Bayesian calibration of differentiable simulators

Arnau Quera-Bofarull, Joel Dyer

Anisoara Calinescu, Michael Wooldridge

BioInference conference — Oxford, June 2023

www.arnau.ai

Differentiable Simulators

$$f : \mathbb{R}^m \longrightarrow \mathbb{R}^n$$

$$(J_f)_{ij} = \frac{\partial f_i}{\partial x_j}$$

Differentiation of Computer Programs

Numerical differentiation

$$\frac{\partial f}{\partial x} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

Inaccurate

Expensive $\mathcal{O}(m \times n)$

$$f : \mathbb{R}^m \longrightarrow \mathbb{R}^n$$

Differentiation of Computer Programs

Symbolic differentiation

$$f(x) = (2x + \sin(x))x^2(x+3)(x+5)$$

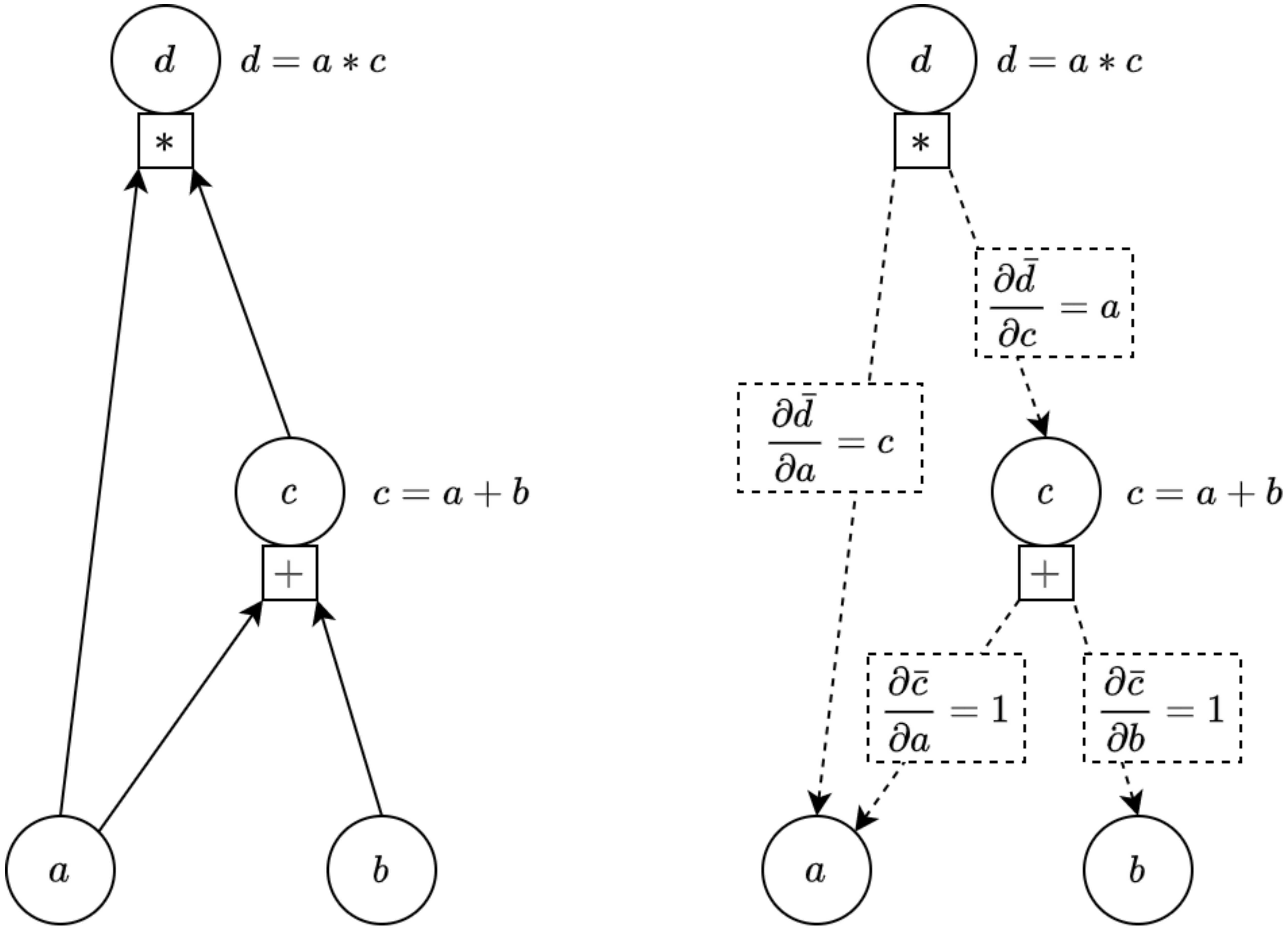
$$f'(x) = x(2x(45 + 32x + 5x^2) + x(15 + 8x + x^2)\cos(x) + (30 + 24x + 4x^2)\sin(x))$$

Exact

Expensive

Differentiation of Computer Programs

Automatic differentiation (AD)



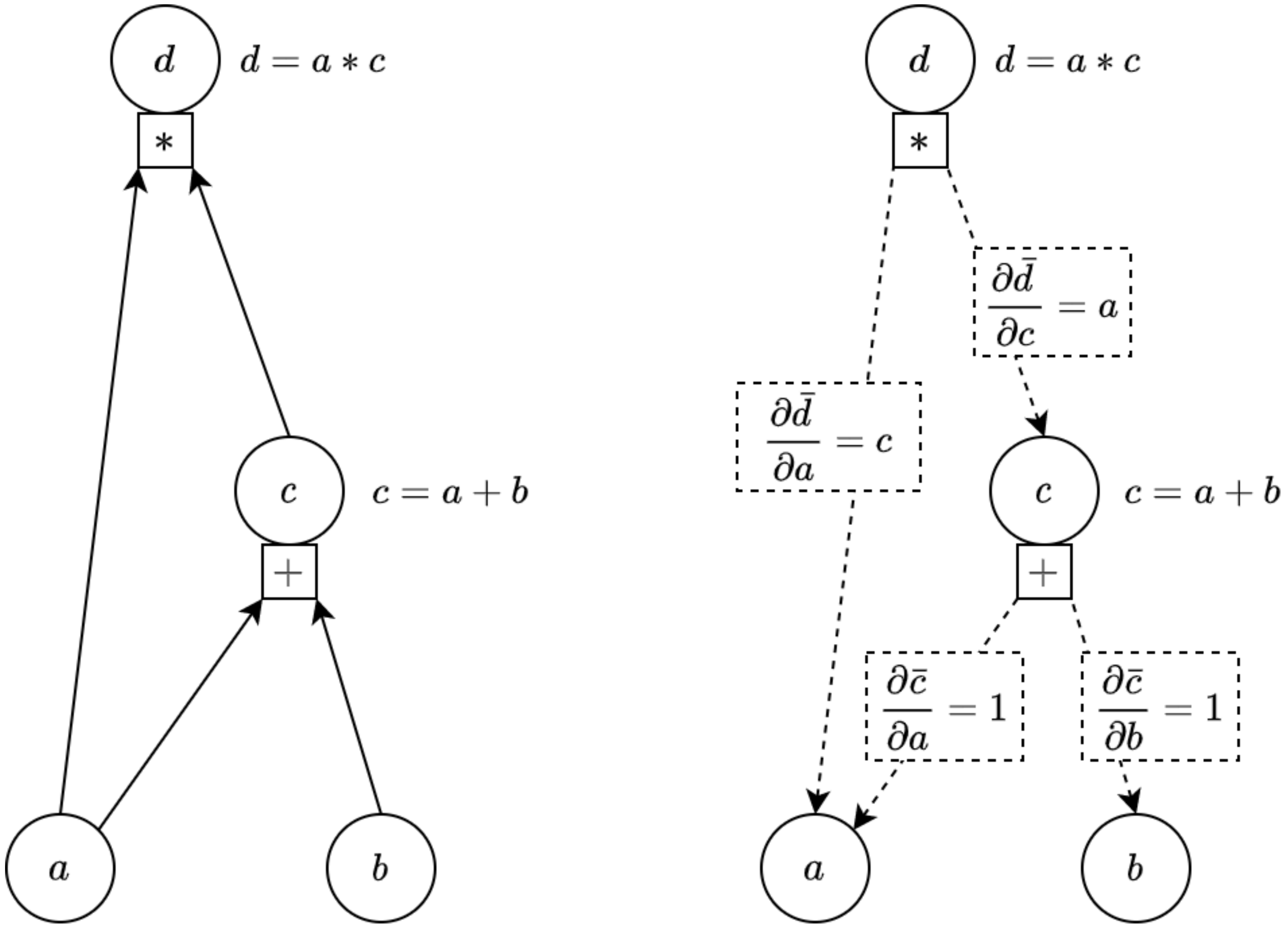
Exact

Fast $\mathcal{O}(\min(m, n))$

sidsite.com

$$f : \mathbb{R}^m \longrightarrow \mathbb{R}^n$$

Forward vs Reverse mode AD

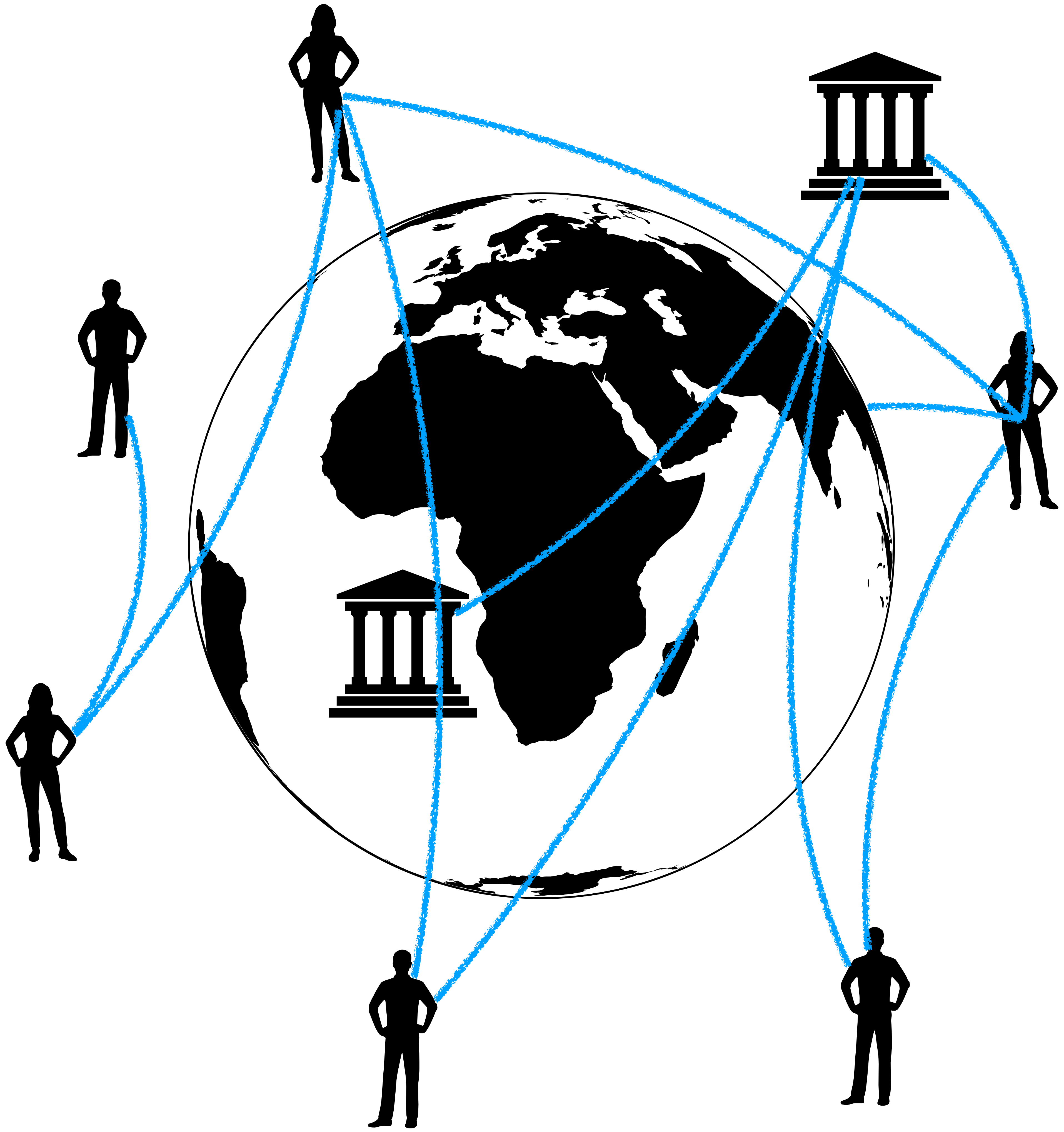


sidsite.com

$$f : \mathbb{R}^m \longrightarrow \mathbb{R}^n$$

Forward ~ m
Reverse ~ n

Agent-Based Models



- Expensive to calibrate
- Difficult to interpret / validate

Differentiable Agent-Based Models

Can ABMs be made differentiable?

Stochasticity?

Con

Yes, they can

ory?

Meaningful gradients?

Case study: the JUNE epidemiological model

- JUNE is a 1:1 epi model of England (56 million agents)

- GradABM-JUNE is its differentiable implementation (PyTorch).

	Simulation
JUNE	50 hours
GradABM-June (GPU)	5 seconds

Tensorisation enables scalability to millions (billions?) of agents

References:

arnau.ai/talks

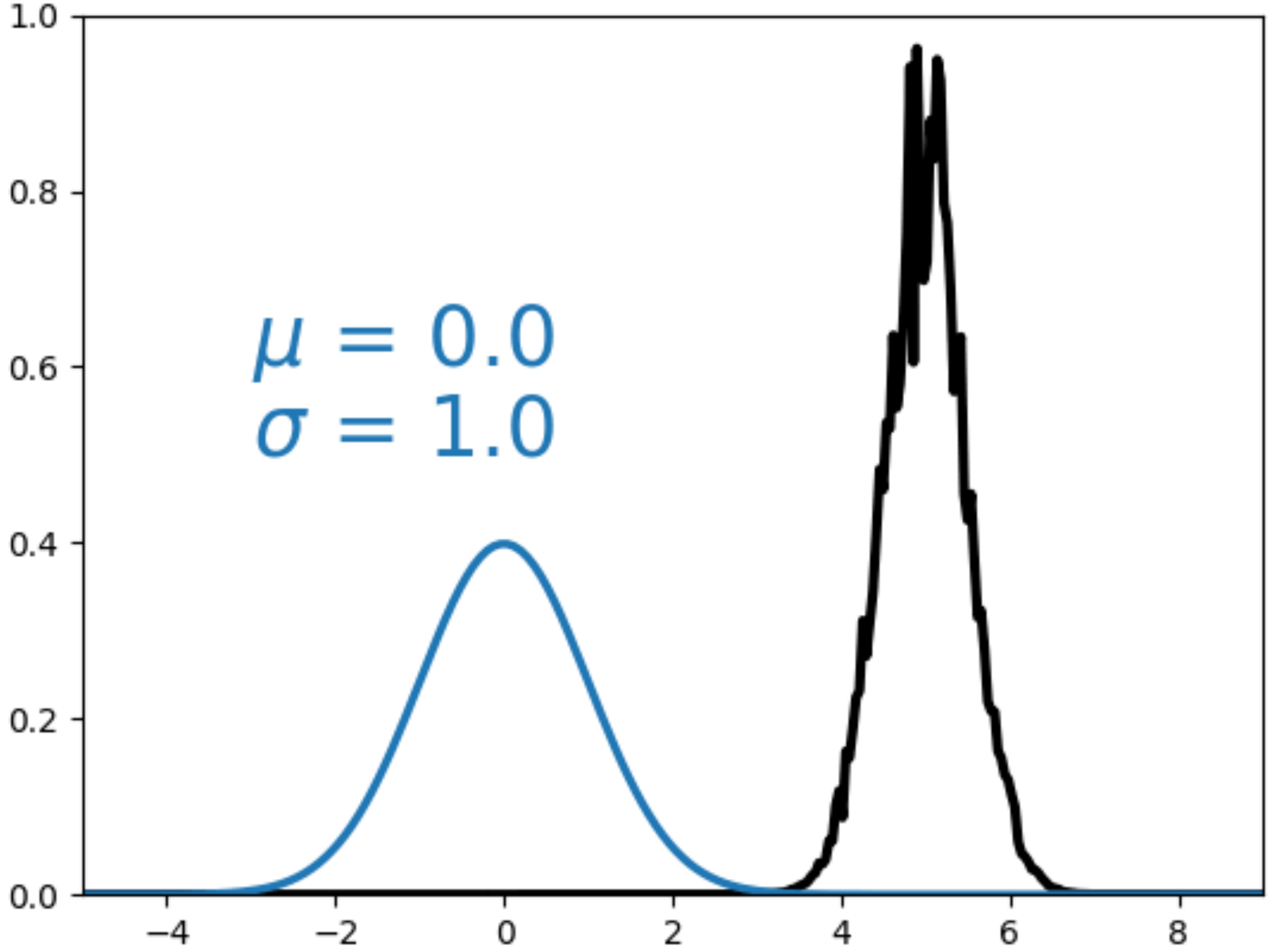
Case study: the JUNE epidemiological model

- We can use generalized variational inference for calibration

	Simulation	Calibration (No UQ)	Bayesian Calibration
JUNE	50 hours	-	100k hours
GradABM-June (GPU)	5 seconds	20 minutes	8 hours

Differentiability enables fast and accurate model calibration

Variational Inference:
Bayesian inference as an optimisation problem

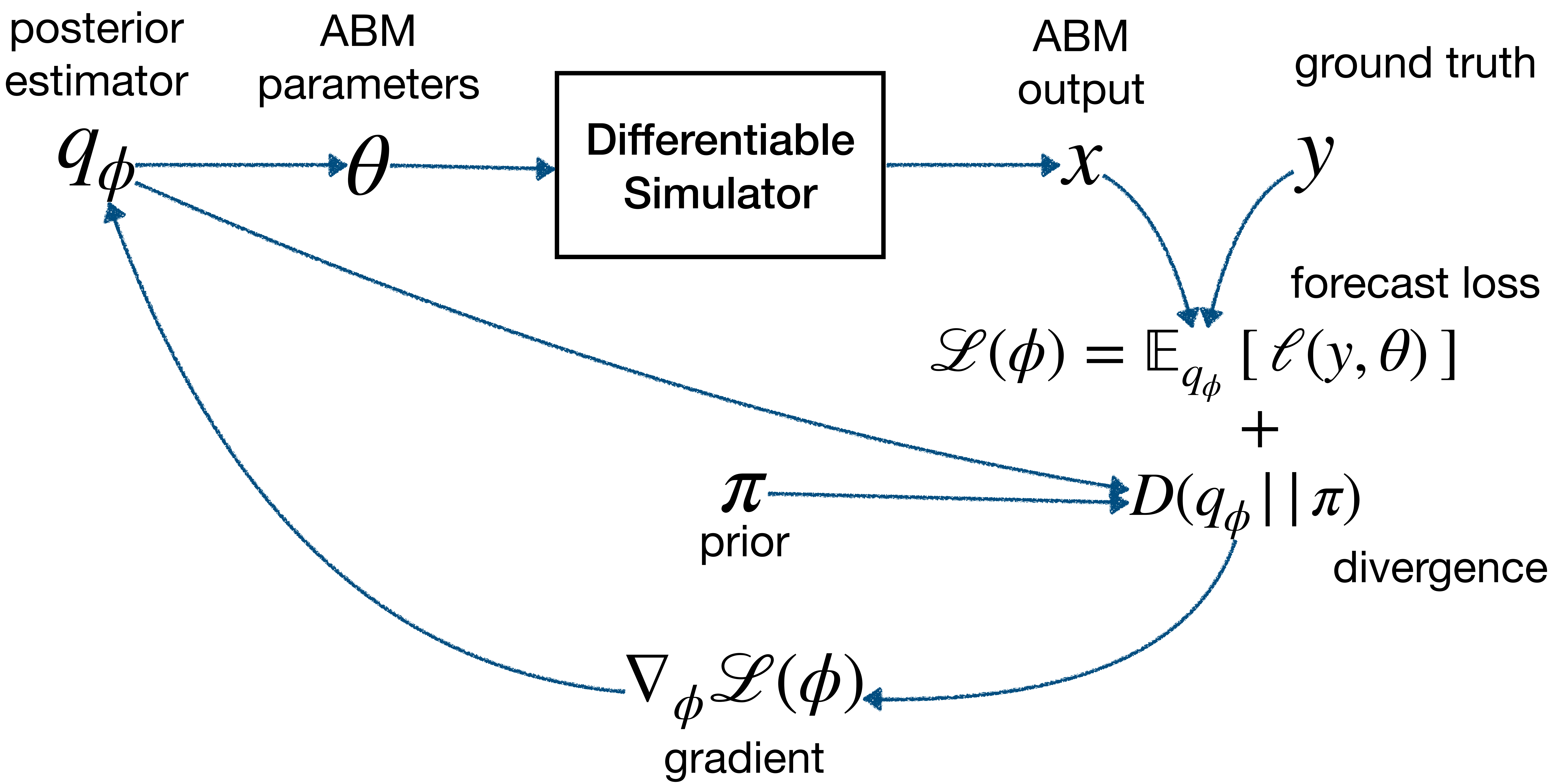


1. Assume posterior can be approximated by a family of distributions

2. Optimise for optimal parameters

Generalised Variational Inference

Knoblauch et al., (2022)



Gradients: path-wise vs score

- Gradient-assisted calibration algorithms need

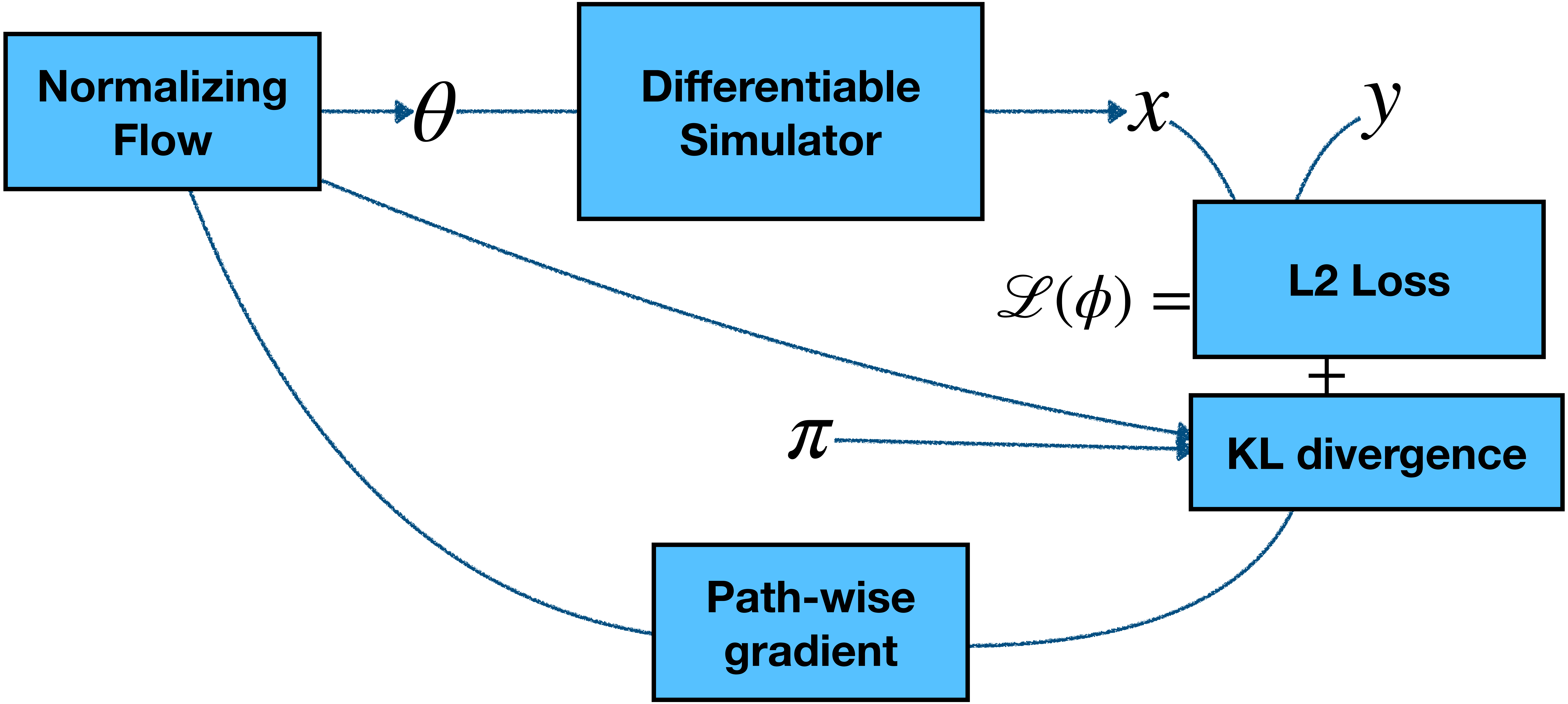
$$\nabla_{\phi}\underbrace{\mathbb{E}_{p(\theta\,;\,\phi)}}_{\text{posterior estimator}}\;\underbrace{\big[f(\theta)\big]}_{\text{simulator}}$$

- Two ways of obtaining the gradient:

1. Differentiating the measure (**score-based gradient**)
2. Differentiating the simulator (**path-wise gradient**)

Typically **path-wise gradient has (much) lower variance** (see Mohamed (2019))

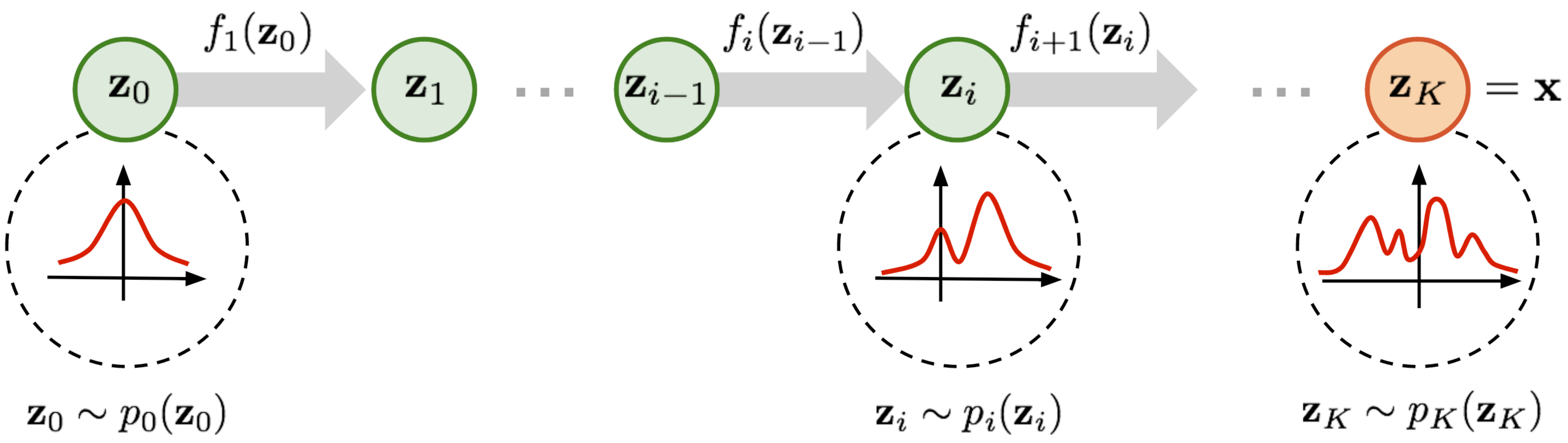
Generalised Variational Inference

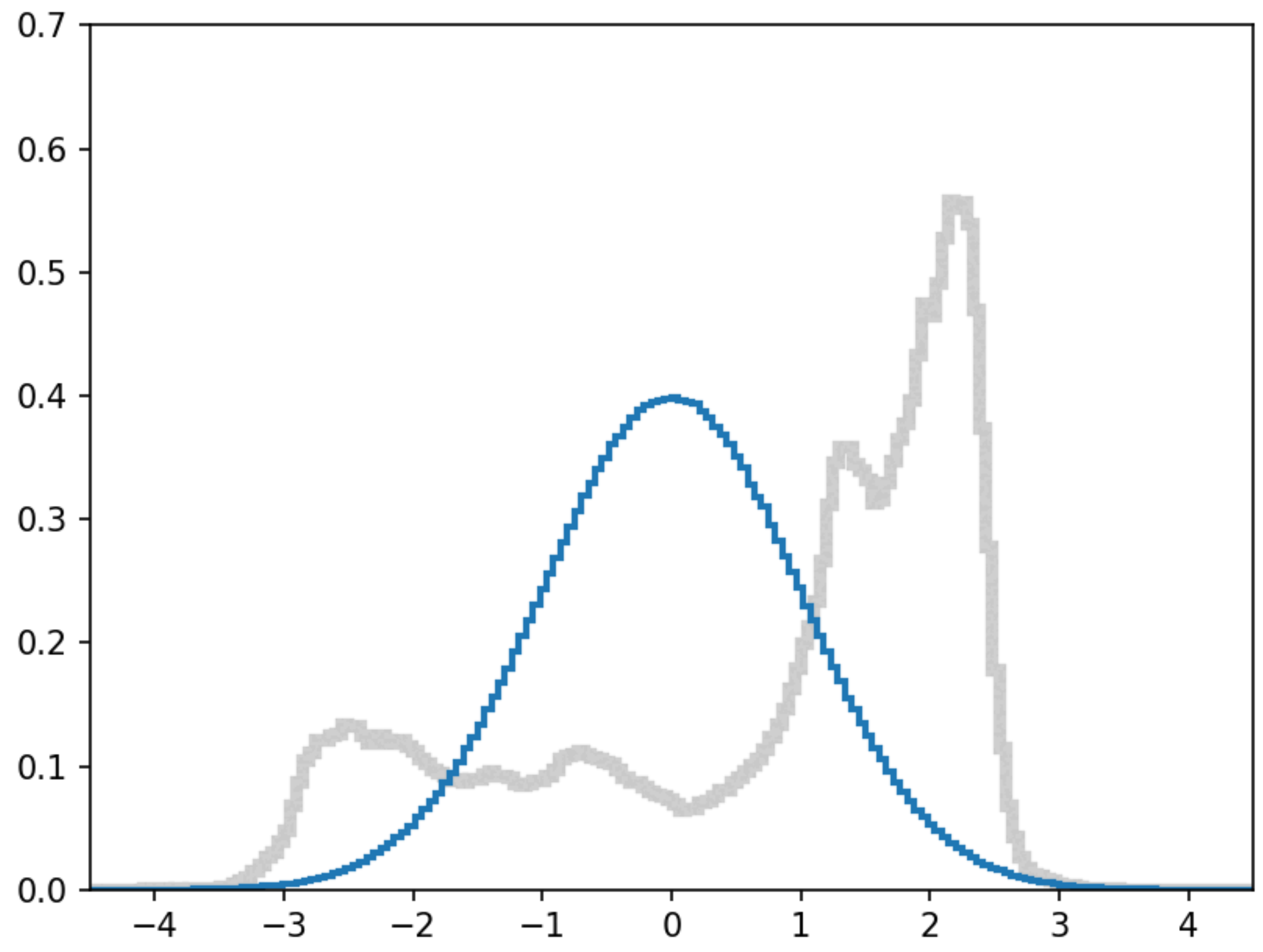


Normalizing Flows

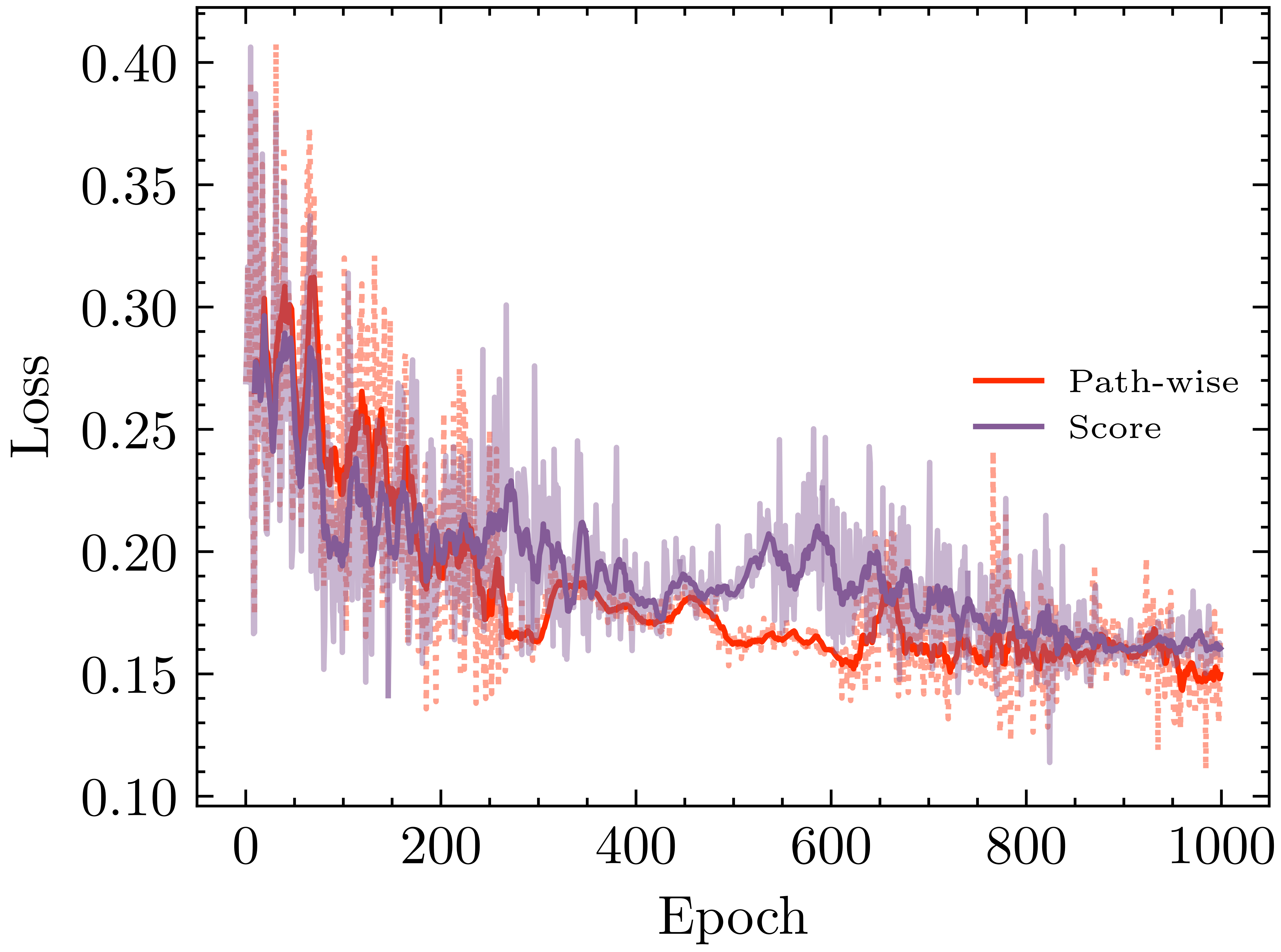
What do we choose for q?

Image credit: Lilian Weng





Gradient Horizon Problem



Model does not train!

Gradient Horizon Problem

Suppose we run an ABM with parameters θ for 4 time-steps

x_1	x_2	x_3	x_4

$$\frac{\mathrm{d}x_3}{\mathrm{d}\theta} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial \theta} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial \theta}$$

$$\frac{\mathrm{d}x_2}{\mathrm{d}\theta} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial \theta} + \frac{\partial f}{\partial \theta}$$

$$x_1 = f(\theta)$$

$$x_2 = f(x_1, \theta)$$

$$+ \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial x_1}\frac{\partial x_1}{\partial \theta}$$

$$x_3 = f(x_1\ x_2, \theta)$$

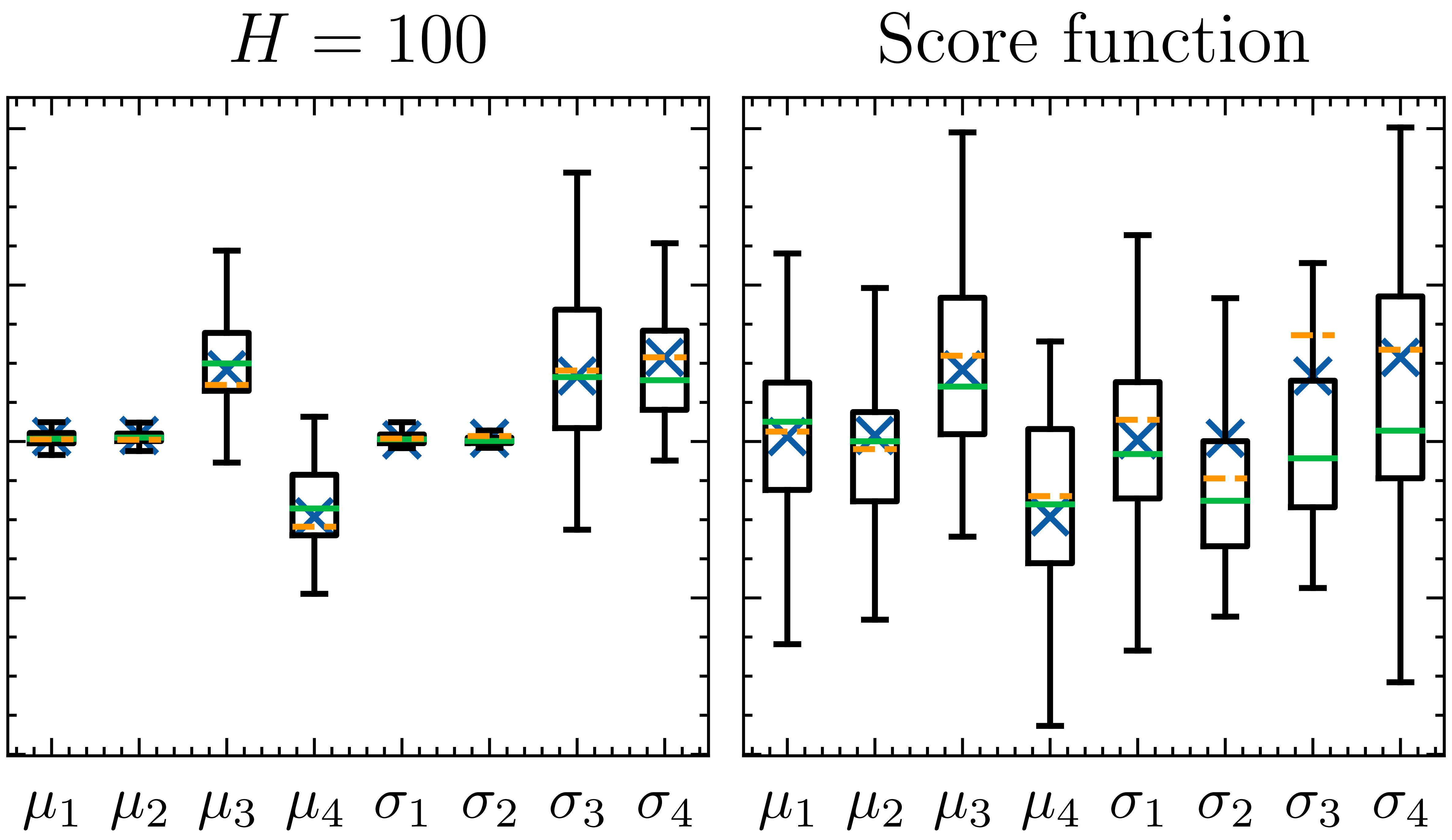
$$x_4 = \dots$$

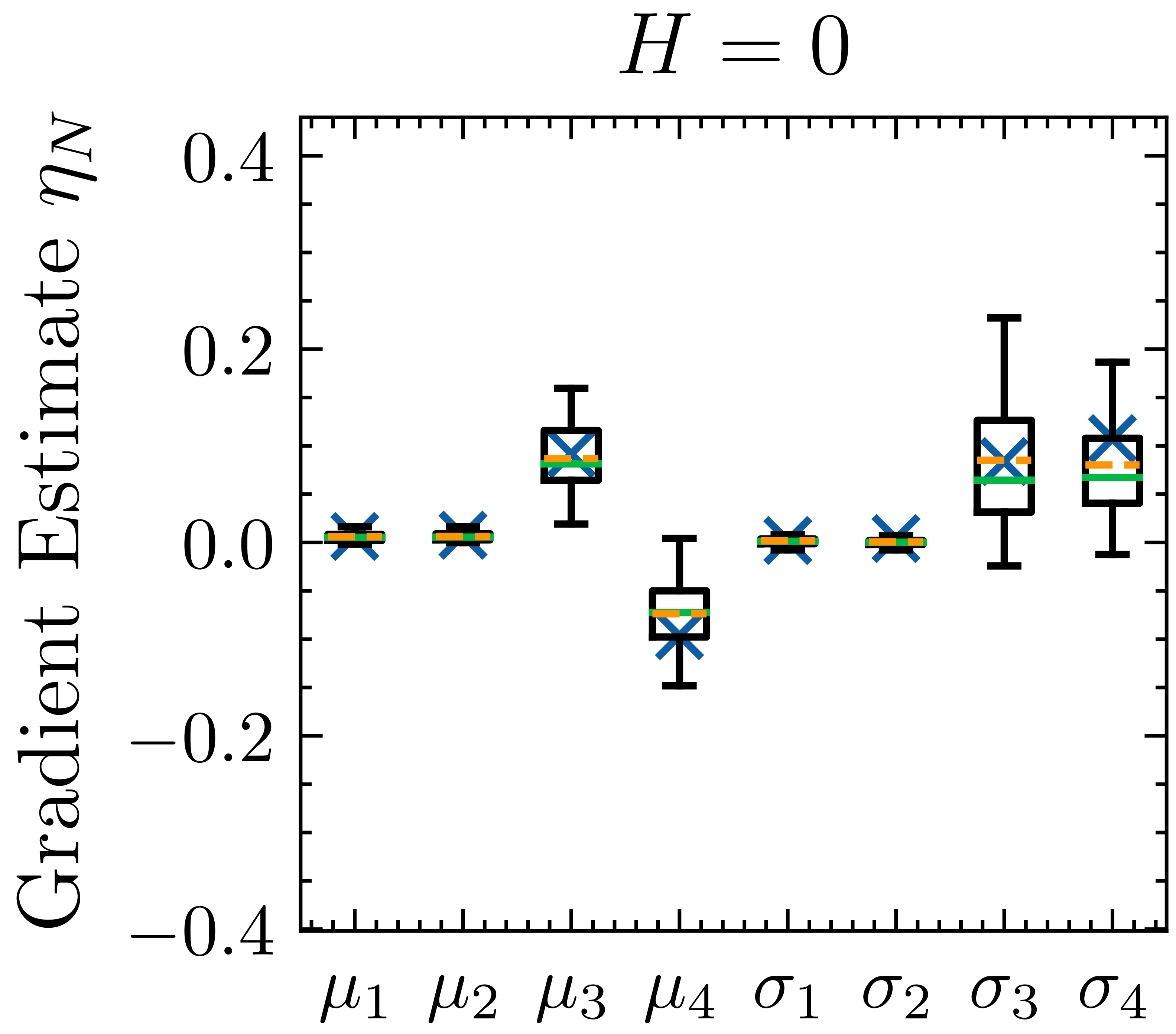
$$+ \frac{\partial f}{\partial \theta}$$

Gradient Horizon Problem

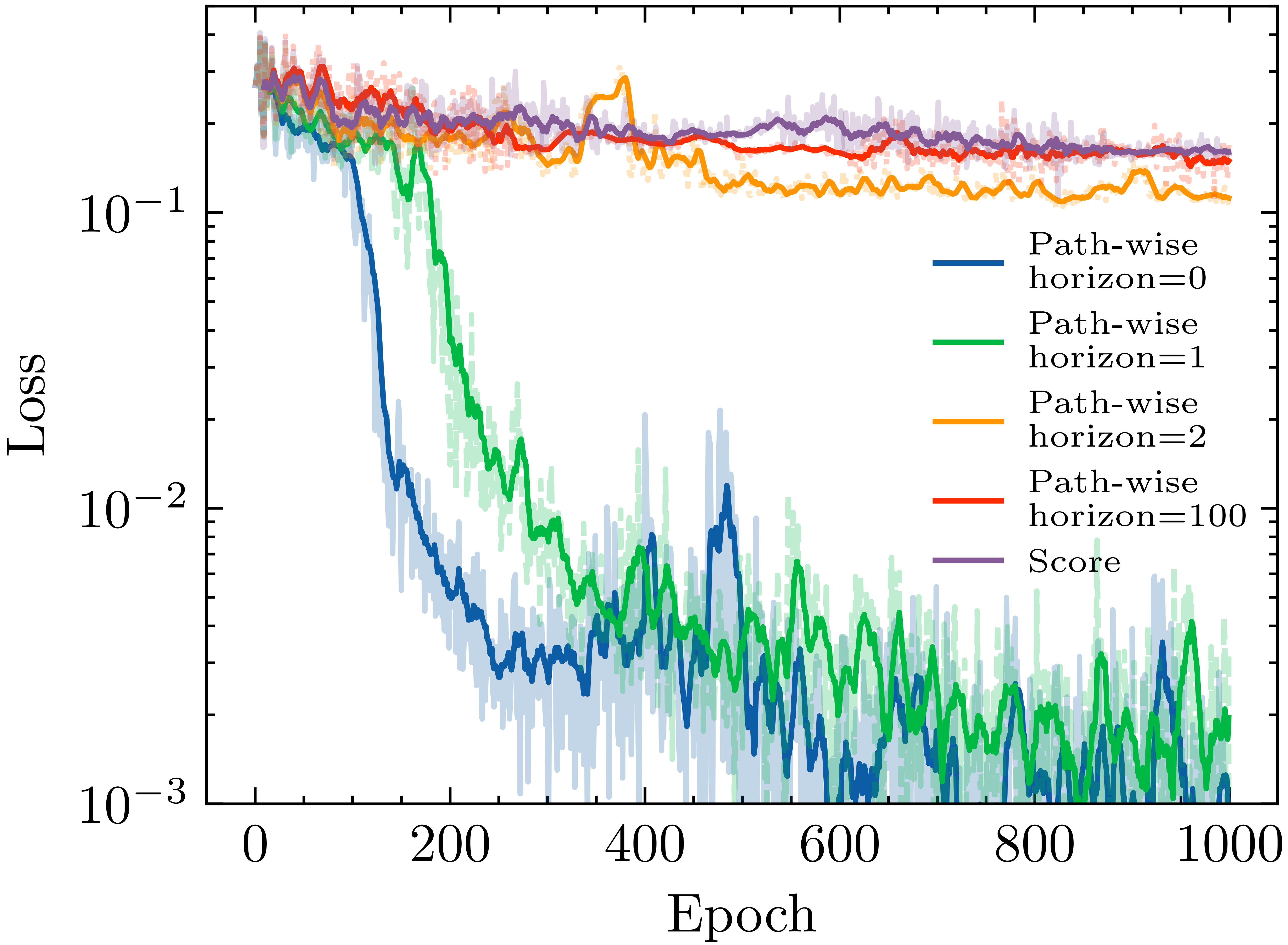
$$\frac{\mathrm{d}x_3}{\mathrm{d}\theta} = \cancel{\frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial \theta}} + \cancel{\frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial \theta}}$$

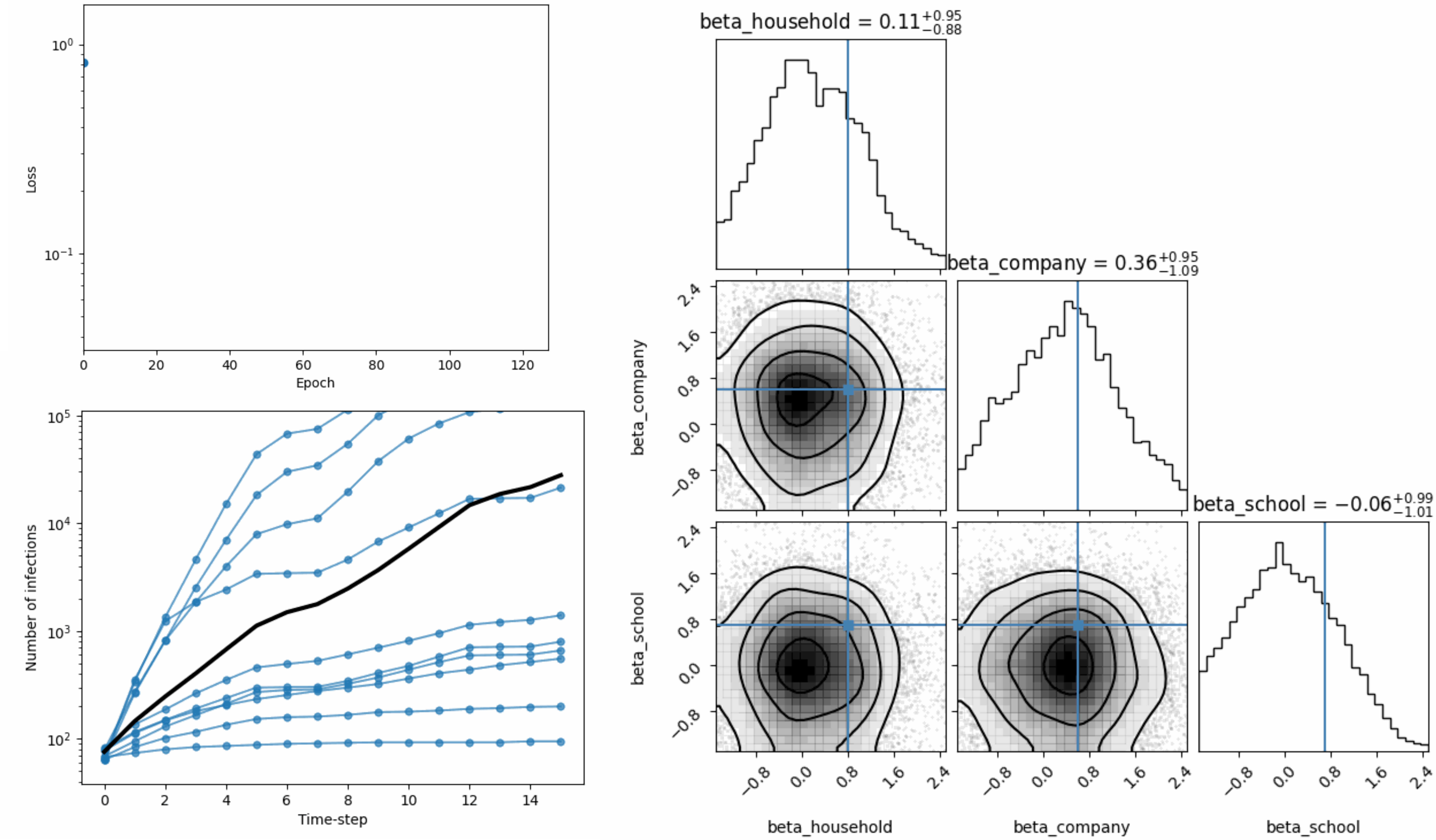
Truncating the gradient reduces variance!





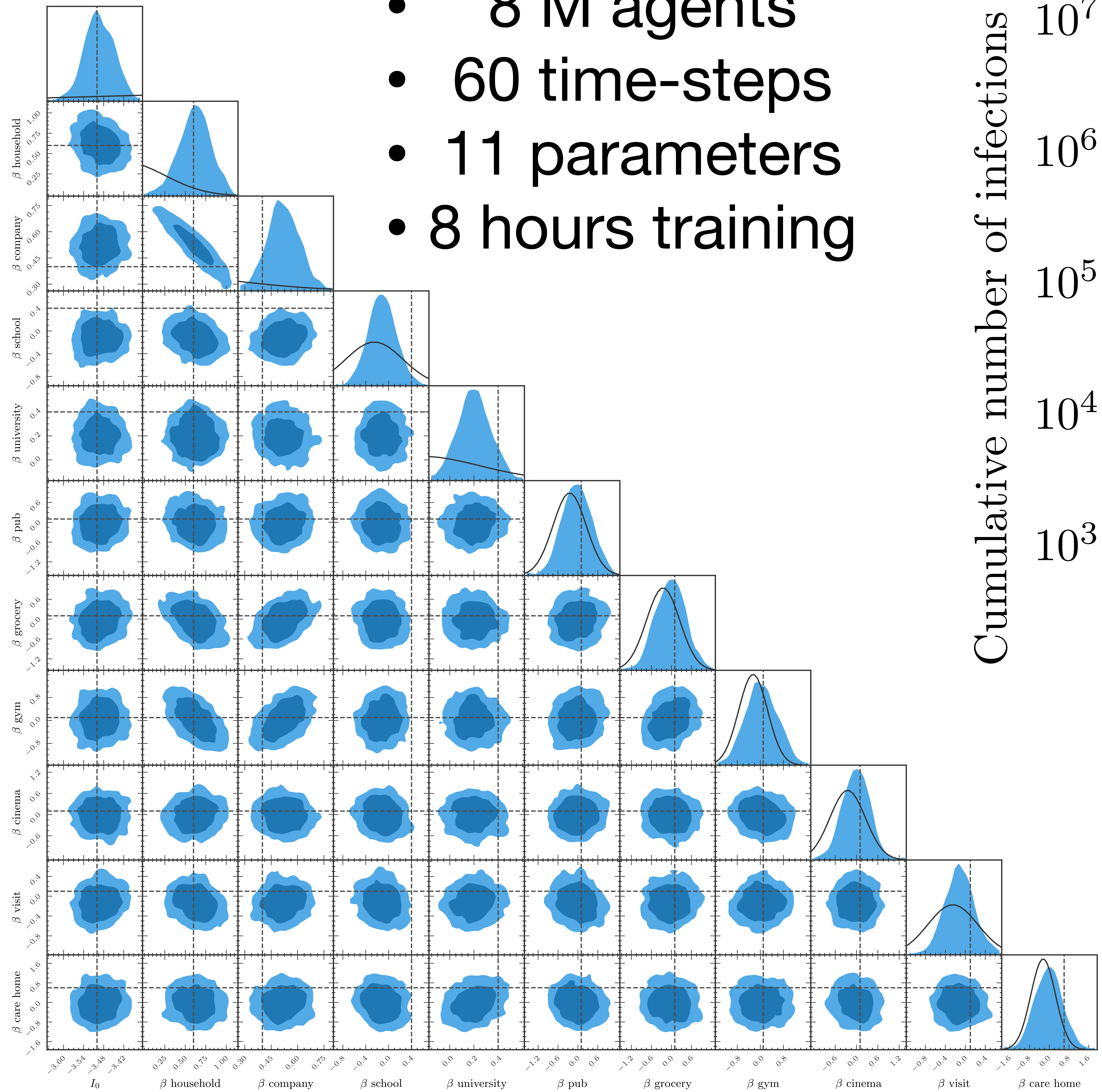
Gradient Horizon Problem

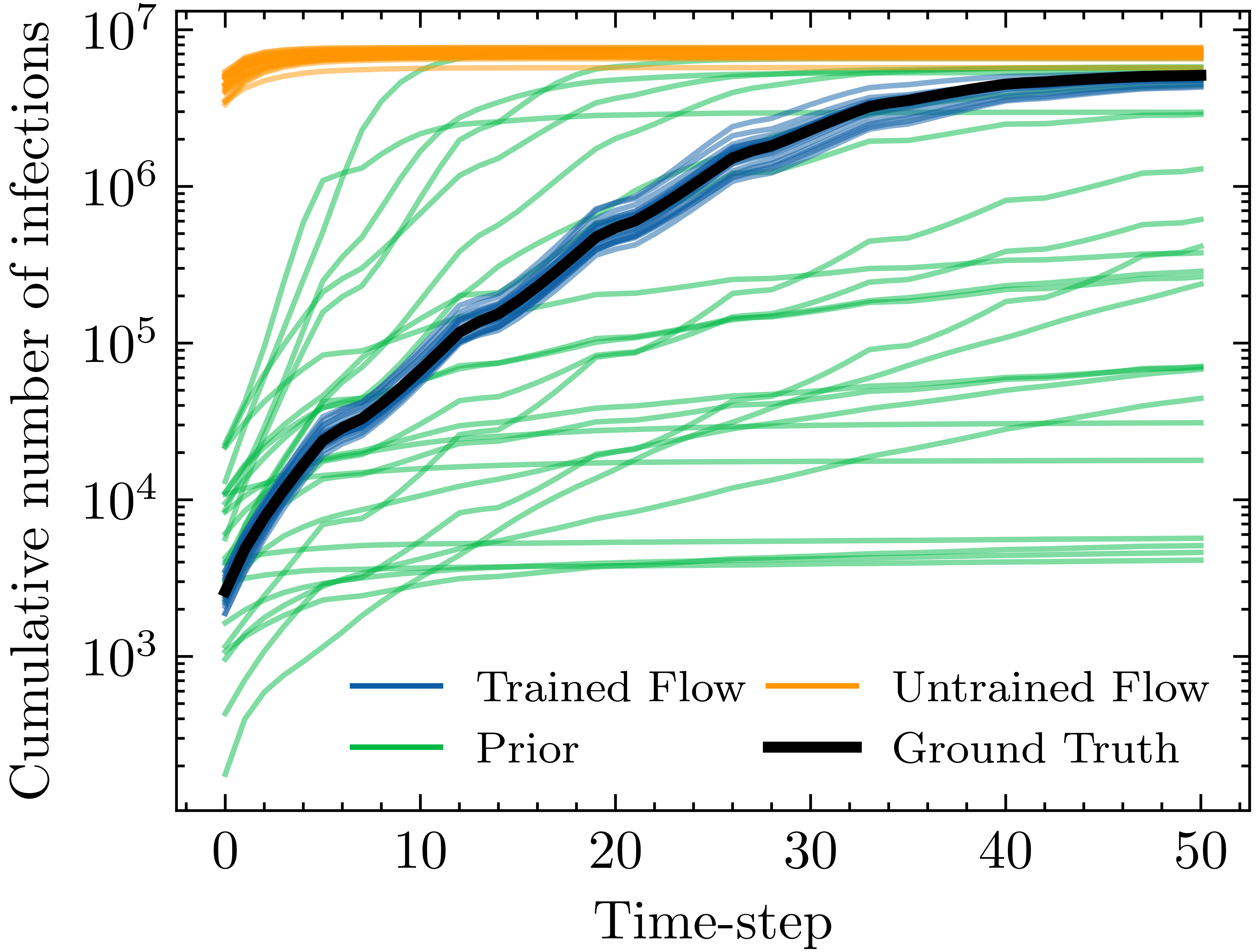




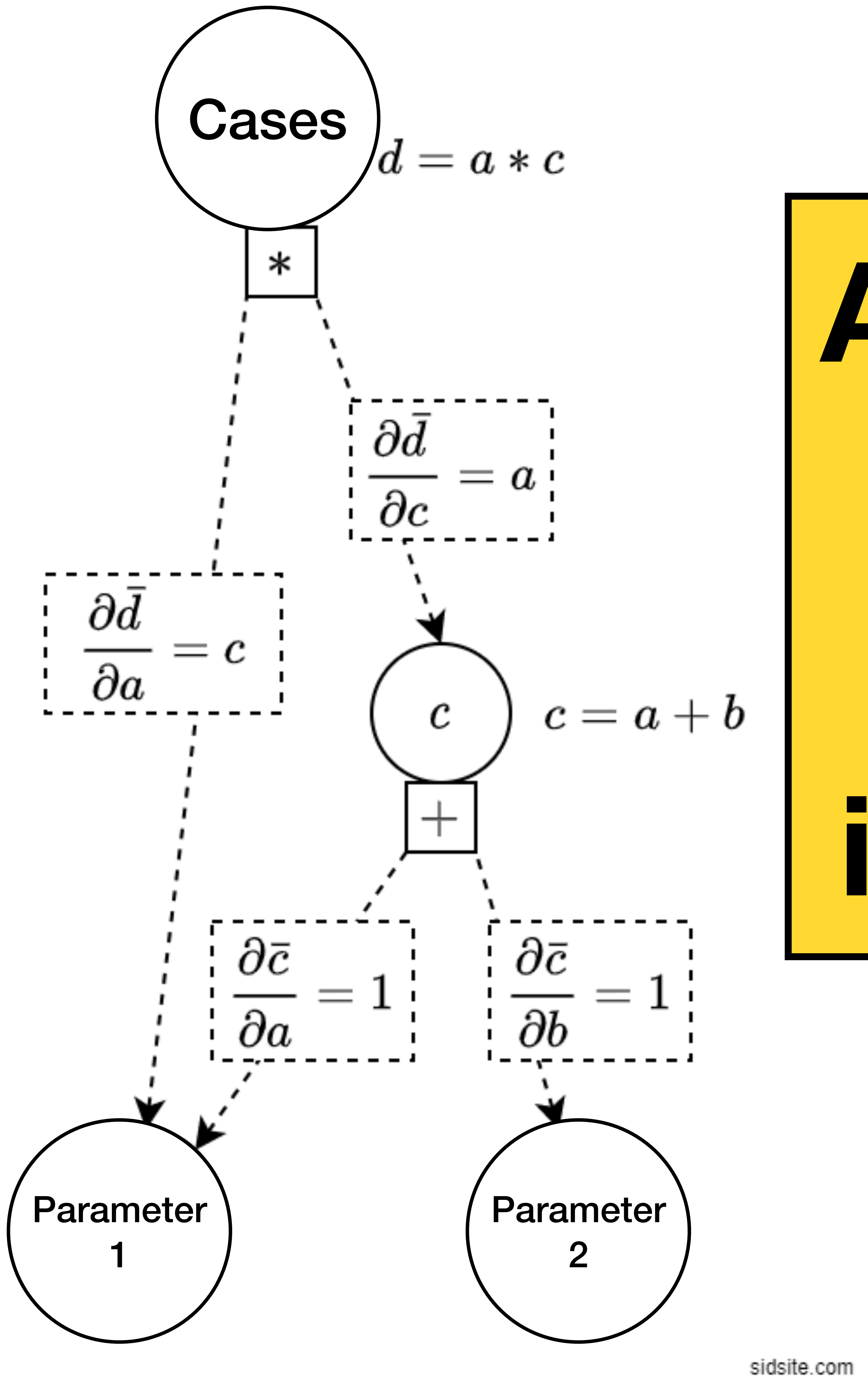
Loss
Epoch
Number of infections
Time-step
beta_household = $0.11^{+0.95}_{-0.88}$
beta_company = $0.36^{+0.95}_{-1.09}$
beta_school = $-0.06^{+0.99}_{-1.01}$
beta_company
beta_school
beta_household
beta_company
beta_school

- 8 M agents
- 60 time-steps
- 11 parameters
- 8 hours training





Sensitivity Analyses



AD performs sensitivity analyses with a single simulation run, independent of # of parameters!

The impact of uncertainty on predictions of the CovidSim epidemiological code

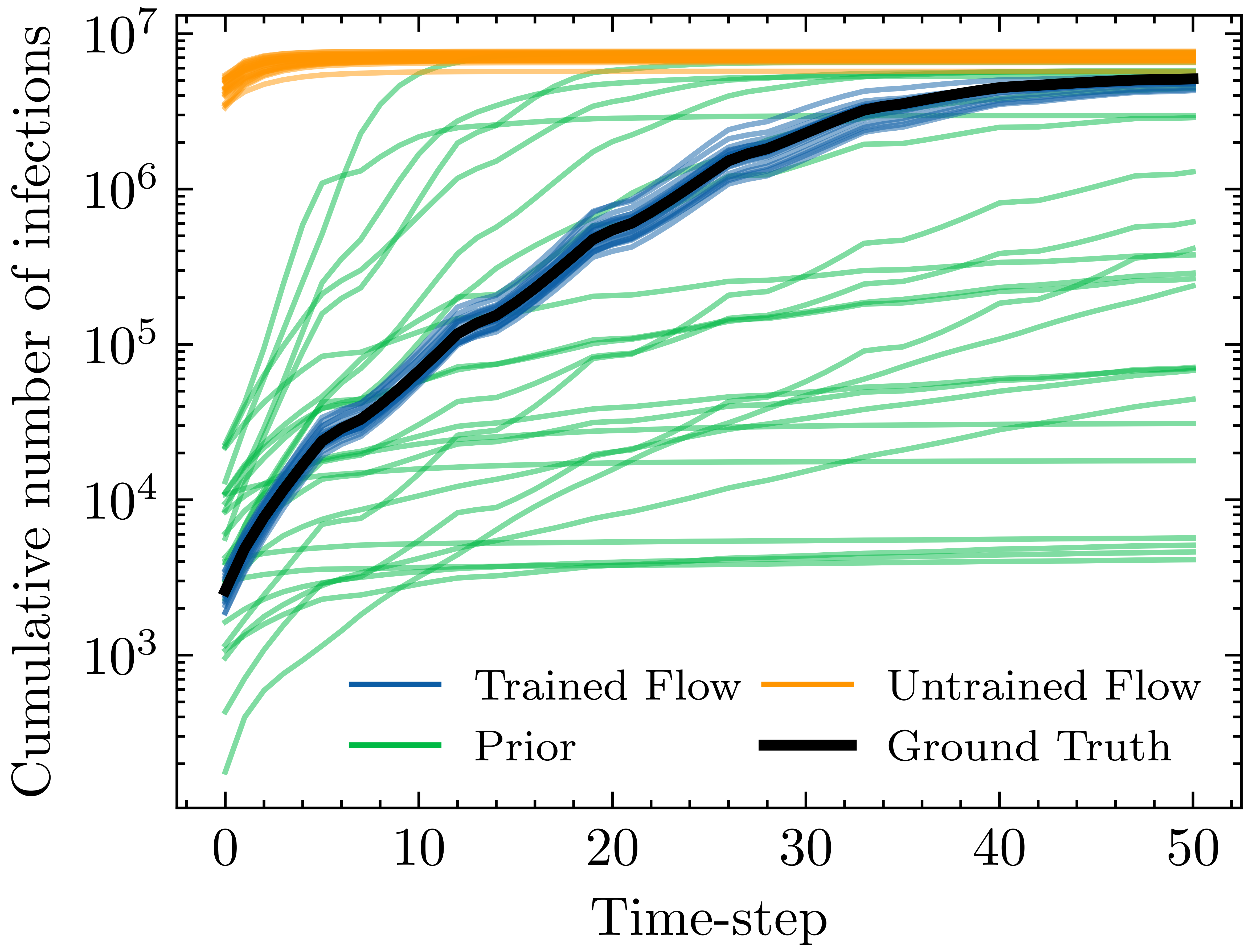
Wouter Edeling[1], Hamid Arabnejad[2], Robbie Sinclair[3], Diana Suleimenova[2], Krishnakumar Gopalakrishnan[3], Bartosz Bosak[4], Derek Groen[2], Imran Mahmood[2], Daan Crommelin[1,5] and Peter V. Coveney[3,6] ✉

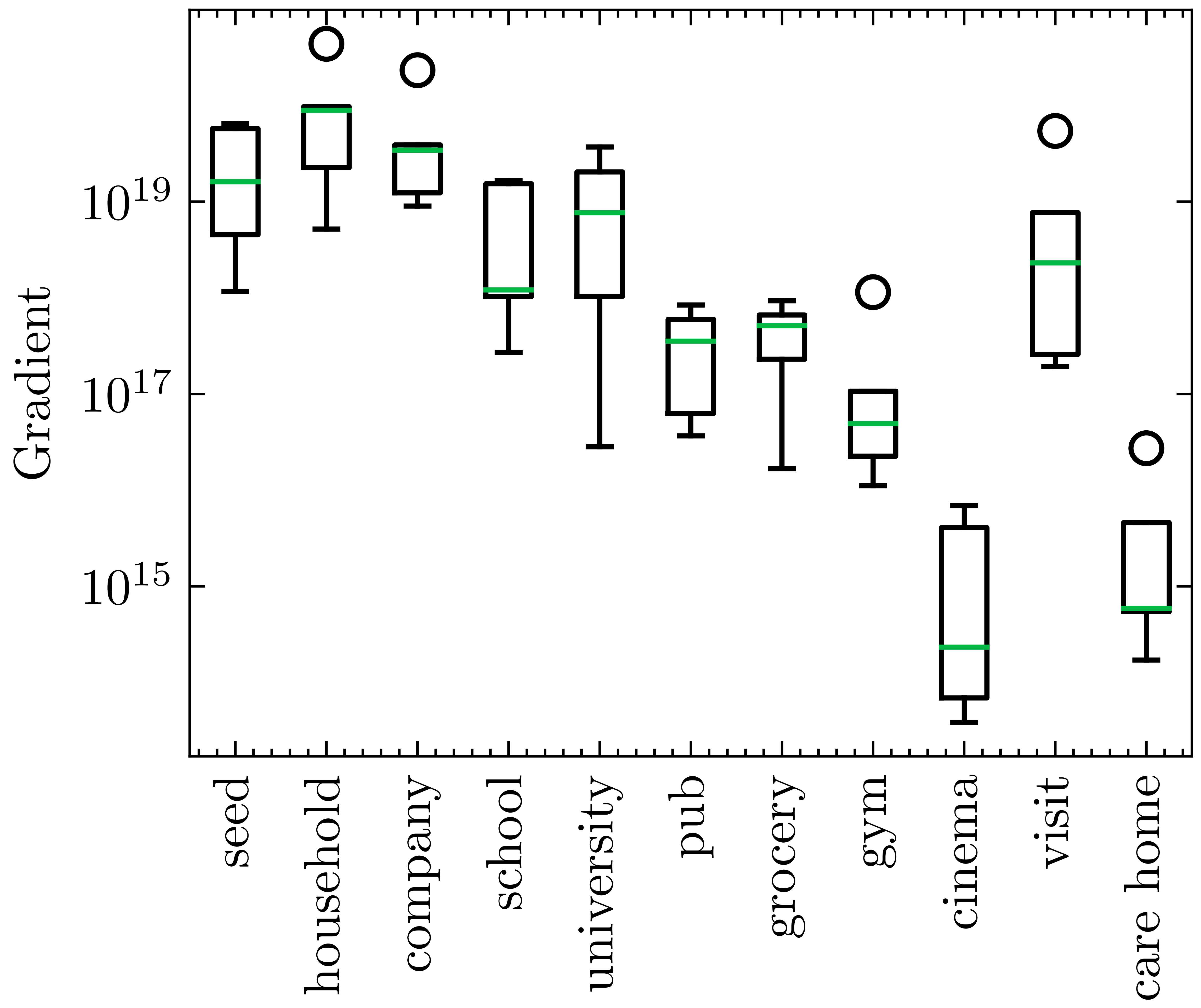
Epidemiological modelling has assisted in identifying interventions that reduce the impact of COVID-19. The UK government relied, in part, on the CovidSim model to guide its policy to contain the rapid spread of the COVID-19 pandemic during March and April 2020; however, CovidSim contains several sources of uncertainty that affect the quality of its predictions: parametric uncertainty, model structure uncertainty and scenario uncertainty. Here we report on parametric sensitivity analysis and uncertainty quantification of the code. From the 940 parameters used as input into CovidSim, we find a subset of 19 to which the code output is most sensitive—imperfect knowledge of these inputs is magnified in the outputs by up to 300%. The model displays substantial bias with respect to observed data, failing to describe validation data well. Quantifying parametric input uncertainty is therefore not sufficient: the effect of model structure and scenario uncertainty must also be properly understood.

Ensemble execution. Consequently, through the use of adaptive methods we make the uncertainty analysis of CovidSim tractable, but our analysis nevertheless required us to perform thousands of runs, each with its own unique set of input parameters. Specifically, we used the Eagle supercomputer at the Posnan

Reverse-mode AD independent of number of parameters!

Sensitivity Analyses





Conclusions

1. Bayesian approaches to calibrating ABMs have numerous benefits

2. ABMs can be made differentiable even with discrete randomness and control flow

3. Diff simulators + Bayesian inference (via Normalizing Flows) promising route to calibrate large-scale ABMs efficiently

Paper + slides: www.arnau.ai/talks

Conclusions

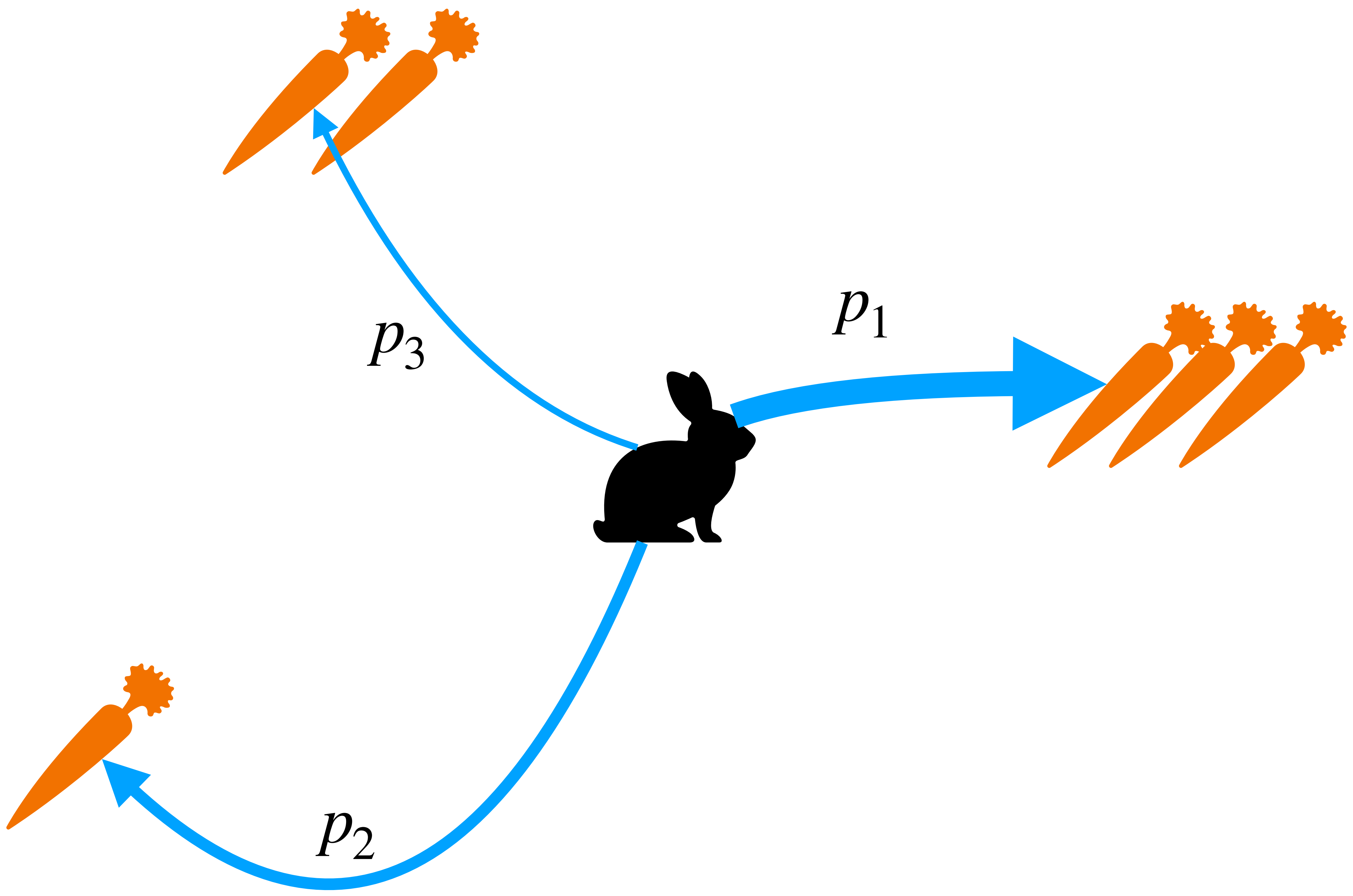
Differentiable agent-based models enable:

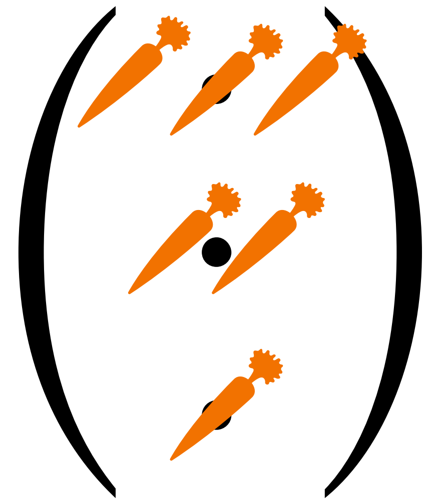
1. Fast **simulation** via tensorisation.
2. Fast and accurate Bayesian **calibration** via gradients.
3. Fast and accurate **sensitivity analyses** via gradients.

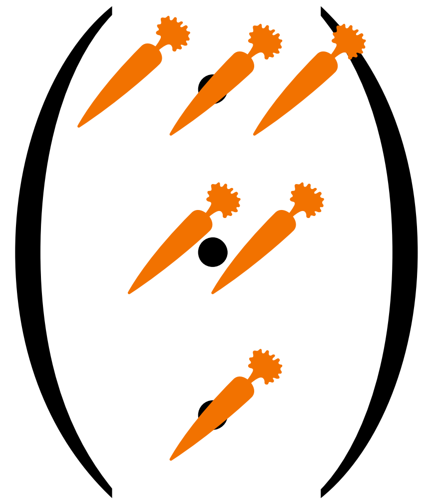
Papers + slides: arnau.ai/talks

Backup slides

Challenge 1: Differentiable Control Flow



$x' = \text{Argmax}$ 

$x' = \text{Softmax}$ 

Challenge 2: Differentiable Stochasticity

Continuous Variables

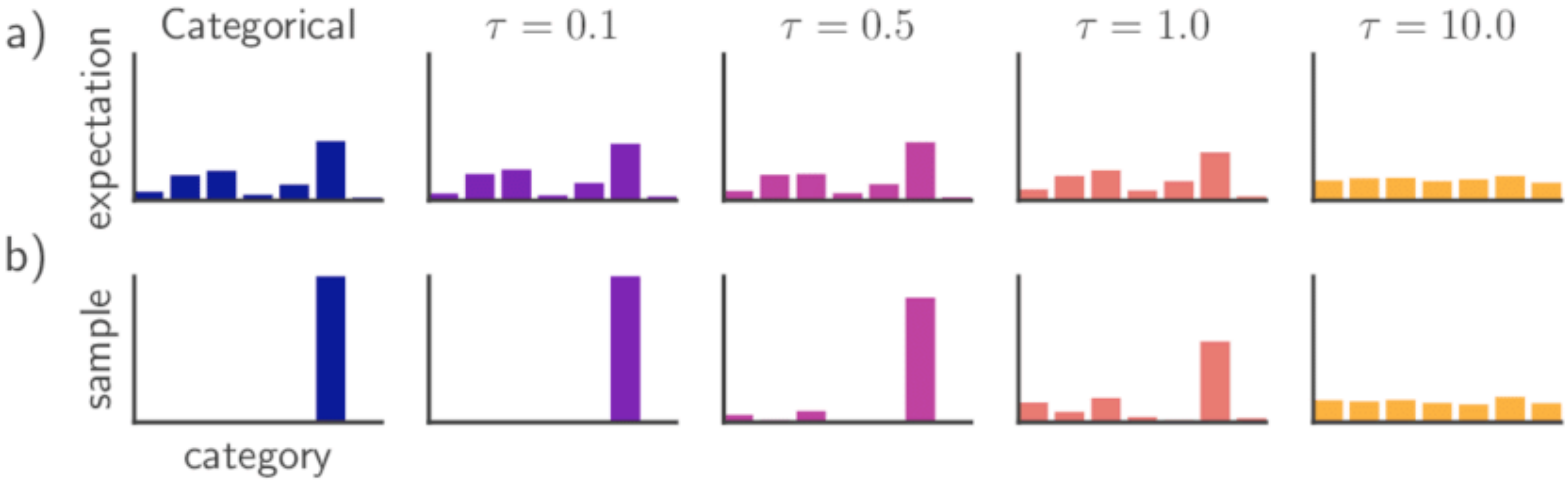
$$x \sim \mathcal{N}(\mu, \sigma) \quad \Longleftrightarrow \quad x = \mu + \sigma\, r \quad r \sim \mathcal{N}(0,1)$$

$$\frac{\mathrm{d}x}{\mathrm{d}\mu} = 1 \quad \frac{\mathrm{d}x}{\mathrm{d}\sigma} = r$$

Challenge 2: Differentiable Stochasticity

Discrete Variables

- Gumbel-Softmax



Jang et al. (2016)