Robust strategies to address the uncertainty of the response variable in Optimal Experimental Design

Sergio Pozuelo-Campos
Mariano Amo-Salas   **Víctor Casero-Alonso**

University of Castilla-La Mancha
Spain

Usual Assumption in OED Theory

$$E[y] = \mu = \eta(x;\theta)$$

$$y \sim \mathcal{N}(\mu, \sigma^2) \quad \text{with constant } \sigma^2$$

Usual Assumption in OED Theory

$$E[y] = \mu = \eta(x;\theta)$$

~~$y \sim \mathcal{N}(\mu, \sigma^2)$ with constant σ^2~~ ✂

Fisher Information Matrix (for a single-point)

$$I(x;\theta) = -E\left[\frac{\partial^2 \log \overbrace{d(y;\eta(x;\theta))}^{\text{pdf}}}{\partial\theta_i \partial\theta_j}\right] = \nu(\eta(x;\theta)) \underbrace{f(x;\theta)}_{\frac{\partial\eta(x;\theta)}{\partial\theta}} f(x;\theta)^T$$

Fisher Information Matrix (for a single-point)

$$I(x;\theta) = -E\left[\frac{\partial^2 \log \overbrace{d(y;\eta(x;\theta))}^{\text{pdf}}}{\partial\theta_i\partial\theta_j}\right] = \nu(\eta(x;\theta))\underbrace{f(x;\theta)}_{\frac{\partial\eta(x;\theta)}{\partial\theta}}f(x;\theta)^T$$

$$-E\left[\frac{\partial^2 \log d(y;\eta(x;\theta))}{\partial\eta(x;\theta)^2}\right]$$

Elemental Information Matrix (EIM)

(Atkinson et al. 2014)

Fisher Information Matrix (for a single-point)

$$I(x;\theta) = -E\left[\frac{\partial^2 \log \overbrace{d(y;\eta(x;\theta))}^{\text{pdf}}}{\boxed{\partial\theta_i\partial\theta_j}}\right] = \nu(\eta(x;\theta))\underbrace{f(x;\theta)}_{\frac{\partial\eta(x;\theta)}{\partial\theta}} f(x;\theta)^T$$

$$-E\left[\frac{\partial^2 \log d(y;\eta(x;\theta))}{\boxed{\partial\eta(x;\theta)^2}}\right]$$

Elemental Information Matrix (EIM)

(Atkinson et al. 2014)

OED for Any Probability Distribution

Design

$$\xi = \left\{ \begin{matrix} x_1 & \ldots & x_q \\ w_1 & \ldots & w_q \end{matrix} \right\} \in \Xi, \quad \sum_{i=1}^{q} w_i = 1,$$

FIM for a Design

$$M(\xi;\theta) = \int_{x \in \mathcal{X}} I(x;\theta)\xi(x)dx$$

Criterion Function

$$\Phi[M(\xi;\theta)]$$

Optimal Design

$$\xi^* = \arg \underset{\xi \in \Xi}{\text{mín}}\, \Phi[M(\xi;\theta)]$$

OED for Any Probability Distribution

Design

$$\xi = \left\{ \begin{matrix} x_1 & \dots & x_q \\ w_1 & \dots & w_q \end{matrix} \right\} \in \Xi, \quad \sum_{i=1}^{q} w_i = 1,$$

FIM for a Design

$$M(\xi;\theta) = \int_{x\in\mathcal{X}} I(x;\theta)\xi(x)dx$$

Criterion Function

$$\Phi[M(\xi;\theta)]$$

Optimal Design

$$\xi^* = \mathop{\mathrm{arg\,mín}}_{\xi\in\Xi} \Phi[M(\xi;\theta)]$$

D-Optimality

$$\Phi[M(\xi;\theta)] = \log\det M^{-1}(\xi;\theta)$$

Sensitivity Function

$$\varphi(x,\xi;\theta) = m - \nu(\eta(x;\theta))f(x;\theta)^T M^{-1}(\xi;\theta)f(x;\theta)$$

Heteroscedastic Normal Distribution

$$y_i \sim \mathcal{N}(\mu, \underbrace{\sigma_i^2}_{\mathrm{Var}[y] = kE[y]^{2r}})$$

Heteroscedastic Normal Distribution

$$y_i \sim \mathcal{N}(\mu, \underbrace{\sigma_i^2}_{\text{Var}[y] = kE[y]^{2r}})$$

k	r	Var[y]		Distribution
1	0	1	$\rightarrow$	Homoscedastic Normal
1	0.5	$E[y]$	$\sim$	Poisson
1	1	$E[y]^2$	$\sim$	Exponential
$1/\alpha$	1	$E[y]^2/\alpha$	$\sim$	Gamma

(D-)Efficiency analysis

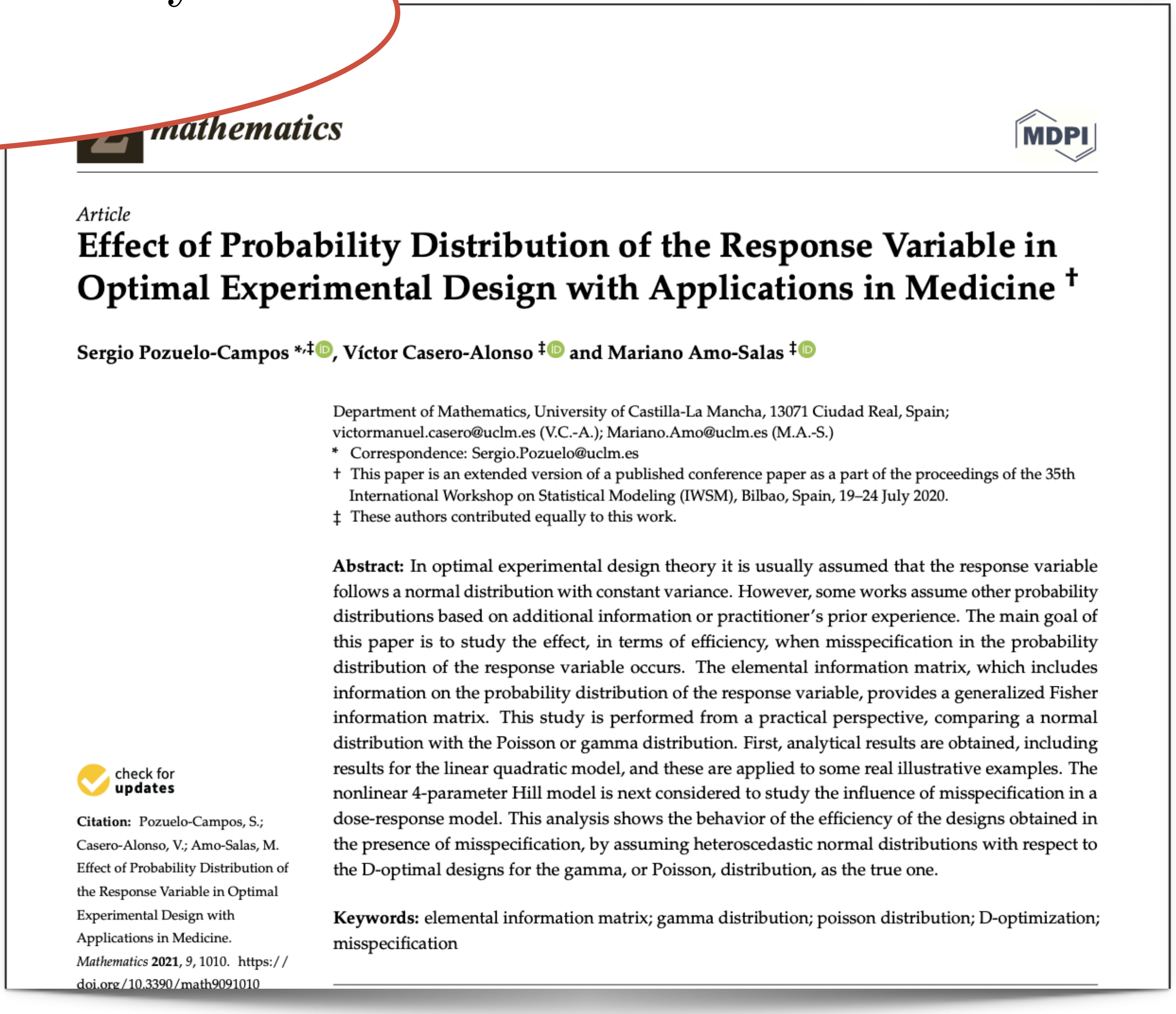
$$\mathrm{eff}_D(\xi_A^*|\xi_R^*) = \left(\frac{\det M_R(\xi_A^*)}{\det M_R(\xi_R^*)}\right)^{1/m}$$

(D-)Efficiency analysis

$$\mathrm{eff}_D(\underset{\substack{\uparrow\\ \textbf{Assumed}\\ \textbf{pdf}}}{\xi_A^*}\,|\,\underset{\substack{\uparrow\\ \textbf{Real}\\ \textbf{pdf}}}{\xi_R^*}) = \left(\frac{\det M_R(\xi_A^*)}{\det M_R(\xi_R^*)} \right)^{1/m}$$

I do not doubt about probability distribution

Be careful!!



mathematics

MDPI

Article

Effect of Probability Distribution of the Response Variable in Optimal Experimental Design with Applications in Medicine †

Sergio Pozuelo-Campos *,‡, **Víctor Casero-Alonso** ‡ and **Mariano Amo-Salas** ‡

Department of Mathematics, University of Castilla-La Mancha, 13071 Ciudad Real, Spain; victormanuel.casero@uclm.es (V.C.-A.); Mariano.Amo@uclm.es (M.A.-S.)
* Correspondence: Sergio.Pozuelo@uclm.es

† This paper is an extended version of a published conference paper as a part of the proceedings of the 35th International Workshop on Statistical Modeling (IWSM), Bilbao, Spain, 19–24 July 2020.

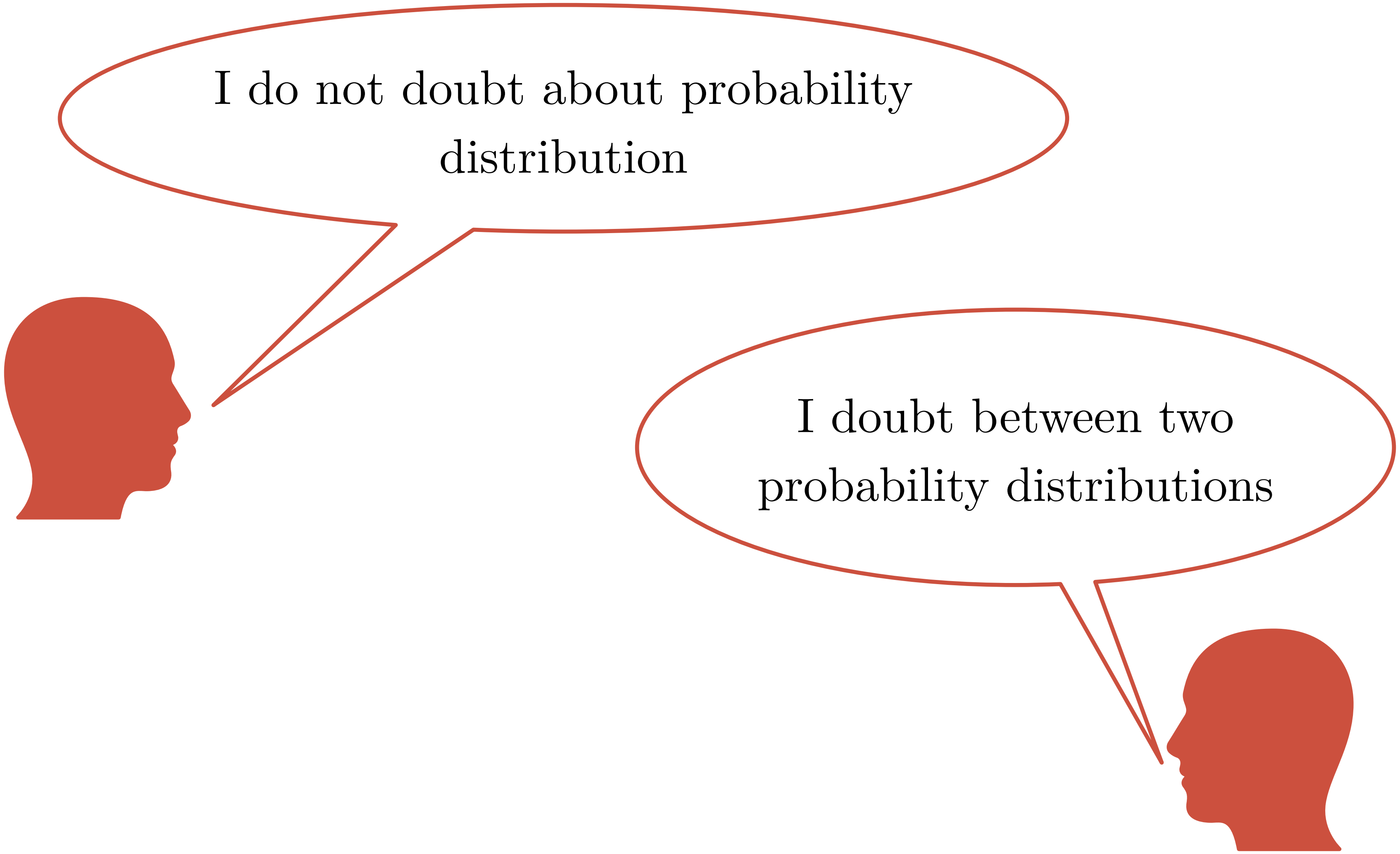
‡ These authors contributed equally to this work.

Abstract: In optimal experimental design theory it is usually assumed that the response variable follows a normal distribution with constant variance. However, some works assume other probability distributions based on additional information or practitioner's prior experience. The main goal of this paper is to study the effect, in terms of efficiency, when misspecification in the probability distribution of the response variable occurs. The elemental information matrix, which includes information on the probability distribution of the response variable, provides a generalized Fisher information matrix. This study is performed from a practical perspective, comparing a normal distribution with the Poisson or gamma distribution. First, analytical results are obtained, including results for the linear quadratic model, and these are applied to some real illustrative examples. The nonlinear 4-parameter Hill model is next considered to study the influence of misspecification in a dose-response model. This analysis shows the behavior of the efficiency of the designs obtained in the presence of misspecification, by assuming heteroscedastic normal distributions with respect to the D-optimal designs for the gamma, or Poisson, distribution, as the true one.

Keywords: elemental information matrix; gamma distribution; poisson distribution; D-optimization; misspecification

check for updates

Citation: Pozuelo-Campos, S.; Casero-Alonso, V.; Amo-Salas, M. Effect of Probability Distribution of the Response Variable in Optimal Experimental Design with Applications in Medicine. *Mathematics* **2021**, *9*, 1010. https://doi.org/10.3390/math9091010


I do not doubt about probability distribution
I doubt between two probability distributions

4 robust alternatives
to minimize the loss of efficiency

3-parameters Hill model

$$E[y] = \frac{Vx^n}{x^n + K^n}$$

$n = 1 \quad \longrightarrow \quad$ Michaelis-Menten model

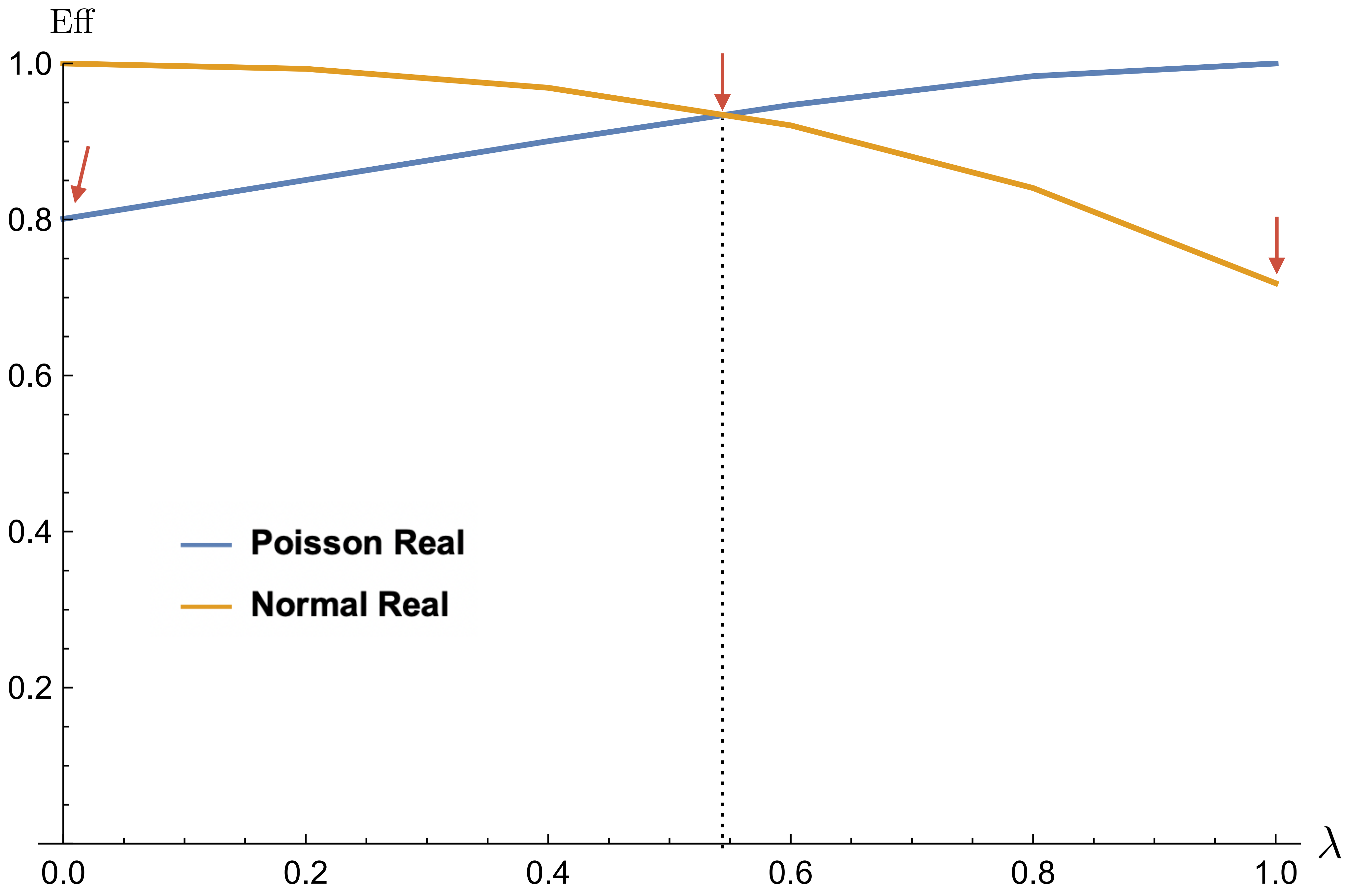
I. Compound designs

$$\left(\mathrm{eff}_A(\xi|\xi^*)\right)^{\lambda} \cdot \left(\mathrm{eff}_B(\xi|\xi^*)^{(1-\lambda)}\right)$$

$$\lambda \in [0,1]$$

Both D-optimal criterion,
each for a different probability distribution

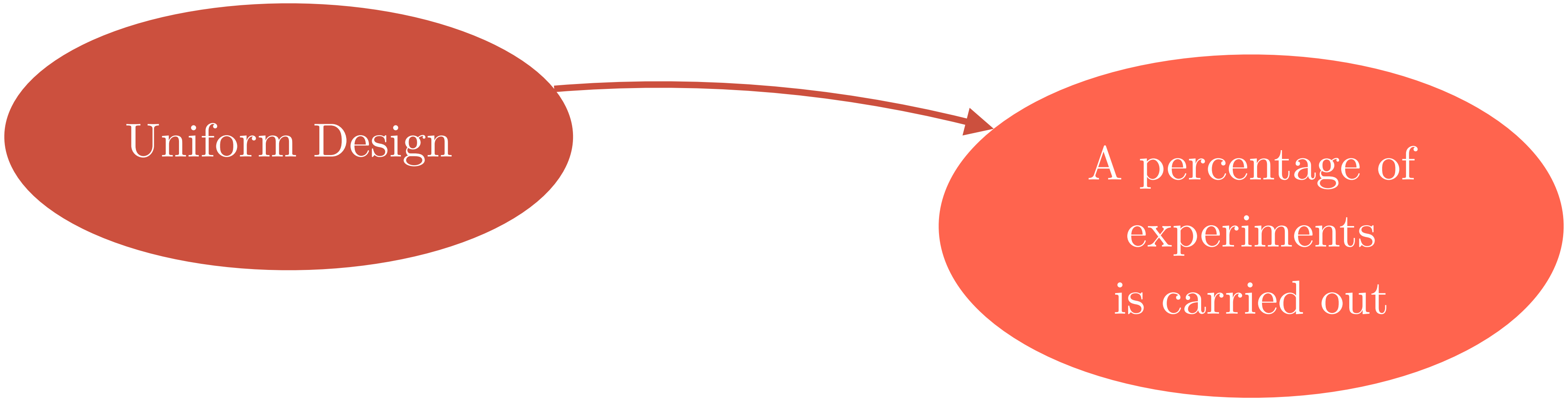
I. Compound designs



II. Multistage designs

Uniform Design

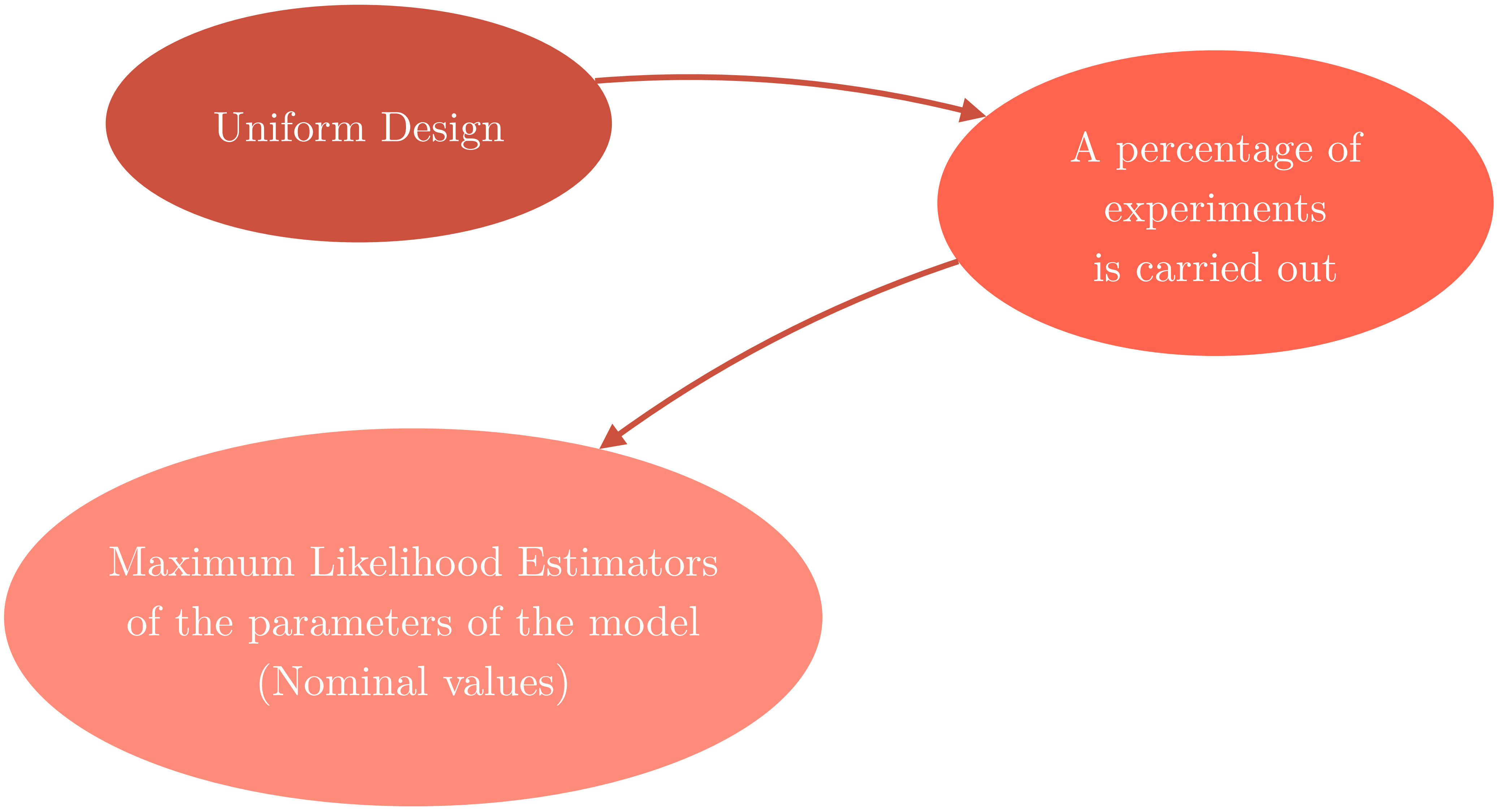
II. Multistage designs



Uniform Design

A percentage of
experiments
is carried out

II. Multistage designs

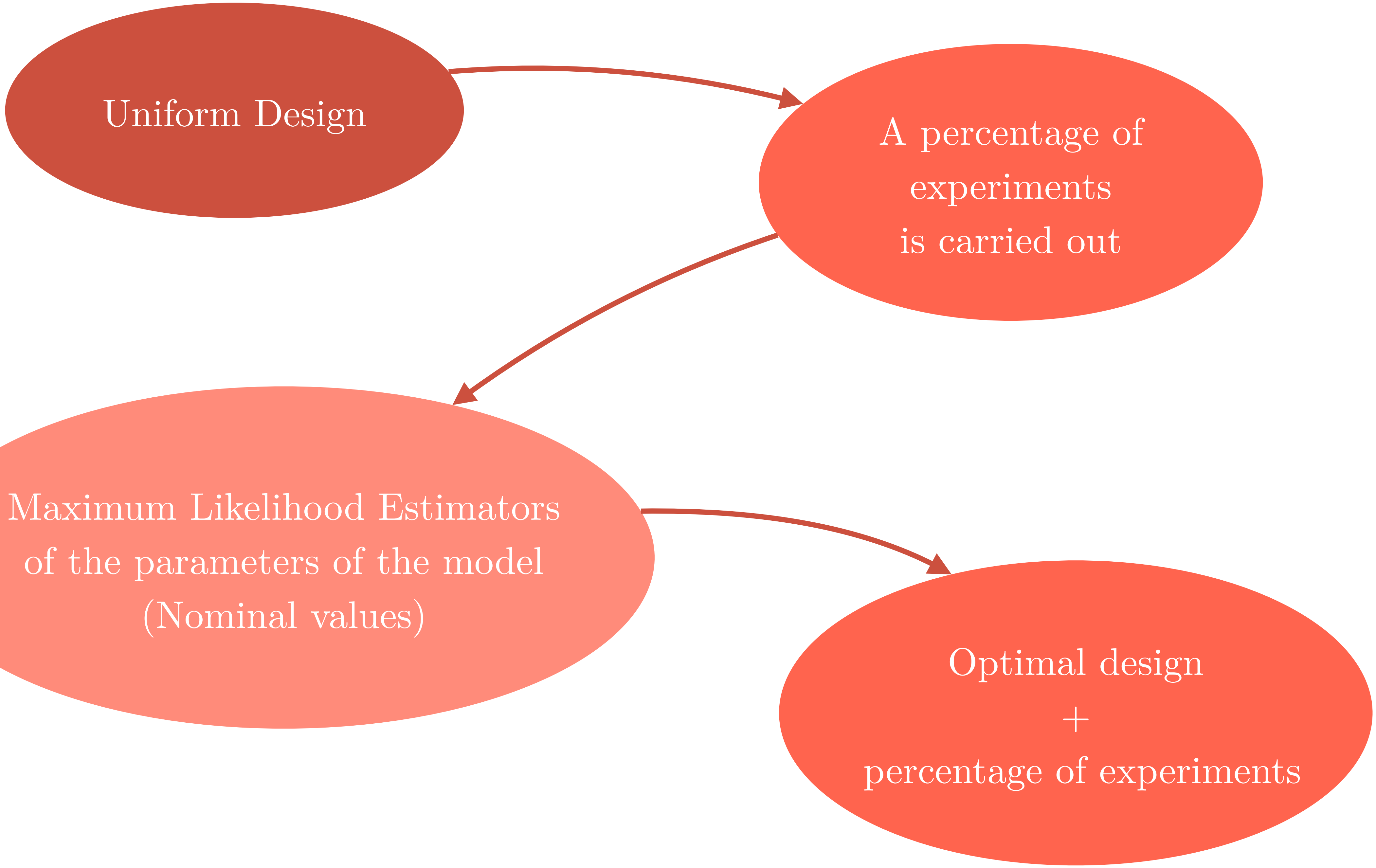


Uniform Design

A percentage of
experiments
is carried out

Maximum Likelihood Estimators
of the parameters of the model
(Nominal values)

II. Multistage designs



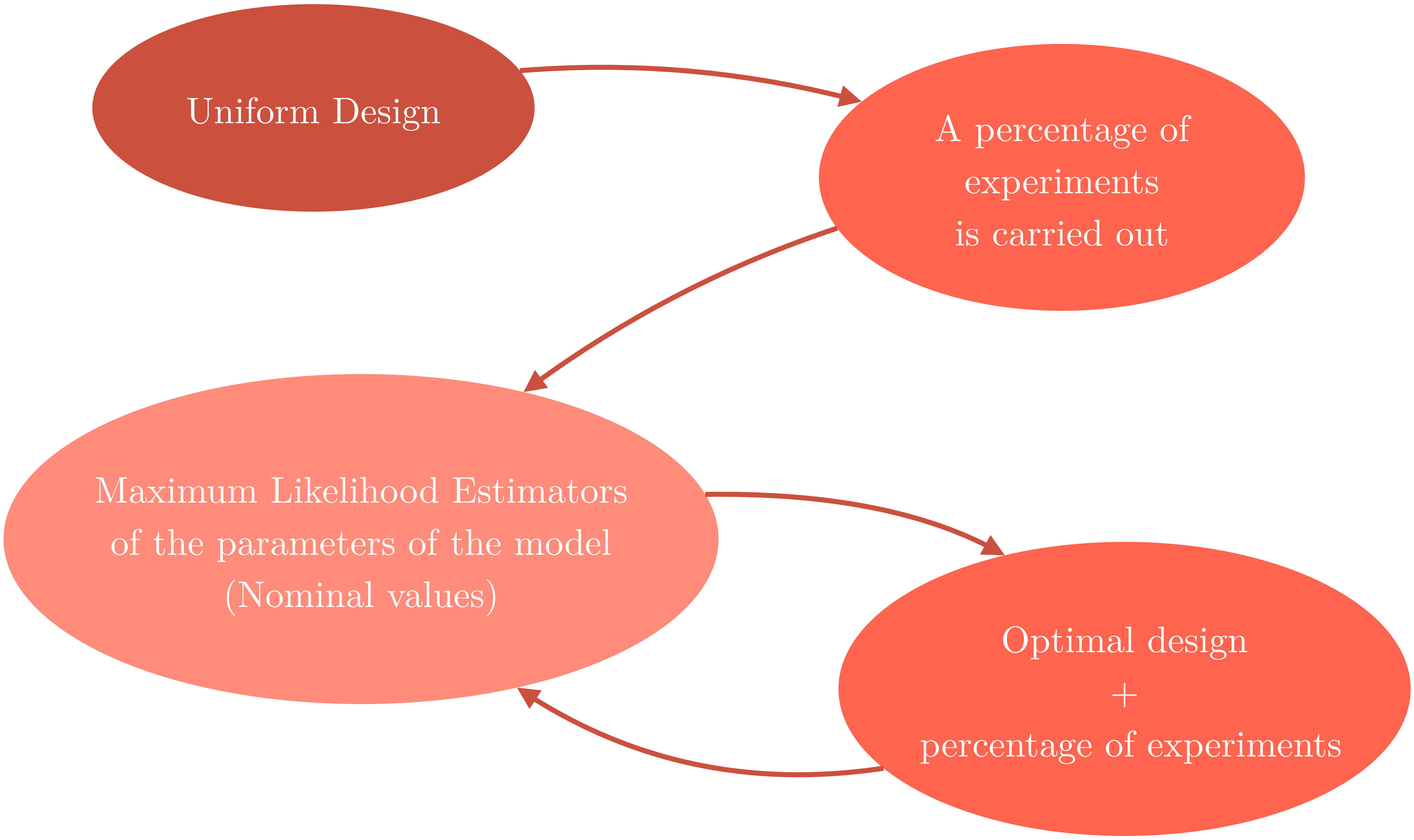
Uniform Design

A percentage of
experiments
is carried out

Maximum Likelihood Estimators
of the parameters of the model
(Nominal values)

Optimal design
+
percentage of experiments

II. Multistage designs



Uniform Design

A percentage of
experiments
is carried out

Maximum Likelihood Estimators
of the parameters of the model
(Nominal values)

Optimal design
+
percentage of experiments

III. Design based on MqLE

(Maximum quasi-likelihood estimators)

$$\mathrm{Var}[y_i] = \phi v(E[y_i])$$

III. Design based on MqLE

(Maximum quasi-likelihood estimators)

$$\mathrm{Var}[y_i] = \phi v(E[y_i])$$

$$g(x) = v(E[y_i])^{(-1/2)} \frac{\partial \eta(x;\theta)}{\partial \theta}$$

III. Design based on MqLE

(Maximum quasi-likelihood estimators)

$$\mathrm{Var}[y_i] = \phi v(E[y_i])$$

$$g(x) = v(E[y_i])^{(-1/2)} \frac{\partial \eta(x;\theta)}{\partial \theta}$$

$$\Delta(\theta;\xi) = \frac{1}{\phi} \sum_i g(x_i) g(x_i)^T$$

Information Matrix
Based on MqLE

III. Design based on MqLE

(Maximum quasi-likelihood estimators)

$$\mathrm{Var}[y_i] = \phi v(E[y_i])$$

$$g(x) = v(E[y_i])^{(-1/2)} \frac{\partial \eta(x;\theta)}{\partial \theta}$$

$$\Delta(\theta;\xi) = \frac{1}{\phi} \sum_i g(x_i) g(x_i)^T$$

Information Matrix
Based on MqLE

↓

D-optimization

IV. Multiple Objetive Annealing Algorithm

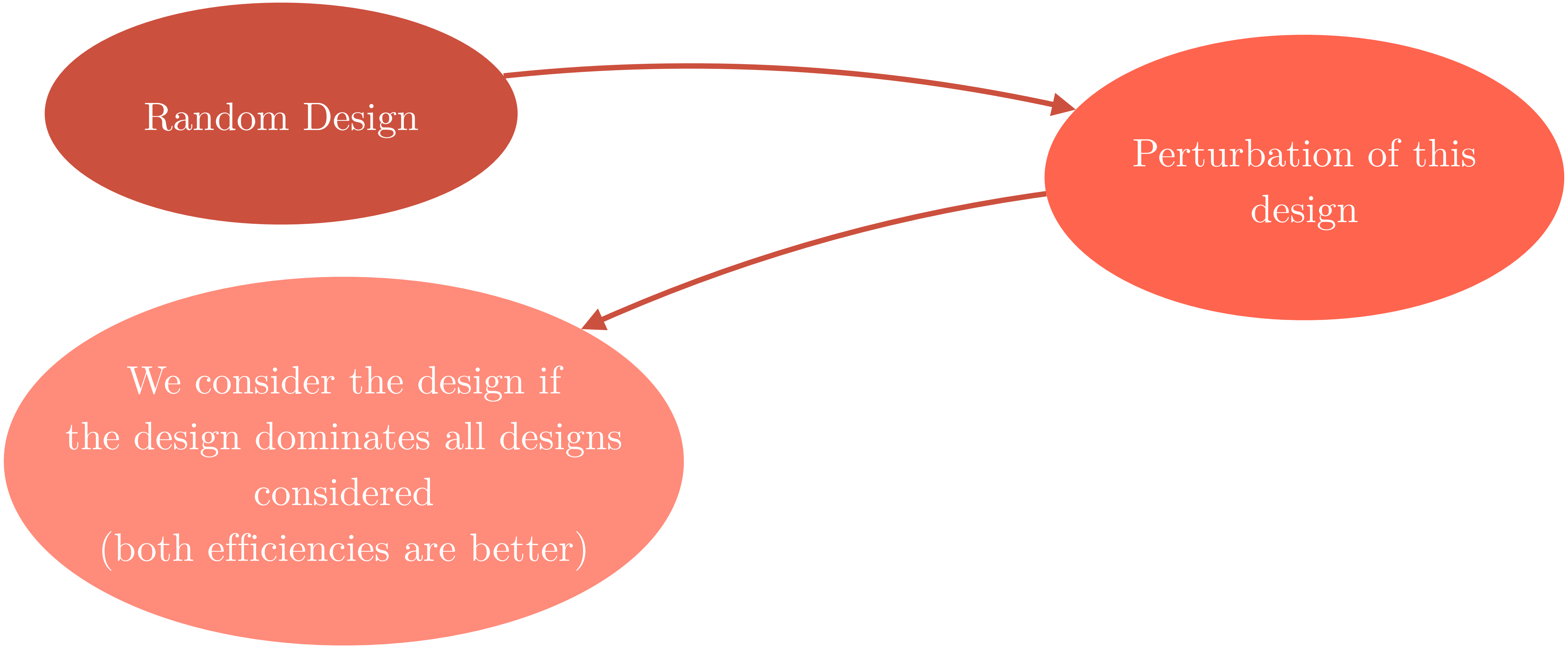
$$\begin{tikzcd}
 & \max \operatorname{eff}(\xi|\xi_{\mathcal{A}}) \\
\xi \arrow[ur] \arrow[dr] & \text{(Objective functions)} \\
 & \max \operatorname{eff}(\xi|\xi_{\mathcal{B}})
\end{tikzcd}$$

IV. Multiple Objetive Annealing Algorithm

Random Design

IV. Multiple Objective Annealing Algorithm

IV. Multiple Objective Annealing Algorithm

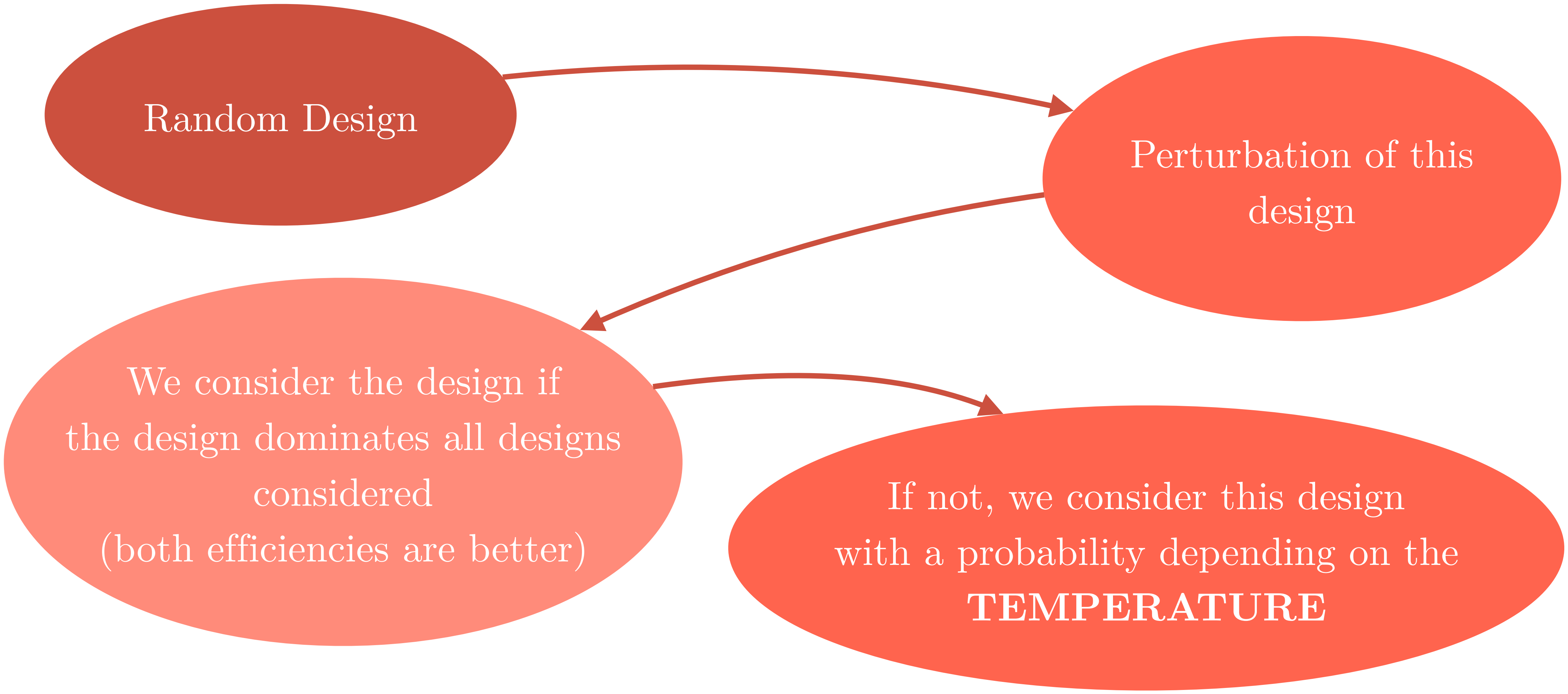


Random Design

Perturbation of this design

We consider the design if
the design dominates all designs considered
(both efficiencies are better)

IV. Multiple Objetive Annealing Algorithm



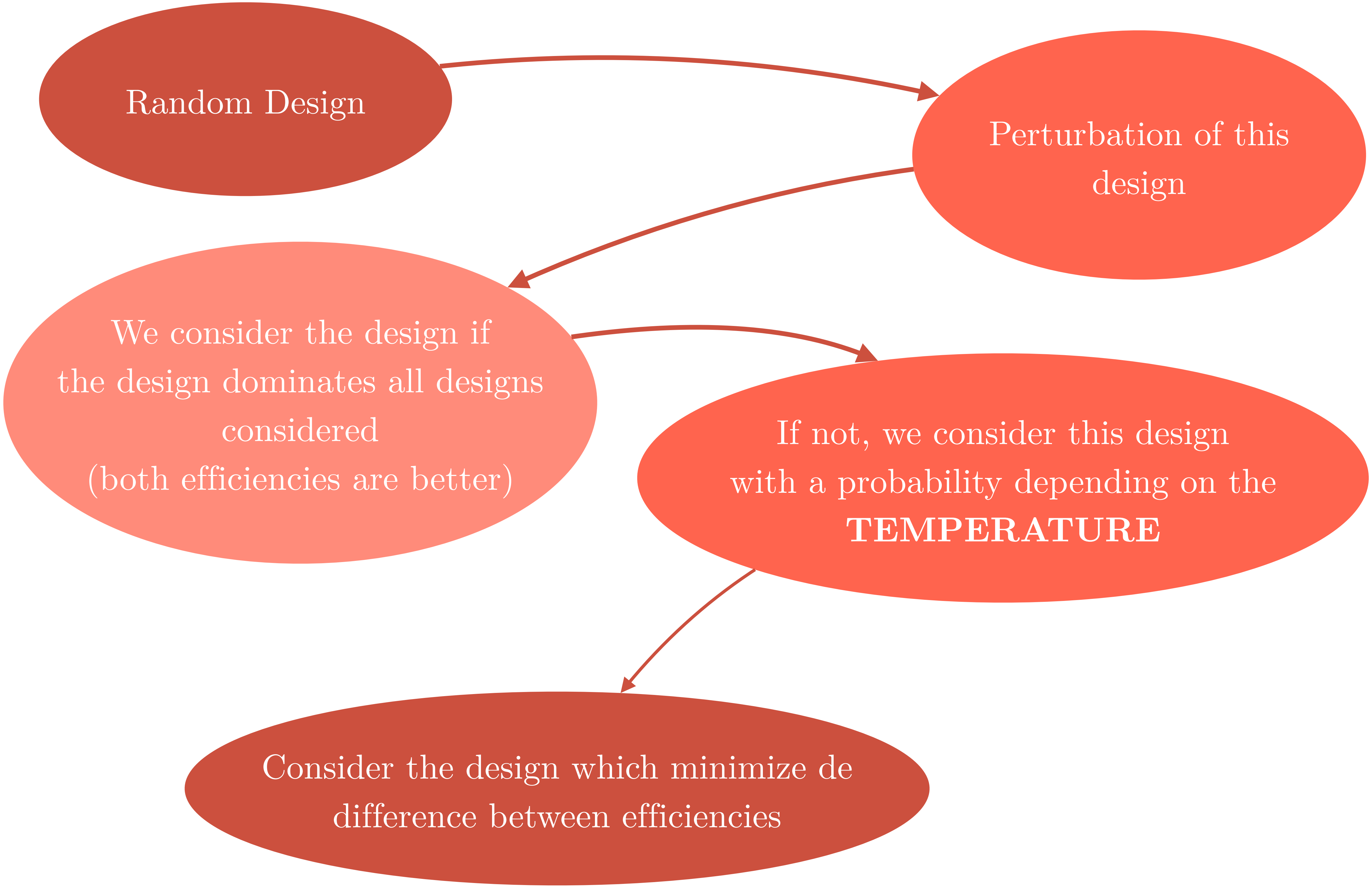
Random Design

Perturbation of this design

We consider the design if the design dominates all designs considered (both efficiencies are better)

If not, we consider this design with a probability depending on the **TEMPERATURE**

IV. Multiple Objective Annealing Algorithm



Random Design

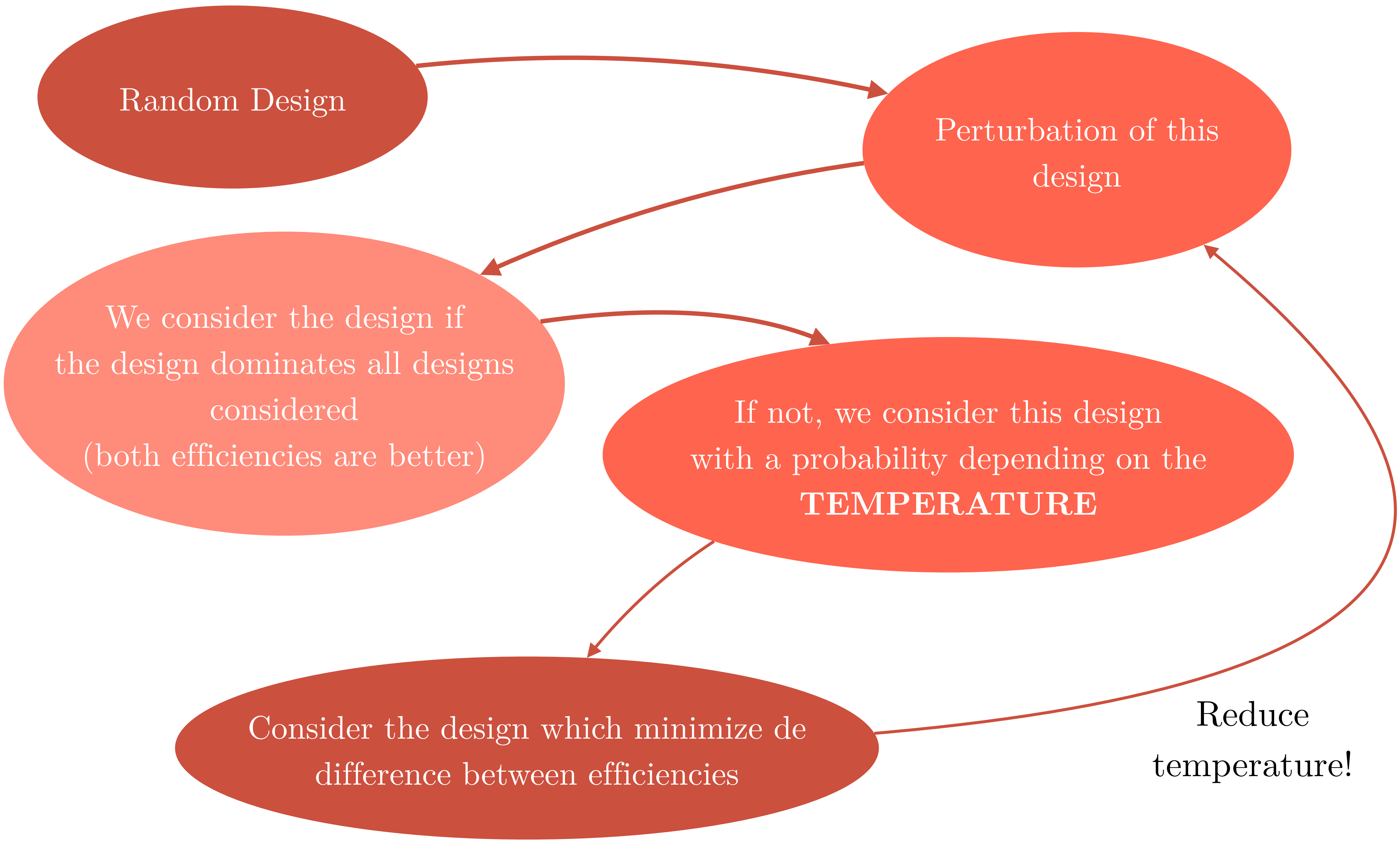
Perturbation of this design

We consider the design if the design dominates all designs considered
(both efficiencies are better)

If not, we consider this design with a probability depending on the **TEMPERATURE**

Consider the design which minimize de difference between efficiencies

IV. Multiple Objetive Annealing Algorithm



Random Design

Perturbation of this design

We consider the design if the design dominates all designs considered
(both efficiencies are better)

If not, we consider this design with a probability depending on the **TEMPERATURE**

Consider the design which minimize de difference between efficiencies

Reduce temperature!

Efficiencies of different alternatives

Poisson Vs Normal **heteroscedastic**

$$\mathrm{eff}(\xi_{\mathcal{N}}|\xi_{\mathcal{P}}) = 0.7596 \qquad \mathrm{eff}(\xi_{\mathcal{P}}|\xi_{\mathcal{N}}) = 0.8144$$

	I	**II**	**III**	**IV**
Poisson	0.9093	0.9444	1	0.9377
Heteroscedastic normal	0.9023	0.8801	0.8144	0.8847

Efficiencies of different alternatives

Poisson Vs Normal **homoscedastic**

$\text{eff}(\xi_{\mathcal{N}}|\xi_{\mathcal{P}}) = 0.8008$ $\text{eff}(\xi_{\mathcal{P}}|\xi_{\mathcal{N}}) = 0.7183$

	I	**II**	**III**	**IV**
Poisson	0.9357	0.5704	0.8008	0.9352
Homoscedastic normal	0.9353	0.8953	1	0.9352

Efficiencies of different alternatives

Poisson
Vs
Normal heteroscedastic

$$\text{Var}[y] = E[y]$$

	I	II	III	IV
Poisson	0.9093	0.9444	1	0.9377
Heteroscedastic normal	0.9023	0.8801	0.8144	0.8847

Poisson
Vs
Normal homoscedastic

	I	II	III	IV
Poisson	0.9357	0.5704	0.8008	0.9352
Homoscedastic normal	0.9353	0.8953	1	0.9352

“Future” work...

Thank you very much for your attention!