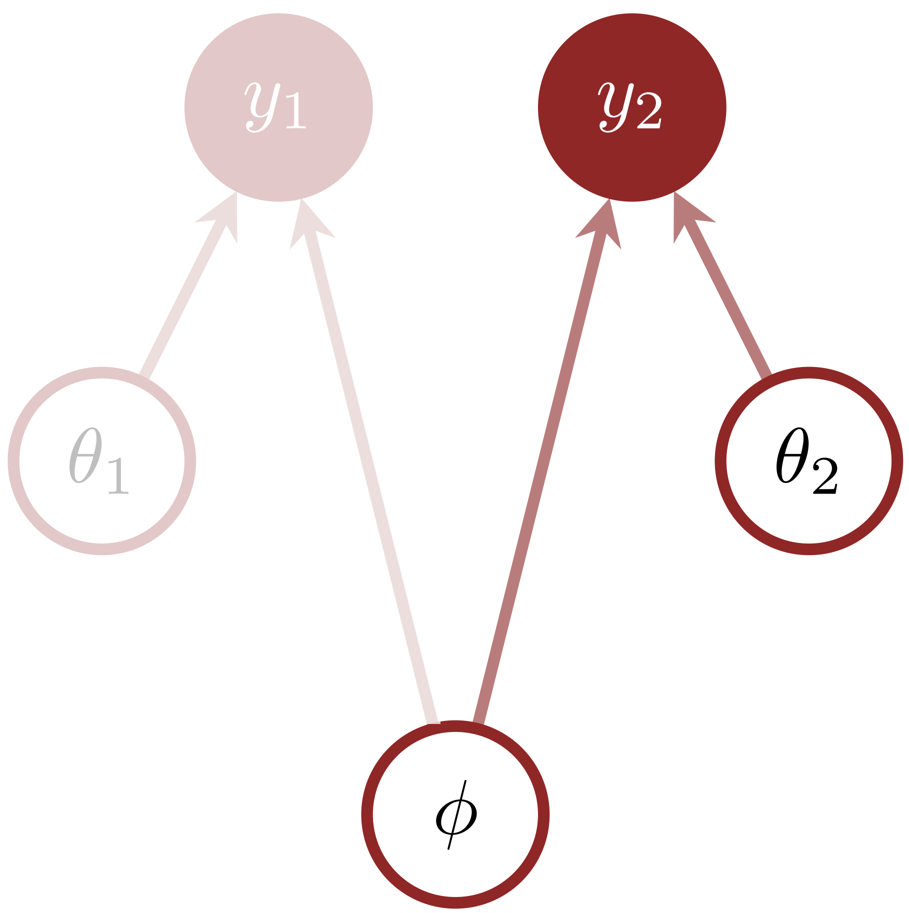
$$\pi(\phi \mid \tilde{y}_1) = \int \mathrm{d}\theta_1 \, \pi(\theta_1, \phi \mid \tilde{y}_1)$$

This prior model is then updated into the final posterior distribution using the second observation.



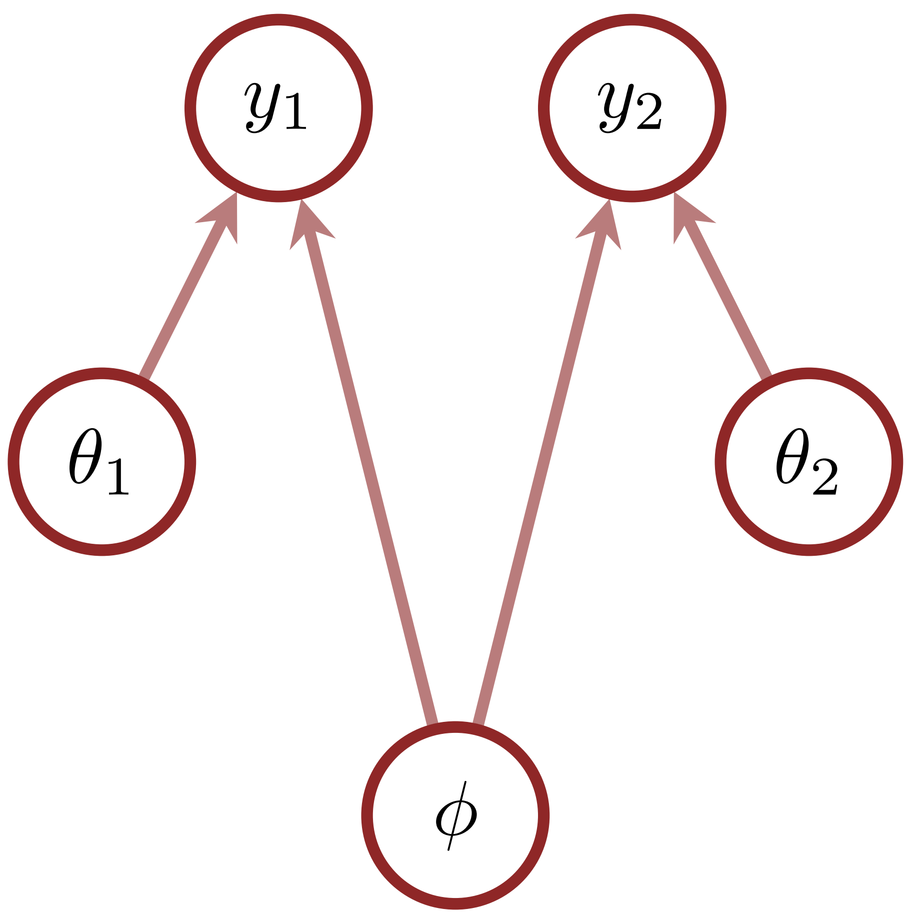
$$
\pi(\theta_2, \phi \mid \tilde{y}_1, \tilde{y}_2) \propto \pi(\tilde{y}_2 \mid \theta_2, \phi) \, \pi(\theta_2) \, \pi(\phi \mid \tilde{y}_1)
$$

While this is all technically correct the marginal prior model can be infeasible to implement in practice.



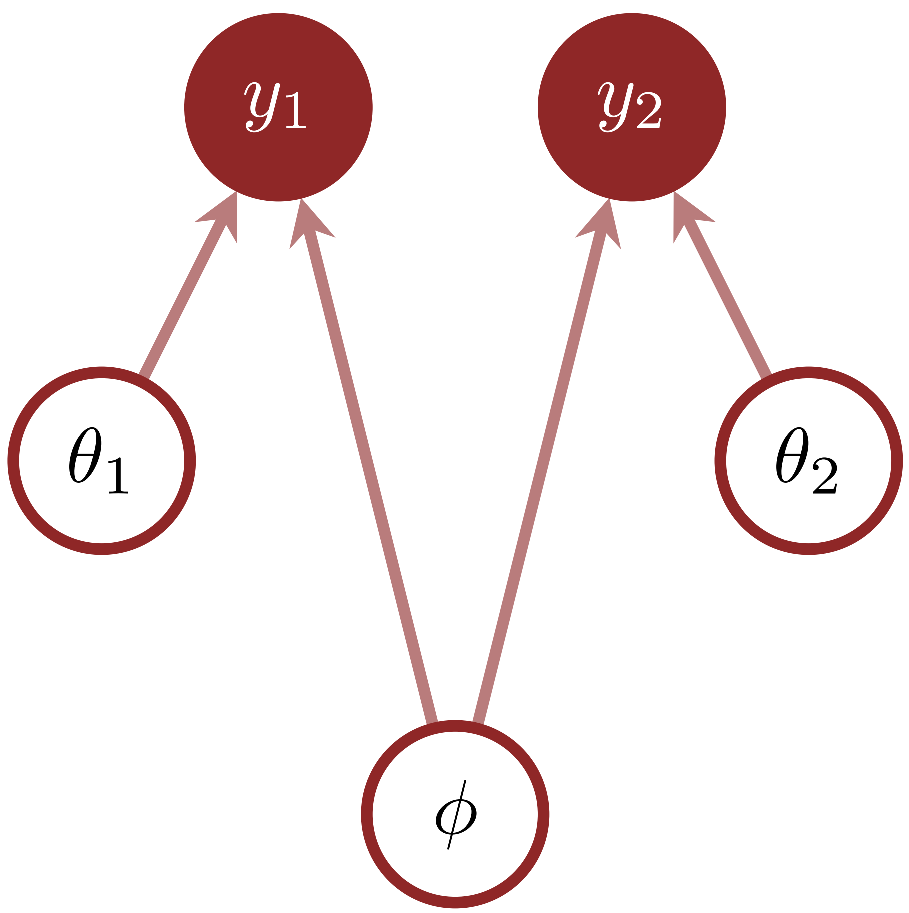
$$\pi(\theta_2, \phi \mid \tilde{y}_1, \tilde{y}_2) \propto \pi(\tilde{y}_2 \mid \theta_2, \phi) \, \pi(\theta_2) \, \pi(\phi \mid \tilde{y}_1)$$

Evaluating the joint model on both observations returns the final posterior distribution in one implementable step.



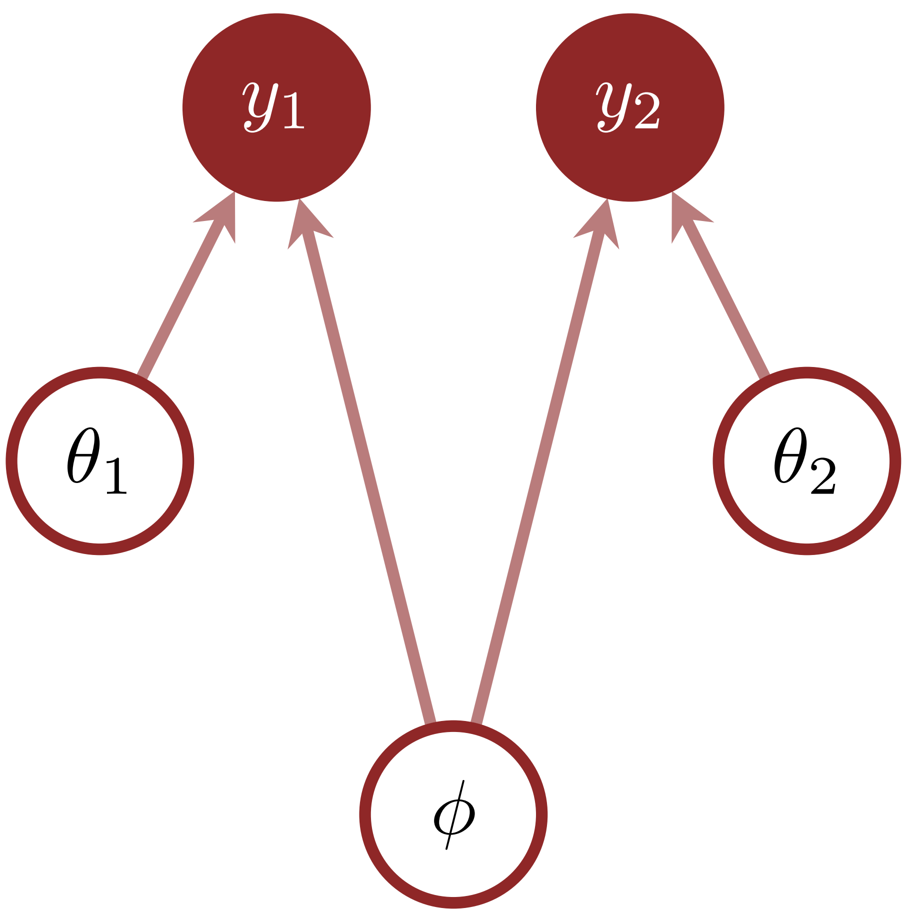
$$\pi(y_1, y_2, \theta_1, \theta_2, \phi)$$

Evaluating the joint model on both observations returns the final posterior distribution in one implementable step.



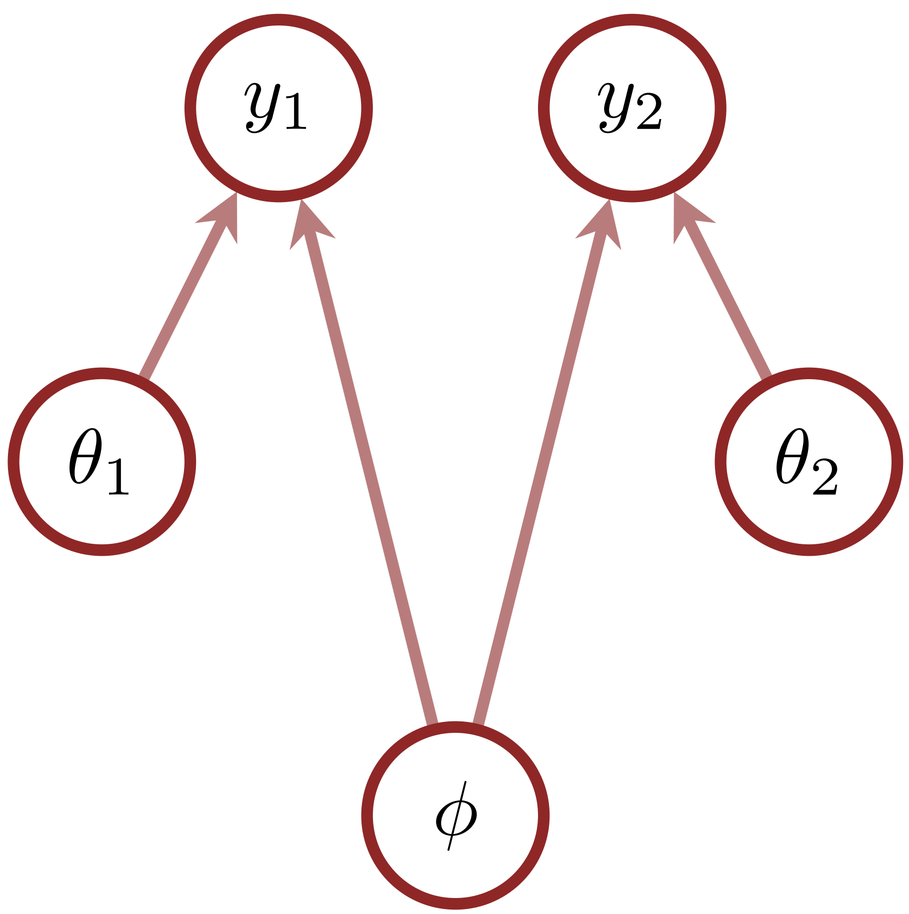
$$\pi(\tilde{y}_1, \tilde{y}_2, \theta_1, \theta_2, \phi)$$

Evaluating the joint model on both observations returns the final posterior distribution in one implementable step.



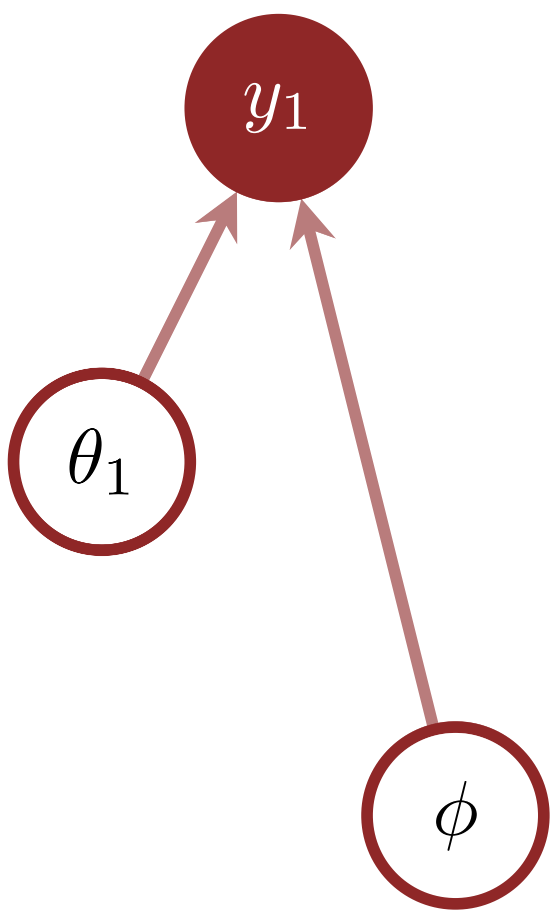
$$\pi(\theta_1, \theta_2, \phi \mid \tilde{y}_1, \tilde{y}_2) \propto \pi(\tilde{y}_1, \tilde{y}_2, \theta_1, \theta_2, \phi)$$

As before these generalized predictions can be achieved sequentially, at least in theory.



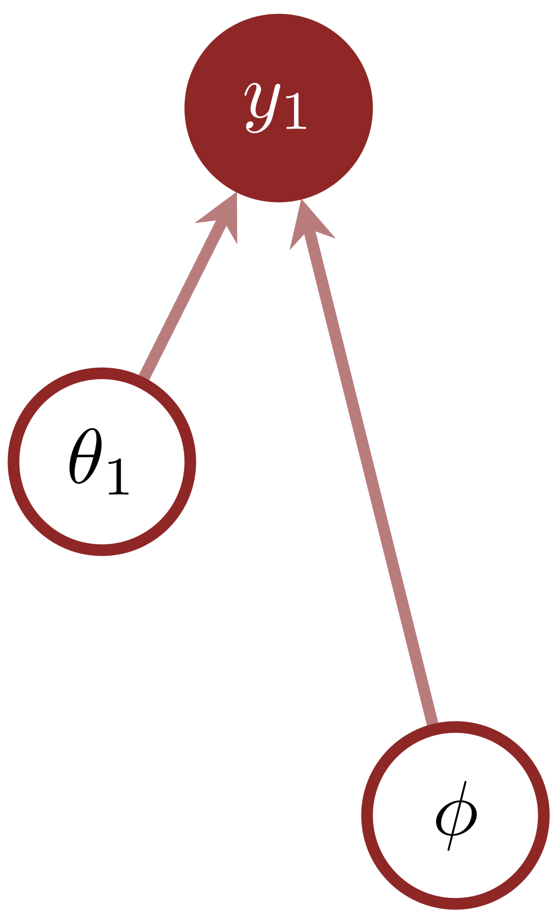
$$
\begin{aligned}
\pi(y_1, y_2, \theta_1, \theta_2, \phi) = \; & \pi(y_1 \mid \theta_1, \phi) \\
& \pi(y_2 \mid \theta_2, \phi) \\
& \pi(\theta_1) \, \pi(\theta_2) \, \pi(\phi)
\end{aligned}
$$

A model of the first data generating process informs a posterior distribution for the relevant parameters.



$$\pi(\theta_1, \phi \mid \tilde{y}_1) \propto \pi(y_1 \mid \theta_1, \phi) \, \pi(\theta_1) \, \pi(\phi)$$

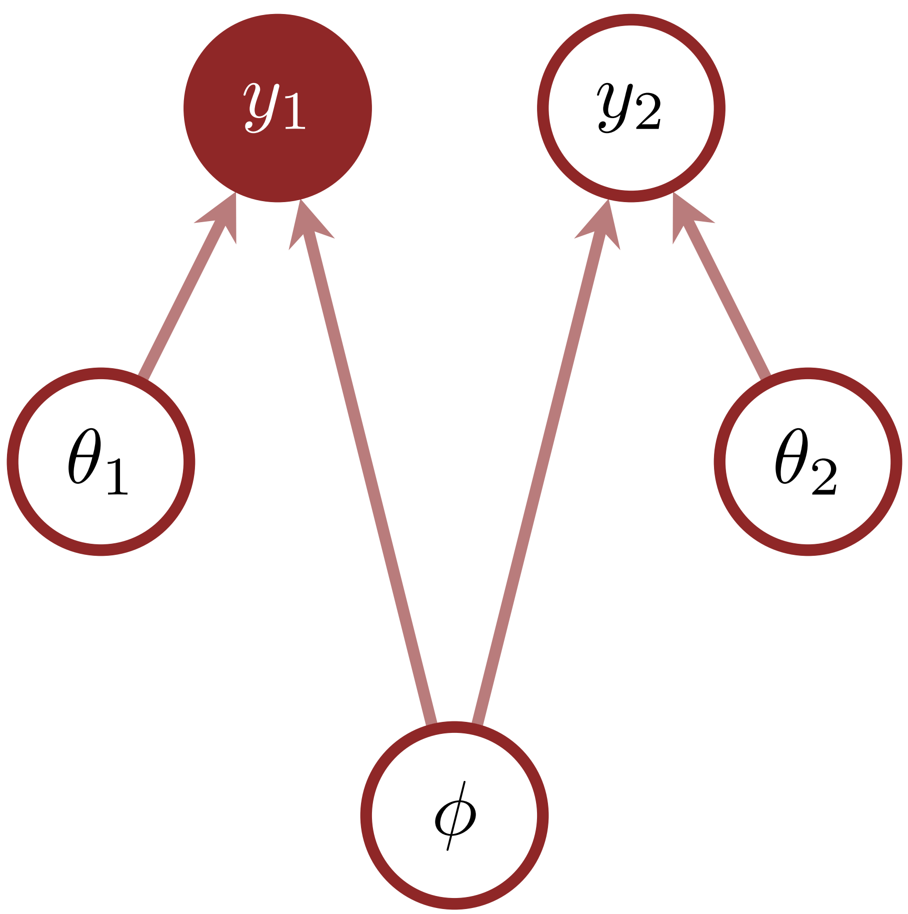
Marginalizing this posterior distribution informs the shared parameter that we need to make predictions.



$$\pi(\theta_1, \phi \mid \tilde{y}_1) \propto \pi(y_1 \mid \theta_1, \phi) \, \pi(\theta_1) \, \pi(\phi)$$

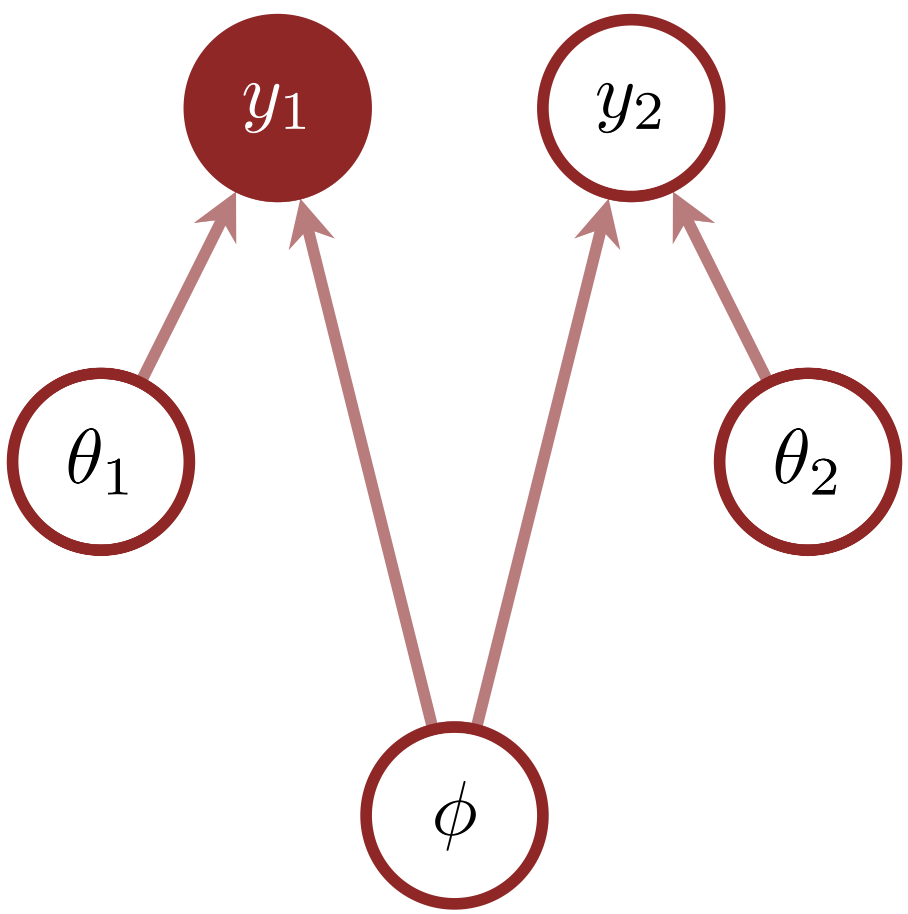
$$\pi(\phi \mid \tilde{y}_1) = \int \mathrm{d}\theta_1 \, \pi(\theta_1, \phi \mid \tilde{y}_1)$$

Finally we make predictions by averaging the second observational model against the marginalized posterior.



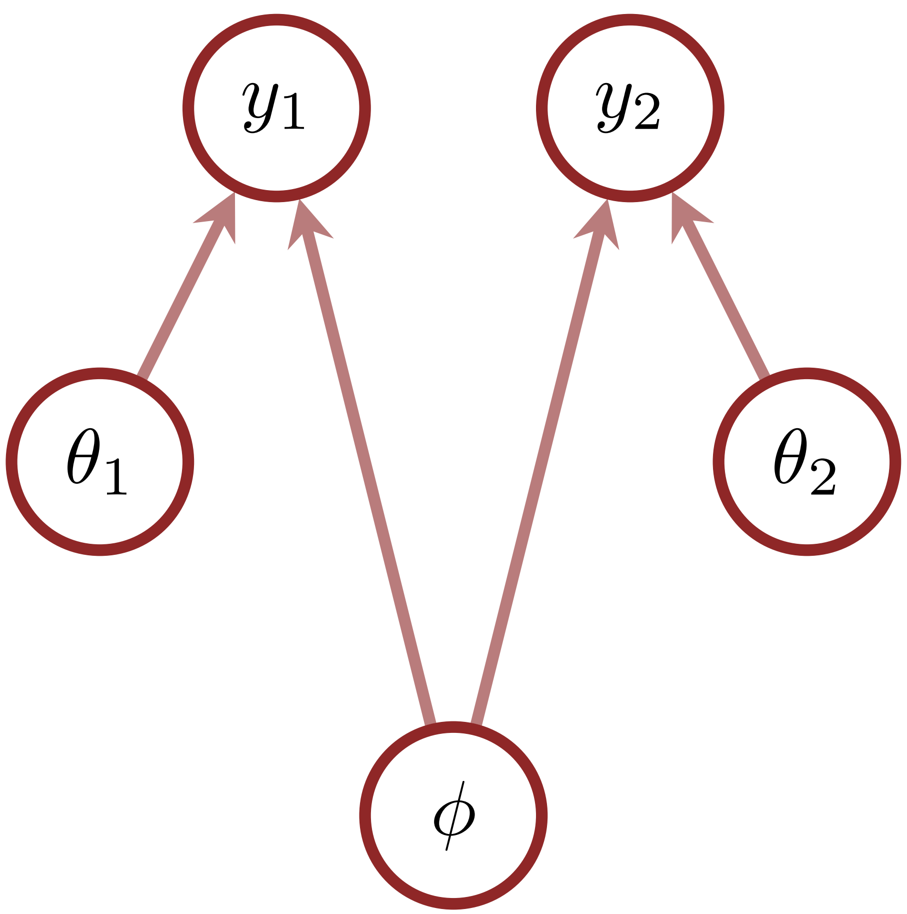
$$
\begin{aligned}
\pi(y_2 \mid \tilde{y}_1) \\
&= \int \mathrm{d}\theta_2 \, \mathrm{d}\phi \, \pi(y_2 \mid \theta_2, \phi) \, \pi(\theta_2) \, \pi(\phi \mid \tilde{y}_1)
\end{aligned}
$$

Once again the practicality of this sequential approach is limited by the feasibility of the intermediate terms.



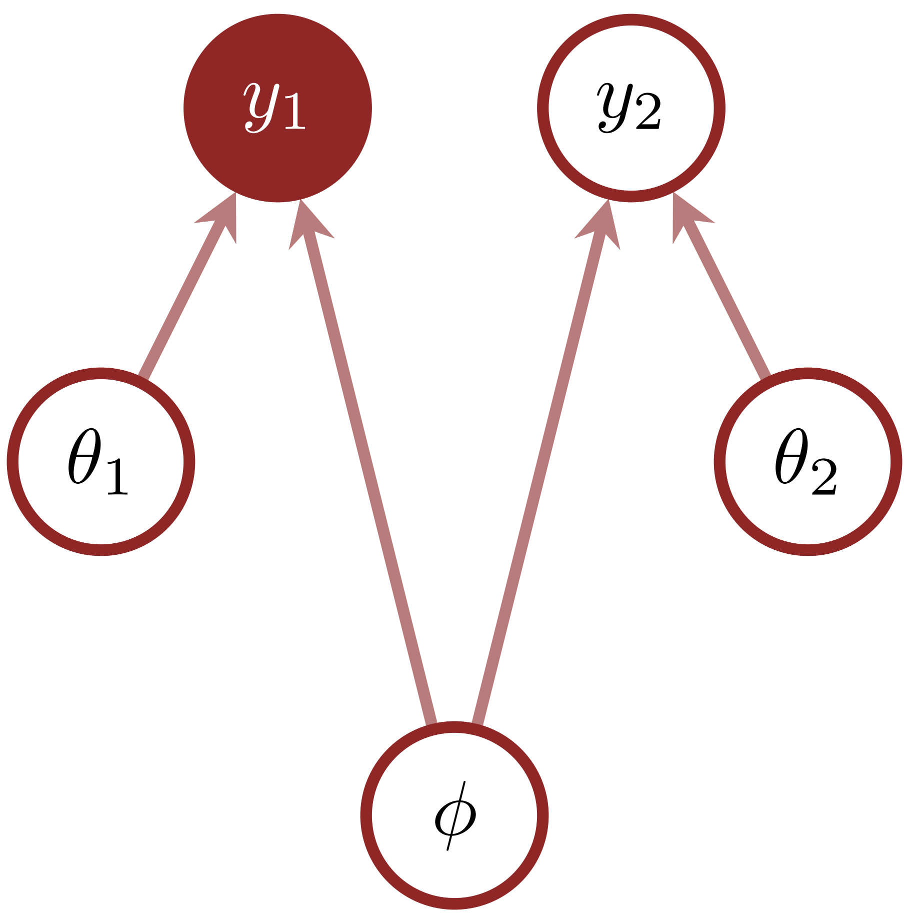
$$\pi(y_2 \mid \tilde{y}_1) = \int \mathrm{d}\theta_2 \, \mathrm{d}\phi \, \pi(y_2 \mid \theta_2, \phi) \, \pi(\theta_2) \, \pi(\phi \mid \tilde{y}_1)$$

We don't need to construct those intermediate terms, however, when working with the joint model directly.



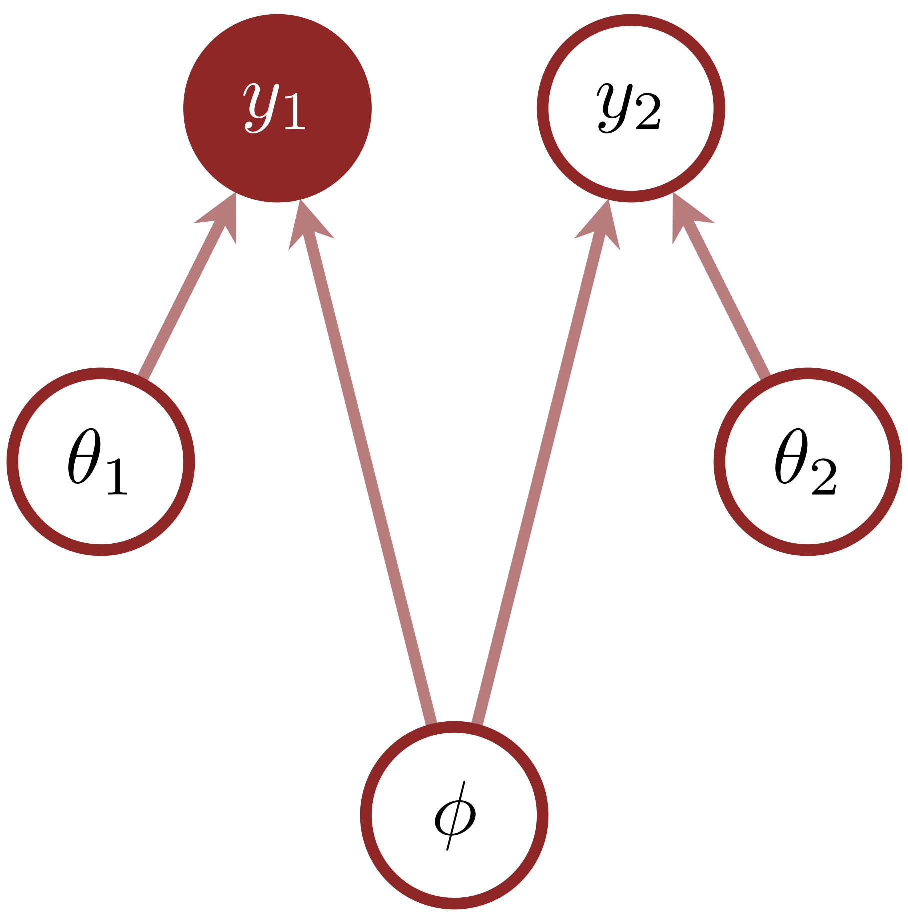
$$\pi(y_1, y_2, \theta_1, \theta_2, \phi)$$

We don't need to construct those intermediate terms, however, when working with the joint model directly.



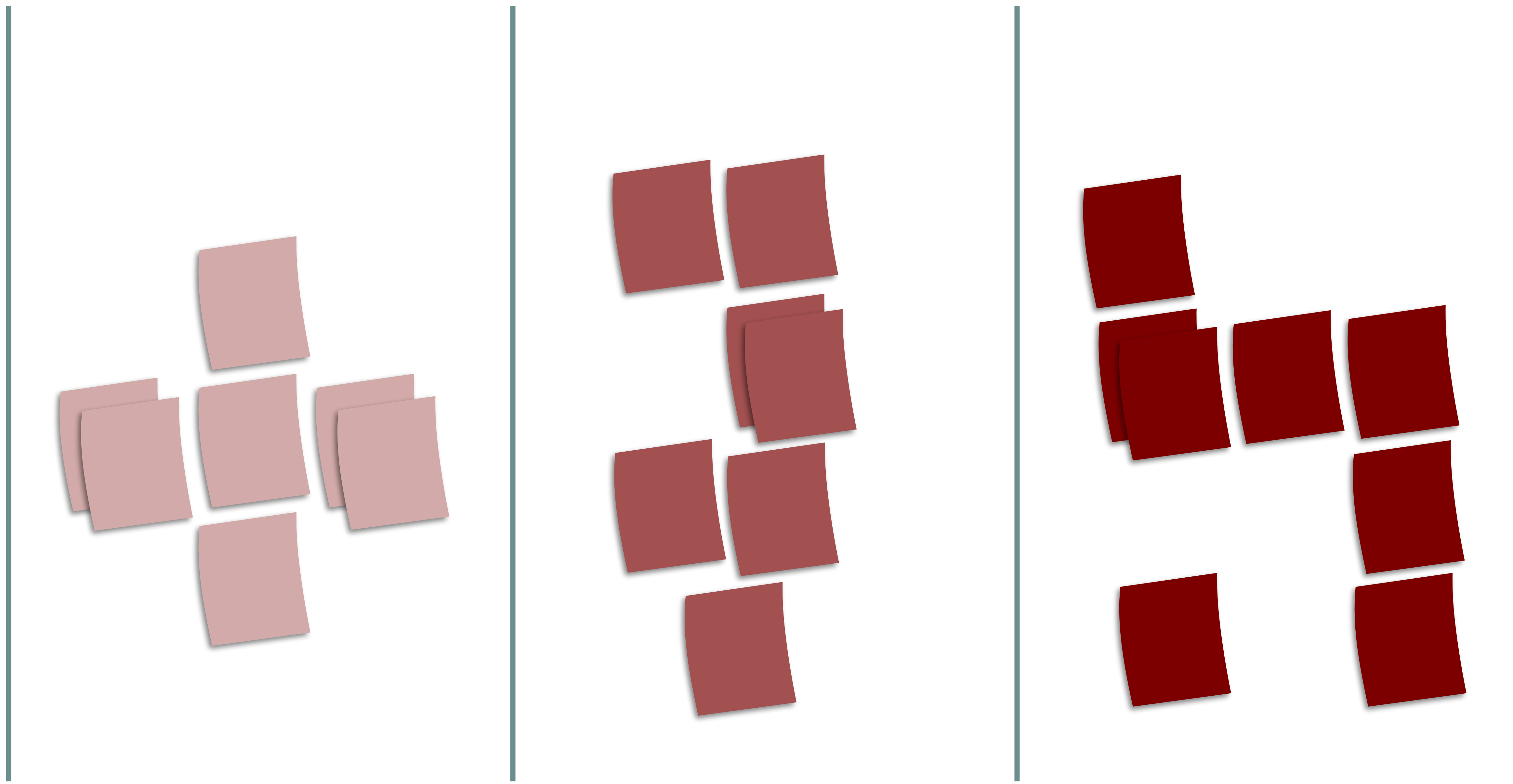
$$\pi(\tilde{y}_1, y_2, \theta_1, \theta_2, \phi)$$

We don't need to construct those intermediate terms, however, when working with the joint model directly.



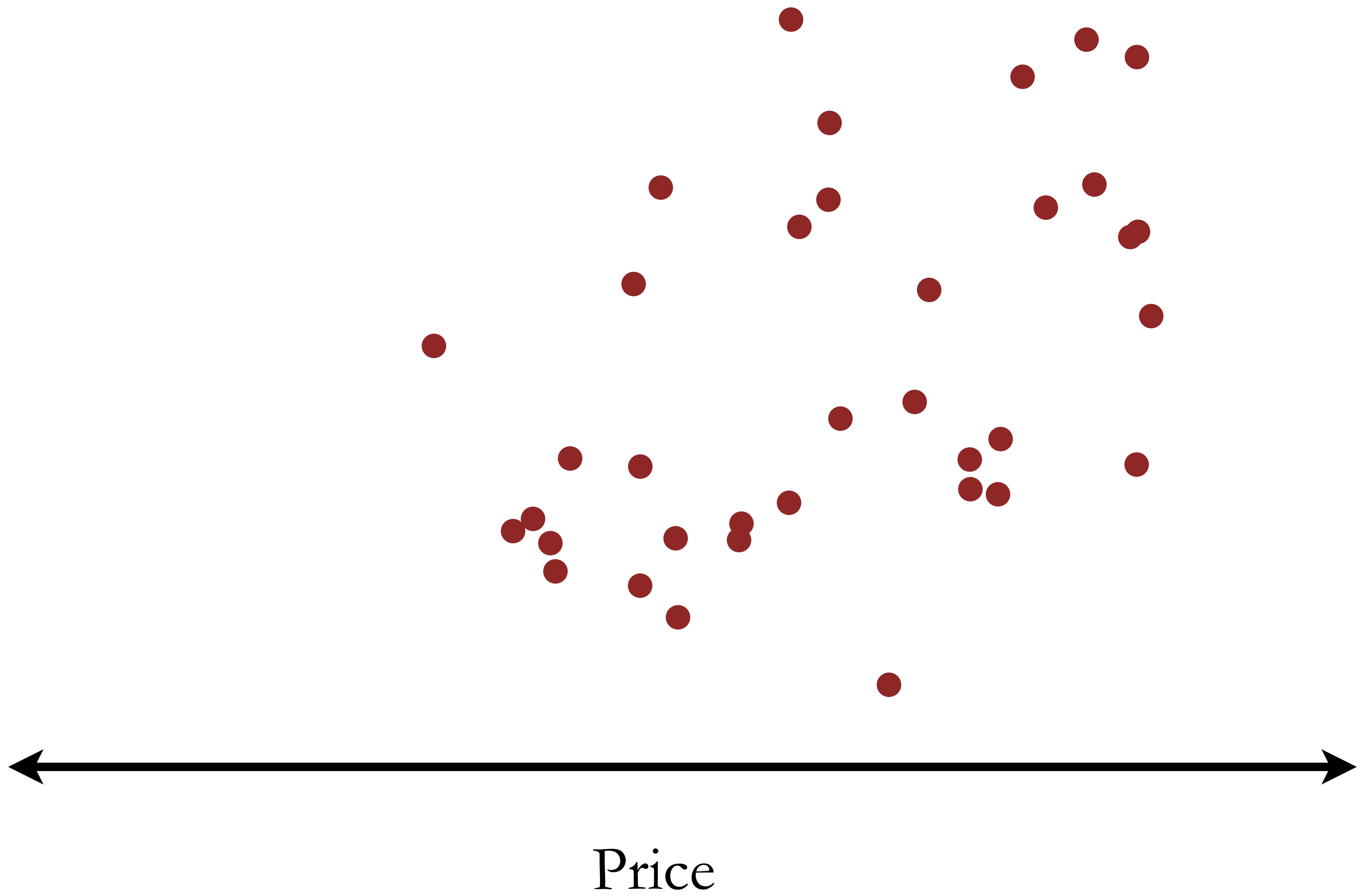
$$
\begin{aligned}
\pi(y_2 \mid \tilde{y}_1)
&= \int \mathrm{d}\theta_1 \, \mathrm{d}\theta_2 \, \mathrm{d}\phi \, \pi(\tilde{y}_1, y_2, \theta_1, \theta_2, \phi)
\end{aligned}
$$

Advanced Narrative Techniques

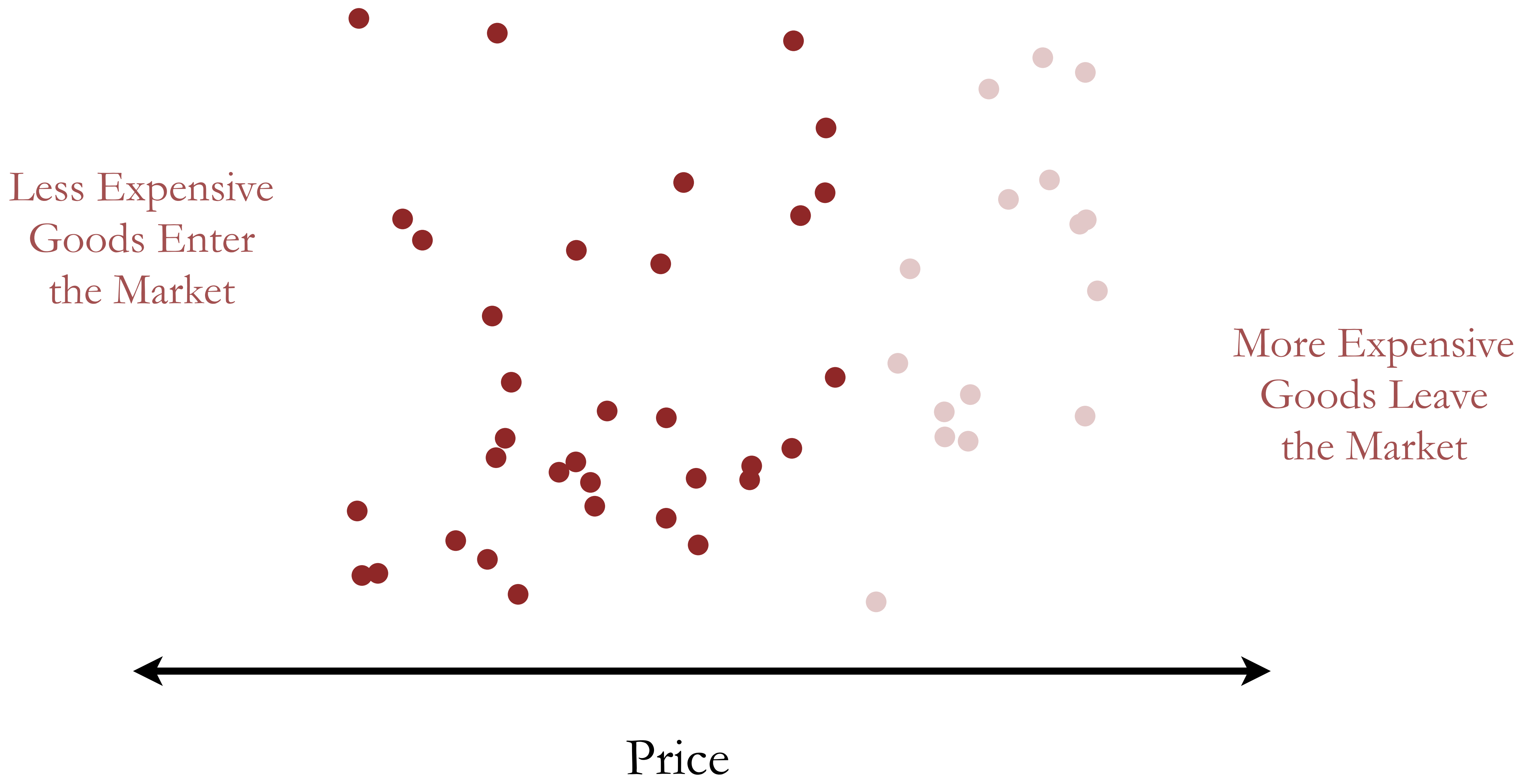


“Causal” “Inference”

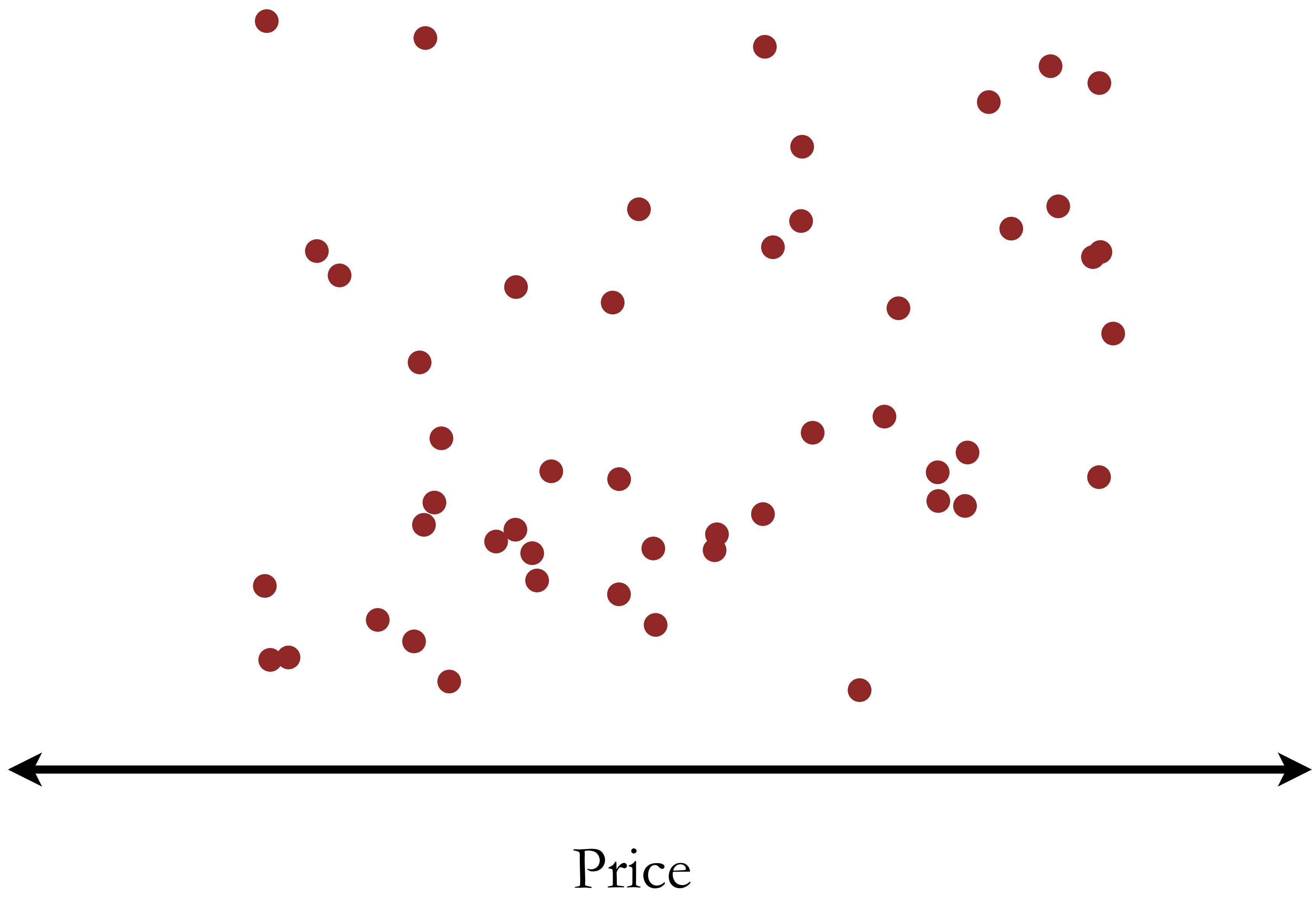
Data generating processes often change for reasons that are far beyond our control.



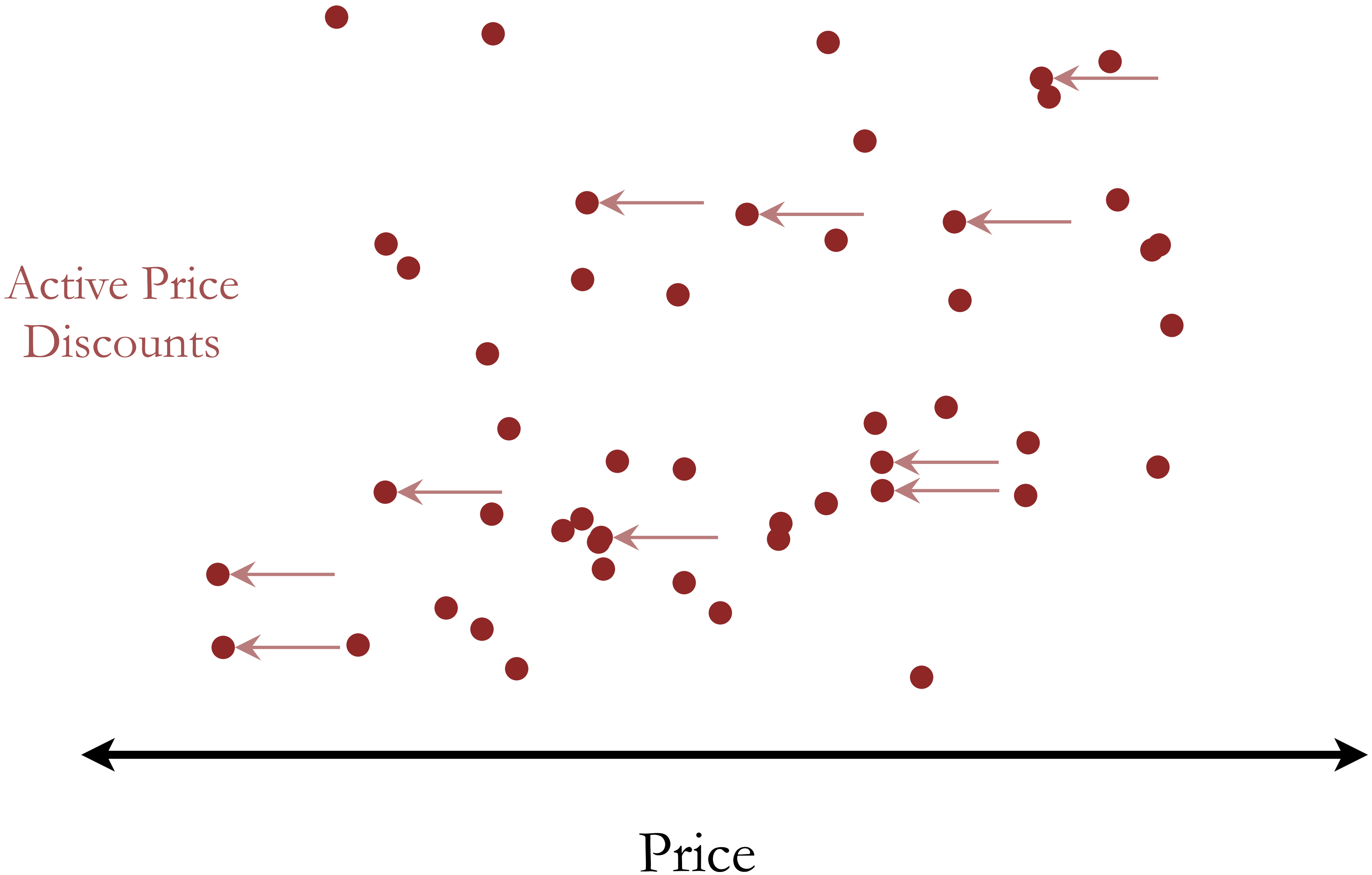
Data generating processes often change for reasons that are far beyond our control.



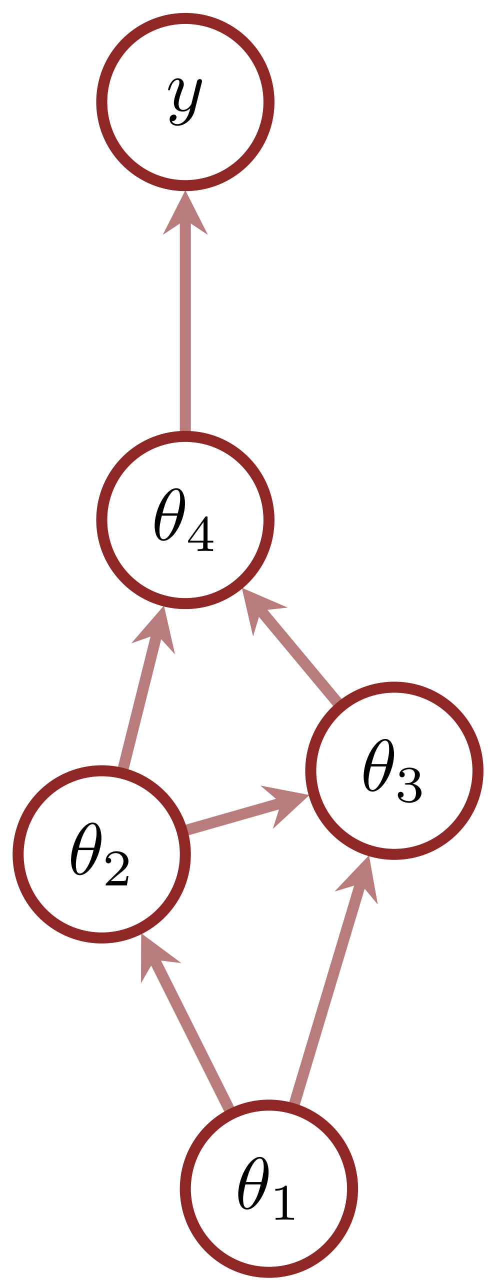
Often, however, we are interested in quantifying what happens when we actively *intervene*.



Often, however, we are interested in quantifying what happens when we actively *intervene*.

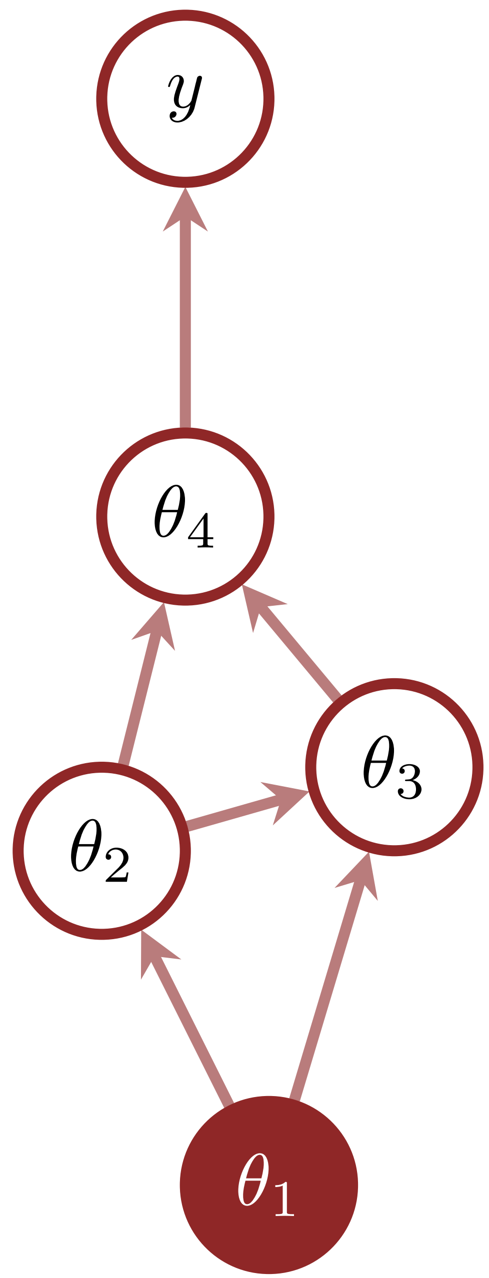


Interventions can sometimes be modeled by conditioning on an initial model, but not always.



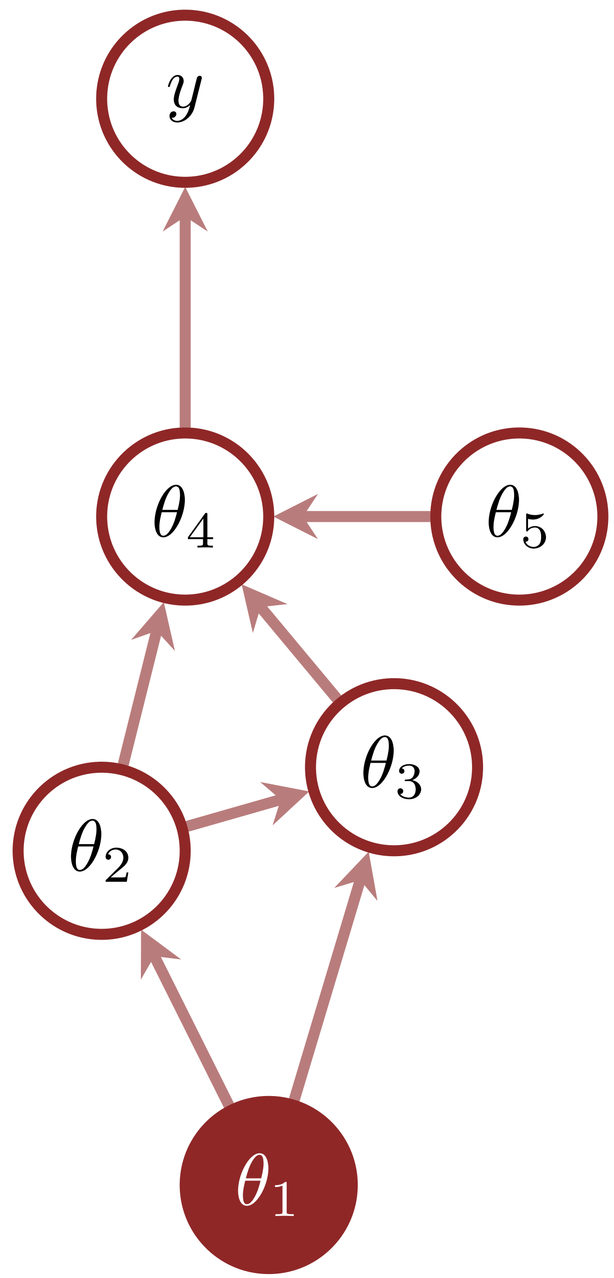
Before Intervention

Interventions can sometimes be modeled by conditioning on an initial model, but not always.



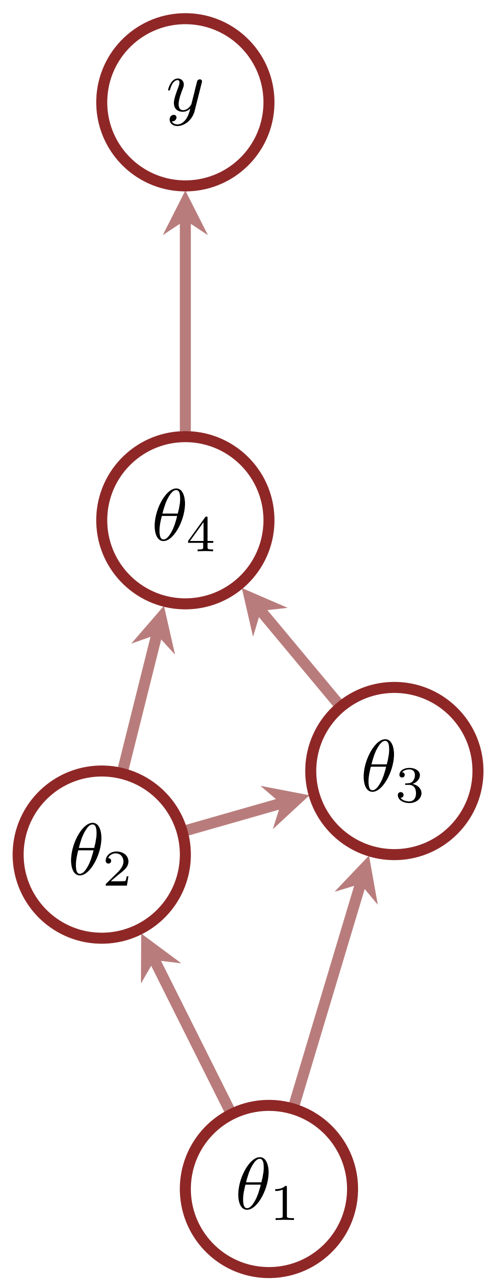
Intervention 1

Interventions can sometimes be modeled by conditioning on an initial model, but not always.



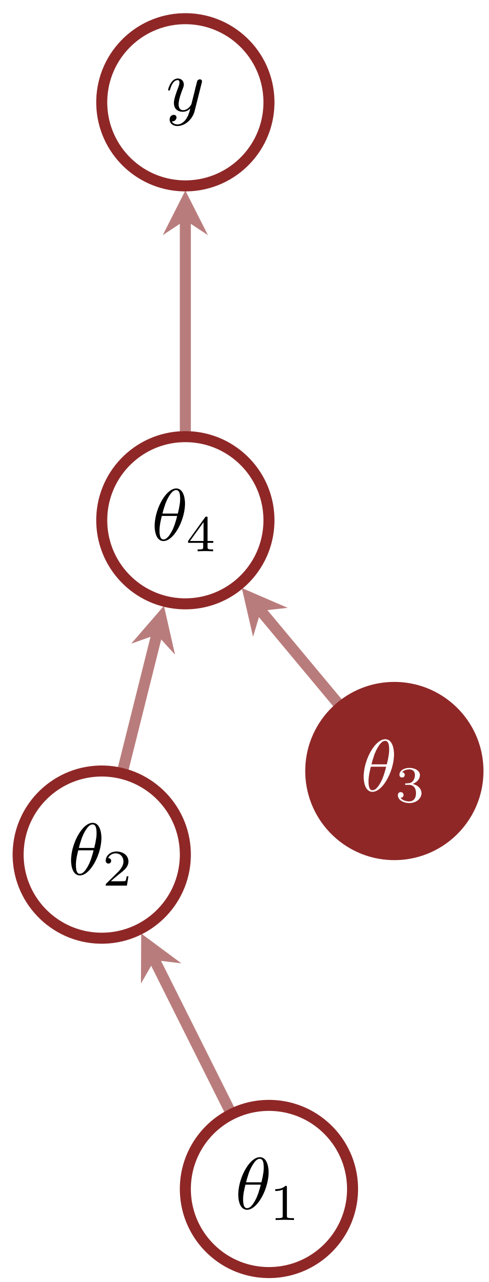
Intervention 2

Interventions that target an intermediate parameter, and obstruct the data generating process, require special care.



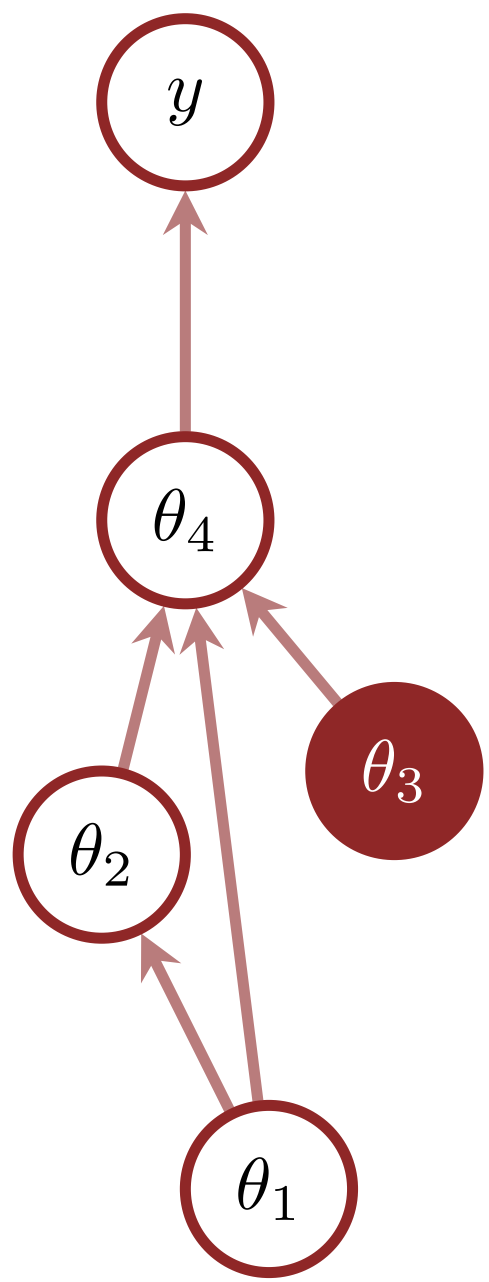
Before Intervention

Interventions that target an intermediate parameter, and obstruct the data generating process, require special care.



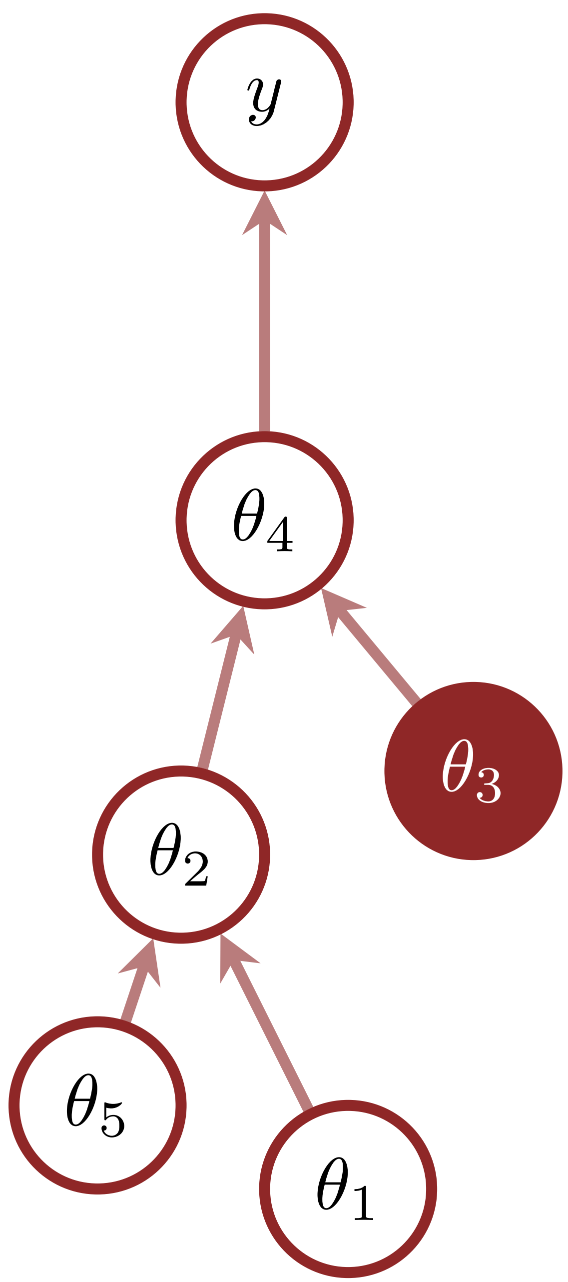
Intervention 3

Interventions that target an intermediate parameter, and obstruct the data generating process, require special care.



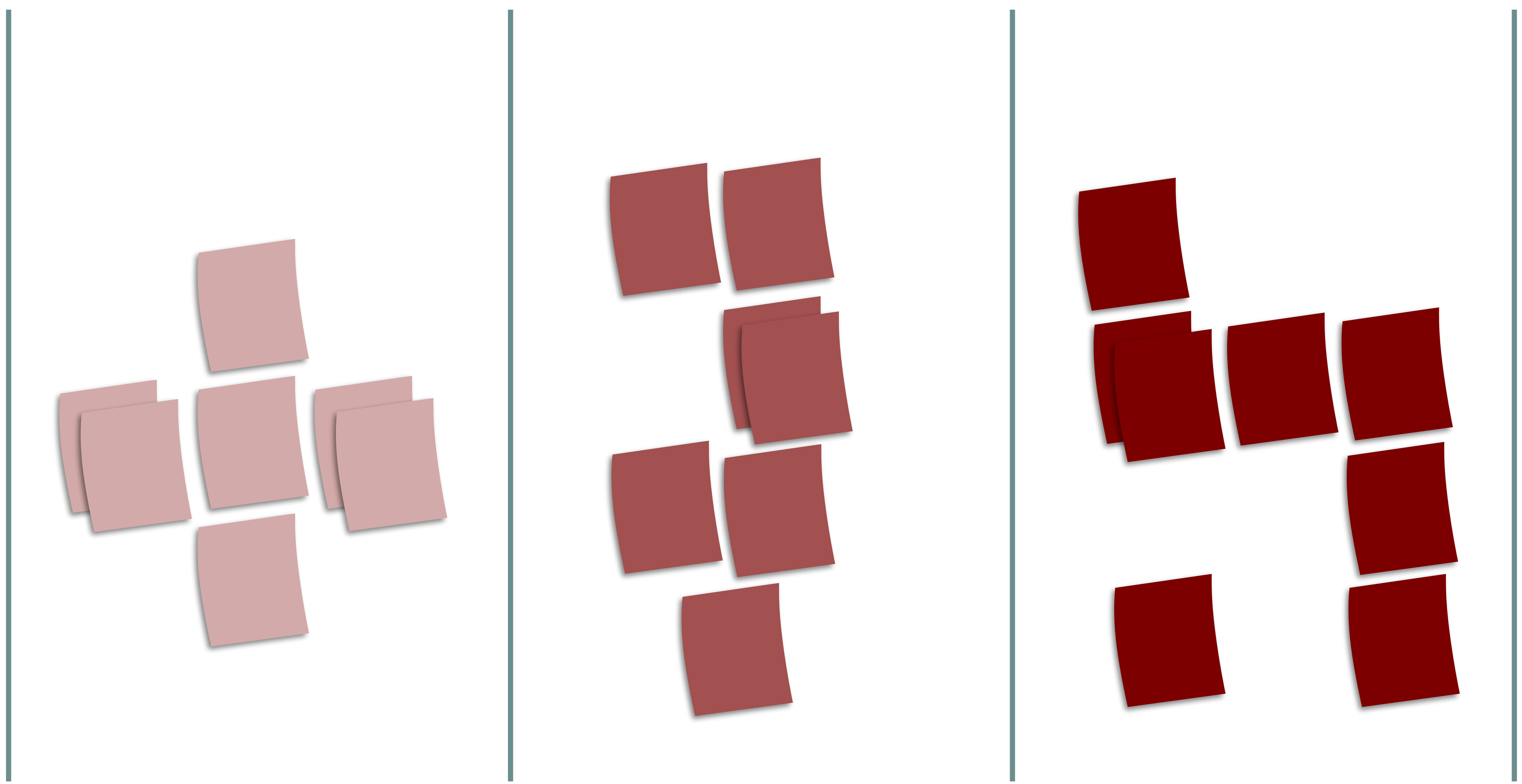
Intervention 4

Interventions that target an intermediate parameter, and obstruct the data generating process, require special care.



Intervention 5

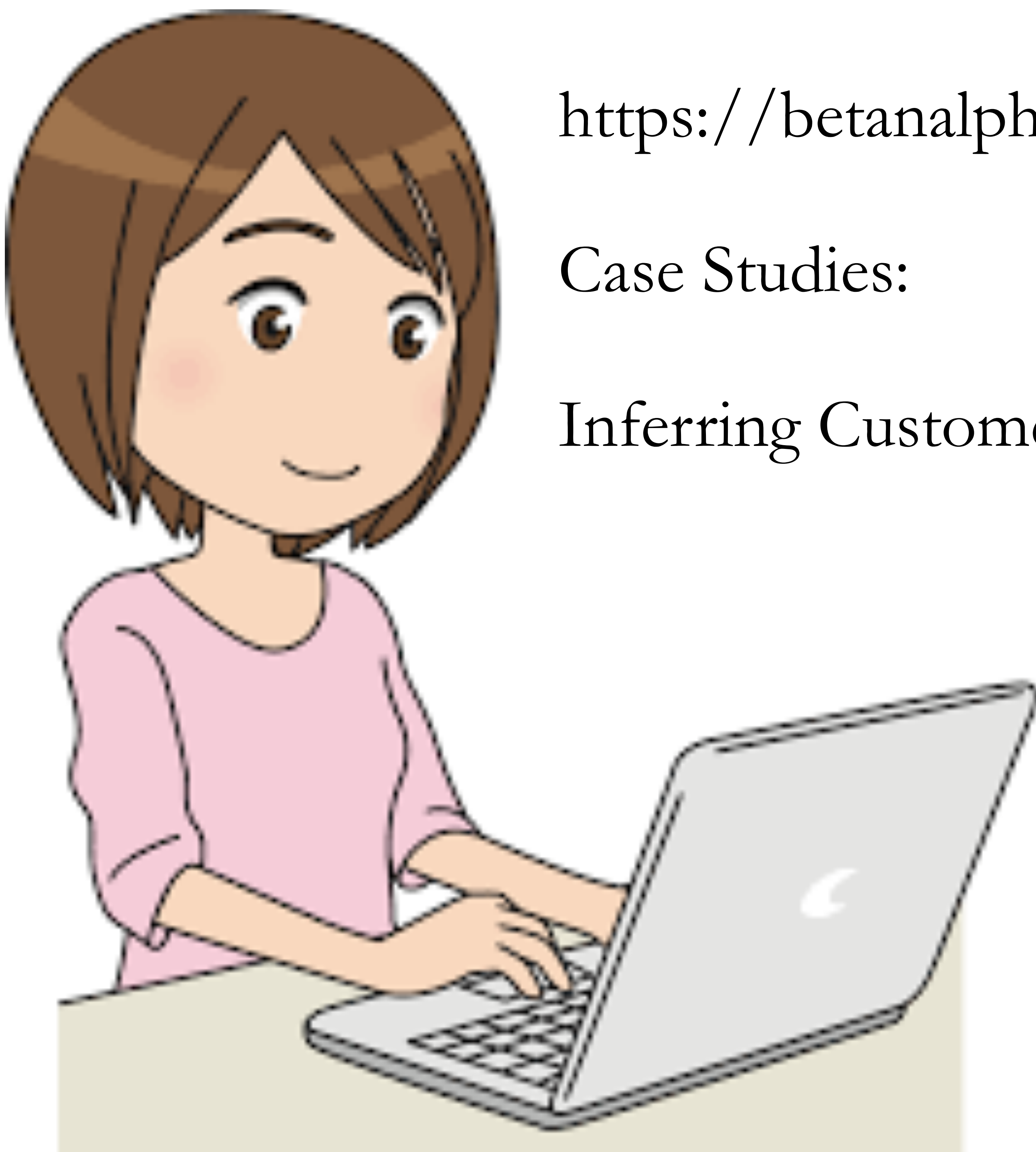
Advanced Narrative Techniques



Unreliable Narrators

As with any model a useful narratively generative model has to be carefully validated against our domain expertise.

Story Time



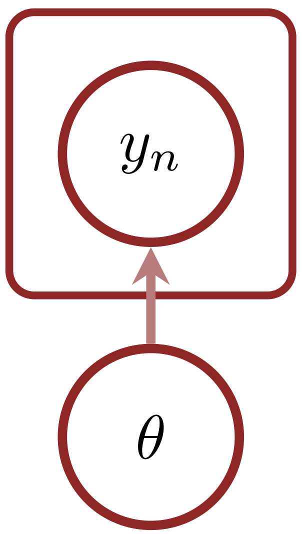
https://betanalpha.github.io/writing/

Case Studies:

Inferring Customer Conversion HTML PDF

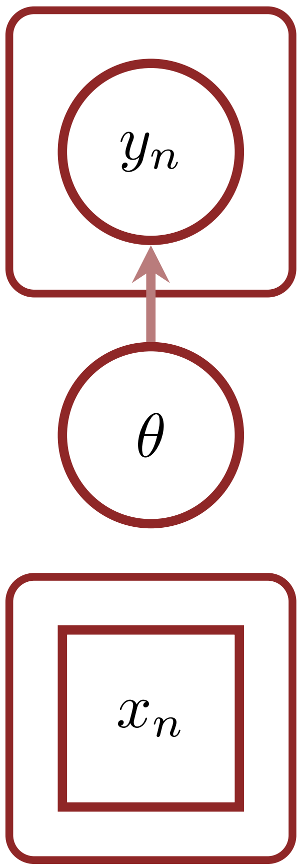
To start the discussion let's consider how visits to a company website convert to some sort of sale.

We start with a simple model where all customers have the same, homogeneous conversion probability.



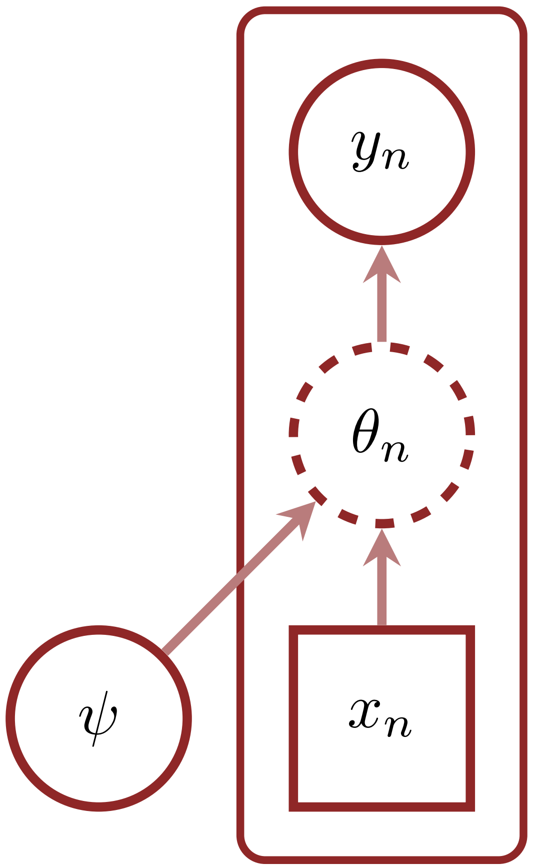
$$\pi(y_1, \ldots, y_N, \theta) = \left[\prod_{n=1}^{N} \text{Bernoulli}(y_n \mid \theta) \right] \cdot \pi(\theta)$$

Can we better model the conversion process if certain characteristics of each visitor are available?



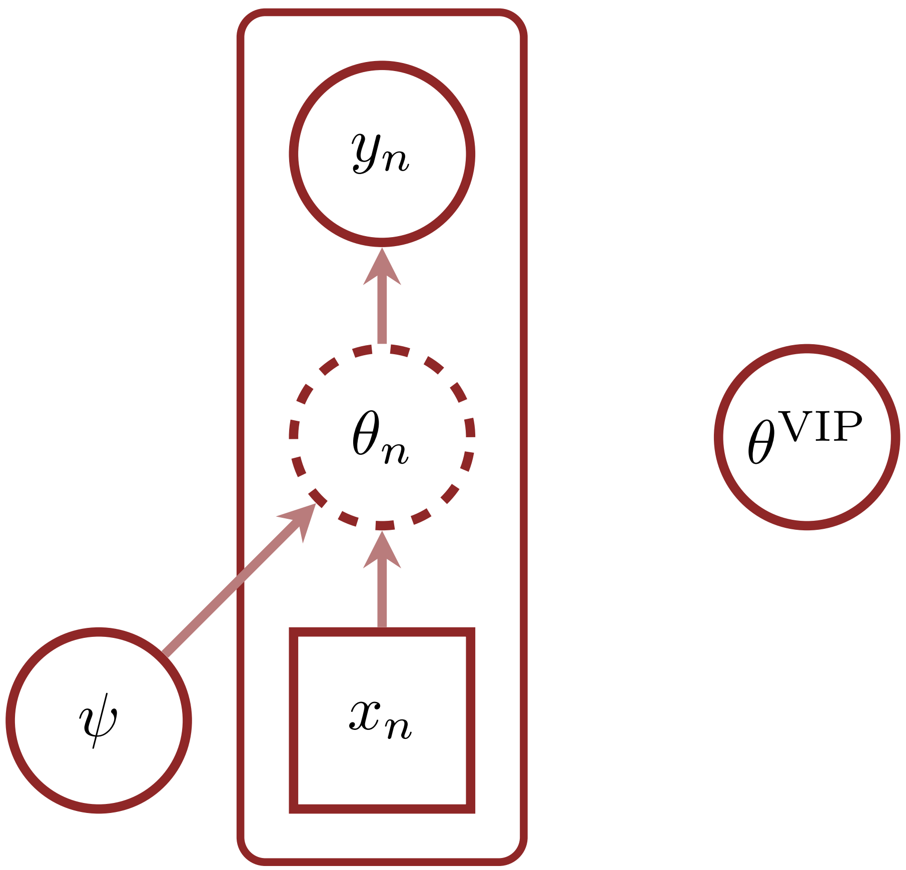
$$\pi(y_1, \ldots, y_N, \theta) = \left[\prod_{n=1}^{N} \text{Bernoulli}(y_n \mid \theta) \right] \cdot \pi(\theta)$$

For example we can deterministically construct a conversion probability from individual characteristics.



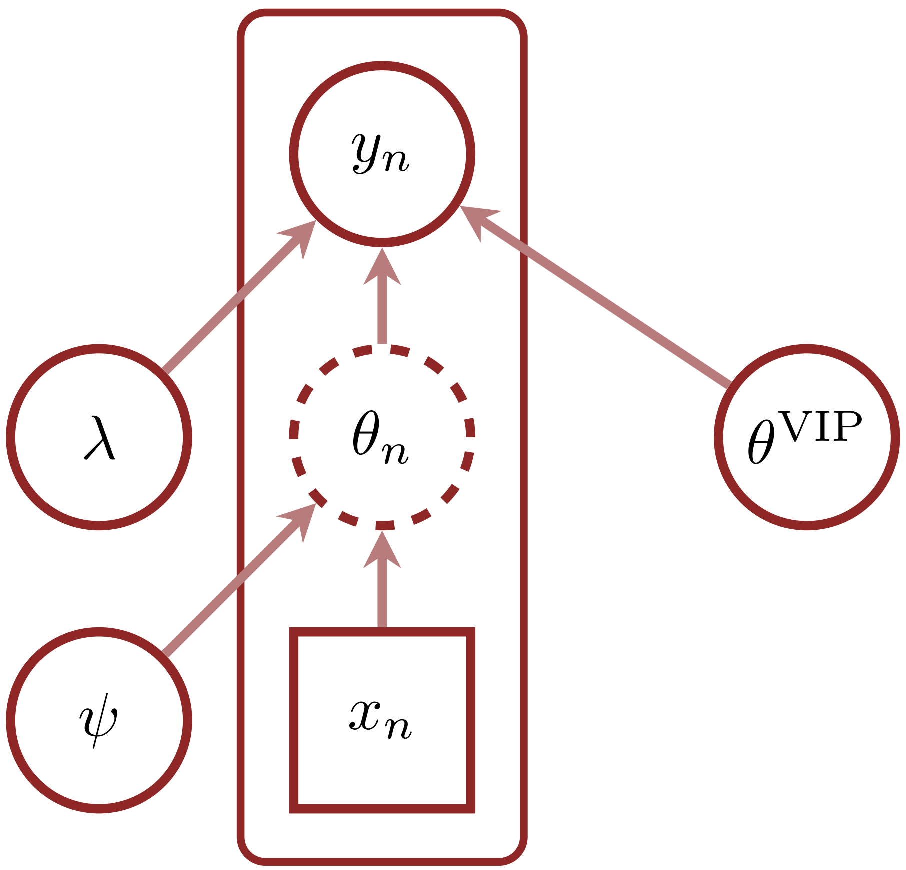
$$
\pi(y_1, \ldots, y_N, \psi; x_1, \ldots, x_N) = \left[\prod_{n=1}^{N} \text{Bernoulli}(y_n \mid \theta(x_n, \psi)) \right] \cdot \pi(\psi)
$$

The variation in this reconstruction, however, may not be able to account for different types of visitors.



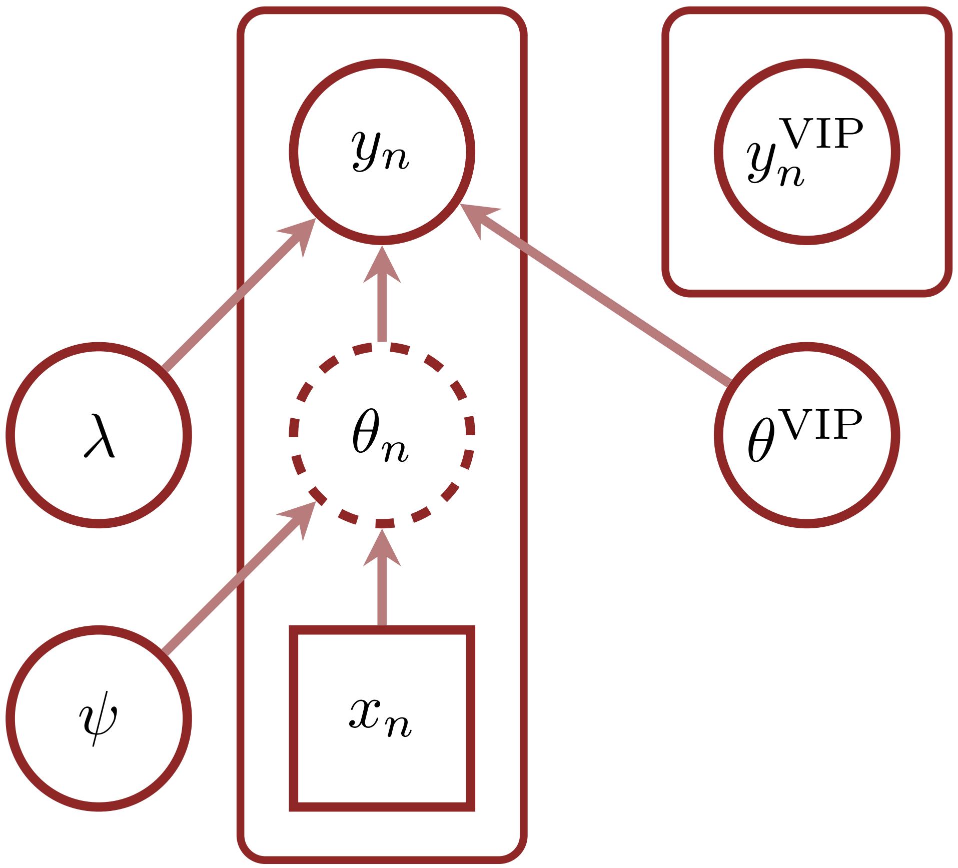
$$
\begin{aligned}
&\pi(y_1, \ldots, y_N, \psi; x_1, \ldots, x_N) \\
&\quad = \left[\prod_{n=1}^{N} \text{Bernoulli}(y_n \mid \theta(x_n, \psi)) \right] \cdot \pi(\psi)
\end{aligned}
$$

To account for the possibility that a visitor might be a loyal customer we could introduce a mixture model.



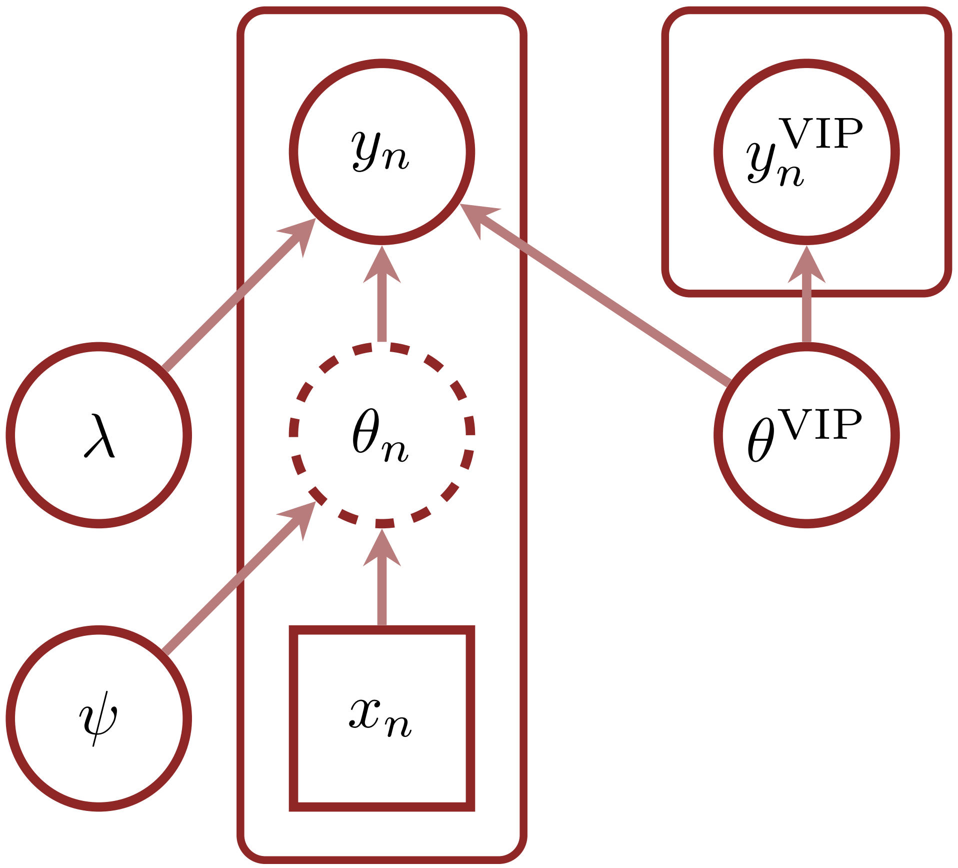
$$
\begin{aligned}
&\pi(y_1, \ldots, y_N, \psi, \theta^{\text{VIP}}, \lambda; x_1, \ldots, x_N) \\
&\quad = \Bigg[\prod_{n=1}^{N} \quad \lambda \cdot \text{Bernoulli}(y_n \mid \theta(x_n, \psi)) \\
&\quad\quad\quad + (1 - \lambda) \cdot \text{Bernoulli}(y_n \mid \theta^{\text{VIP}}) \Bigg] \\
&\quad\quad \cdot \pi(\psi) \, \pi(\theta^{\text{VIP}}) \, \pi(\lambda)
\end{aligned}
$$

Data from only loyal customers can be very helpful in differentiating the possible visitor behaviors.



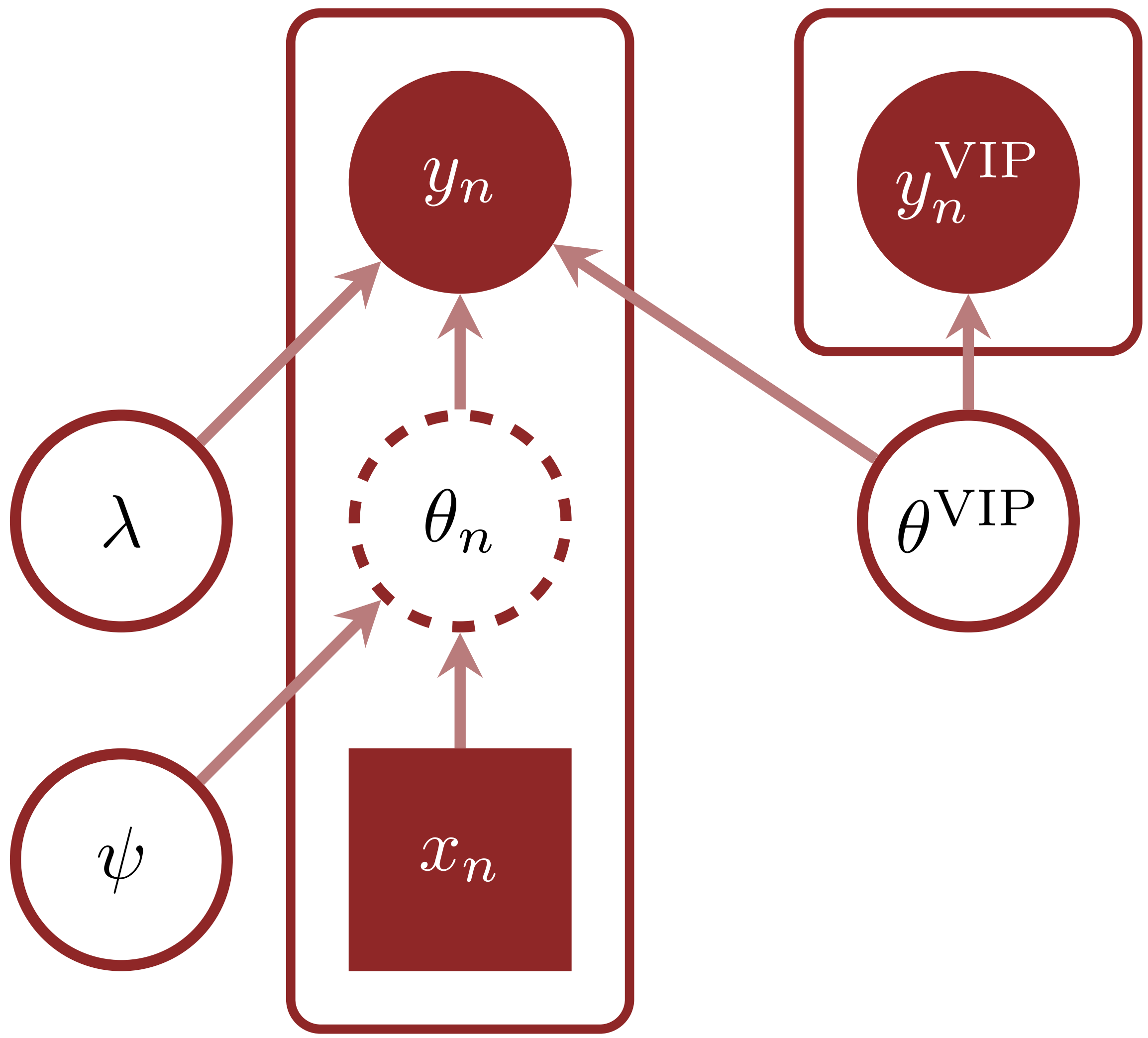
$$
\begin{aligned}
\pi(y_1, \ldots, y_N, \psi, \theta^{\mathrm{VIP}}, \lambda; x_1, \ldots, x_N)
&= \Bigg[\prod_{n=1}^{N} \quad \lambda \cdot \mathrm{Bernoulli}(y_n \mid \theta(x_n, \psi)) \\
&\qquad\qquad + (1 - \lambda) \cdot \mathrm{Bernoulli}(y_n \mid \theta^{\mathrm{VIP}}) \Bigg] \\
&\quad \cdot \pi(\psi) \, \pi(\theta^{\mathrm{VIP}}) \, \pi(\lambda)
\end{aligned}
$$

Incorporating these observations into the existing model ensures that they consistently inform our inferences.

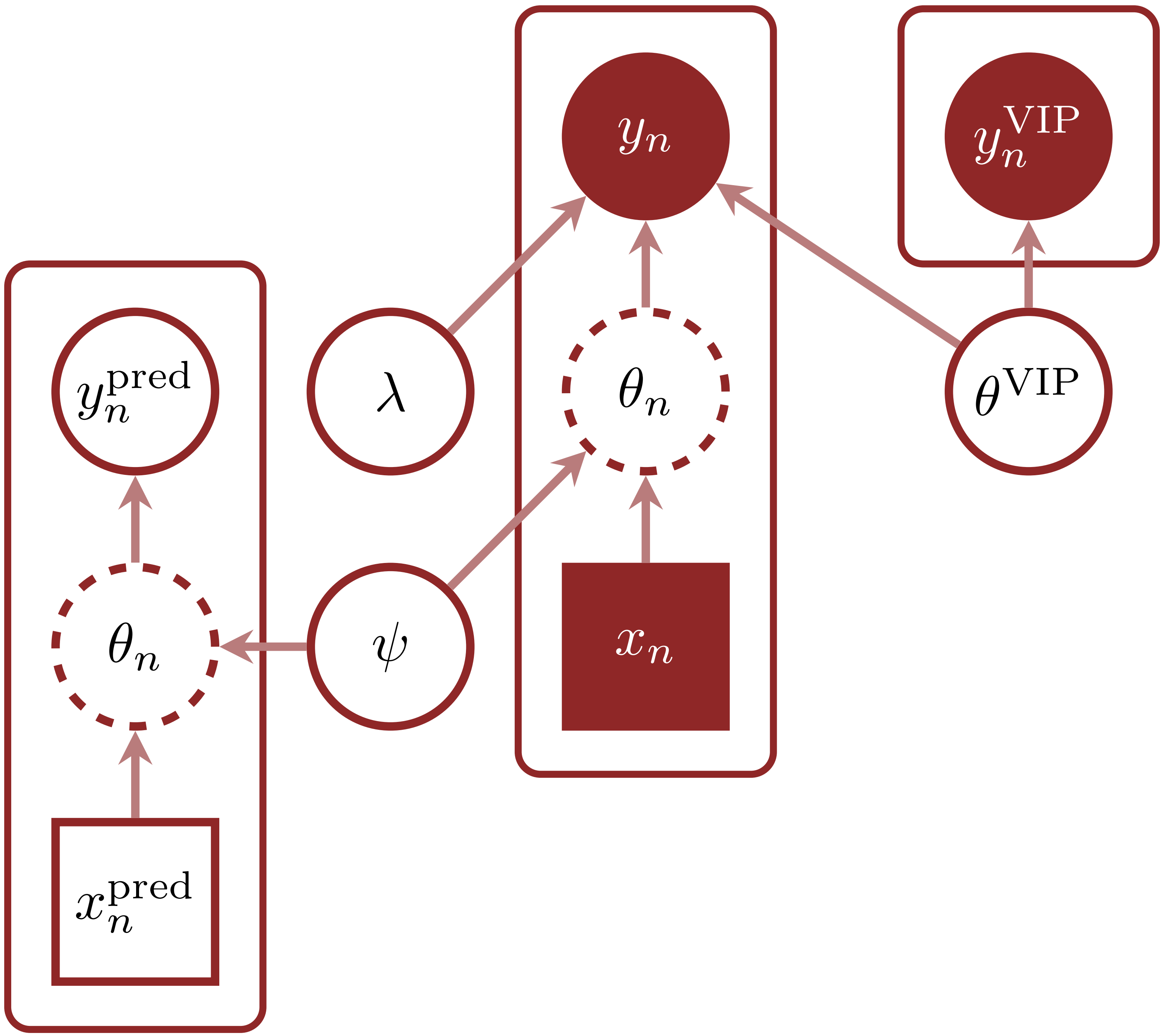


$$
\begin{aligned}
&\pi(y_1, \ldots, y_N, y_1^{\mathrm{VIP}}, \ldots, y_{N^{\mathrm{VIP}}}^{\mathrm{VIP}}, \psi, \theta^{\mathrm{VIP}}, \lambda; x_1, \ldots, x_N) \\
&\quad = \left[\prod_{n=1}^{N} \quad \lambda \cdot \mathrm{Bernoulli}(y_n \mid \theta(x_n, \psi)) \right. \\
&\qquad\qquad \left. + (1 - \lambda) \cdot \mathrm{Bernoulli}(y_n \mid \theta^{\mathrm{VIP}}) \right] \\
&\qquad \cdot \left[\prod_{n=1}^{N^{\mathrm{VIP}}} \mathrm{Bernoulli}(y_n^{\mathrm{VIP}} \mid \theta^{\mathrm{VIP}}) \right] \\
&\qquad \cdot \pi(\psi) \, \pi(\theta^{\mathrm{VIP}}) \, \pi(\lambda)
\end{aligned}
$$

Once we have isolated the desired behavior we can use it to inform predictions and decisions.



Once we have learned the desired behavior we can use it to inform any predictions and decisions of interest.



Once we have learned the desired behavior we can use it to inform any predictions and decisions of interest.