

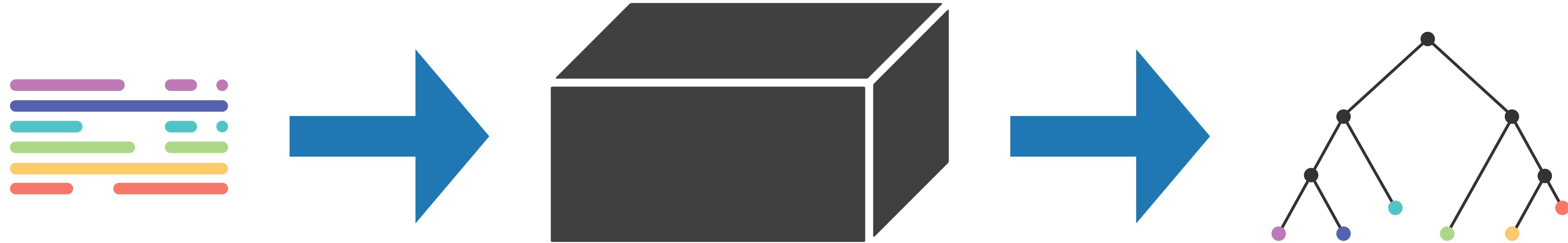
Troubleshooting



JŪLIJA PEČERSKA
SQUAMISH, 2023

FEW REMINDERS

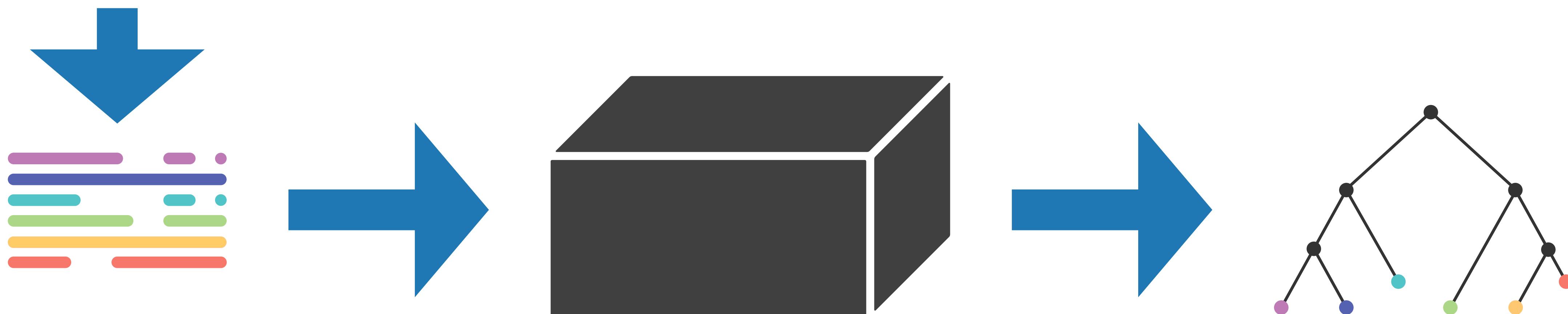
BAYESIAN PHYLOGENETICS IS NOT MAGIC



FEW REMINDERS

BAYESIAN PHYLOGENETICS IS NOT MAGIC

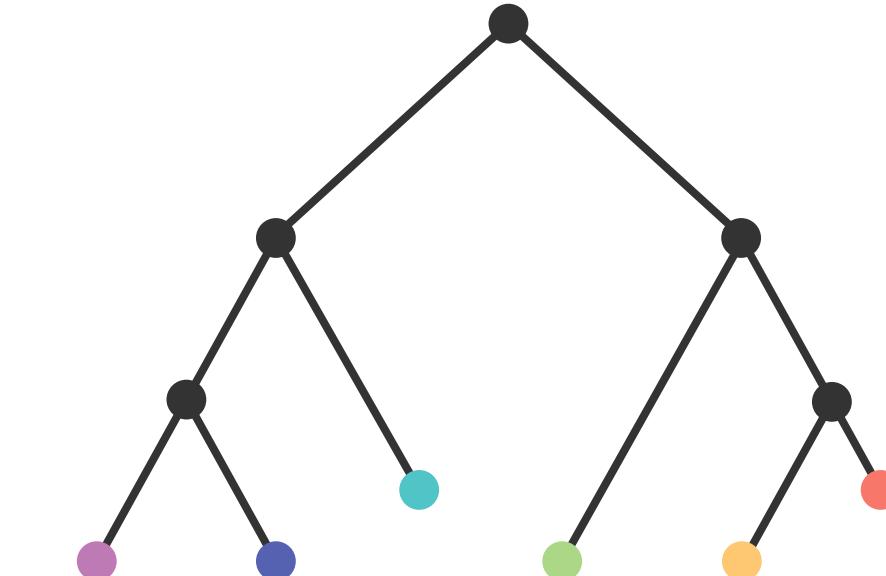
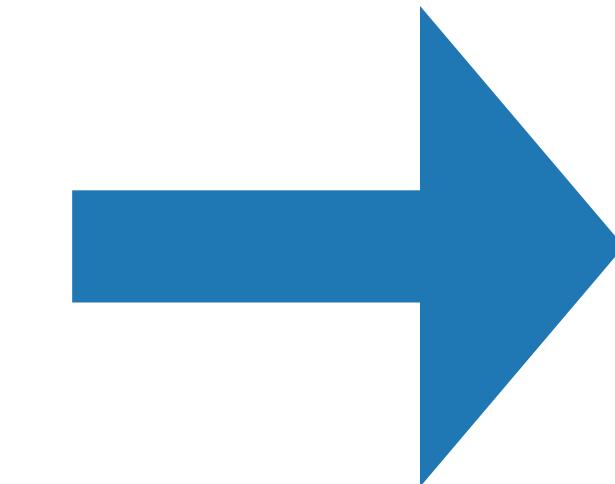
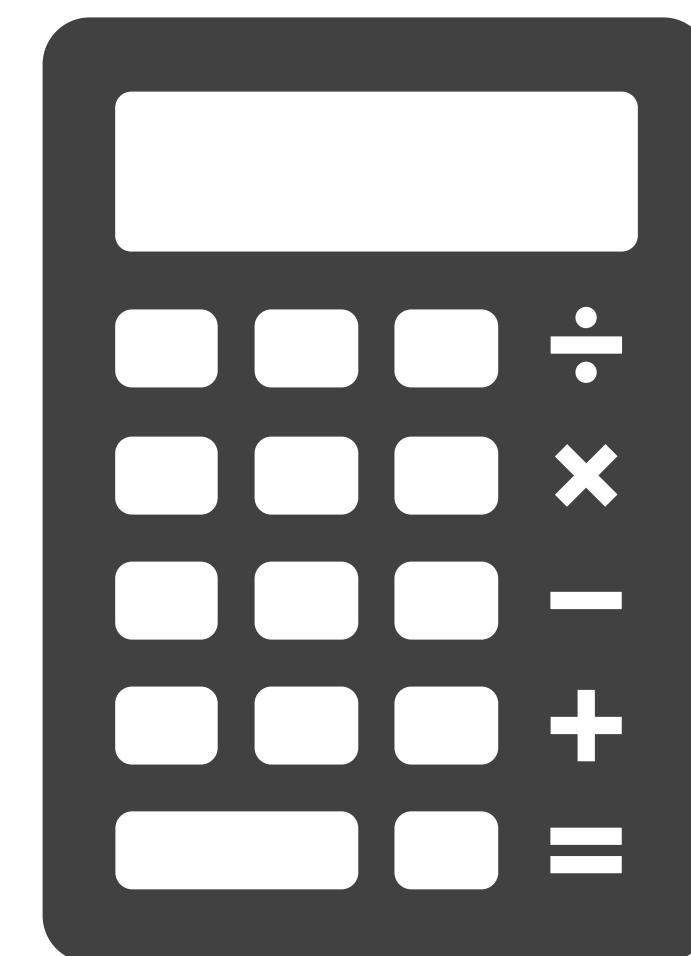
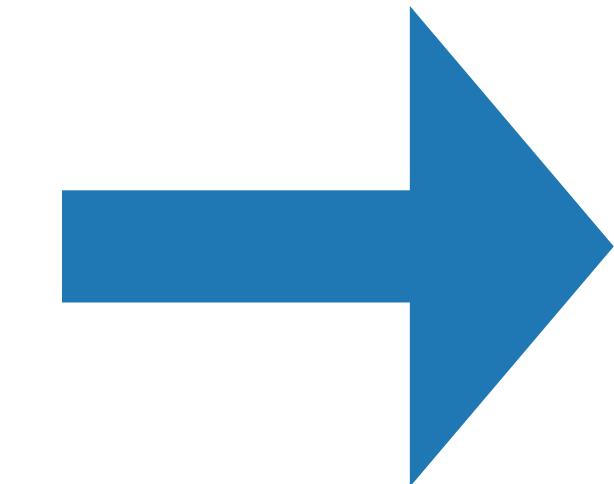
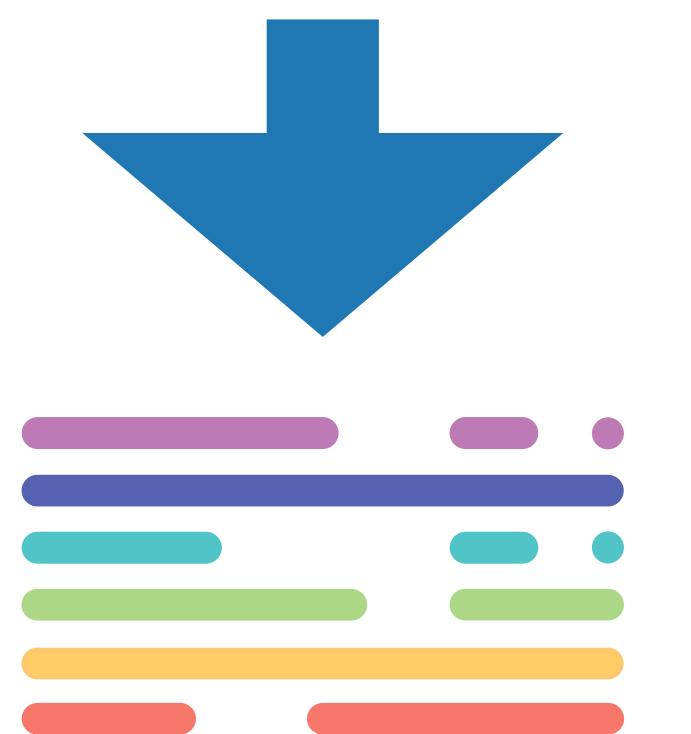
Data gathering and preprocessing



FEW REMINDERS

BAYESIAN PHYLOGENETICS IS NOT MAGIC

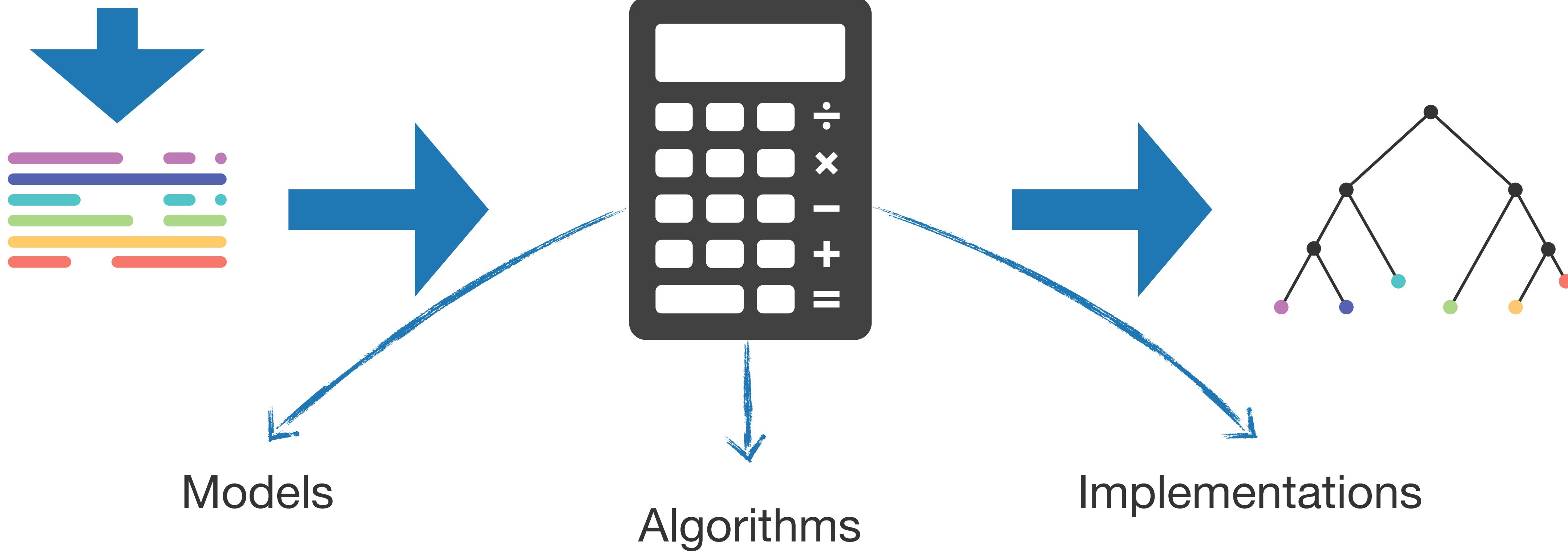
Data gathering and
preprocessing



FEW REMINDERS

BAYESIAN PHYLOGENETICS IS NOT MAGIC

Data gathering and preprocessing



REMINDER #1

MODELS & THE BAYES THEOREM

$$P(E \circ O \circ \text{grid} \circ O | \equiv) = \frac{P(\equiv | E \circ O \circ \text{grid} \circ O) P(E | \circ O) P(\circ | O) P(O)}{P(\equiv)}$$

*Figure adapted from du Plessis and Stadler, 2015,
Getting to the root of epidemic spread with phylodynamic analysis of genomic data*

- Modelling choices are important:
 - Models define the processes we want to describe (kinda);
 - Priors define parameter spaces;
 - Models and priors are intertwined.

REMINDER #2

ALGORITHMS & BAYESIAN MCMC

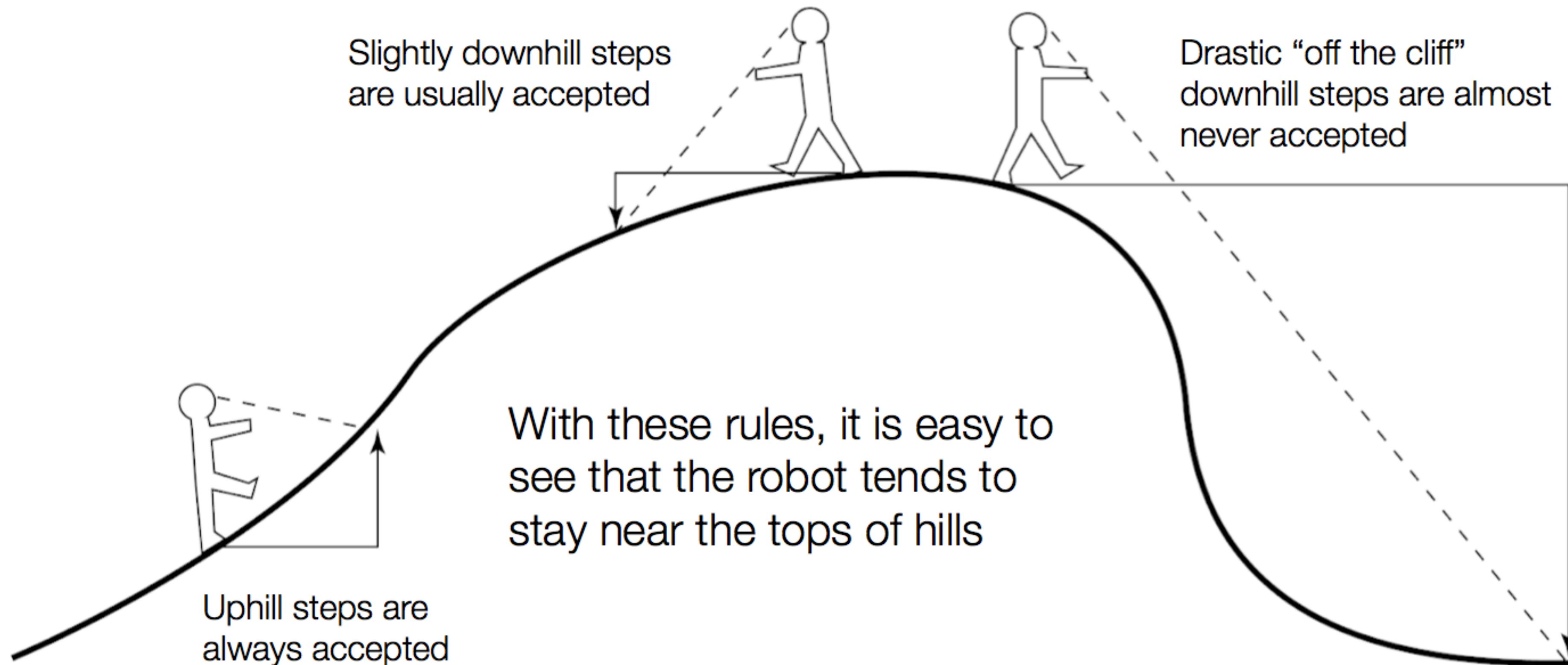


Figure courtesy of Paul O. Lewis

REMINDER #3

IMPLEMENTATIONS & RNG

- True randomness:



REMINDER #3

IMPLEMENTATIONS & RNG

- True randomness:



What's a seed?



REMINDER #3

IMPLEMENTATIONS & RNG

- True randomness:



- Pseudo randomness:

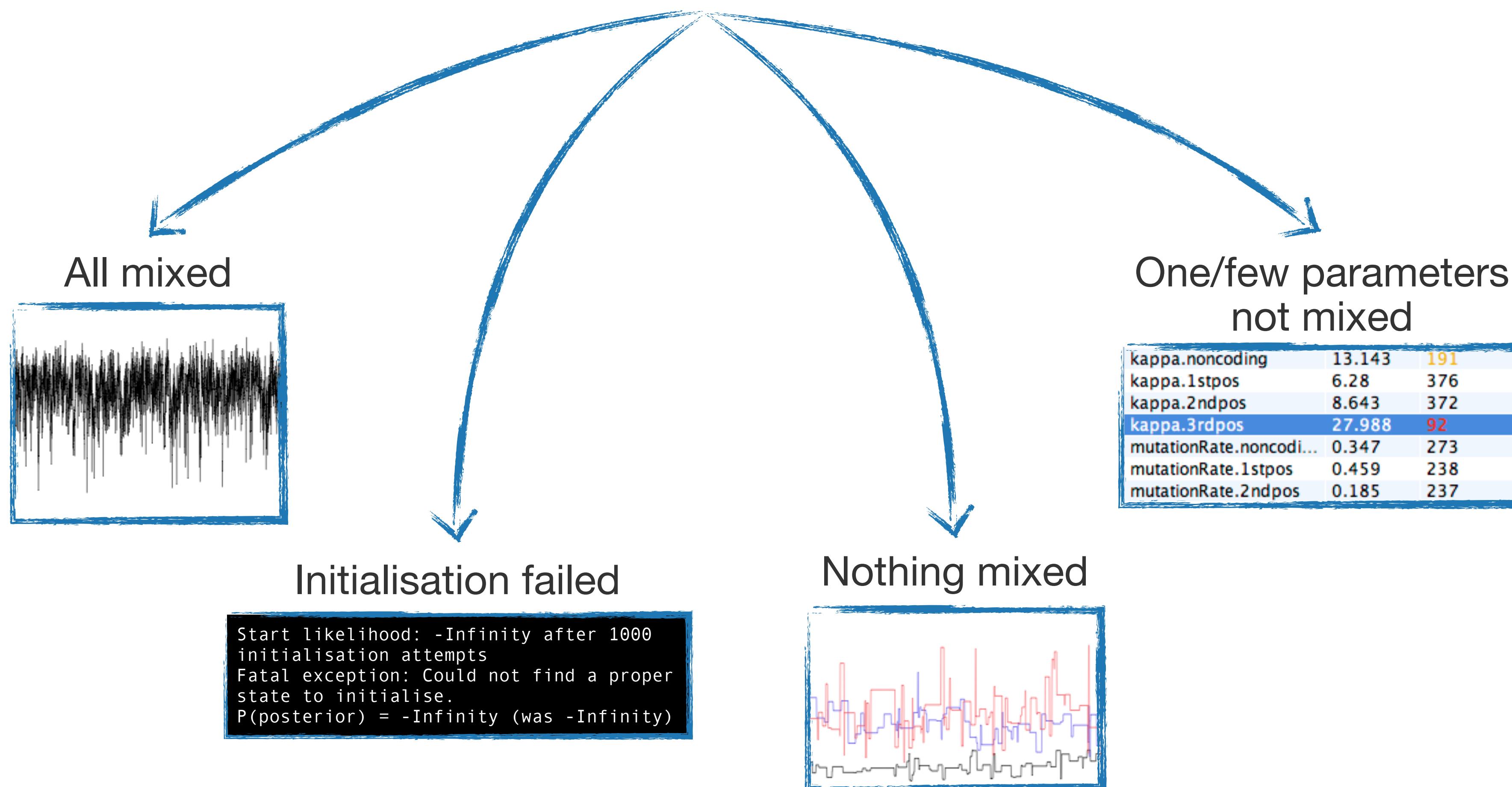
$$X_{n+1} = (aX_n + c) \% m$$

- X_0 is the seed for the PRNG.

Changing the seed is **NOT** going to help.

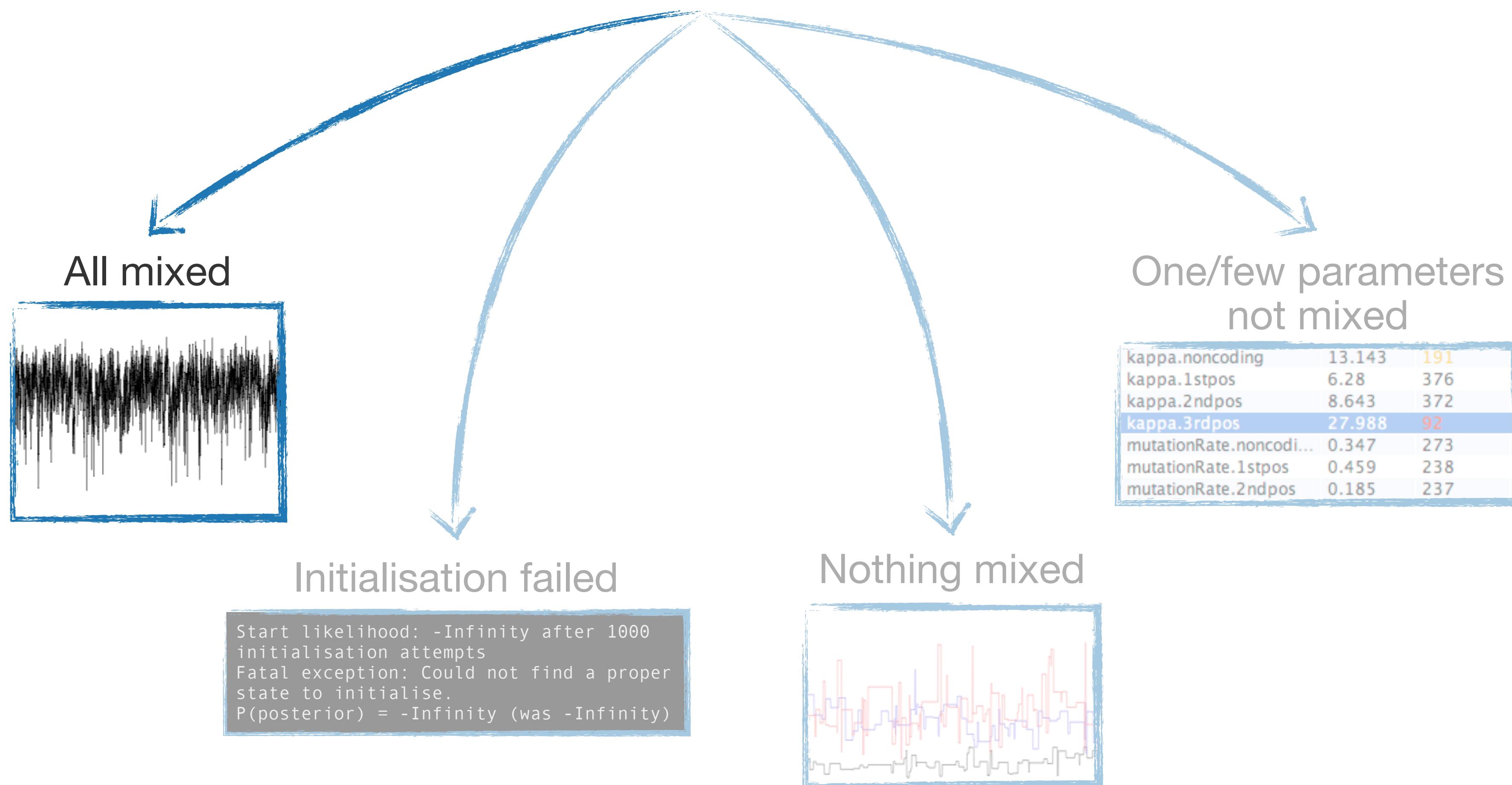
POSSIBLE SCENARIOS

THE MANY FATES OF A BEAST2 ANALYSIS



POSSIBLE SCENARIOS

THE MANY FATES OF A BEAST2 ANALYSIS



EVERYTHING MIXED

TIME TO CELEBRATE?

Good news:

Bayesian analysis always gives an answer!

Bad news:

We need to figure out if it is a meaningful one.



EVERYTHING MIXED BUT DOES IT MAKE SENSE?

- Sample from the prior:
 - Is your prior what you think it is?
- Plot posteriors and priors together:
 - Did we learn anything from the data?
- Perturb priors:
 - Are your results robust to the choices of priors?

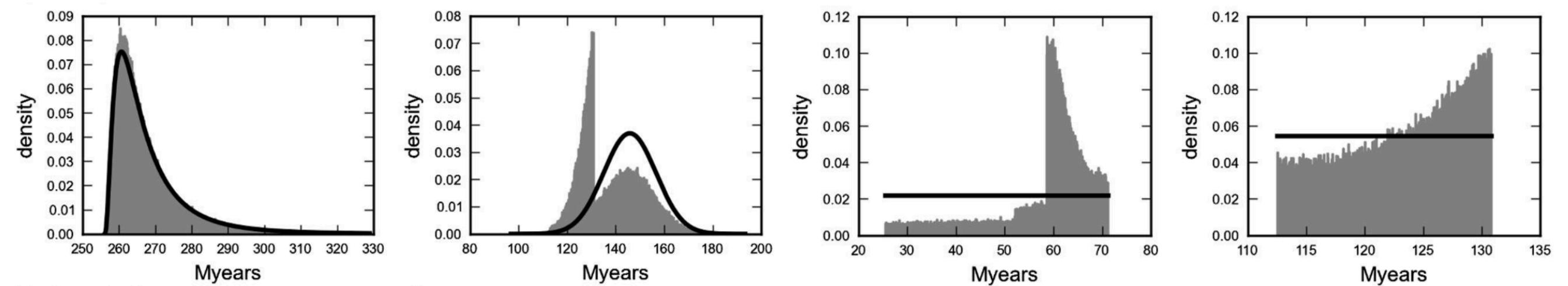
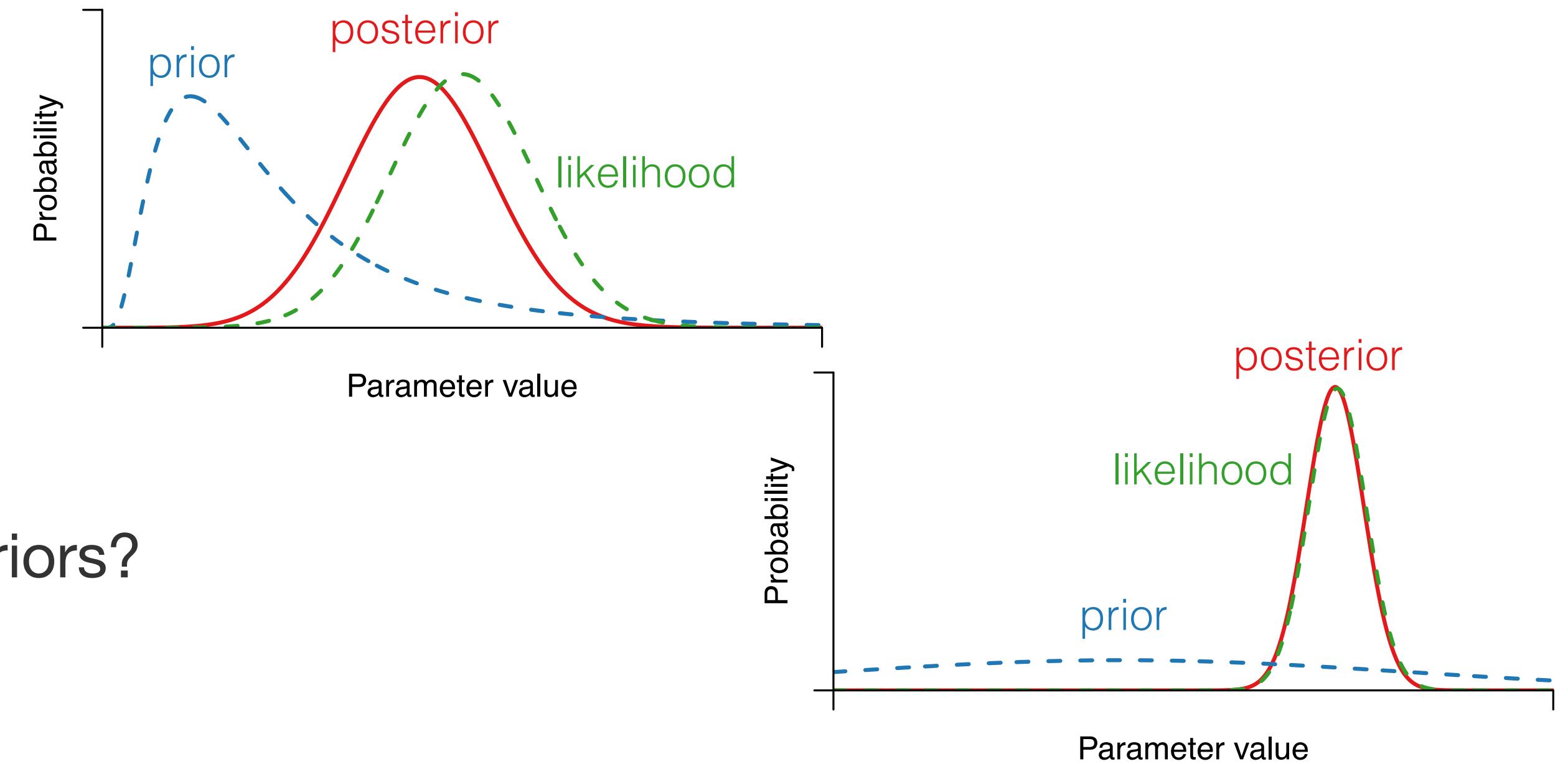


Figure adapted from Heled and Drummond 2012



EVERYTHING MIXED AND MAKES SENSE, WHAT NOW?

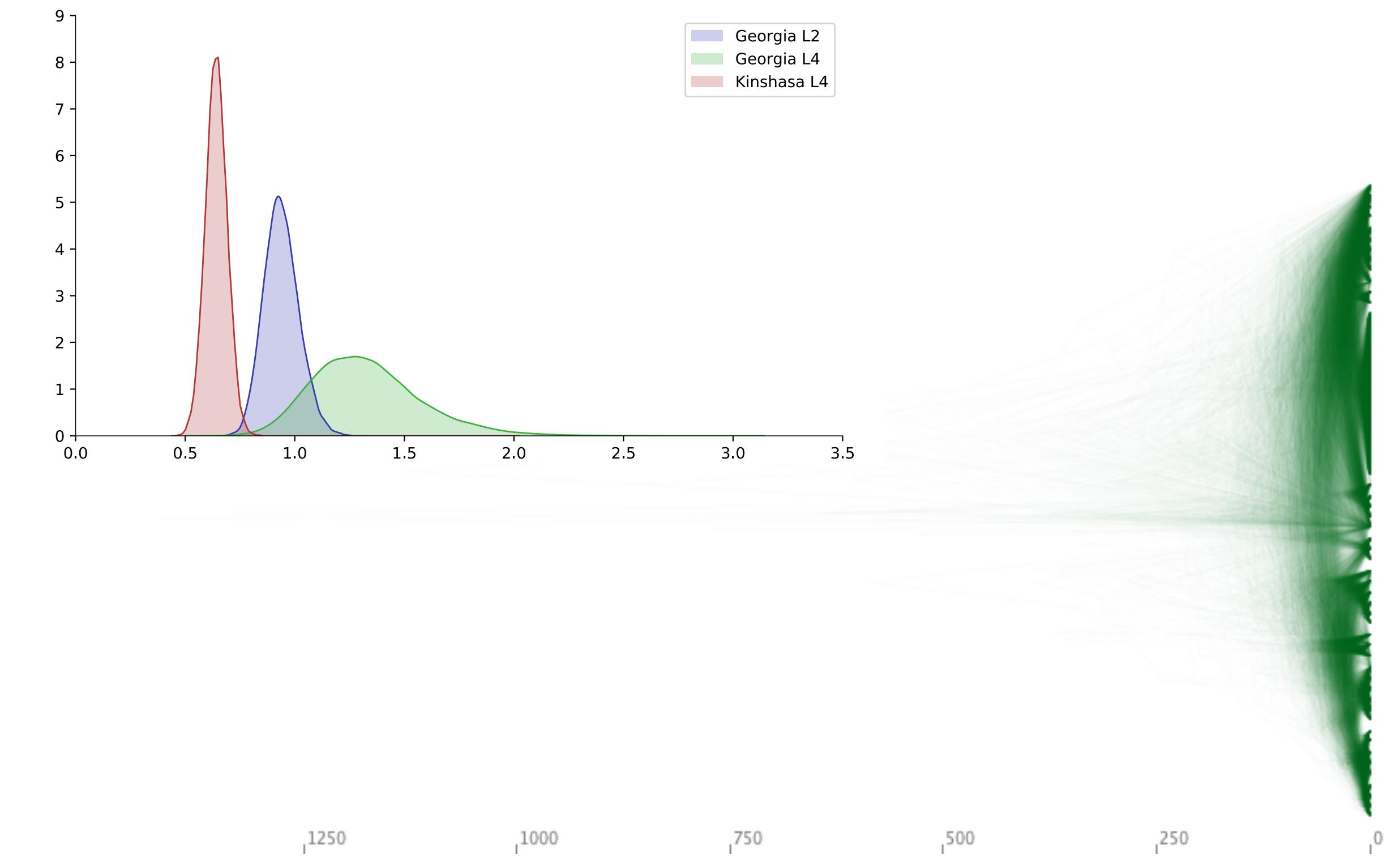
We want certainty, but a Bayesian answer is how uncertain we are.

Collect more data?

Collect a different type of data?

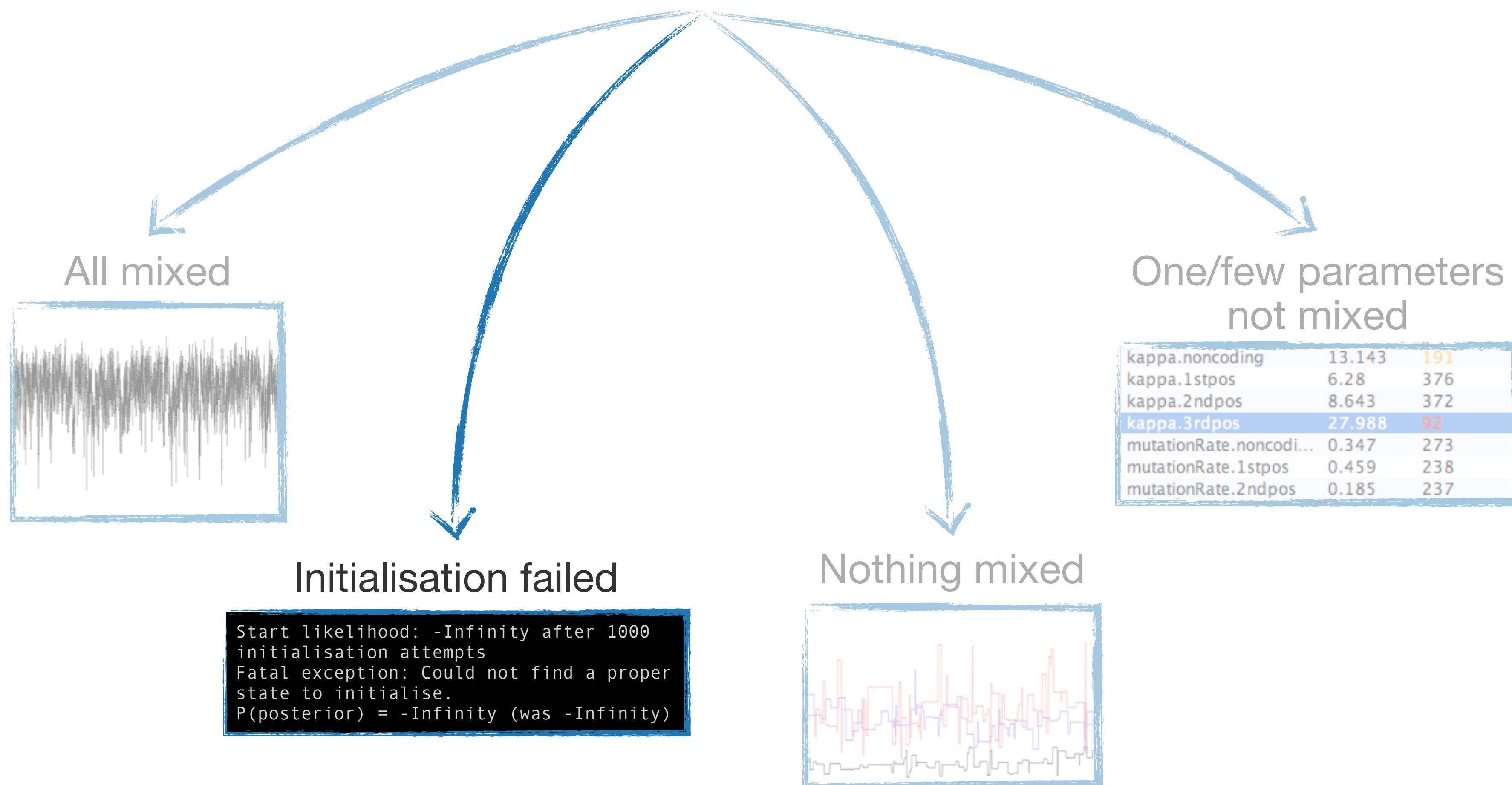
Embrace the uncertainty;

Report the uncertainty!



POSSIBLE SCENARIOS

THE MANY FATES OF A BEAST2 ANALYSIS



INITIALISATION FAILED

```
Start likelihood: -Infinity after 1000 initialisation attempts
Fatal exception: Could not find a proper state to initialise. Perhaps try another seed.

P(posterior) = -Infinity (was -Infinity)
P(prior) = -Infinity (was -Infinity)
    P(BDMM) = -Infinity (was -Infinity)
    P(R0Prior) = -0.5586849541070393 (was -0.5586849541070393)
    P(rPrior) = -11.46042136866474 (was -11.46042136866474)
    P(rateMatrixPrior) = -0.14088025499381485 (was -0.14088025499381485)
    P(samplingProportionPrior) = -10.049507225748343 (was -10.049507225748343)
    P(becomeUninfectiousRatePrior) = -0.7811241751317991 (was -0.7811241751317991)
java.lang.RuntimeException: Could not find a proper state to initialise. Perhaps try another seed.

        at beast.core.MCMC.run(Unknown Source)
        at beast.app.BeastMCMC.run(Unknown Source)
        at beast.app.beastapp.BeastMain.<init>(Unknown Source)
        at beast.app.beastapp.BeastMain.main(Unknown Source)
        at beast.app.beastapp.BeastLauncher.main(Unknown Source)

Fatal exception: Could not find a proper state to initialise. Perhaps try another seed.

BEAST has terminated with an error. Please select QUIT from the menu.
```

PARAMETER PRIOR IS -INFINITY

Example: `P(rateMatrixPrior) = -Infinity (was -Infinity)`

Possible solutions:

$$\log(P) = \text{-infinity} \quad \text{---} \quad P = 0$$

- ✗ Change seed;
- ✓ Adjust initial conditions;
- ✓ Check for incompatible priors;
- ✓ Check for underflow (too low values);
- ✓ Increase initialisation attempt number;
- ✓ Talk to BEAST2 creators or tammers.

MODEL PRIOR IS -INFINITY

Example: $P(\text{BDMM}) = -\text{Infinity}$ (was $-\text{Infinity}$)

Possible solutions:

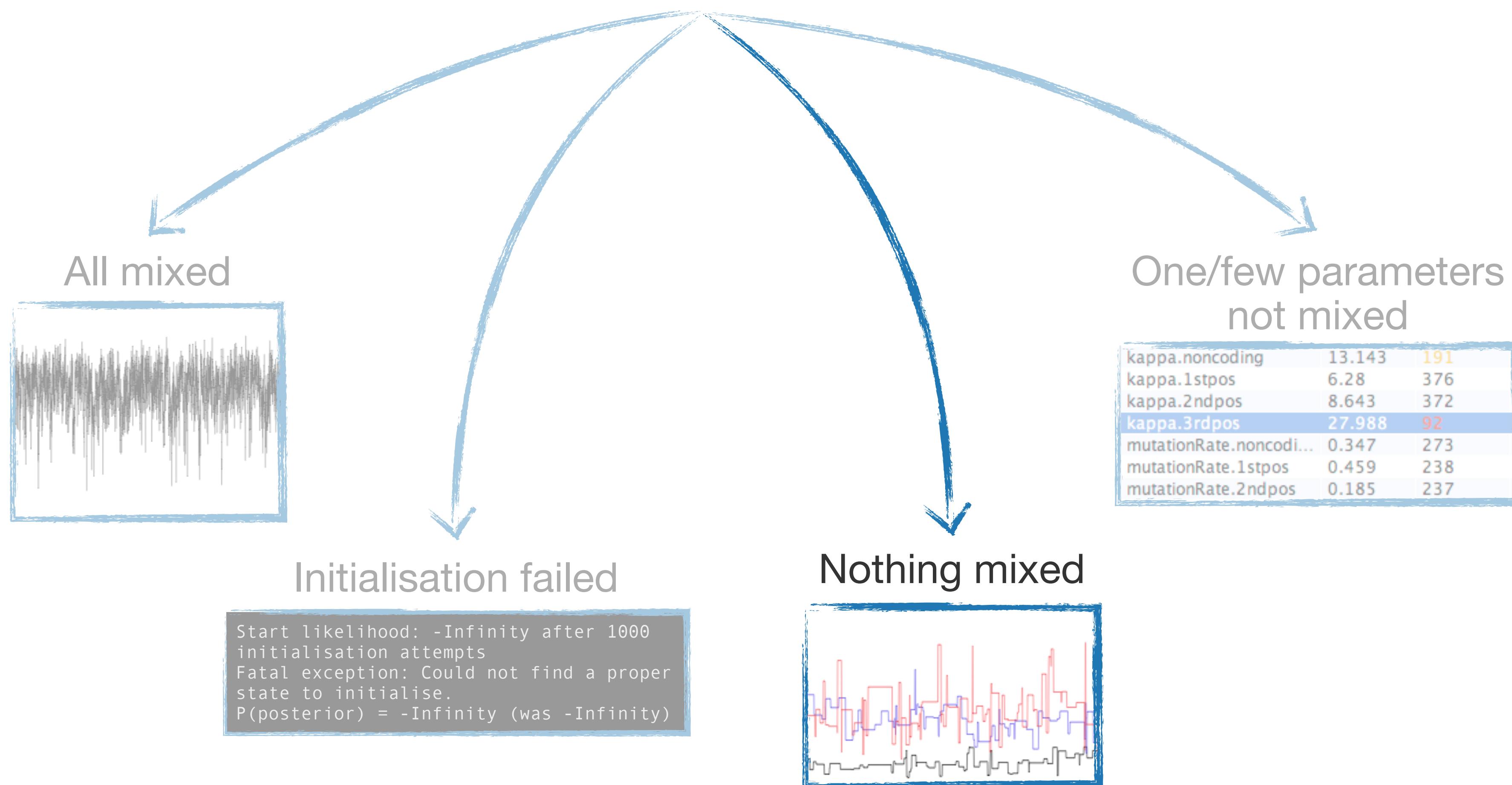
- ✗ Change seed;
- ✓ Check model:
 - Model cannot handle that much data;
 - Data incompatible with the process you're trying to impose on it.
- ✓ Talk to BEAST2 creators or tammers.

OTHER INITIALISATION ERRORS

- The BEAUTi and the BEAST versions do not match
 - Always check that you're using matching versions
- A package the run needs is missing
 - Install the missing packages through BEAUTi
- XML error
 - Run BEAST with -validate to validate the XML
- And others!

POSSIBLE SCENARIOS

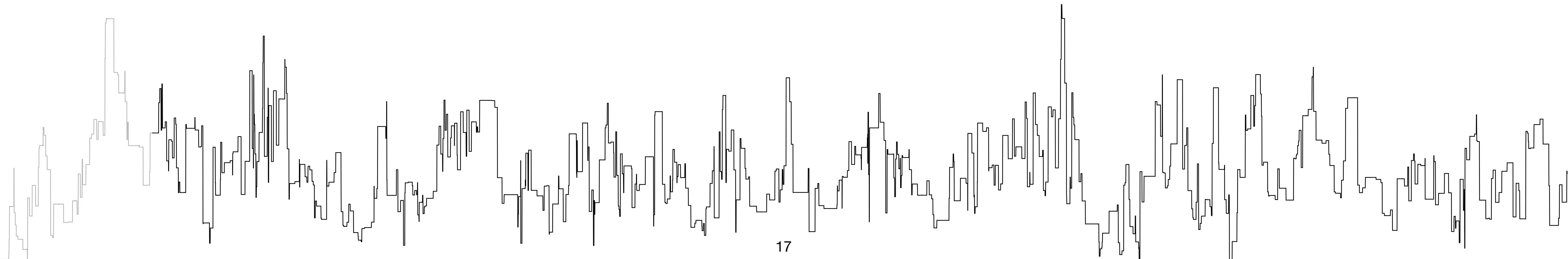
THE MANY FATES OF A BEAST2 ANALYSIS



NOTHING MIXED

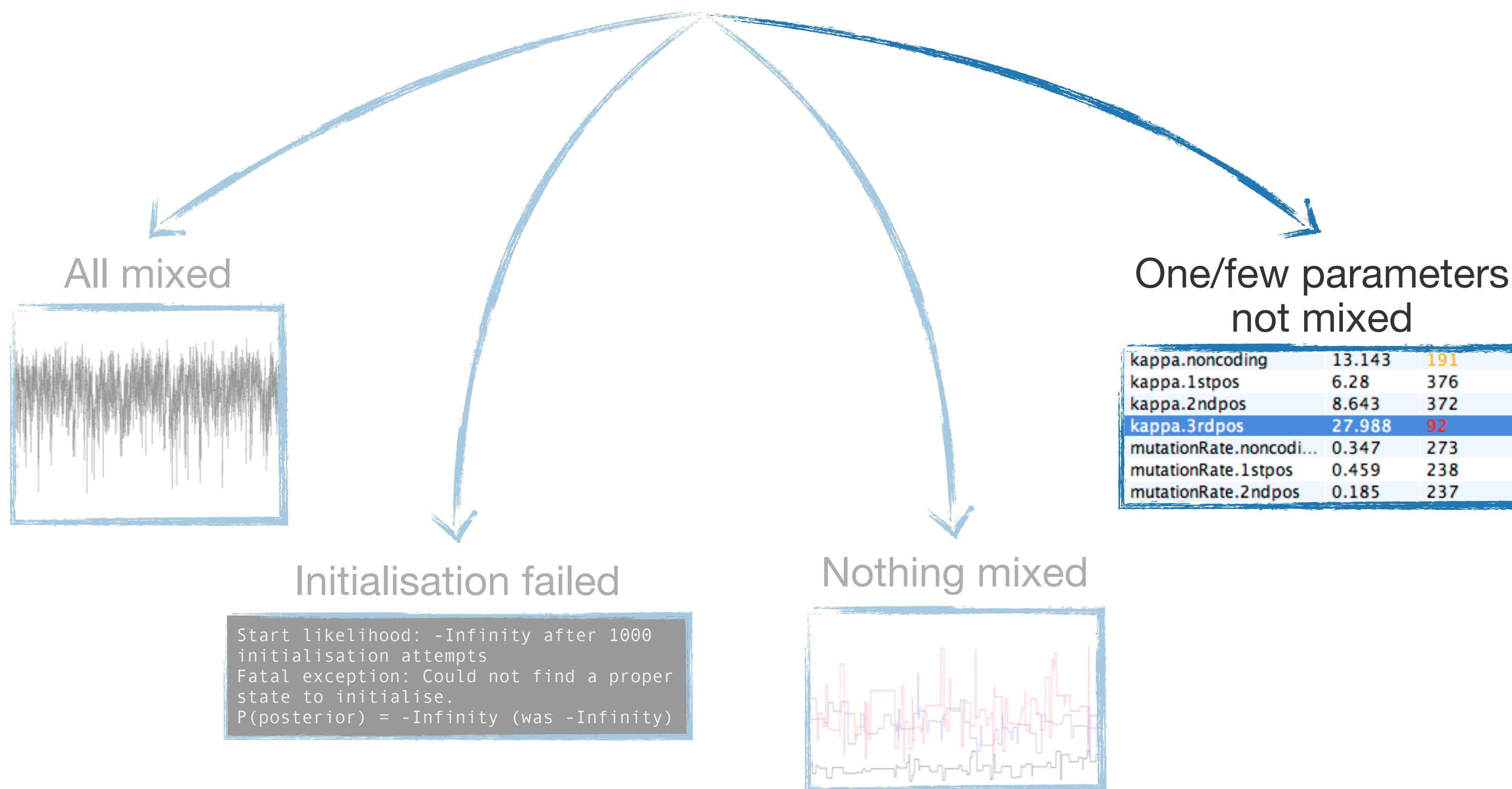
Possible solutions:

- ✓ Increase chain length;
- ✓ Run multiple independent chains;
- ✓ Increase sampling frequency (if ACT permits);
- ✓ Check for parameter identifiability;
- ✓ Check for gross model misspecification.



POSSIBLE SCENARIOS

THE MANY FATES OF A BEAST2 ANALYSIS



A NUMERICAL PARAMETER DIDN'T MIX

Possible solutions:

- ✓ Increase chain length;
- ✓ Tweak the operator weights:
 - ✓ Increase weight for low ESS parameters;
 - ✓ Use updown operator for correlated parameters;
- ✓ Run longer (or combine several independent chains).

BONUS COMPLEXITY

TREE SPACE MIXING

Bad news:

At the moment we can not directly examine the ESS of the trees;

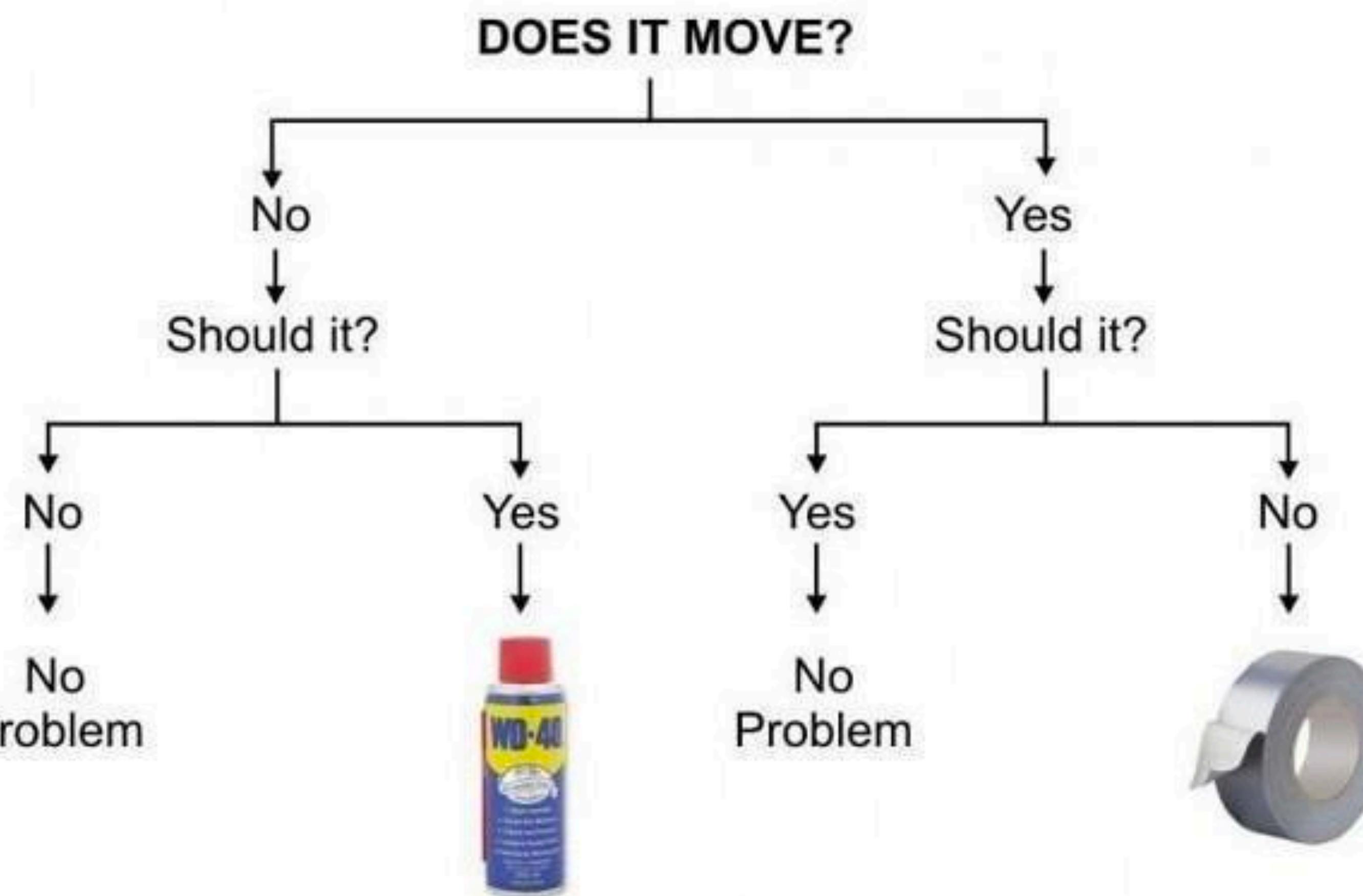
Good news:

Good mixing of the continuous parameters and likelihoods is indicative;

Can also use AWTY to see if the clade probabilities stabilised.

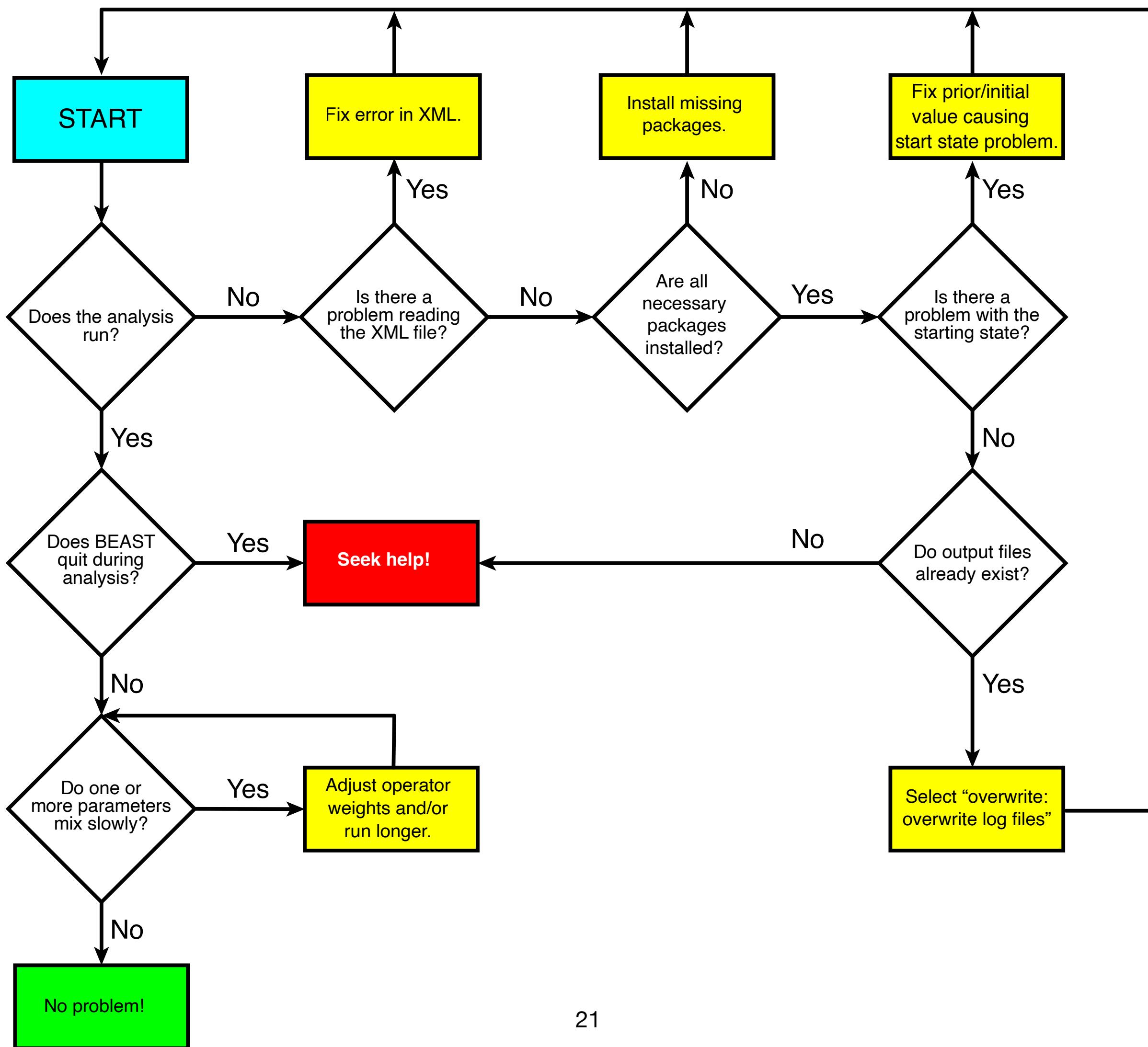
GENERAL PROBLEM-SOLVING FLOWCHART

Engineering Flowchart



GENERAL PROBLEM-SOLVING FLOWCHART

COURTESY OF LOUIS DU PLESSIS



TAKEAWAYS

- Know thy models and priors;
 - Know thy software and its assumptions;
 - Admit to being an expert on your work.
-
- Now let's tame some BEASTs!

