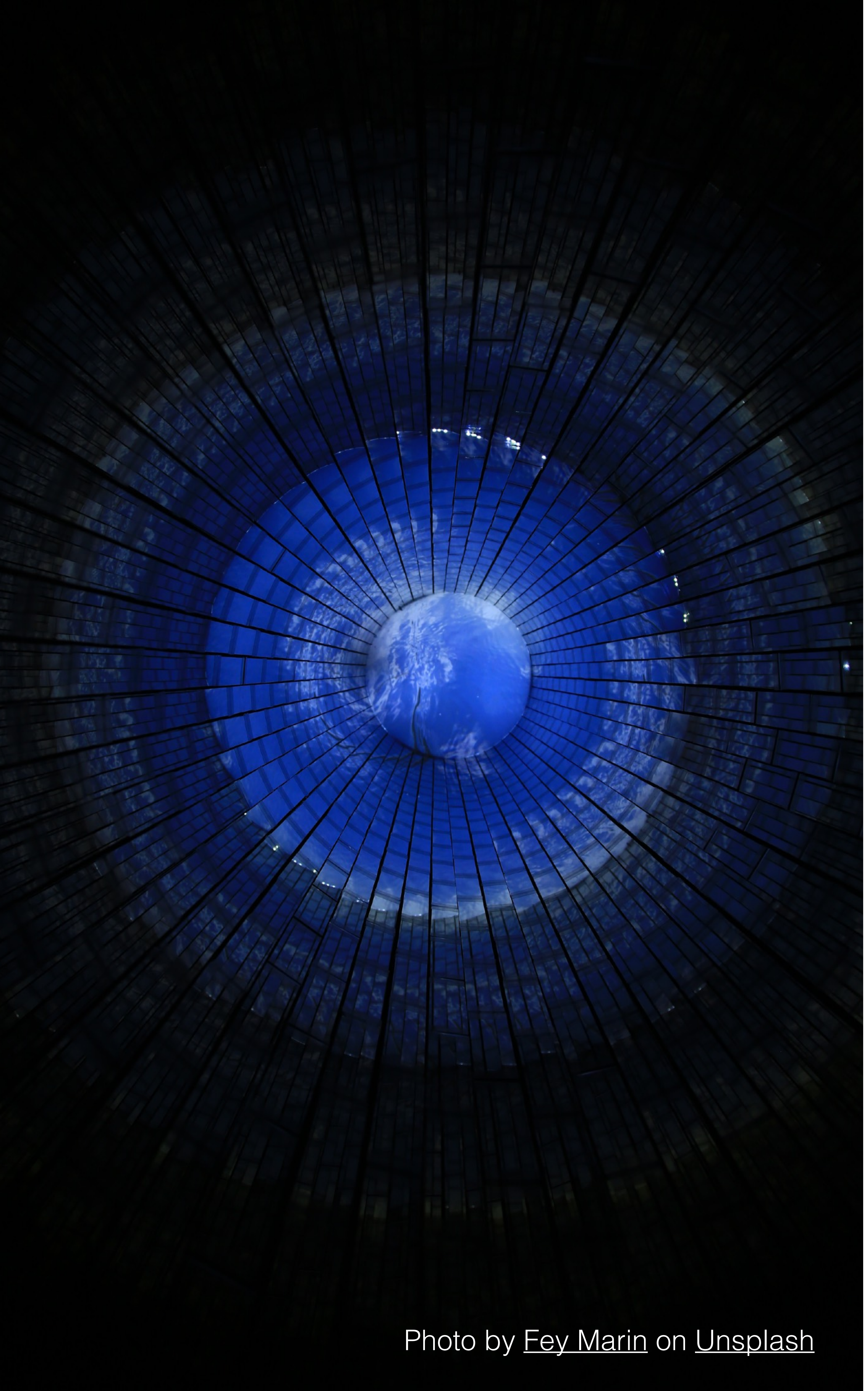
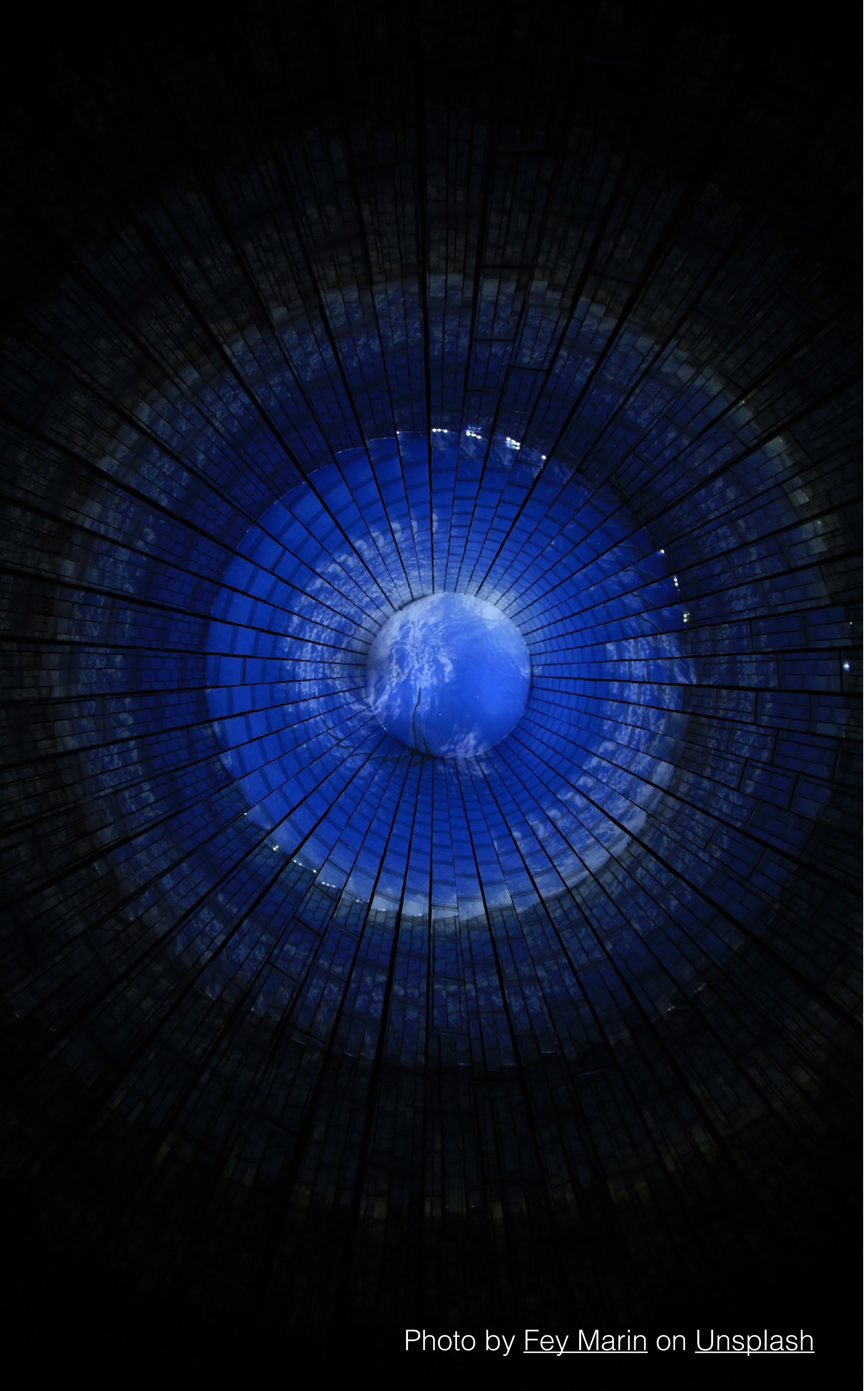


An **improper prior distribution** is any probability density function (or mass function) that is non-negative but does not integrate (or sum) to 1.

Why would we do such a thing?!



Let $X_1, \ldots, X_n \,|\, \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$ with σ^2 fixed and known, and $\pi(\mu) = c > 0$ for all μ. Clearly this prior is improper. And yet we'll verify numerically in the Unit #2 code that the posterior is proper.

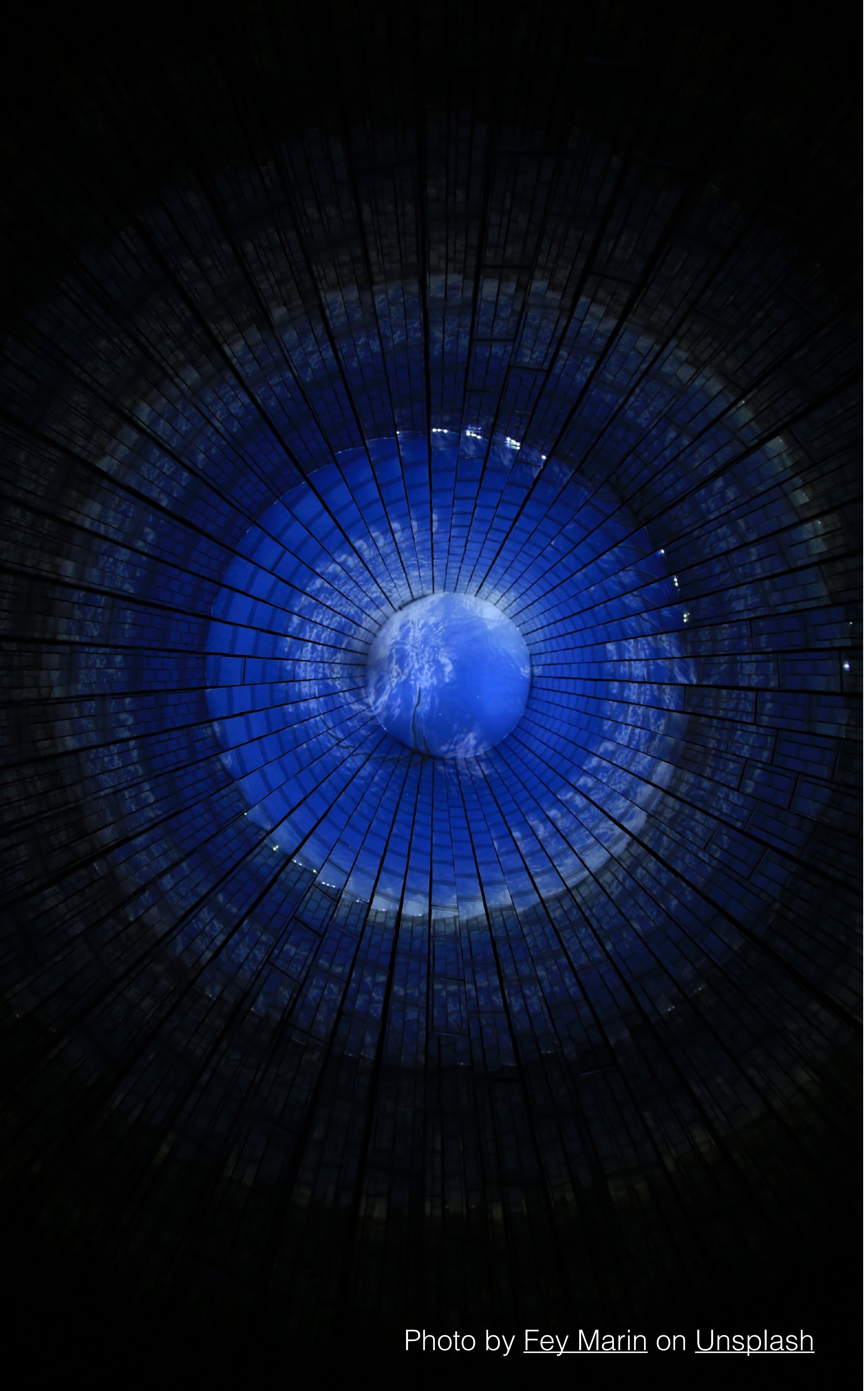


Photo by Fey Marin on Unsplash

Notes on improper priors:

1. One must verify that an improper prior leads to a proper posterior. Otherwise, the Bayesian inference is not valid!

2. It's not clear what an improper prior means in terms of subjective degrees of belief. An improper prior is not a proper pdf, so how can it summarize an individual's uncertainty?

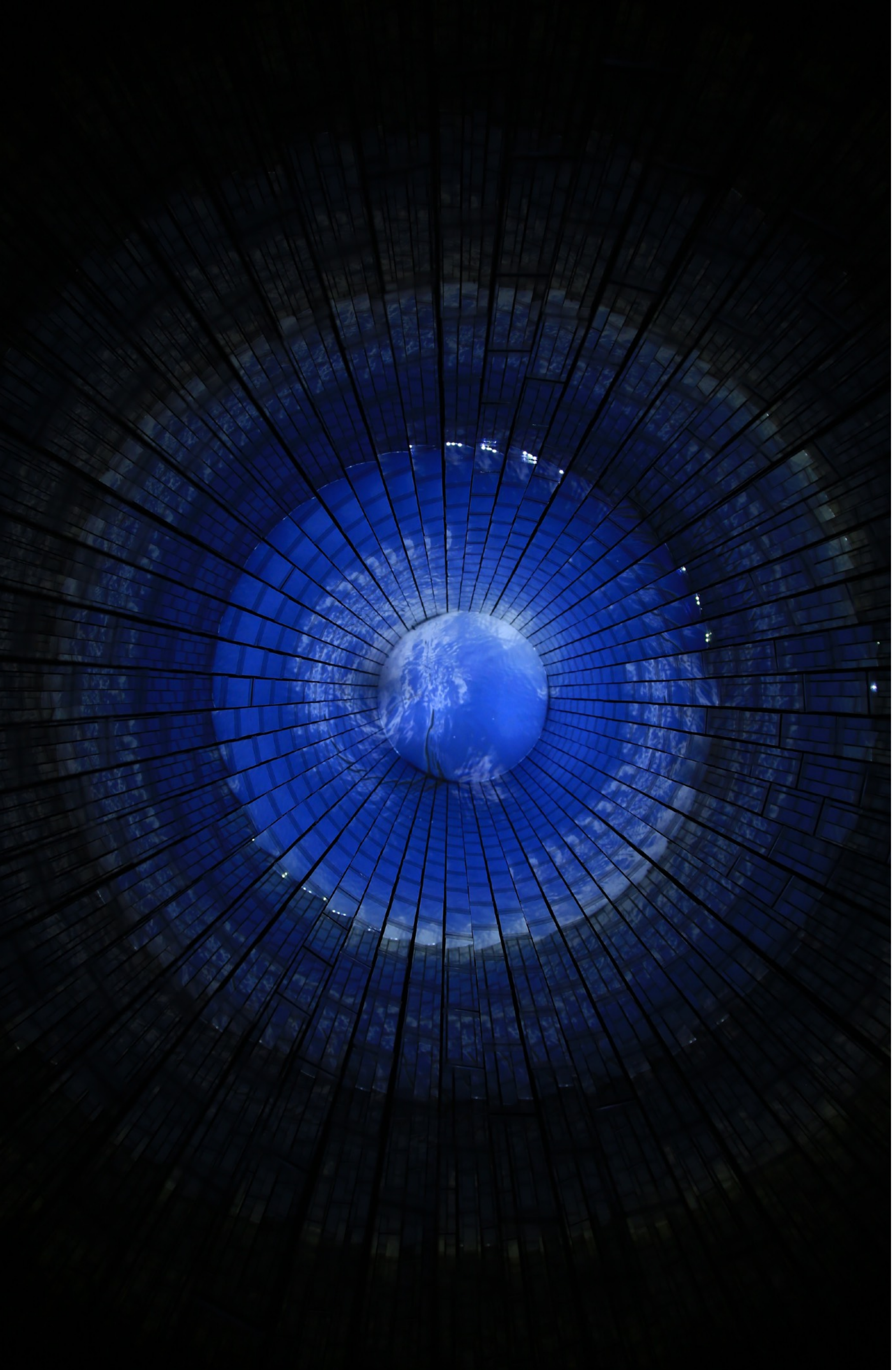


Photo by Fey Marin on Unsplash

What happens when there isn't any prior information?

The **discrete principle of indifference** states that if there are n possible values in the parameter space, each outcome should be assigned the same prior probability, $1/n$.

The **continuous principle of indifference** states that if there is an interval of possible values in the parameter space, say $\theta \in (a, b)$ then the probability of any subinterval $I = (c, d)$ should have equal probability: $P(I) = \dfrac{d - c}{b - a}$.

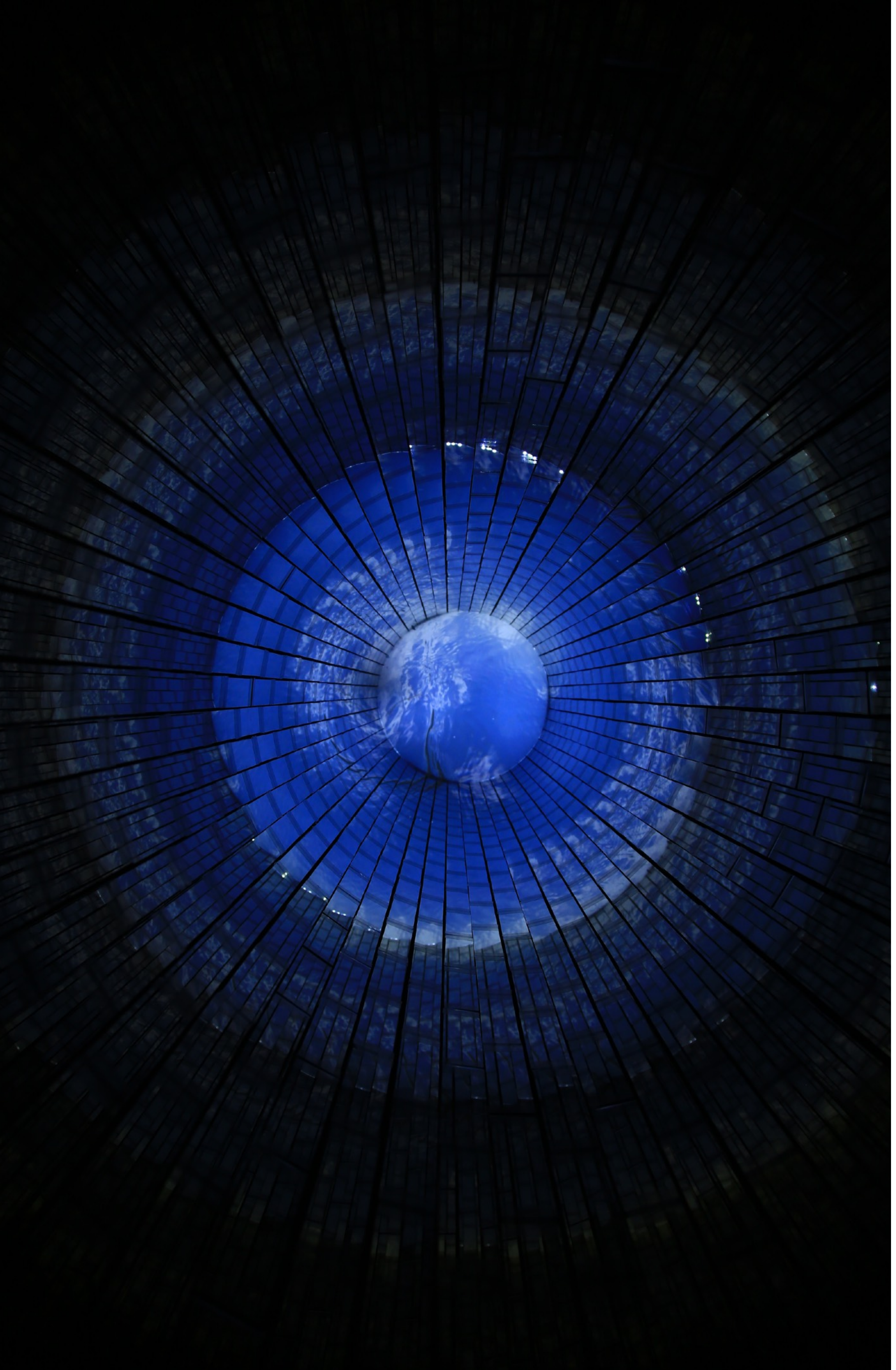
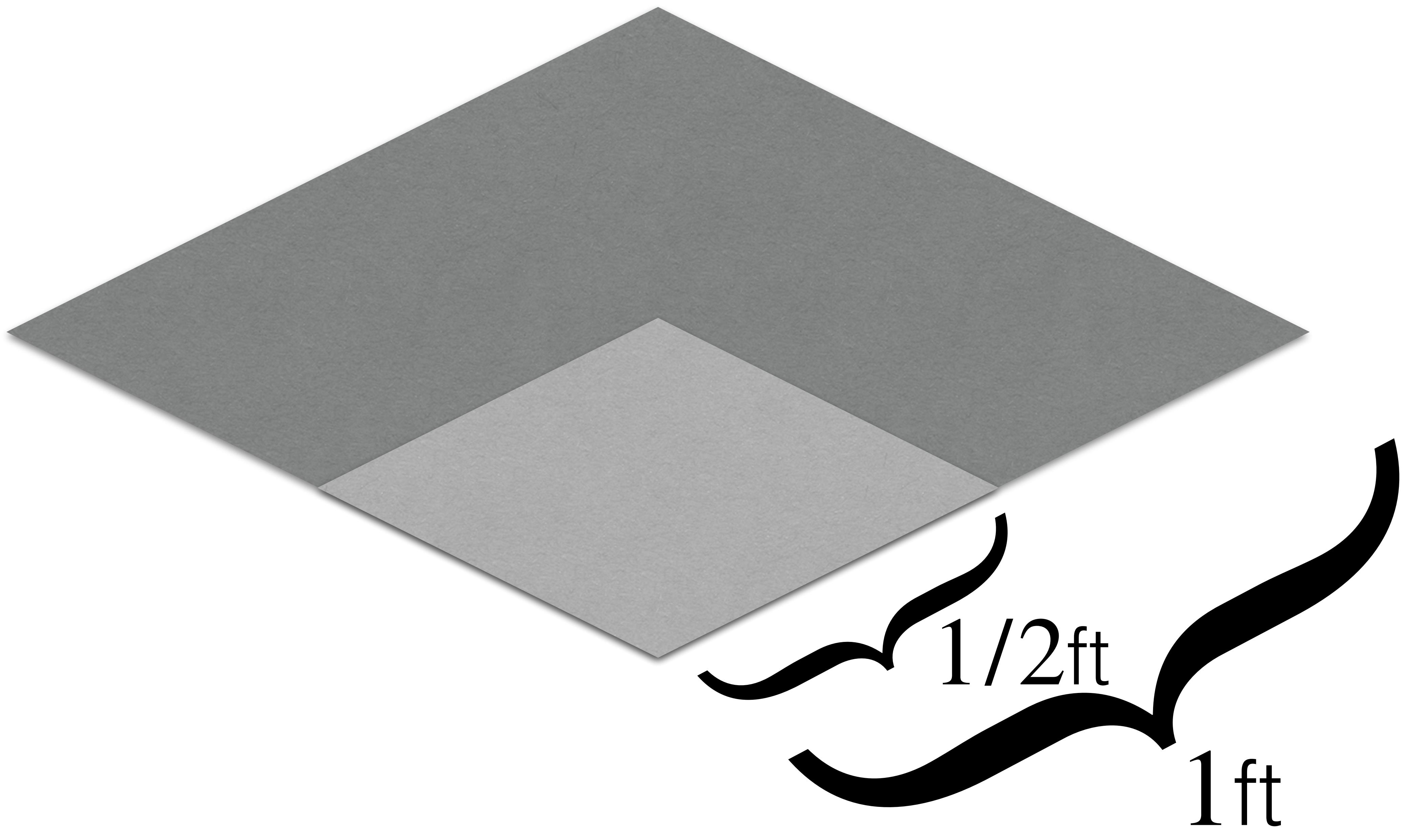


Photo by Fey Marin on Unsplash



1. A factory produces cubes with side-length between 0 and 1 foot; what is the probability that a randomly chosen cube has side-length between 0 and $1/2$ of a foot?

2. A factory produces cubes with face-area between 0 and 1 square-feet; what is the probability that a randomly chosen cube has face-area between 0 and $1/4$ square-feet?

Example from https://stanford.io/40KyxGc

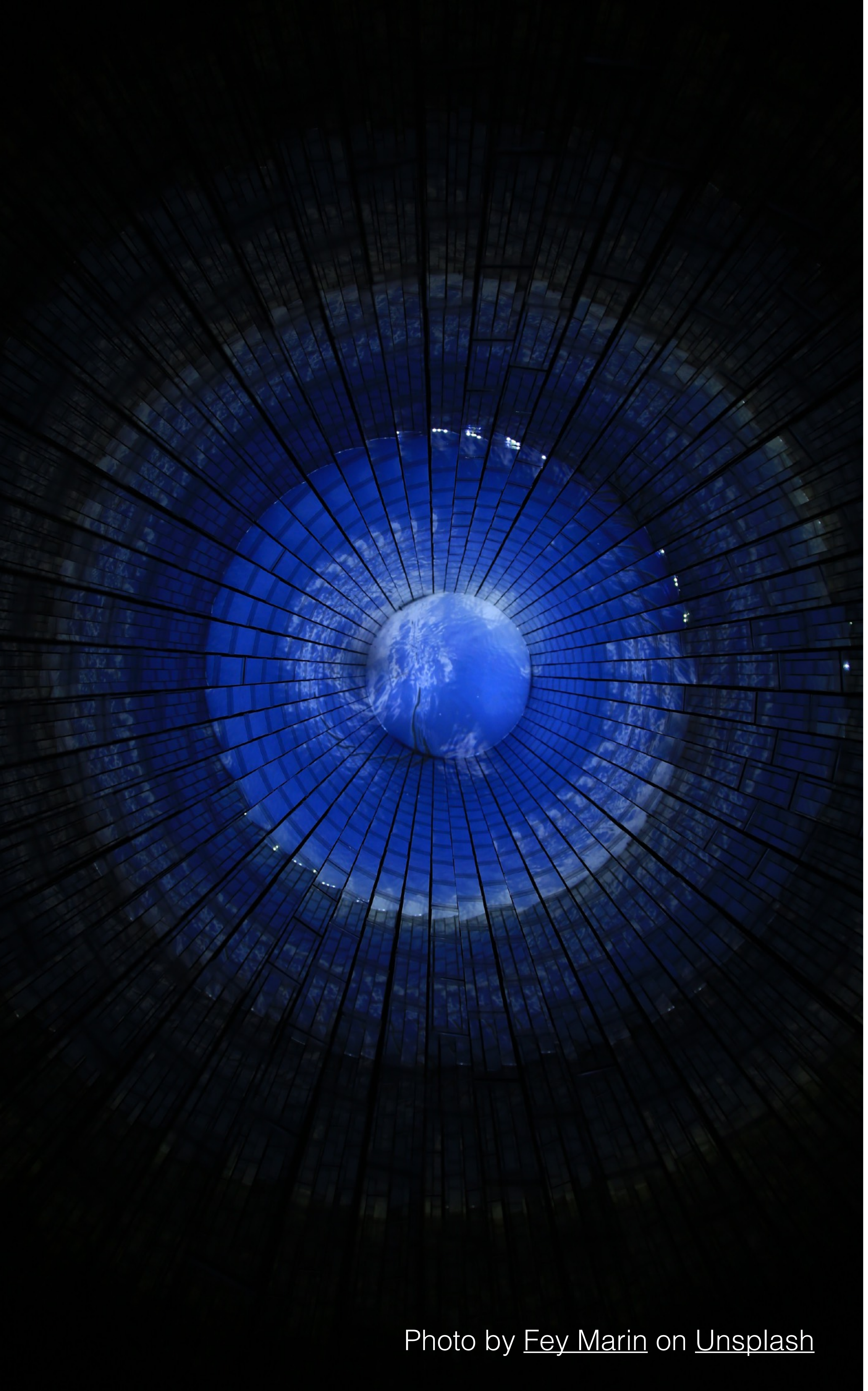
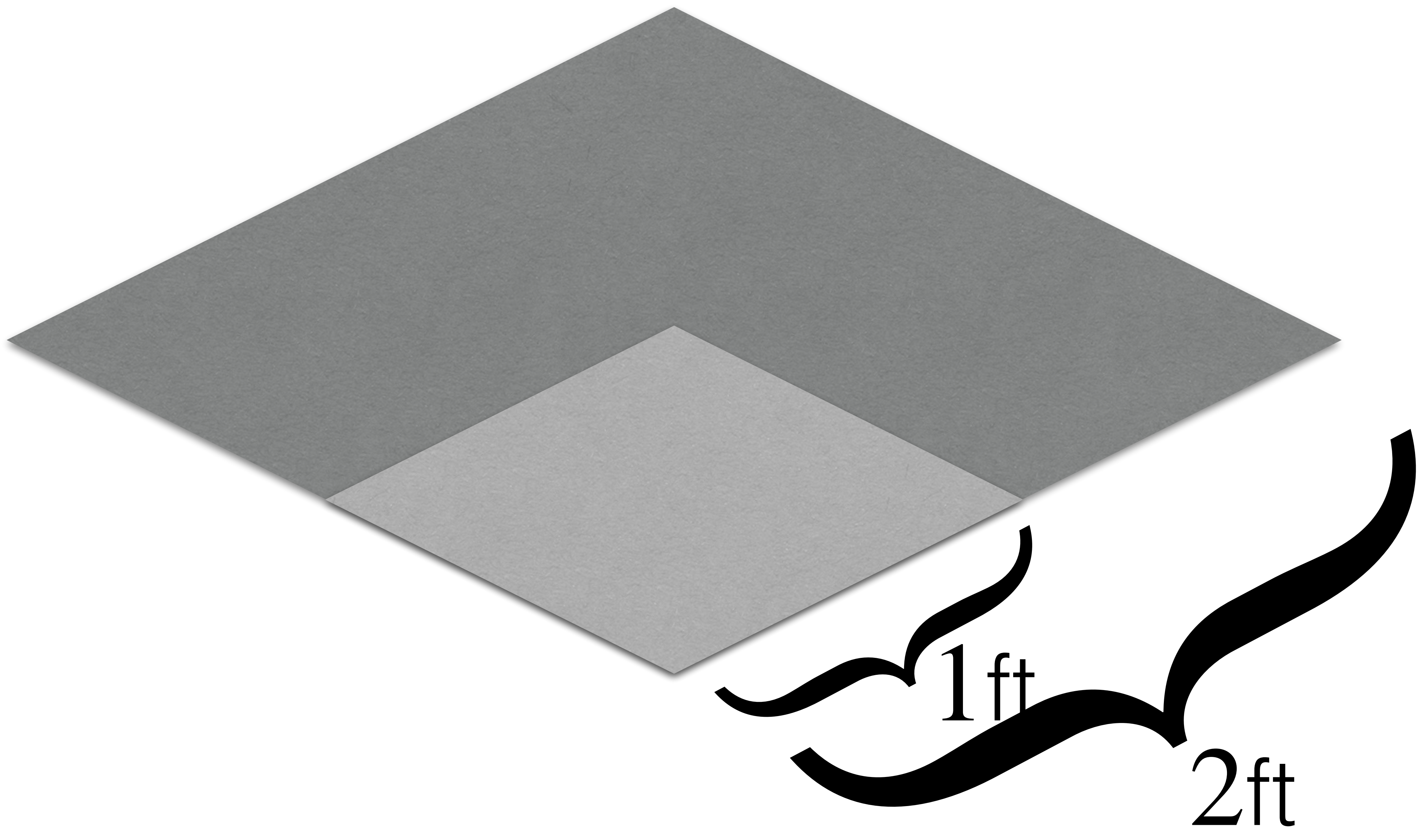


Photo by Fey Marin on Unsplash



The dimension on which we apply the principle of indifference changes the probability assignments. That is, probabilities assigned according to the principle of indifference are not *invariant* to different parameterizations.

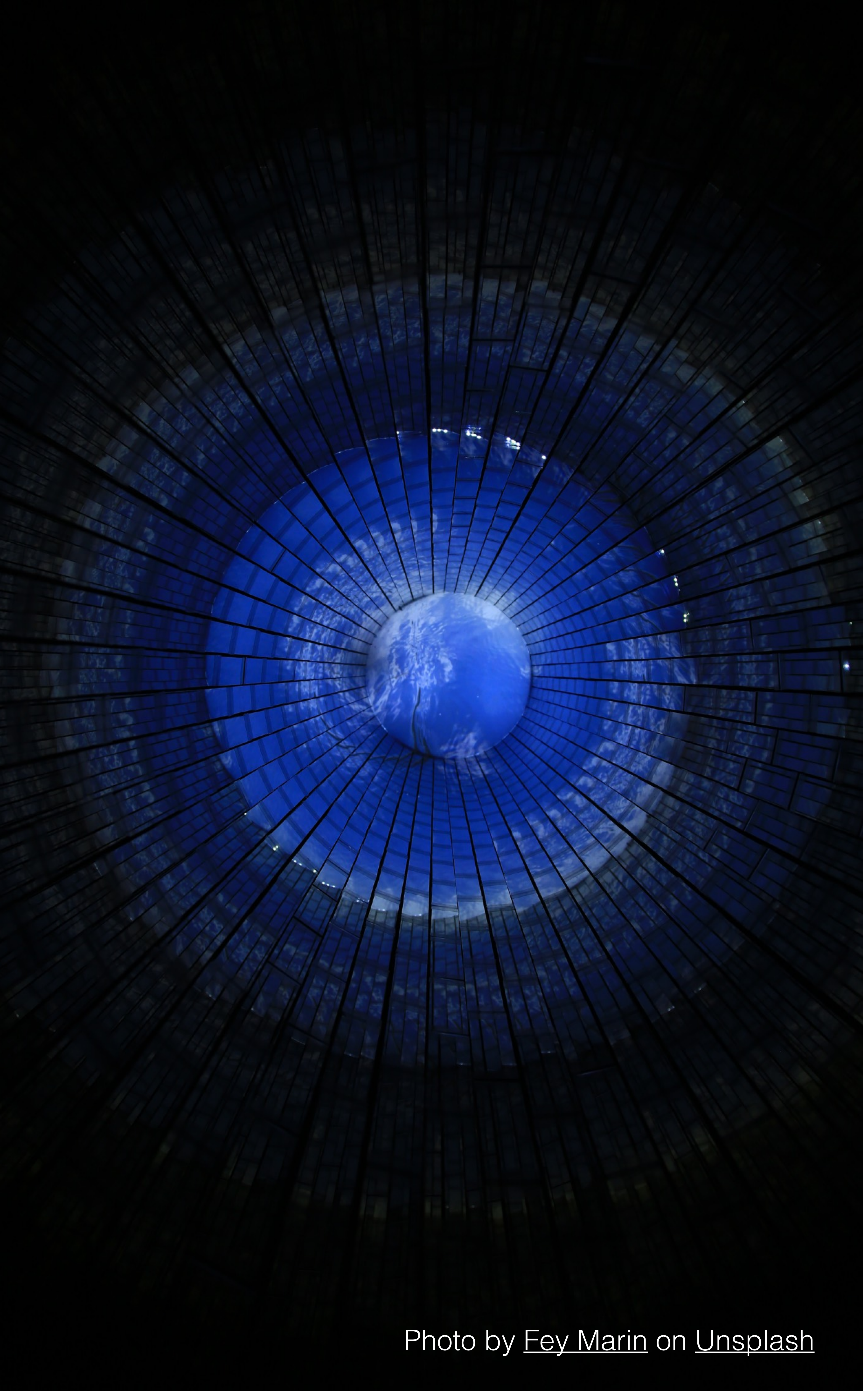
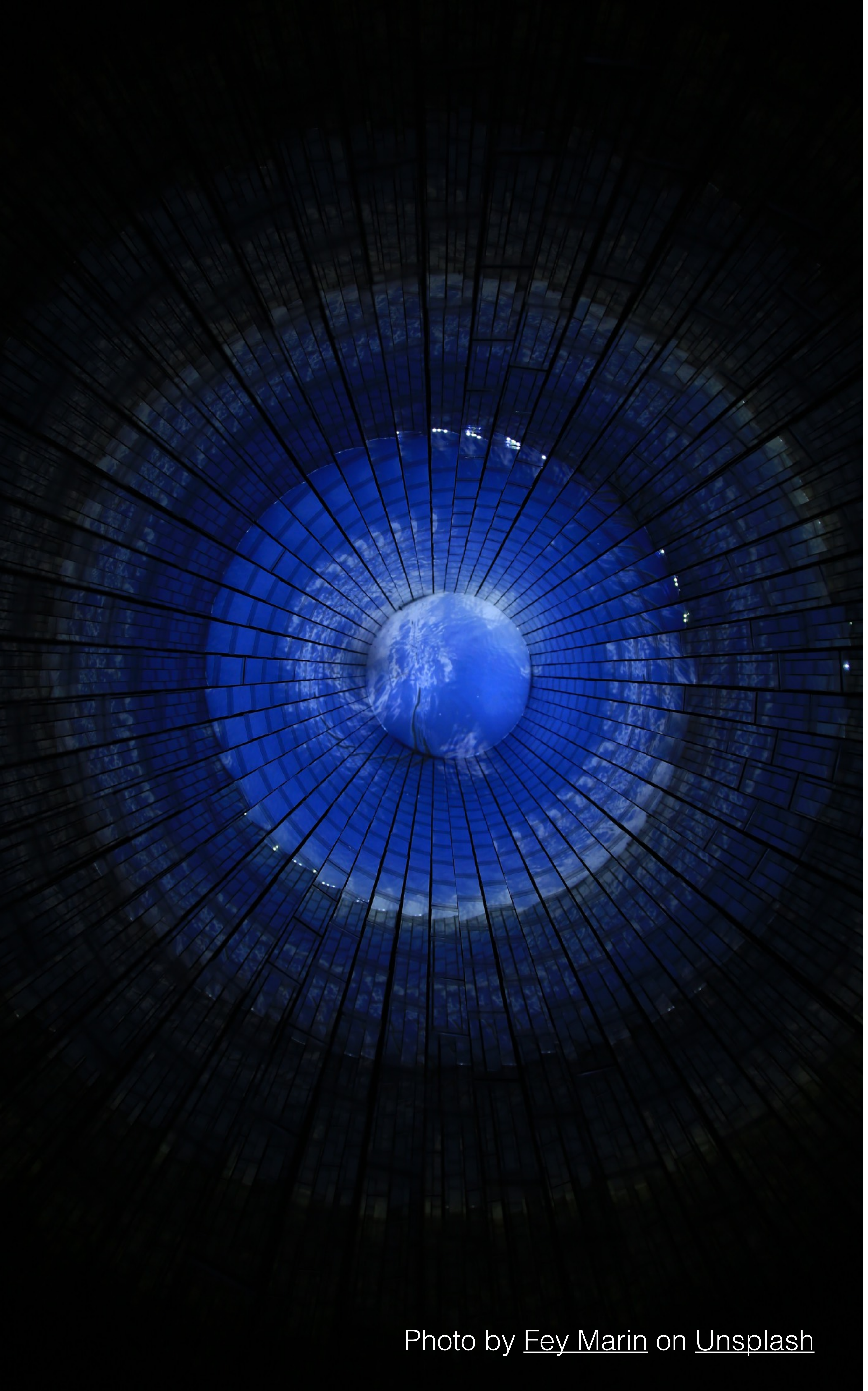


Photo by Fey Marin on Unsplash

Objective priors are an attempt to avoid personal bias and violations of invariance. They may also provide baselines for comparisons for subjective methods.



The **univariate Jeffrey's prior** is defined as any prior proportional to the square root of the Fisher information number: $\pi(\theta) \propto \sqrt{I(\theta)}$.

Result: Jeffrey's prior is invariant to reparameterization.

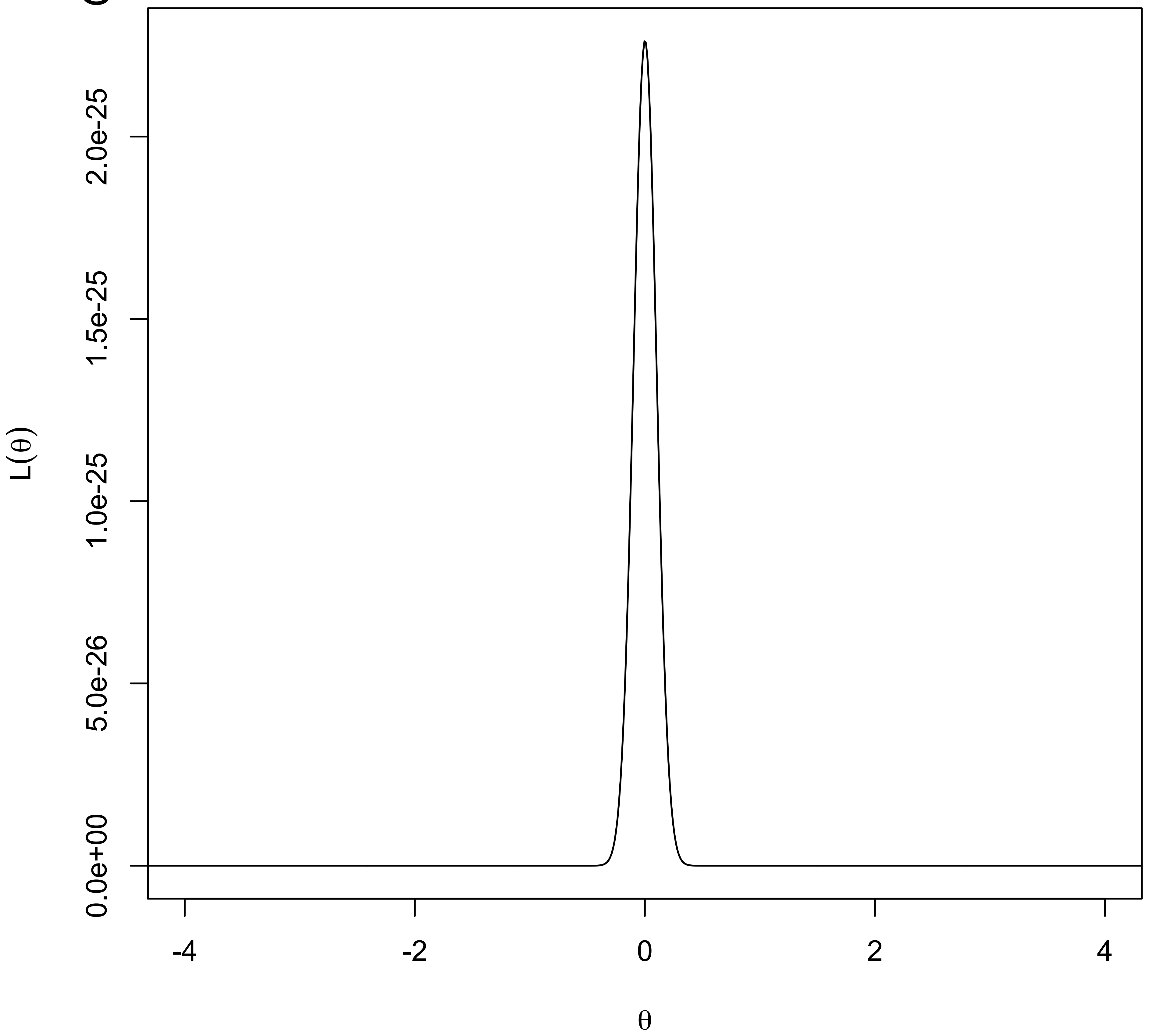
To prove this invariance, we will need to :

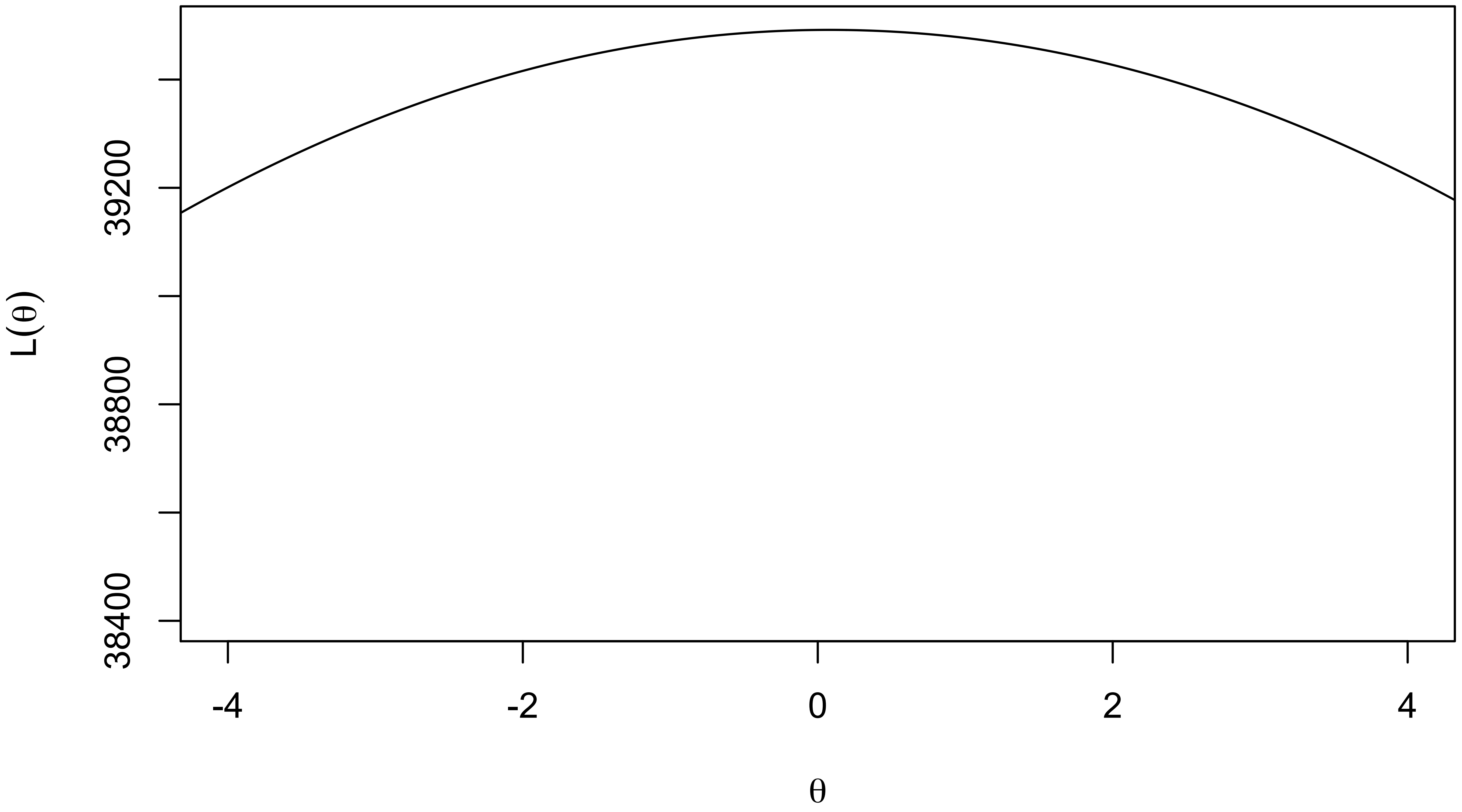
1. Recall how to find pdfs of transformations of random variables (we did this when we derived the inverse gamma distribution a few slides back!)

2. Review Fisher's information

Fisher's Information

The Fisher information number (or matrix, when $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$) tells us how much information a sample of data contains about the unknown parameter If $f(\mathbf{x} \mid \theta)$ is sharply peaked with respect to changes in θ, it will be easier to estimate θ from the data than if $f(\mathbf{x} \mid \theta)$ is shallow, it will be harder.





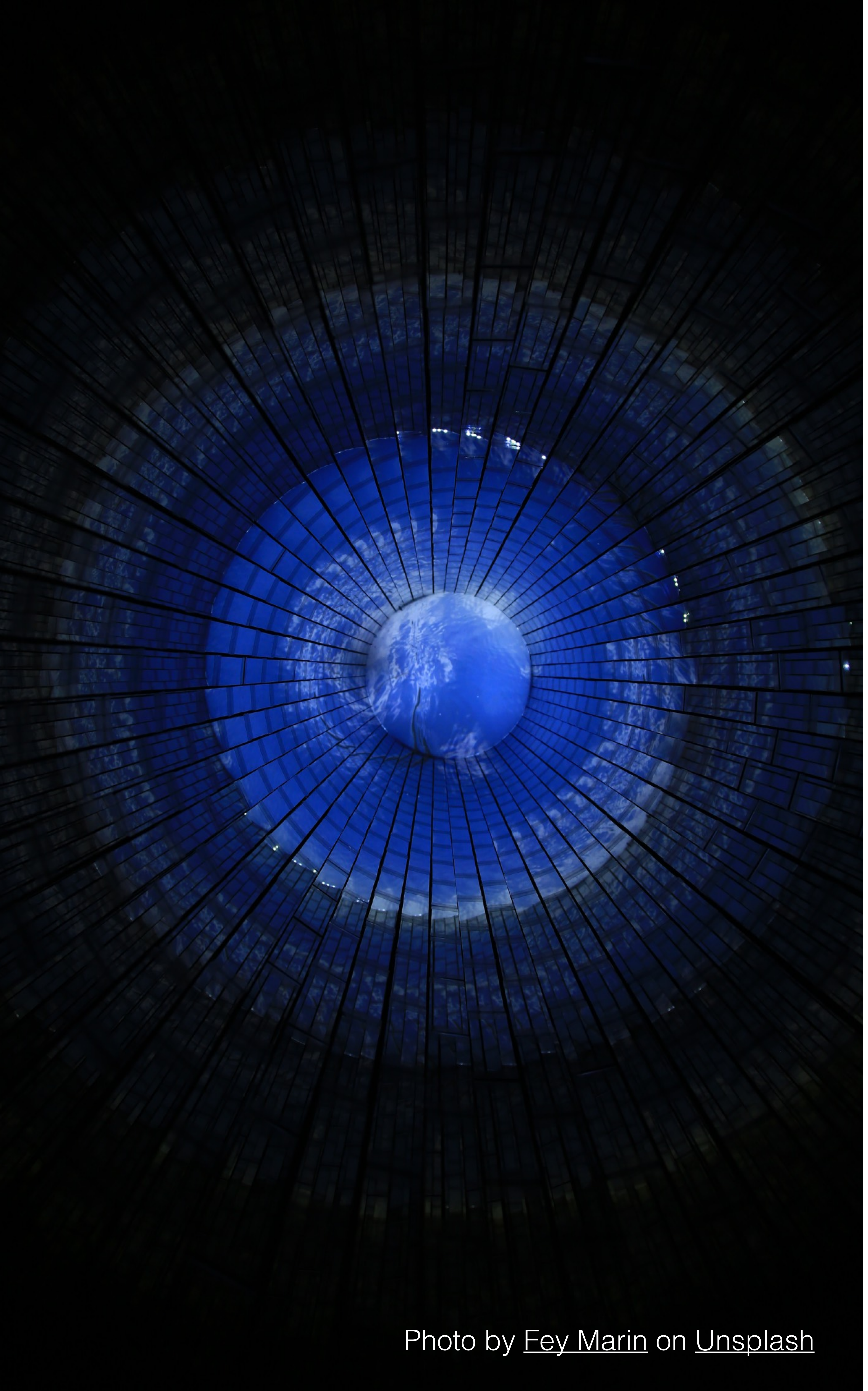


Photo by Fey Marin on Unsplash

The **score** is the derivative (gradient) of the log-likelihood function with respect to the parameter (vector) $\boldsymbol{\theta}$.

Interpretation: The score indicates how sensitive a likelihood function is to its parameter(s) $\boldsymbol{\theta}$.

Theorem: Under regularity conditions, the expectation of the score is zero.

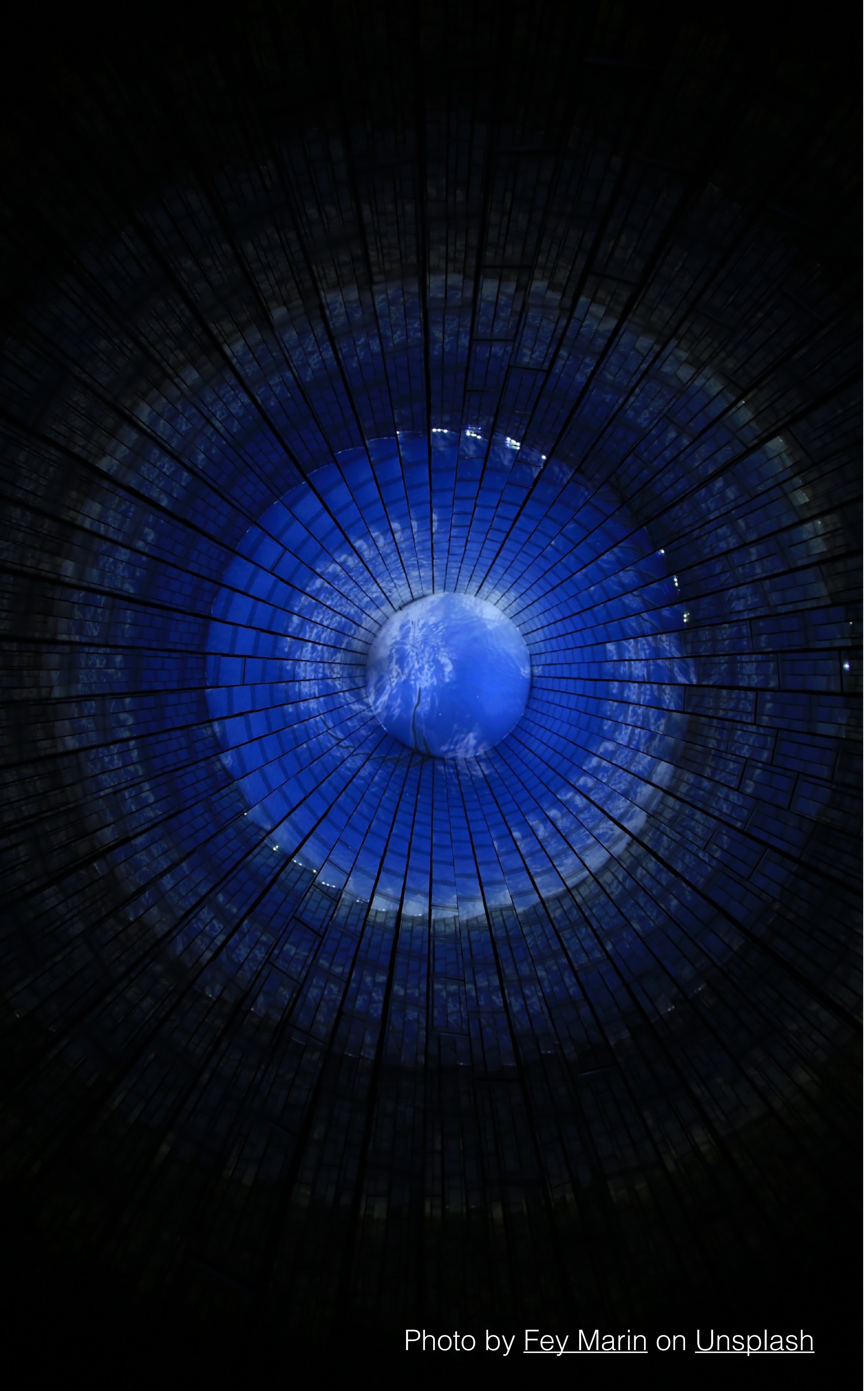
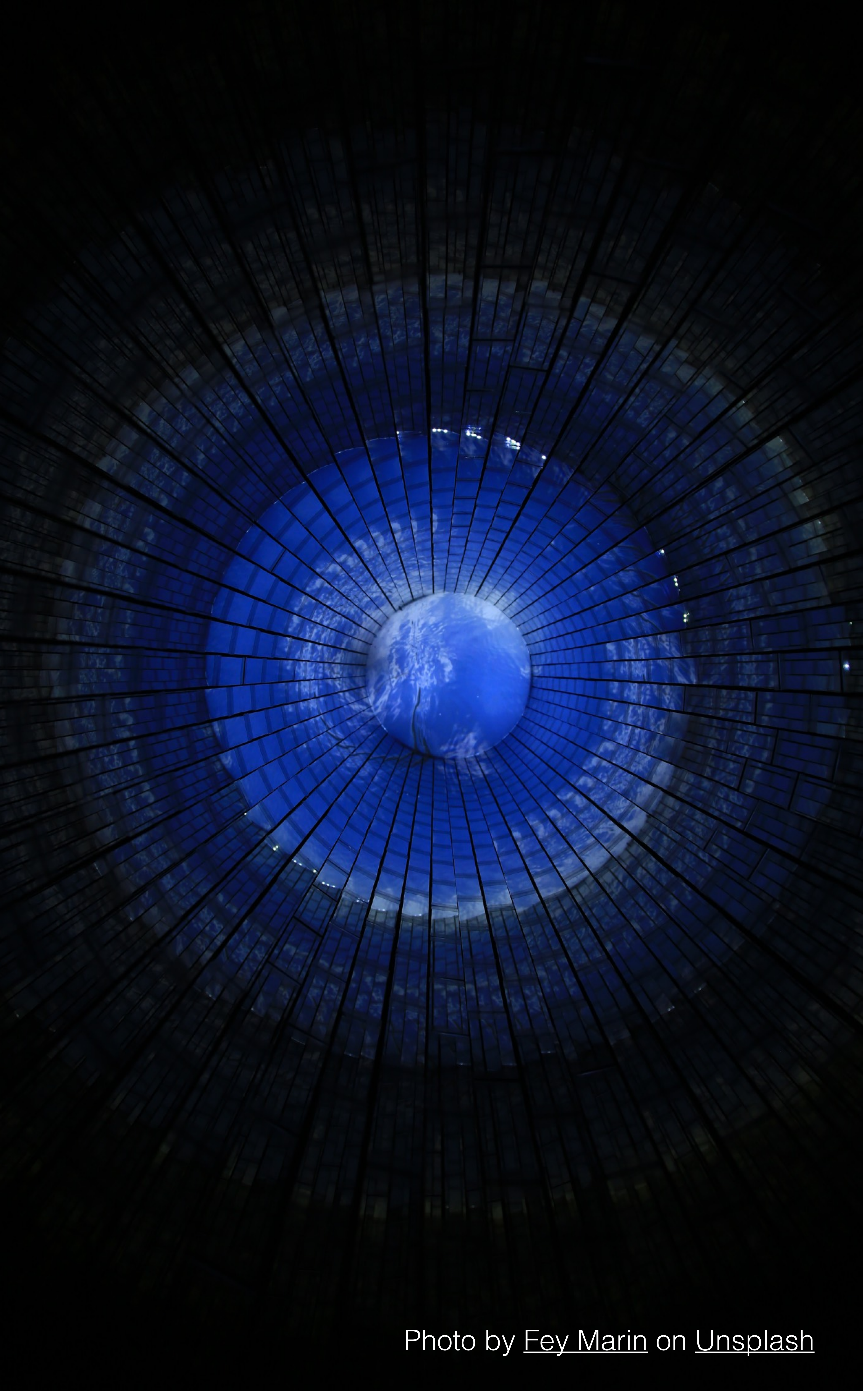


Photo by Fey Marin on Unsplash

The **Fisher Information** is defined as the variance of the score:

Interpretation: The Fisher Information is a measure of the curvature of the graph of the log likelihood. Near the MLE, high Fisher Information suggests that the max is “sharp”; low Fisher Information suggests that the max is “dull”.

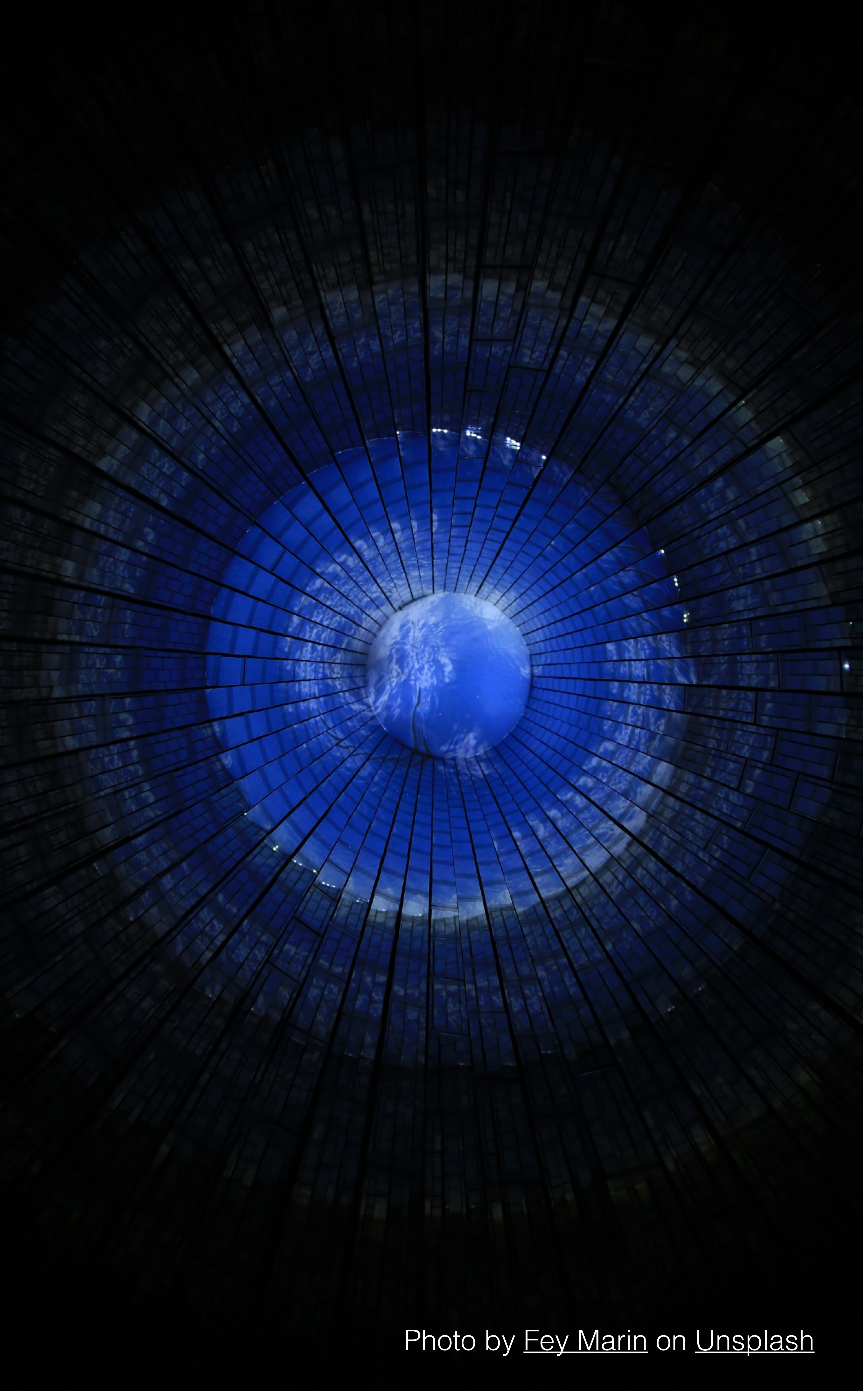
The Fisher information can also be written as:



To show that Jeffrey's prior is invariant to transformation, consider $X_1, \ldots, X_n \overset{iid}{\sim} f(\mathbf{x} \mid \theta)$ and $\pi_\theta(\theta) \propto \sqrt{I_\theta(\theta)}$.

Consider a reparameterization/transformation $\gamma = g(\theta)$. $\pi_\theta(\theta)$ is invariant to transformation if

$\pi_\gamma(\gamma) = \pi_\theta(\theta) \left| \dfrac{d\theta}{d\gamma} \right|$. $\pi_\gamma(\gamma) \propto \sqrt{I_\gamma(\gamma)}$ satisfies this relationship.



Let $Y \,|\, \lambda \sim Poisson(\lambda)$. Derive the Jeffreys' prior for λ.