No-Free-Lunch theorems for reliability analysis

Mohsen Rashki[1*], Matthias G.R. Faes[2]
[1]*Department of Architecture Engineering, University of Sistan and Baluchestan, Zahedan, Iran.*
[2]*Chair for Reliability Engineering, TU Dortmund University, Dortmund, Germany.*

Abstract

In the most of engineering problems, because of the lack of complete information about the structure of the performance function, selection of the optimal approach for efficient reliability analysis is in essence a decision under uncertainty. This issue is investigated in this paper and, by representing reliability methods as search algorithms, No-Free-Lunch theorems (NFL) of search and optimization are used to propose similarly NFL for reliability analysis. Using NFL, this study aims to answer some basic questions about the existence and the selection of optimal reliability methods for black/grey-box problems and proposes a mathematical framework for the application of detection theory in structural reliability. Black and grey-box problems in this context refer to structural reliability problems with respectively no and partial information on the topology of the limit state function. Then, by employing Dempster-Shafer theory of evidence as a generalized Bayesian decision making theorem, a practical "experts-in-the-loop" approach for the selection of an optimal reliability method in uncertain conditions is proposed. To meet this aim, providing a step-by-step solution of some selection problem examples, it is shown that knowledge of several experts can be fused into one all-encompassing knowledge representation to reduce the probability of making an error in the selection of an optimal approach for efficient reliability analysis.

Keywords: Failure probability, optimization, No-Free-Lunch theorems, Human Reliability, Detection Theory, Data Fusion, decision making

INTRODUCTION

Due to many developments in the last decades, structural reliability theory is considered as a rational tool for analysis and design of engineering systems in the presence of uncertainties. Most applications of structural reliability theory consist of performing safety analysis, as well as determining the probability and mode of failure of engineering systems. Nonetheless, the method can also be used to estimate the occurrence of rare events, perform the estimation of uncertain model quantities or aid decision making processes under uncertainty; see e.g., (Bartsoen et al., 2023; Ghasemi et al., 2019).

Many structural reliability algorithms have been developed for probability estimation. These can be roughly classified into two main groups: A) Design point-based reliability methods, that often decouple probability estimation into optimization and reliability phases, and B) one-stage simulation approaches that often estimate probability by the combination of reliability and optimization process using random sampling (Valdebenito et al., 2010). Considering this classification, ongoing discussions between supporters of each category are being had to this date. These discussions stem from the underlying beliefs that methods from category (A) outperform methods from category (B) in term of efficiency (see next section for a definition), or vice versa, that all methods from category (A) are by default less worthy than those of category (B).

In this context, considering many alternative approaches for solving an in-hand problem, reliability analysts are often faced with some fundamental and important questions:

- Among many available reliability approaches, which one is the best (in terms of efficiency) for solving a problem without any information about the structure of the problem (known as black-box problems)?

- Does any approach exist that presents high efficiency for solving all types of reliability problems?
- How to select an optimal approach for solving a problem when some information about the structure of the problem is available (defined here as a grey-box problem)?
- How to select an optimal approach for efficiently solving a problem when there is conflict among several practitioners about picking the right algorithm from a set of alternatives?

The main contribution of this paper is to study these questions and provide a mathematical framework to answer the proposed basic questions. For this purpose, in the next section, we present reliability algorithms as a search process. Then, we show that No-Free-Lunch theorems (NFL) in search and optimization (D. Wolpert & Macready, 1996, 1997) can be adapted for application in probability estimation and reliability analysis. It is clarified that the reliability analysis is a human-in-the-loop process and the optimization skills, reliability knowledge, sensations, emotions and capabilities of the practitioner in decision-making are also a part of a reliability problem. Further, using human reliability analysis, we develop an application of signal detection theory in structural reliability analysis. Then, the application of decision-making methods in selection of optimal reliability method is studied. Based on the available information of a problem and comments of experts, different decision-making approaches (e. g., Theory of evidence (P. Li & Wei, 2019), Analytic hierarchy process (AHP) (Ataei et al., 2013), Multiple criteria decision analysis (MCDM) (G.-D. Li et al., 2007), Bayesian inference using maximum entropy (Gull, 1988) and etc.) may be used for solving grey-box problems. Also, considering the versatility of the theory of evidence for application in selection problems, evidence theory is adapted to use in the selection of optimal algorithm. We finally provide some discussions and conclusions. As a final note to this paper, we do acknowledge that it does not give a complete answer to the question: “which algorithm is the best among a set of alternatives". It rather is intended to serve as new view-point to deal with this pressing question, and aims at giving some practical tools to aid in the case-specific selection of an optimal algorithm for the problem at hand.

RELIABILITY METHODS AS SEARCH ALGORITHMS

Let $F \subset \mathcal{R}^n$ be the failure domain of a system with standard normal random variables. Then, the failure probability of the system under consideration can be presented as a multinormal integral as follows:

$$\mathrm{P(F)} = \int_{F=\{g(\boldsymbol{u})\leq 0\}} \varphi_n(\boldsymbol{u})d\boldsymbol{u}, \tag{1}$$

where $g(.)$ is the performance function, $F = \{g(\boldsymbol{u}) \leq 0\}$ represents the failure of the system and, $\varphi_n(.)$ is the probability density function (PDF) of the random variables $\boldsymbol{U}$. Here, a point that is located in the limit state surface $g(\boldsymbol{u}) = 0$ and for which $\varphi_n(\boldsymbol{u})$ is maximal, is defined as the design point $\boldsymbol{u}^*$ (generally known as most probable point of failure, MPP, if it is in failure domain *F*) and its distance to the origin is known as reliability index β (e.g., $\beta = |\boldsymbol{u}^*|$). To solve the problem introduced in Eq. (1), two theories have been developed:

- For efficiently solving the proposed integral, Asymptotic approximation theory represents the design point as $\boldsymbol{u}^* = \beta\boldsymbol{u}$, where $|\boldsymbol{u}| = 1$, and proves: if for a problem there are *k* design points $\boldsymbol{u}_i$, i=1, …, *k* given with $|\boldsymbol{u}_i| = 1$, then the probability content is concentrated in the asymptotically neighborhoods of design points and we have:

$$\mathrm{P}(\beta\mathrm{F}) \sim \sum_{i=1}^{k} P\big(\beta(D_i^{\varepsilon} \cap F)\big), \quad \beta \longrightarrow \infty, \tag{2}$$

where $D_i^{\varepsilon} = \{|\boldsymbol{u} - \boldsymbol{u}_i| < \varepsilon\}$ are balls with radius ε around the design points $\boldsymbol{u}_1, \dots, \boldsymbol{u}_k$ (Breitung, 1994, 2021). According to asymptotic approximation theory, robust algorithms are essentially the same in one elementary step, namely, they require to find a global minimal distance point in domain F since the probability content is concentrated around it. Employing $c(\boldsymbol{u}) = |\boldsymbol{u}|$, this point can be determined as optimization process as follows:

$$\underset{\boldsymbol{u} \in F}{\arg\min}\, c(\boldsymbol{u}), \tag{3}$$

which is a standard constrained optimization problem.

- Parallel to the asymptotic theory, a set of methods have been developed to solve the problem using a different perspective. If we consider S as the safety of system, the failure probability can be presented as $\mathrm{P(F)} = 1 - \mathrm{P(S)}$. Rather than evaluating the probability weight of the entire failure region, the proposed approaches obtain some information about the failure and safe domain (usually by random sampling) and then, they provide an approximation about the failure probability by combination of results using postprocessing (e. g., using expectation theorem and/or Bayesian computations (Dang, Valdebenito, et al., 2022; Rashki, 2021a)). To find the upper bound of safety probability, similar to the previous approach, these methods also need to find the failure domain but they do not require to know the accurate location of design points and their exploring often lead to have few information about the neighborhoods of MPPs.

From the preceding discussion, it can be deduced that, to provide a proper probability estimation for both strategies, a search process should be included in reliability analysis. This statement is further investigated in the following subsections with more details.

Categorization of reliability algorithms based on search behaviors

Considering the proposed explanations, in this subsection we explain how existing reliability methods use search/optimization process in probability evaluation.

Decoupled reliability approaches

In the structural reliability literature, there are some well-known reliability approaches that estimate the failure probability by employing optimization and reliability phases as two stand-alone processes. In these methods, which are called decoupled approaches in this study, it does not matter how the design points (or MPPs) in the optimization phase are determined. According to the asymptotic theory, they only use design points to approximate the failure probability in a separate reliability phase. The original versions of first/second order reliability methods (FORM/SORM) (Breitung, 1984; Hasofer & Lind, 1974), line sampling (Schuëller et al., 2004), design point-based importance sampling (Bourgund, 1986), beta-sphere subset simulation (BESS) (Rashki, 2021a) and radial basis importance sampling (RBIS) (Harbitz, 1986) are examples of such decoupled reliability methods.

For the case of simple problems (i.e., linear and moderate nonlinear performance functions with one MPP), it is possible to use a gradient-based optimization algorithm (which often initializes the search process from the origin) with high fidelity to find design points for reliability analysis. Having design points in hand, they provide an estimation of failure probability with a very high efficiency, which is the reason that these approaches often known as efficient reliability approaches. However, for complicated problems, and problems with several important failure regions (such as first-excursion probability problems), this tactic may give erroneous results and lead to misleading conclusions, if a solution can be even obtained at all. Therefore, for such problems, the search process should be started from several random initial points to assure the accuracy of results which reduces the efficiency of algorithm. Please note that we use the term "important failure regions" to indicate that design points may be meaningless as a proxy for calculating the failure probability in high dimensional spaces or strongly nonlinear problems. With "important failure regions" we indicate those regions in parameter space that contribute most significantly to the calculation of the failure probability.

The key drawback of these approaches is that we have not any statistical analyses regarding the correctness and accuracy of probability estimation since the reliability calculations are completely relying on the accuracy of the optimization phase. Furthermore, in the case where the optimization problem is not convex, there is no guarantee regarding the accuracy of the optimization results. Therefore, for the case of very complex or black-box problems, if a design point search process is performed using a few optimizations, the accuracy of obtained results must be always in doubt. In recent years, also applications of meta-heuristic global optimization methods (with relatively high computations costs) have been developed for addressing the discussed issue (Elegbede, 2005; Zhong et al., 2020). Nonetheless, also here, no proof of having obtained the global optimum exists, and hence, the accuracy of the obtained design points must also be doubted. As a final remark, it should be noted that even for relatively uncomplicated problems, it is prudent to run multiple analyses from different starting points in case it cannot be proven a priori whether the optimization problem is convex.

Random search-based reliability approaches

In some reliability analysis approaches, in contrast with decoupled methods, probability computations and search process are merged together. As described earlier, these approaches often explore the safe region and find the important failure domains of the problem (i.e., neighborhoods of MPPs) by a sequence of random sampling and then, using the information obtained during the search process (for both S and F domain), they estimate the desired probability of failure using an ensuing post-processing step.

The most popular approach that estimates the failure probability by random sampling, is the crude Monte Carlo simulation (MCS). MCS works by drawing random samples according to φ_n and mathematically will converge to the proper solution with a probability of one given sufficient samples. The MCS generator searches the entire space and does not consistently ignore any region, ensuring convergence without any prior assumptions on the topology of the limit state surface or its dimension (Zabinsky, 2009). As mentioned, MCS collects information about the safe domain by random sampling. Once sufficient samples in the failure domain are collected, it uses the obtained information during a search process to approximate the failure probability using post processing (e. g., by combining the information obtained from both safe and failure domain) as $\mathrm{P(F)} = \frac{n_f}{n_f+n_s} = \frac{n_F}{N}$, where n_f and n_s are the number of samples generated in respectively the failure F and safety S domains and $N = n_f + n_s$. Also, similar formulations exist for assessing the variance on this estimator. When MCS estimates the failure probability with suitable variance, one may find that the information on the important failure domain, as provided by algorithm, is limited. Conversely, the information about the safe domain is statistically nearly complete.

MCS is generally efficient when the failure domains are located very close to the origin since few samples N are generated to obtain a sufficiently large n_f. Rather, for problems where the failure domains are located far from the origin, crude MCS is computationally extremely costly. To address this drawback, different random sampling MCS methods have been developed (the so-called variance reduction methods), including, e.g., directional simulation (Melchers, 1990), subset simulation (Au & Beck, 2001), weighted average simulation (Rashki et al., 2012), sequential importance sampling (Papaioannou et al., 2016), soft Monte Carlo (Rashki, 2021c), directional importance sampling (Misraji et al., 2020), Multilevel Monte Carlo Simulation (Callens et al., 2022), sequential space conversion methods (Rashki, 2021b), and Bayesian Optimization methods (Dang, Wei, et al., 2022), among many other approaches. All of these approaches can be also presented as search algorithms that explore the safe domain and seek for (areas around the) MPPs to provide an approximation of the failure probability.

For example, traditional subset simulation uses a modified random walk algorithm to explore the safe domain according to φ_n and find important failure regions. The method generates initial random samples around the origin and then, considering certain portion of generated samples as seeds, it uses a modified Metropolis Hastings algorithm to search the parameter space to find important failure domains within several subsets. Once the method has found neighborhoods of MPPs (e. g., using m search subsets in the safe domain), the method estimates the desired probability by postprocessing. Hereto, it uses the combination of results obtained from failure and safe domain as $\mathrm{P(F)} = \mathrm{P}(F_1)\prod_{i=1}^{m-1}\mathrm{P}(F_{i+1}|F_i)$, where $\mathrm{P}(F_\mathrm{i})$ can be obtained by counting the number of conditional random search points located in the factitious failure domain (Au & Beck, 2001). In this formulation, $\prod_{i=1}^{m-1}\mathrm{P}(F_{i+1}|F_i)$ is the information obtained from exploring in the safety domain S.

As a result, in these approaches, instead of computing the failure probability by only employing the failure region (e. g., design points and MPPs), a considerable part of the estimation is obtained from the information of safe/failure domain, obtained in random search process. Evidently, when the location of design points can be obtained in an efficient manner (e. g., efficient gradient-based optimization algorithm), random search may not be a clever approach for estimating small failure probabilities. The same conclusion can be found in (Breitung, 2021) and (Breitung, 2022) with more details. Note that this statement is limited to low to moderate dimensional problems. Indeed, in non-compressible high-dimensional spaces, due to their geometry and topology, the design point in and of itself might not be a useful proxy for calculating the failure probability. Also, in these cases, the value of a potential physical interpretation of a design point is questionable.

Efficiency of algorithms

Efficiency is an objective parameter that is often used to compare the performance of two (or more) algorithms together. For the case of structural reliability analysis, the efficiency of an estimator $\hat{\mu}$ for a parameter μ, denoted $\mathrm{Eff}(\hat{\mu})$ can be defined as (Breitung, 2021):

$$\mathrm{Eff}(\hat{\mu}) = [\mathrm{MSE}(\hat{\mu}) \times Co(\hat{\mu})]^{-1}, \tag{4}$$

where $\mathrm{MSE}(\hat{\mu}) = \mathrm{var}(\hat{\mu}) + (\mathrm{E}[\hat{\mu}] - \mu)^2$ is the mean-square error of $\hat{\mu}$, $(\mathrm{E}[\hat{\mu}] - \mu)$ is a term denoting the bias of the estimator and $Co(\cdot)$ is a term denoting the expected function call associated with computing $\hat{\mu}$ (e.g., the number of samples in a Monte Carlo estimator). In other words, when two unbiased estimators $\hat{\mu}_1$ and $\hat{\mu}_2$ have the same function call, then the estimator with the smallest variance is to be selected. In the case of a simple MC estimator, the efficiency is proportional to the number of samples n. In this case, it is given as $\mathrm{Eff}(\hat{\mu}) = 1/[\mathrm{MSE}(\hat{\mu}) \times n]$. In case the estimator is biased, the asymptotic efficiency as $n \to \infty$ has to be taken into account. In this paper, the efficiency is considered as a main feature of the "optimal algorithm", compared with other alternative algorithms, for efficiently solving reliability problems.

<u>NO-FREE-LUNCH THEOREMS IN RELIABILITY ANALYSIS (NFLR)</u>

According to the explanations proposed in previous section and as discussed in (Breitung, 2021) and (Breitung, 2022), one may conclude that for small to medium sized problems, both decoupled approaches and random search-based simulation methods are searching neighborhoods of MPPs to approximate the probability of failure. Building on this result, we aim to translate the original No-Free-Lunch (NFL) theorems in search and optimization (see (D. Wolpert & Macready, 1996, 1997)) towards reliability analysis to answer the basic questions proposed in introduction.

NFLR #1: There is no-free-lunch for black-box reliability problems.

According to the NFL theorem, for a black-box problem with no prior knowledge about the function at hand, the information collected with the data sample is not helpful in guiding the search in which direction is better to explore next (Serafino, 2013), (D. H. Wolpert, 2021). As a result, all algorithms that search for an extremum of a cost function perform exactly the same when averaged over all possible cost functions. Therefore, when we consider reliability algorithms as mostly consisting of a searching process, no reliability algorithm is preferable to others in term of function calls (e. g., no decoupled reliability method is preferable to a random search approach and vice versa.) (D. Wolpert & Macready, 1997). Hence, "Algorithm A outperforms algorithm B in term of efficiency" is a misleading statement since if an algorithm performs well on a certain class of problems, then, according to NFL, it necessarily will perform poorly on the set of all remaining problems.

Therefore, for each newly proposed reliability algorithm, including several efficiently solved examples, the authors should specify in their publication for what set of problems the algorithm is tailored, and which of those are considered in analysis. We even would argue that there is value in highlighting to which types of examples the method does *not* work efficiently.

Proof. As a first step, we exclude a part of the un-important space that has very small effect on the failure probability estimation (e.g., space with failure probability less than 10^{-20}). By doing this, we can reduce the infinite physical probability space to a finite search space X (for instance, bounding the physical space of two random variable to [$L^{-1}(10^{-20})$ $L^{-1}(1-10^{-20})$] where L^{-1} is inverse cumulative density function of the random variable). Based on this, we can show, as presented in the previous section, that at the core of any robust reliability approach lies an optimization algorithm α which functions based on mapping some previously visited set of samples d^x to a single new sample in X, as:

$$\alpha : d \in \mathcal{D} \rightarrow \{x | x \notin d^x\}, \tag{5}$$

in which $\mathcal{D}$ denotes the space of all (m-sized) samples and α is determistic in the sense of every sample map to a unique new point [29]. The NFL theorem says that for any pair of algorithms α_1 and α_2, the overall performance of algorithms over all possible cost functions C with uniform probability density is equal to:

$$\sum_C P\left(d_m^y | C, m, \alpha_1\right) = \sum_C P\left(d_m^y | C, m, \alpha_2\right), \tag{6}$$

where $P\left(d_m^y | C, m, \alpha_i\right)$ is the performance of algorithm *i* after m iterations and d_m^y is the associated cost. The mathematical proof of Eq. (6) can be found in Ref (D. Wolpert & Macready, 1997).

Remark. The number of function calls, associated in phase two of decoupled reliability methods (i.e., the post-processing) is not considered in this section.

NFLR #2: Reliability analysis is a human-in-the-loop process and the importance of human reliability for efficient analysis is equal to all knowledge about the in-hand problem.

As shown in Fig 1, this study emphasizes that reliability analysis is a human-in-the-loop process and hence, human reliability analysis plays a main role for achieving optimal results for a grey-box problem (i.e., a problem where *some* knowledge on the limit state function is available). The selection of the optimal algorithm requires both proper information, as well as a thorough understanding about the geometry of the considered performance functions. On top, it requires expertise of the analyst to use an unbiased estimator to match with the available information of the problem. Especially concerning the last point, different researchers often select different approaches to solve the same problem, based on differences in background knowledge, sensations, memories, emotions and the availability and/or implementation complexity of codes. This often leads to obtaining diverse results for the same problem, including the determined value of the failure probability and its variance and/or the required number of function evaluations. This issue is schematically illustrated in Fig 2.

The NFL says when a practitioner fails to incorporate the information of the problem into the optimization algorithm, even for simple problems with complete information about the structure of the problem, there is no-free-lunch in reliability analysis (D. Wolpert & Macready, 1997). For instance, no matter how much information is available for an in-hand problem, there is no free lunch in the analysis when the analyst believes that his/her favorite algorithm is always the best solution for solving all types of reliability problems. Due to this predetermined mindset, for such analyst, there is no difference between a grey and black box problems.

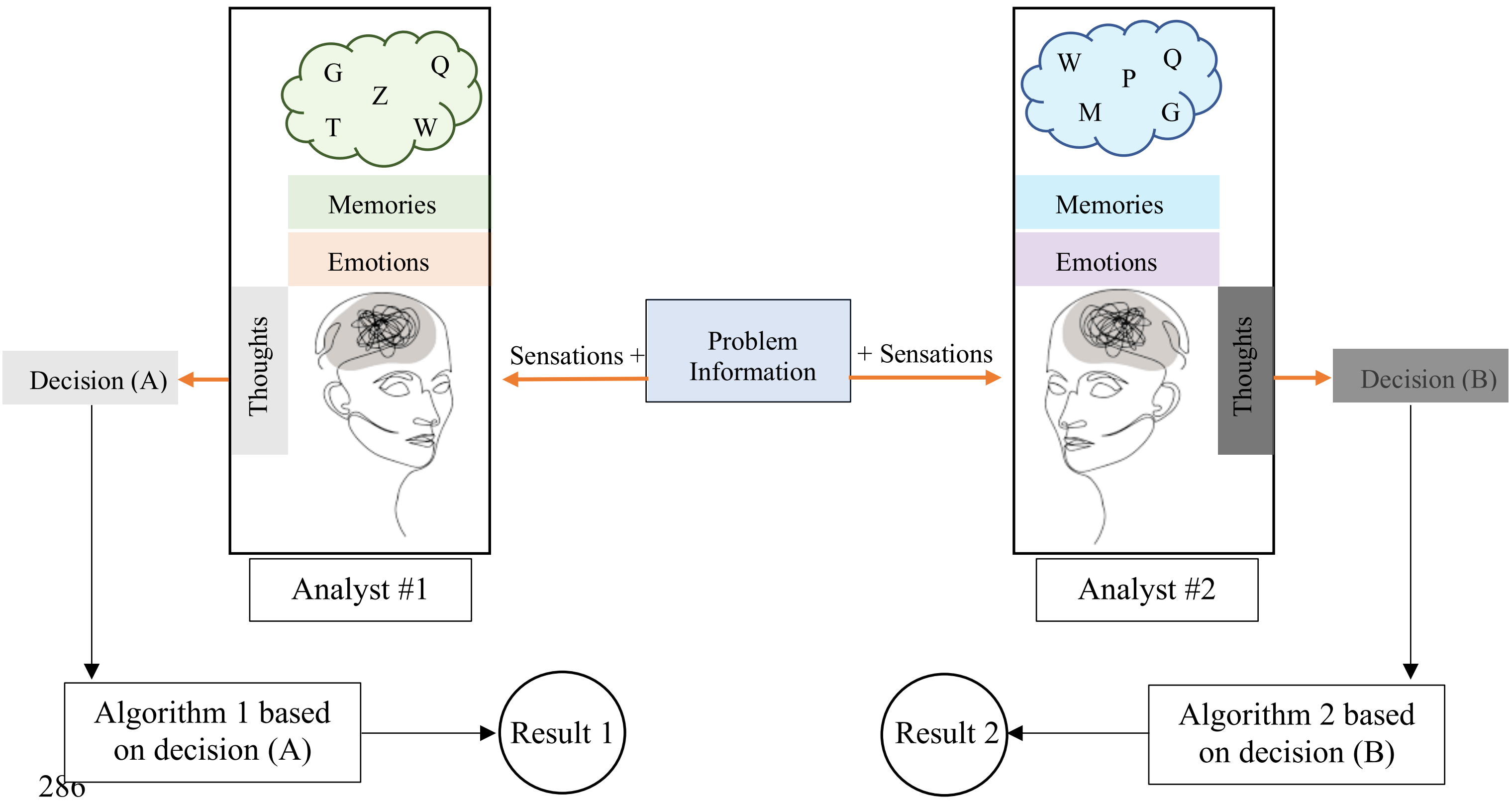


Fig 1. Decision making to solve a problem under uncertainty: Different sensation, emotions and memories results in using different algorithms to solve the problem

Proof. For the space X of all cost functions, NFL represents the probability of obtaining a certain d_m^y by algorithm α after m iteration as follows (D. Wolpert & Macready, 1997):

$$P\left(d_m^y \middle| m, \alpha\right) = \sum_C P\left(d_m^y \middle| m, \alpha, C\right) P(C), \tag{8}$$

In which $\mathrm{P}(C)$ is the prior probability that the optimization problem in hand has cost function C. By defining vectors $\vec{v}_{d_m^y, m, \alpha} \equiv P\left(d_m^y \middle| m, \alpha, C\right)$ and $\vec{s} \equiv \mathrm{P}(C)$, the NFL theorem represents a geometric representation of Eq. (8) as follows:

$$P\left(d_m^y \middle| m, \alpha\right) = \vec{\boldsymbol{v}}_{d_m^y, m, \alpha} \cdot \vec{\boldsymbol{s}}\,, \tag{9}$$

which means: to get the desired behavior, prior $\vec{\boldsymbol{s}}$ (i. e., all knowledge about C) must match or aligned with algorithm α (D. Wolpert & Macready, 1997). According to the NFL theorem, this need for matching is necessary for an algorithm to perform well in practice on specific problems. Otherwise, there is no such thing as free-lunch in reliability analysis.

Eq. (8) can be described in a different way: In this equation, $\mathrm{P}(C)$ can be also seen as a statement concerning the practitioners' choice of optimization algorithms (D. Wolpert & Macready, 1997). For solving a certain problem C, practitioner A will have a list of alternative algorithms $\alpha_{\boldsymbol{k}}, k = \{1,2, \dots, N\}$ in mind. In this perspective, $\alpha_{\boldsymbol{k}}$ will be a discrete random variable and its possible outcomes, denoted by V, is the set of alternative reliability approaches including the optimal approach $\{\alpha_1, \alpha_2, \dots, \alpha_i, \dots, \alpha_N\}$. The proposed implementation reveals that to get desired behavior, having knowledge about C and alternative algorithms $\alpha_{\boldsymbol{k}}$, practitioner A should select the optimal algorithm α_i, match with C, with the probability of one $P(V = \alpha_i|A) = 1$. However, considering inadequate background, experience, knowledge, sensations, emotions and capabilities in decision-making, practitioner A may fail to select the optimal algorithm α_i even for a simple problem with complete information about the in-hand problem. Here, $P(V = \alpha_i|A)$ that is the probability of employing the optimal approach α_i from practitioner A represents the human reliability in decision making under uncertainty and can be presented as $R(A) = P(V = \alpha_i|A)$.

Human reliability is the probability of humans conducting specific tasks (e. g., decision making in this study) with satisfactory performance and focuses on estimating the human error probability. The human error is the opposite of human reliability and basically is described as follows (Calixto, 2016):

$$\mathrm{P(HE)} = \frac{\text{Number of errors}}{\text{Number of errors opportunities}} \tag{10}$$

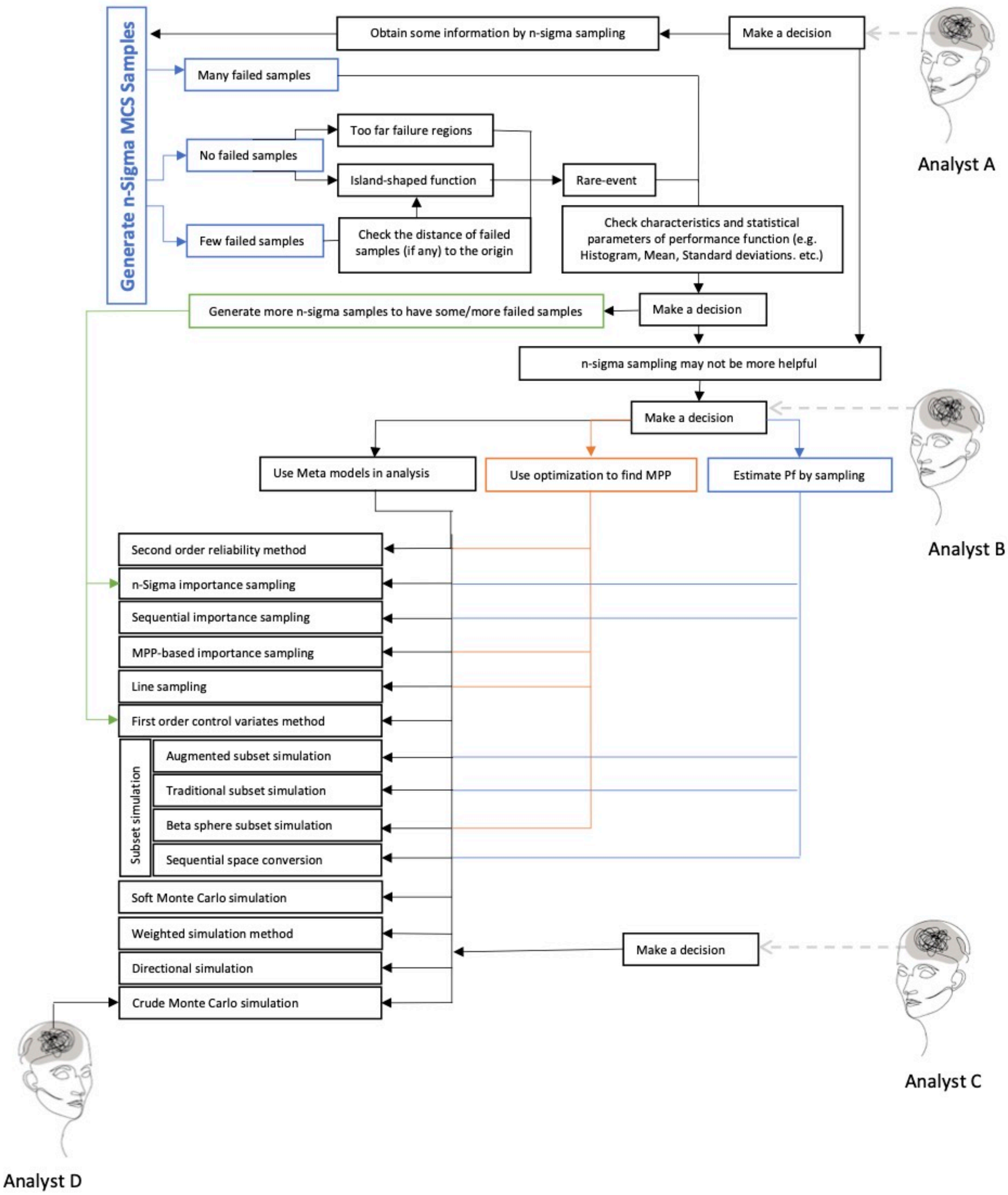
where P(HE) denoted the human error probability. According to the proposed implementations, we have P(HE)= $1 - R(A)$.

DETECTION THEORY IN STRUCTURAL RELIABILITY ANALYSIS

The importance of human reliability analysis in decision making was highlighted in the previous section. In this section, we connect it to detection theory as a well-developed theory for analysis and decision making under uncertainty. To meet this aim, we classify reliability problems into two broad categories: Type I) problems with simple linear and moderate nonlinear geometry whose important failure domains can be accurately determined by efficient gradient-based algorithms (e.g., initialized from origin) and, Type II) complex geometrical problems whose important failure domains should be found by random search algorithms (e.g., problems in non-linear dynamics). For a problem with incomplete structural information, determining the category of the problem and choosing an optimal search approach is a decision under uncertainty. Here, considering the available information, the practitioner should make a decision:

- Decision A: The failure domains can be correctly determined by efficient gradient-based algorithms, or
- Decision B: The failure domains should be searched by robust-but-costly random search algorithms.

337



338 Figure 2. Selection of a reliability method based on different attitudes: A) Having a scenario for
339 selection of optimal algorithm, B) Using sensation to determine the optimal algorithm, C)
340 Selecting an algorithm by chance, and D) Employing crude Monte Carlo for all problems.

Therefore, the main challenge in reliability analysis appears in this step: The practitioner should make a proper decision and select an algorithm that matches with the available information on the function. Otherwise, they will pay a too high cost for the analysis (no-free-lunch).

We aim at objectifying this statement by resorting to Signal detection theory (SDT). SDT migrated from radar/communication fields to medicine (Tiwari et al., 2021) and psychology (Kingdom & Prins, 2016), is a theory that investigates this issue in a mathematical way. Consider a situation where a decision-making person is faced with a single stimulus (signal) that is either faint (i.e., the strength of the signal is low), or confusing (i.e., the stimulus contradicts prior information). Based on this stimulus, the person must decide whether the signal is there or not. In fact, the situation is even more potentially confusing in case there are multiple other, uninteresting, stimuli that are similar to the original signal (we define these from now on as 'noise'). In particular, we are interested in two precise situations:

- The signal is present, and the person identifies it as signal or noise; we define these respectively as hits and misses (See Fig. 3 A, where the green color indicates a correct decision and the color red indicates an incorrect decision).
- The signal is absent, and the person identifies it as signal or noise; we define these as false alarms or correct rejections (See Fig. 3 B).

Based on these definitions, a confusion matrix (as shown in Figure 3) can be constructed based on the following definitions (Singh et al., 2021):

- True positive rate (TP): the total number of correct results or predictions when the actual class was positive (i.e., "hits").
- False positive rate (FP): the total number of wrong results or predictions when the actual class was positive (i.e., "misses")

- True negative rate (TN): the total number of correct results or predictions when the actual class was negative (i.e., “false alarms”)
- False negative rate (FN) is the total number of wrong results or predictions when the actual class was negative (i.e., “correct rejections”)

By considering the former human reliability test (decision about the selection of algorithm) as a binary decision task, the same confusion matrix (Singh et al., 2021) and probability outcomes for all problem-decision (correspond to signal-response and/or positive-negative) combinations can be obtained as shown in Fig 4.

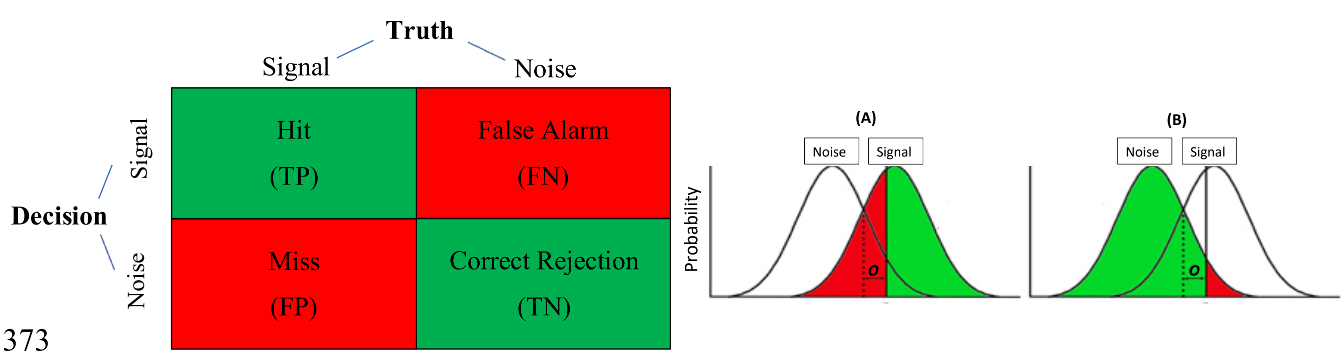


Fig 3. Confusion matrix and probability outcomes: (A) Proportions of hits and misses under signal distribution (*o* is biased), (B) Proportions of false alarms and correct rejections represented under the noise distributions (Anderson, 2015)

		Problem: Type I	Problem: Type II
Decision	Type I	$\frac{n_{(A)}}{N_{(A)}} = 1 - \phi(-\beta_I)$	$\frac{n_{(A)}}{N_{(B)}} = \phi(-\beta_{II})$
Decision	Type II	$\frac{n_{(B)}}{N_{(A)}} = \phi(-\beta_I)$	$\frac{n_{(B)}}{N_{(B)}} = 1 - \phi(-\beta_{II})$

Fig 4. Simulating confusion matrix for reliability analysis

Here, considering the two main categories of problems presented in this section, we divide the former defined P(HE) into two types: The first one, denoted as $P\left(D_{(B)}|T_I\right)$, represents the probability that decision B ($D_{(B)}$) is made for problem type I (T_I):

$$P\left(D_{(B)}|T_I\right) = \frac{n_{(B)}}{N_{(A)}}. \tag{11}$$

Accordingly, the probability that decision B ($D_{(B)}$) is made for problem type I (T_I) is presented as:

$$P\left(D_{(A)}|T_{II}\right) = \frac{n_{(A)}}{N_{(B)}}. \tag{12}$$

An example of how one can estimate these probabilities is illustrated in Fig 5. The test may be conducted by gathering a number of benchmark problems and asking a practitioner to determine the type of the problem. Then, the proposed outcomes can be mapped to a decision space using normal distributions (Fig 6) and the obtained results provide the opportunity of employing developments in detection theory in structural reliability analysis (Fig 7).

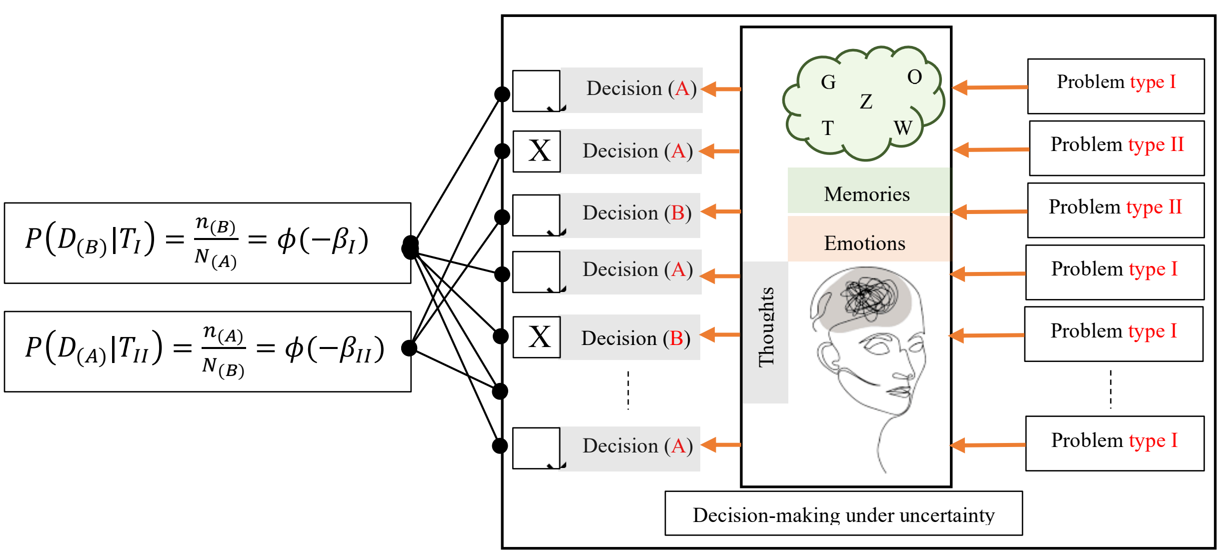


Fig 5. Human reliability test in decision making under uncertainty

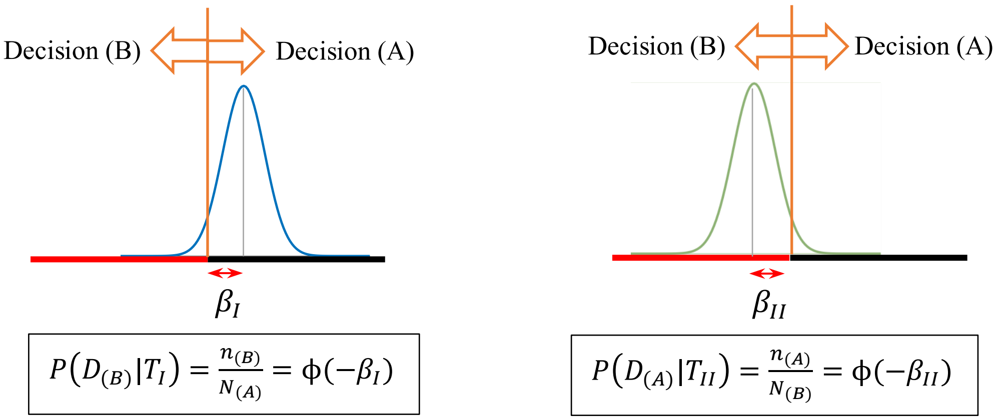


Fig 6. Mapping human reliability test to decisions space

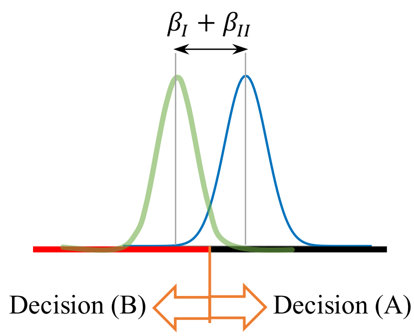


Fig 7. Representing the selection of the optimal reliability method as a detection problem

As a result, a portion of SDT that is developed for differentiating a person/classifier ability to discriminate the presence and absence of a stimulus (Colloca, 2013), can be used as a framework to assess the capabilities of the practitioner for selecting the correct reliability/optimization analysis algorithms. The most frequently used performance metrics for classification according to these values are accuracy (ACC), precision (P), sensitivity (Sn), specificity (Sp), F-score and Matthew Correlation Coefficient (MCC) values that can be calculated as follows (Al-Turjman & Deebak, 2020; Singh et al., 2021):

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \tag{13}$$

$$P = \frac{TP}{TP+FP} \tag{14}$$

$$Sn = \frac{TP}{TP+FN} \tag{15}$$

$$Sp = \frac{TN}{TN+FP} \tag{16}$$

$$F - score = 2 \times \frac{P \times Sn}{P+Sn} \tag{17}$$

$$MCC = \frac{subtract\ (TP \times TN, FP \times FN) P \times Sn}{\sqrt{(FP+FN) \times (TP+TN) \times (TP+FN) + (TP+FP)}} \tag{18}$$

As another straightforward application, the *detectability index* (DI) (de Valk et al., 1981) provides a practical tool for conclusions about the decision-making capabilities of the practitioner for efficiently solving reliability problems:

$$\text{DI} = -\phi^{-1}\left(P\left(D_{(A)}|T_{II}\right)\right) - \phi^{-1}\left(P\left(D_{(B)}|T_{I}\right)\right) \tag{19}$$

$$= \beta_I + \beta_{II} \ ,$$

in which DI can be considered as an index to separate experts among the other analysts for efficiently solving reliability problems. For instance, as shown in Fig 8, we can find that DI is close/equal to zero for analysts that: always use their favorite approach in analysis (Figs 8-A, and 8-B) or select a method by mere chance (Fig 8-C). Clearly, we should avoid to employ such analysts for reliability analysis. On the other hand, how to use the evaluation of experts to obtain optimal results in analyzes is investigated in the next section.

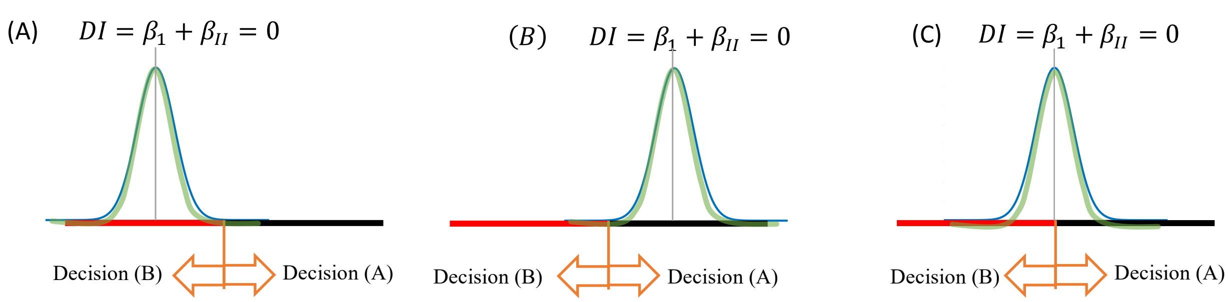


Fig 8. Zero detectability index for analyst whom only uses: A) gradient-based approaches in analysis (i.e., $P\left(D_{(B)}|T_I\right) = 0, P\left(D_{(A)}|T_{II}\right) = 1$), B) Random search approaches in analysis (i.e.,

$P(D_{(B)}|T_I) = 1, P(D_{(A)}|T_{II}) = 0$) or C) select algorithms by the chance (i.e., $P(D_{(B)}|T_I) = P(D_{(A)}|T_{II}) = 50\%$

EVIDENCE THEORY FOR THE SELECTION OF OPTIMAL ALGORITHM

As discussed in NFLR#2, for efficiently solving a reliability problem, a practitioner should select an optimal algorithm that matches to the information of the problem. In this section, a potential procedure for selection of such an optimal algorithm is suggested. Generally, the choice of an optimal algorithm would be a function of the dimension of the in-hand problem, the potential range of failure probability, robustness/efficiency of the alternative algorithms for solving nonlinear problems, and their complexity (in terms of hyper-parameters and the nature of search process). However, in a realistic engineering context, selecting an optimal algorithm among different alternatives may be difficult and we might be faced with the following confusing situation. For an in-hand problem, different expert practitioners suggest different approaches as being the optimal algorithm for the problem under consideration. Sometimes, we may even find that this conflict in the evaluation of experts, instead of decreasing the uncertainty, increases our doubt about the selection of potential optimal algorithm in reliability analysis.

Here, considering the selection of the optimal algorithm as a decision problem under uncertainty, one may use different approaches to reduce error in analysis. For this purpose, as mentioned in the introduction, one may adapt MCDM or AHP for the proposed selection problem or use Bayesian inference in analysis. Using linguistic assessments, experts may evaluate the potential performance of a set of selected reliability algorithms for different criteria (e. g., dimension of the problem, nonlinearity of LSF and etc.). Then, MCDM or AHP may be used for ranking algorithms according to the existing decision-making methods to select optimal method. In this section, by employing the Dempster-Shafer theory of evidence (Dempster, 1968), we offer a potential solution for this issue. Besides the simplicity, compared to MCDM or AHP, a main advantage of this theory is capability of solving decision making problems considering human reliability noting that this approach already presented promising performance for solving selection problems and reliability analysis (P. Li & Wei, 2019), (Rakowsky, 2007).

In essence, Dempster-Shafer evidence theory allows us to combine information from multiple independent sources, even when the corresponding beliefs are conflicting. Based on these sources of information, the Dempster-Shafer rule combines the included information into a measure of belief that highlights portions of evidence that align, while downplaying those portions that conflict (Frittella et al., 2020).

Following this approach, a practical solution for upgrading classic reliability analysis into an experts-in-the-loop process is proposed that increases the probability of choosing an optimal approach among alternatives for efficient reliability analysis.

Data fusion in decision level

In a general engineering context, *sensors* are devices that collect selected pieces of information from the environment in which they operate, which is subsequently used to infer the state of the system under consideration (Gros, 1997). However, to include the reliability and completeness of the information obtained, multiple sensors might need to be used to overcome operating range limitations. In this context, data fusion is a very useful tool to aggregate the information coming from multiple sources. In the domain of sensor fusion, it has been shown numerously that increasing the number of sensors can lead to a significant reduction in error (See Fig 9) (Gros, 1997). Note that the definition of “error” is in this context highly problem-dependent.

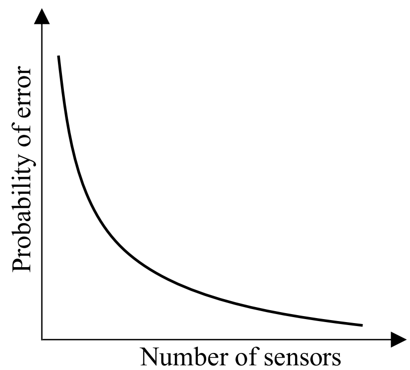


Fig 9. Probability of errors versus number of sensors (Gros, 1997)

To know how fusion theory helps us to efficiently select the right tools to solve reliability problems, it is enough to consider the evaluations of practitioners as the data obtained from the sensors. Using this approach, we can present our original selection problem (See Fig 1) as a data fusion problem (See Fig 10) where experts play the role of sensors and evidence theory plays the role of fusion center to combine different pieces of knowledge into a single knowledge base to derive a sound conclusion in a space with reduced uncertainty. In the following subsections, this theory is adapted for solving selection problems with step-by-step explanations.

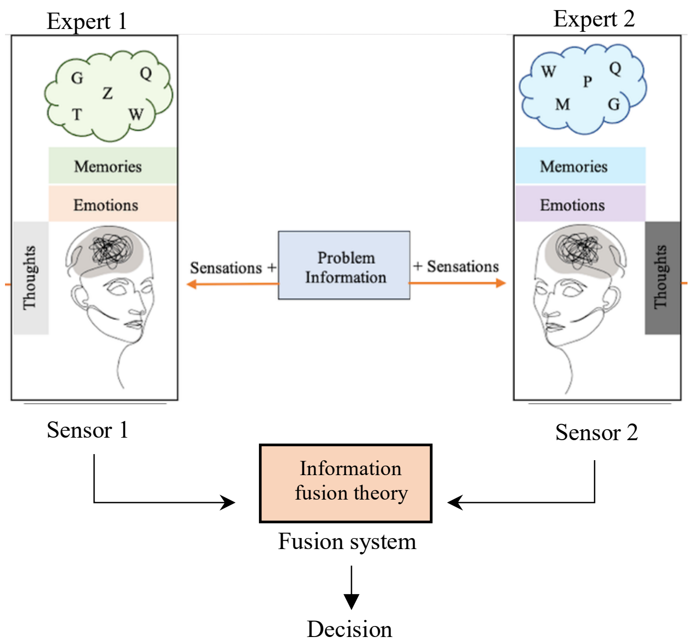


Fig 10. Fusing knowledge of two experts by fusion theory to make a decision about optimal reliability approach

Selection of optimal algorithm by Dempster-Shafer theory of evidence

Dempster-Shafer theory allows us to integrate data coming from several independent sources, even when these data are incomplete. In a sense, it can also be regarded as a general extension of Bayesian theory (Delavar & Sadrykia, 2020; Dempster, 1968). As mentioned before, based on these sources of information, the Dempster-Shafer rule combines the included information into a measure of belief that highlights portions of evidence that align, while downplaying those portions that conflict (Frittella et al., 2020). A brief description of the combination rule of this theory is presented here (for more details, see (Dempster, 1968; Gros, 1997; Delavar and Sadrykia, 2020)). Considering a set of hypotheses, called frame of discernments $\ominus = \{H_1, H_2, \ldots, H_N\}$, a mass function (also known as basic probability assignment, BPA) can be defined for any subset of frame of discernment having the following properties:

$$m: P(\ominus) \rightarrow [0,1]$$

$$\sum_{A \in P(\ominus)} m(A) = 1, \;\; m(\emptyset) = 0, \tag{20}$$

where each subset $A \in\ominus$ such as $m_j(A) > 0$ is called a focal element of m (Delavar & Sadrykia, 2020; Dempster, 1968). The focal element shows how strongly the existing evidence supports A. DST suggests a rule for combination of two mass functions m_1 and m_2 to yield a new mass function with decreased uncertainty:

$$m(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1-k} \text{ when } A \neq \emptyset \qquad k = \sum_{B \cap C = \emptyset} m_1(B)\, m_2(C), \tag{21}$$

where B and C are two focal elements of m_1 and m_2. In this formulation, k is a normalization constant that measures the level of conflict between the two input belief functions in which, $k = 0$ represents the absence of conflict between m_1 and m_2, and $k = 1$ implies complete inconsistency between m_1 and m_2 (Lepskiy, 2013; Yager, 1987). Here, to ease the understanding of the application of DST in reliability analysis, we present a step-by-step solution of the method, illustrated with examples.

Assume that for the given problem $g = f(u)$, we wish to find an optimal reliability approach among three potential optimal alternatives A, B and C. We ask two practitioners to assign a score (e. g., from 1 to 100, considering the items described in the beginning of this section) or a mass probability to each method and tabulate results as presented in Table 1. In this table, for instance, $Score_{1,B}$ corresponds to the score of method B from believe of expert #1. To present the problem in the form of DST, we should convert scores to mass probabilities. Therefore, once each method is scored, result may be normalized as follows (See Tables 1 and 2):

$$S_{i,A} = \frac{\mathrm{Score}_{i,A}}{\mathrm{Score}_{i,A} + \mathrm{Score}_{i,B} + \mathrm{Score}_{i,C}}. \tag{22}$$

Then, we have $S_{i,A} + S_{i,B} + S_{i,C} = 1.$ For instance, Table 2 says from belief of Expert #1, there is $S_{1,A}$, $S_{2,A}$, $S_{3,A}$ percentage chance for methods A, B and C respectively, to be optimal algorithms for solving in-hand reliability problem.

Table 1. Scoring three methods for solving a certain reliability problem using the knowledge of two experts

	method A	method B	method C
Expert #1	$\mathrm{Score}_{1,A}$	$\mathrm{Score}_{1,B}$	$\mathrm{Score}_{1,C}$
Expert #2	$\mathrm{Score}_{2,A}$	$\mathrm{Score}_{2,B}$	$\mathrm{Score}_{2,C}$

Table 2. Presenting optimal selection problem in the form of Dempster-Shafer's theory

	method A	method B	method C
Expert #1	$S_{1,A}$	$S_{1,B}$	$S_{1,C}$
Expert #2	$S_{2,A}$	$S_{2,B}$	$S_{2,C}$

Then, a Dempster combination matrix (DCM), required for computations is then constructed as follows:

$$\mathrm{DCM} = \text{Expert \#1} \begin{array}{c} \\ \\ S_{1,A} \\ S_{1,B} \\ S_{1,C} \end{array} \begin{array}{c} \text{Expert \#2} \\ \begin{array}{ccc} S_{2,A} & S_{2,B} & S_{2,C} \end{array} \\ \begin{bmatrix} S_{1,A}\cdot S_{2,A} & S_{1,A}\cdot S_{2,B} & S_{1,A}\cdot S_{2,C} \\ S_{1,B}\cdot S_{2,A} & S_{1,B}\cdot S_{2,B} & S_{1,B}\cdot S_{2,C} \\ S_{1,c}\cdot S_{2,A} & S_{1,c}\cdot S_{2,B} & S_{1,c}\cdot S_{2,c} \end{bmatrix} \end{array} \tag{23}$$

Finally, the parameters presented in Eq. (21) can be computed as follows:

$$\mathrm{K}=S_{1,A}\cdot S_{2,B}+S_{1,A}\cdot S_{2,C}+S_{1,B}\cdot S_{2,A}+S_{1,B}\cdot S_{2,C}+S_{1,C}\cdot S_{2,A}+S_{1,C}\cdot S_{2,B}$$

$$m_A=\frac{S_{1,A}\cdot S_{2,A}}{1-K},$$

$$m_B=\frac{S_{1,B}\cdot S_{2,B}}{1-K},$$

$$m_C=\frac{S_{1,c}\cdot S_{2,c}}{1-K},$$

$$m_A+m_B+m_C=1 \qquad (24)$$

A) Example of DST method selection taking evaluations of two experts

Here, assume that scoring is performed for three approaches and the results are normalized as presented in Table 3.

Table 3. The belief of two experts regarding mass probabilities of three alternative reliability algorithms (A, B and C)

	method A	method B	method C
Expert #1	0.20	0.35	0.45
Expert #2	0.10	0.40	0.50

Having evaluations of two experts in hand, the following DCM can be obtained by multiplying the mass functions of evidences as follows:

$$DCM = \begin{array}{c} \\ 0.1 \\ 0.4 \\ 0.5 \end{array} \begin{array}{c} \begin{array}{ccc} 0.2 & 0.35 & 0.45 \end{array} \\ \begin{bmatrix} 0.020 & 0.035 & 0.045 \\ 0.080 & 0.140 & 0.180 \\ 0.100 & 0.175 & 0.225 \end{bmatrix} \end{array}$$

In the proposed matrix, we have the main diagonal of DCM as $P_{AA}=0.1 \times 0.2 = 0.020$, $P_{BB}=0.4 \times 0.35 = 0.140$ and $P_{CC}=0.5 \times 0.45 = 0.225$. Besides, the constant factor K can be estimated as all component of CM except main diagonal components:

K=0.035+0.045+0.080+0.180+0.100+0.175=0.615,

Then, a new up-dated probability for each alternative would be as follows:

$m_A = \frac{P_{AA}}{1-K} = \frac{0.020}{1-0.615} = 0.052,$

$m_B = \frac{P_{BB}}{1-K} = \frac{0.140}{1-0.615} = 0.36,$

$m_C = \frac{P_{CC}}{1-K} = \frac{0.225}{1-0.615} = 0.58,$

The result is tabulated in Table 4. We can find that by combining the knowledge of two practitioners:

A) the optimality probability of method A (from 20 and 10) is reduced to 5.2%,

B) the optimality probability of method B (from 35 and 40) is turned into 36%,

C) the optimality probability of method C (from 45 and 50) is increased to 58%.

Considering the obtained results, one may conclude that Method C is the optimal approach among the three proposed alternatives.

Table 4. DST decision fusion result for two experts regarding method A, B and C

Methods	A	B	C
Fusion result	5.2%	36%	58%

B) Example of DST method selection taking evaluations of four experts

In this example, we increased the number of practitioners from two to four and the result of the evaluations is tabulated in Table 5.

Table 5. The belief of four experts regarding mass probabilities of three alternative reliability algorithms (A, B and C)

	method A	method B	method C
Expert #1	0.20	0.35	0.45
Expert #2	0.10	0.40	0.50
Expert #3	0.30	0.20	0.50
Expert #4	0.20	0.30	0.50

Using DST, the evaluations of four experts are merged together and the updated probabilities are presented in Table 6. The result shows that even though the optimality probability of method C from view point of each expert is about 50%, the theory of evidence reduced uncertainty and we can consider about 85% chance for method C as optimal approach for solving the in-hand reliability problem. This issue is schematically represented in Fig 11.

Table 6. DST decision fusion result for four experts regarding method A, B and C

Methods	A	B	C

Fusion result	1.5%	14%	85%

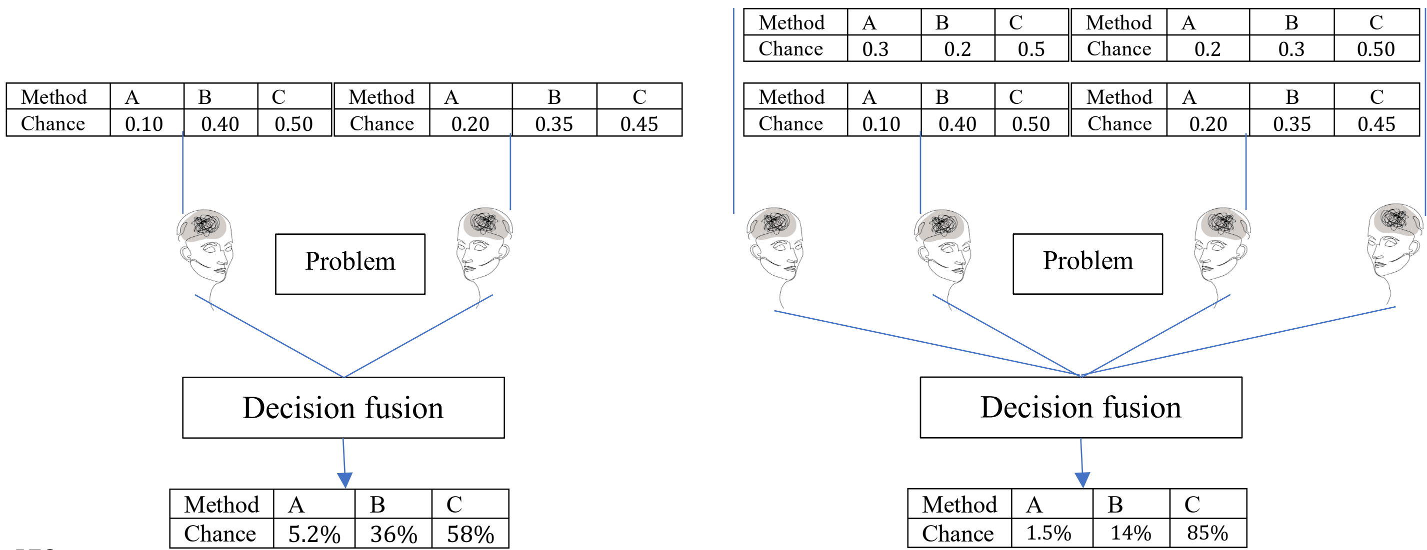


Fig 11. Reducing the uncertainty in decision making by increasing numbers and fusing evaluations of practitioners

Selection of optimal algorithm by Yager's method of fusion

In the previous solution, it is assumed that practitioners are reliable and they have not any doubt in their evaluations regarding the competency of three alternative methods. However, in practice, the human reliability of practitioners is not 100% (i. e., $R < 100\%$). On the contrary, practitioners often have errors in their beliefs or may have some doubts in their evaluations. Even, we may find huge conflict in their judgement. In decision fusion theory, there are several solutions for solving such problems (Cuzzolin, 2021; Lepskiy, 2013) while in this study, we suggest to apply Yager's rule of fusion to solve problems involving practitioners' uncertainty (Yager, 1987).

Yager suggested applying a reliability factor in fusion which is, in this study, representative of reliability of experts in their evaluation (denoted as $R(\cdot)$ in NFLR#2). For the case of this study, the R index can be obtained by human reliability analysis.

For instance, R_1 =90% means practitioner #1 had 90% successful decisions in distinguishing optimal reliability approach in his/her previous experiences, which is equal to 10% failure probability in proper decision making, i.e., $P(HE)$=10% (or according to evidence theory: for 10%, I have no any idea about the optimal approach for the subject case study). Such reliability levels can for instance be calibrated by means of pre-defined selection questions with known answers (Ayyub, 2001). By considering m'_i as mass functions presented by experts and introducing ground probability mass function q, this approach updates probabilities as follows:

$q(\emptyset) \geq 0,$

$S_i(A) = R_i \times m'_i(A),$

$q(A) = \sum_{\cap A_i = A}[S_1(A_1) \times S_2(A_2) \times \ldots \times S_i(A_i) + (1 - R_i) \times S_i],$

$m(A) = \frac{q(A)}{1-q(\emptyset)}.$ (25)

The step-by-step application of combination of probabilities with expert uncertainty (Yager, 1987) is as follows: Let the evaluation of two practitioners regarding three alternative approaches A, B and C be as shown in Table 7. The main difference of this table with Table 2 is that we additionally consider the reliability of practitioners R_i in their analysis.

Table 7. The evaluation of two experts with uncertainty in decision making

Expert/Method	method A	method B	method C	Reliability of expert
Expert #1	$S'_{1,A}$	$S'_{1,B}$	$S'_{1,C}$	R_1
Expert #2	$S'_{2,A}$	$S'_{2,B}$	$S'_{2,C}$	R_2

Having mass probability of each method according to evaluation of experts with reliability R_i, Yager suggested a modification on the assigned probability masses as:

$$S_{i,j}=R_j \cdot S'_{i,j} \quad (26)$$

Then, adapts probability outcomes as shown in Table 8.

Table 8. Adapting evaluation of experts according to Yager's approach

Expert/Method	method A	method B	method C	Expert failure probability
Expert #1	$S_{1,1}$	$S_{1,2}$	$S_{1,3}$	EF_1=1-R_1
Expert #2	$S_{2,1}$	$S_{2,2}$	$S_{2,3}$	EF_2=1-R_2

For the updated probabilities, the updated combination matrix CM then becomes as follows:

$$CM = \text{Expert \#1}\ \begin{array}{c} \\ S_{1,A} \\ S_{1,B} \\ S_{1,C} \\ EF_1 \end{array} \begin{array}{c} \text{Expert \#2} \\ \begin{array}{cccc} S_{2,A} & S_{2,B} & S_{2,C} & EF_2 \end{array} \\ \begin{bmatrix} S_{1,A}\cdot S_{2,A} & S_{1,A}\cdot S_{2,B} & S_{1,A}\cdot S_{2,C} & S_{1,A}\cdot EF_2 \\ S_{1,B}\cdot S_{2,A} & S_{1,B}\cdot S_{2,B} & S_{1,B}\cdot S_{2,C} & S_{1,B}\cdot EF_2 \\ S_{1,c}\cdot S_{2,A} & S_{1,c}\cdot S_{2,B} & S_{1,c}\cdot S_{2,c} & S_{1,C}\cdot EF_2 \\ EF_1\cdot S_{2,A} & EF_1\cdot S_{2,B} & EF_1\cdot S_{2,c} & EF_1\cdot EF_2 \end{bmatrix} \end{array} \qquad (27)$$

Then, the combined probability for each alternative can be estimated as follows:

K=$S_{1,A}\cdot S_{2,B}$+$S_{1,A}\cdot S_{2,C}$+$S_{1,B}\cdot S_{2,A}$+$S_{1,B}\cdot S_{2,C}$+$S_{1,C}\cdot S_{2,A}$+$S_{1,C}\cdot S_{2,B}$

m_A=$\frac{q_A}{1-K}$,

m_B=$\frac{q_B}{1-K}$, (28)

m_C=$\frac{q_C}{1-K}$,

where,

q_A=$S_{1,A}\cdot S_{2,A}$+$S_{1,A}\cdot EF_2$+$EF_1\cdot S_{2,A}$

q_B=$S_{1,B}\cdot S_{2,B}$+$S_{1,B}\cdot EF_2$+$EF_1\cdot S_{2,B}$ (29)

q_C=$S_{1,C}\cdot S_{2,C}$+$S_{1,C}\cdot EF_2$+$EF_1\cdot S_{2,C}$

According to Yagers' method of fusion, the ignorance factor (Ig) is as follows:

$$\text{Ig}=1-m_A+m_B+m_C. \qquad (30)$$

This factor reflects the reliability of practitioners in their final estimation (e. g., large value for Ig are equal to a low reliability of obtained probabilities).

A) Example of Yager method selection taking evaluations of two experts with uncertainty

Consider two practitioners with reliabilities of $R_1 = 75\%$ and $R_2 = 90\%$ and evaluation about the optimality chance of three alternatives as presented in Table 9. Here, one may find a kind of conflict in selection of two practitioners for methods A and C.

Table 9. The evaluation of two experts with uncertainty regarding the optimality probability of three desired reliability methods A, B and C

Expert/Method	method A	method B	method C	Reliability of expert
Expert #1	0.2	0.3	0.5	0.75
Expert #2	0.6	0.3	0.1	0.9

According to the implementations presented in this section, the updated probability table (Table 10) and combination matrix CM would be as follows:

$$\mathrm{CM} = \begin{bmatrix} 0.081 & 0.121 & 0.202 & 0.135 \\ 0.040 & 0.061 & 0.101 & 0.067 \\ 0.013 & 0.020 & 0.034 & 0.022 \\ 0.015 & 0.022 & 0.037 & 0.025 \end{bmatrix}$$

Accordingly, by estimating parameters presented in Eqs. (28) and (29), the final updated probabilities correspond with chance of three alternative as optimal algorithms would be as presented in Table 11 while the ignorance factor of the result is Ig=5%.

Table 10. The updated evaluation matrix for two experts considering uncertainties in decision making

Expert/Method	method A	method B	method C	Failure probability of expert
Expert #1	0.150	0.225	0.375	0.25
Expert #2	0.540	0.270	0.090	0.1

Table 11. Yager's decision fusion result for two experts regarding method A, B and C

Methods	A	B	C
Fusion result	46%	30%	19%

The result of combination, which led to a reduction in uncertainty of the selection, shows that algorithm A with $m(A)$=46% chance, as compared to other approaches could be considered as the optimal algorithm for reliability analysis.

B) Example of Yager's method selection taking evaluations of four experts with uncertainty

In this example, the number of practitioners evaluations involving uncertainty in evaluations is increased to four (See Table 12) and the decision fusion results are presented in Table 13. Result shows that among three alternative algorithms, Method A has 66% chance to be optimal approach among other approaches and considering big differences between chance of methods, the proposed algorithm could be a reliable solution for solving the problem.

Table 12. The evaluation of four experts with uncertainty regarding the optimality probability of three desired reliability methods A, B and C

Expert/Method	method A	method B	method C	Reliability of expert
Expert #1	0.20	0.30	0.50	0.75
Expert #2	0.60	0.30	0.10	0.90
Expert #3	0.50	0.15	0.35	0.85
Expert #4	0.35	0.45	0.20	0.80

Table 13. Yager's decision fusion result for four experts regarding method A, B and C

Methods	A	B	C
Fusion result	66%	20%	14%

DISCUSSION AND TOPICS FOR FUTURE RESEARCH

In this paper, the potential application of different theories for improving the quality of research and analysis in structural reliability analysis are investigated:

- Considering well developed studies in the literature about the presented theories (e. g., Human reliability, detection theory, decision making under uncertainty and data fusion theory) this study highlights that investigation regarding to proposed theories and also progress in related “strategies” in uncertain condition, instead of only considering development in “algorithms”, could be subject of future studies (e. g., see Fig 12).

- The lesson that has to be learned from NFL is that for the practitioner *the correct question is not which algorithm I have to use* but *what is the geometry of the problem fitness (Serafino, 2013)*. Therefore, conducting a kind of meta research (Mikolajewicz & Komarova, 2019) including analysis and discussing about performance function of problems for each specific field would be highly helpful for selecting optimal algorithm for efficient reliability of structures.
- Considering the recent developments in meta-models and machine leaning approaches for efficient reliability analysis, similar researches can be conducted for clarification about some aspects of these approaches that was not investigated/clarified in literature.

It might be argued that applying extensive selection procedures as presented in this paper to answer these questions might be too involved for many every-day reliability problems. Indeed, when the reliability problem is trivial, the losses incurred by having no free lunch might be less than the time spent in selecting the most appropriate reliability method. However, since current computing trends are pointing more and more in the direction of high-dimensional problems (both in terms of uncertain quantities, as in degrees of freedom in the underlying models), the selection of the most appropriate method becomes a pressing issue, which has to be treated with utmost care.

A further critical remark with respect to the work presented in this paper is that there are some criticisms regarding the original NFL theorem (Adam et al., 2019; McDermott, 2020). The authors can add another one to this list by highlighting the importance of size/dimension of the problem in reliability analysis. For an at-hand problem, the size/dimension of the problem is always available.

Knowing this, some approaches are preferable to others in term of performance (e.g., problem dependent algorithms that only developed for solving high dimensional problems and vice versa.).

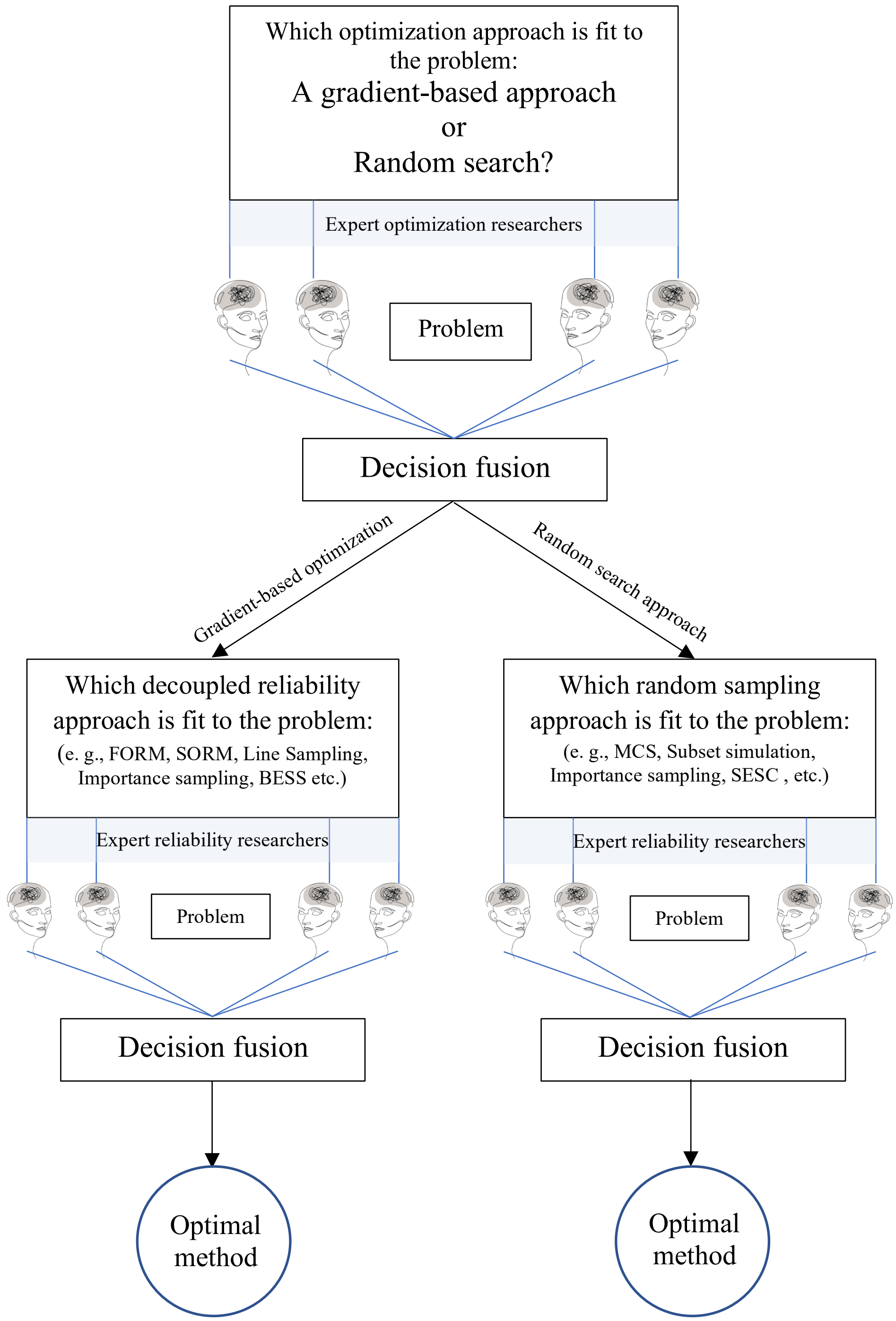


Fig 12. Hierarchical experts-in-the-loop decision fusion strategy by decoupling the skills of practitioners required for optimization and reliability

This specific point is not considered in the proof of original NFL theorem (e. g., one may argue that no purely black-box problem exists in real-world problems). Nonetheless, the authors are of the opinion that this does not nullify the conclusions made in this study; specially in the case of NFLR#2, namely that when a reliability analyst just uses their preferred reliability method for grey-box problems, there is no free lunch in reliability analysis.

CONCLUSION

According to asymptotic approximation theory, search and optimization are the core of robust reliability approaches. In this paper, it shown that the well-known No-Free-Lunch theorems (NFL) in search and optimization can therefore be used to answer basic questions regarding selection of an optimal algorithm for efficient reliability analysis.

It proposed that for the case of problems with no information about the structure of performance function (so called black-boxes), no algorithm is preferable to others in terms of performance. For the case of problems where some information about the reliability problem is available (known as grey-boxes), the proposed adapted NFL theorem shows that the analyst is a key part of the reliability problem, and the importance of human reliability in this case is as all knowledge about the in-hand problem.

It is shown that, parallel to statistical skills, efficient reliability analysis requires expertness of analyst in the field of optimization to distinguish which algorithm is more efficient to properly determine the important failure regions of the problem. Otherwise, even for simple problems, there is no free lunch in reliability analysis. To meet this aim, application of human reliability and detection theory as a mathematical framework for analysis of decision making under uncertainty are investigated.

Once the importance of practitioner and its reliability for efficient probabilistic analysis have been highlighted, we suggested to employ the theory of evidence and data fusion for determining the probabilities of alternative algorithms to be the optimal solution for solving a defined reliability problem. Data fusion shows that increasing the number of sensors (say practitioners), reduces the probability of error in distinguishing the state of system (selection of optimal approach in this study). Having several alternative algorithms as potential optimal approach for solving a certain problem, we show that Dempster-Shafer's theory of evidence (as a well-developed approach for fusing uncertain information) can be used straightforwardly for combining the evaluation of several practitioners regarding alternative algorithms for reducing the probability of error in selection of optimal reliability approach. For easing the understanding of the proposed approach, a step-by-step solution of the proposed experts-in-the-loop approach are illustrated for some examples with different number of practitioners. We finally wrap the developments up in a discussion section.

REFERENCES

Adam, S. P., Alexandropoulos, S.-A. N., Pardalos, P. M., & Vrahatis, M. N. (2019). No Free Lunch Theorem: A Review. In I. C. Demetriou & P. M. Pardalos (Eds.), *Approximation and Optimization: Algorithms, Complexity and Applications* (pp. 57–82). Springer. https://doi.org/10.1007/978-3-030-12767-1_5

Al-Turjman, F., & Deebak, B. D. (Eds.). (2020). *Security in IoT Social Networks* (1st edition). Academic Press.

Anderson, N. (2015). Teaching Signal Detection Theory With Pseudoscience. *Frontiers in Psychology*, *6*, 762. https://doi.org/10.3389/fpsyg.2015.00762

Ataei, M., Shahsavany, H., & Mikaeil, R. (2013). Monte Carlo Analytic Hierarchy Process (MAHP) approach to selection of optimum mining method. *International Journal of Mining Science and Technology*, *23*(4), 573–578. https://doi.org/10.1016/j.ijmst.2013.07.017

Au, S.-K., & Beck, J. L. (2001). Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, *16*(4), 263–277.

Ayyub, B. M. (2001). *Elicitation of expert opinions for uncertainty and risks*. CRC press.

Bartsoen, L., Faes, M. G. R., Andersen, M. S., Wirix-Speetjens, R., Moens, D., Jonkers, I., & Sloten, J. V. (2023). Bayesian parameter estimation of ligament properties based on tibio-femoral kinematics during squatting. *Mechanical Systems and Signal Processing*, *182*, 109525. https://doi.org/10.1016/j.ymssp.2022.109525

Bourgund, U. (1986). *Importance Sampling Procedure Using Design Points, ISPUD: A User's Manual : an Efficient, Accurate and Easy-to-use Multi-purpose Computer Code to Determine Structural Reliability*. Universität Innsbruck, Institut für Mechanik.

Breitung, K. (1984). Asymptotic Approximations for Multinormal Integrals. *Journal of Engineering Mechanics*, *110*(3), 367–366.

Breitung, K. (1994). *Asymptotic Approximations for Probability Integrals* (1st ed.). Springer Berlin, Heidelberg. https://link.springer.com/book/10.1007/BFb0073538

Breitung, K. (2021). SORM, Design Points, Subset Simulation, and Markov Chain Monte Carlo. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, *7*(4), 04021052. https://doi.org/10.1061/AJRUA6.0001166

Breitung, K. (2022). Failure Probability Estimation, a Short History: Concepts and Misconceptions. *Proc. of the 8th Intl. Symp. on Reliability Engineering and Risk Management*. (ISRERM 2022), Hanover, Germany.

Calixto, E. (2016). *Gas and Oil Reliability Engineering: Modeling and Analysis* (2nd edition). Gulf Professional Publishing.

Callens, R., Faes, M., & Moens, D. (2022). Multilevel Quasi-Monte Carlo For Interval Analysis. *International Journal for Uncertainty Quantification*. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2022039245

Colloca, L. (2013). *Placebo and Pain: From bench to bedside*. Elsevier, Oxford, UK.

Cuzzolin, F. (2021). *The Geometry of Uncertainty*. Springer. https://link.springer.com/book/10.1007/978-3-030-63153-6

Dang, C., Valdebenito, M. A., Faes, M. G. R., Wei, P., & Beer, M. (2022). Structural reliability analysis: A Bayesian perspective. *Structural Safety*, *99*, 102259. https://doi.org/10.1016/j.strusafe.2022.102259

Dang, C., Wei, P., Faes, M. G. R., Valdebenito, M. A., & Beer, M. (2022). Parallel adaptive Bayesian quadrature for rare event estimation. *Reliability Engineering & System Safety*, *225*, 108621. https://doi.org/10.1016/j.ress.2022.108621

de Valk, J. P. J., Eijkman, E. G. J., & Vintcent, J. (1981). Detectability index, likelihood ratio and diagnostic output as psychophysical measures for evaluation of medical image processing. *Medical and Biological Engineering and Computing*, *19*(5), 597–604. https://doi.org/10.1007/BF02442774

Delavar, M. R., & Sadrykia, M. (2020). Assessment of Enhanced Dempster-Shafer Theory for Uncertainty Modeling in a GIS-Based Seismic Vulnerability Assessment Model, Case

Study—Tabriz City. *ISPRS International Journal of Geo-Information*, *9*(4), Article 4. https://doi.org/10.3390/ijgi9040195

Dempster, A. P. (1968). A Generalization of Bayesian Inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, *30*(2), 205–247.

Elegbede, C. (2005). Structural reliability assessment based on particles swarm optimization. *Structural Safety*, *27*(2), 171–186. https://doi.org/10.1016/j.strusafe.2004.10.003

Frittella, S., Manoorkar, K., Palmigiano, A., Tzimoulis, A., & Wijnberg, N. (2020). Toward a Dempster-Shafer theory of concepts. *International Journal of Approximate Reasoning*, *125*, 14–25. https://doi.org/10.1016/j.ijar.2020.05.004

Ghasemi, S. H., Kalantari, H., Abdollahikho, S. S., & Nowak, A. S. (2019). Fatigue reliability analysis for medial tibial stress syndrome. *Materials Science & Engineering. C, Materials for Biological Applications*, *99*, 387–393. https://doi.org/10.1016/j.msec.2019.01.076

Gros, X. E. (1997). *NDT Data Fusion*. Wiley.

Gull, S. F. (1988). Bayesian Inductive Inference and Maximum Entropy. In G. J. Erickson & C. R. Smith (Eds.), *Maximum-Entropy and Bayesian Methods in Science and Engineering: Foundations* (pp. 53–74). Springer Netherlands. https://doi.org/10.1007/978-94-009-3049-0_4

Harbitz, A. (1986). An efficient sampling method for probability of failure calculation. *Structural Safety*, *3*(2), 109–115. https://doi.org/10.1016/0167-4730(86)90012-3

Hasofer, A., & Lind, N. (1974). An Exact and Invariant First Order Reliability Format. *Journal of Engineering Mechanics*, *100*(1), 111–121.

Kingdom, F. A. A., & Prins, N. (2016). *Psychophysics: A Practical Introduction* (2nd edition). Academic Press.

Lepskiy, A. (2013). Estimation of Conflict and Decreasing of Ignorance in Dempster-Shafer Theory. *Procedia Computer Science*, *17*, 1113–1120. https://doi.org/10.1016/j.procs.2013.05.142

Li, G.-D., Yamaguchi, D., & Nagai, M. (2007). A grey-based decision-making approach to the supplier selection problem. *Mathematical and Computer Modelling*, *46*(3), 573–581. https://doi.org/10.1016/j.mcm.2006.11.021

Li, P., & Wei, C. (2019). An emergency decision-making method based on D-S evidence theory for probabilistic linguistic term sets. *International Journal of Disaster Risk Reduction*, *37*, 101178. https://doi.org/10.1016/j.ijdrr.2019.101178

McDermott, J. (2020). When and Why Metaheuristics Researchers can Ignore “No Free Lunch” Theorems. *SN Computer Science*, *1*(1), 60. https://doi.org/10.1007/s42979-020-0063-3

Melchers, R. E. (1990). Radial Importance Sampling for Structural Reliability. *Journal of Engineering Mechanics*, *116*(1), 189–203.

Mikolajewicz, N., & Komarova, S. V. (2019). Meta-Analytic Methodology for Basic Research: A Practical Guide. *Frontiers in Physiology*, *10*. https://www.frontiersin.org/articles/10.3389/fphys.2019.00203

Misraji, M. A., Valdebenito, M. A., Jensen, H. A., & Mayorga, C. F. (2020). Application of directional importance sampling for estimation of first excursion probabilities of linear structural systems subject to stochastic Gaussian loading. *Mechanical Systems and Signal Processing*, *139*, 106621. https://doi.org/10.1016/j.ymssp.2020.106621

Papaioannou, I., Papadimitriou, C., & Straub, D. (2016). Sequential importance sampling for structural reliability analysis. *Structural Safety*, *62*, 66–75. https://doi.org/10.1016/j.strusafe.2016.06.002

Rakowsky, U. K. (2007). Fundamentals of the Dempster-Shafer theory and its applications to reliability modeling. *International Journal of Reliability, Quality and Safety Engineering*, *14*(6), 579–601.

Rashki, M. (2021a). Structural reliability reformulation. *Structural Safety*, *88*, 102006. https://doi.org/10.1016/j.strusafe.2020.102006

Rashki, M. (2021b). SESC: A new subset simulation method for rare-events estimation. *Mechanical Systems and Signal Processing*, *150*, 107139. https://doi.org/10.1016/j.ymssp.2020.107139

Rashki, M. (2021c). The soft Monte Carlo method. *Applied Mathematical Modelling*, *94*, 558–575. https://doi.org/10.1016/j.apm.2021.01.022

Rashki, M., Miri, M., & Azhdary Moghaddam, M. (2012). A new efficient simulation method to approximate the probability of failure and most probable point. *Structural Safety*, *39*, 22–29. https://doi.org/10.1016/j.strusafe.2012.06.003

Schuëller, G. I., Pradlwarter, H. J., & Koutsourelakis, P. S. (2004). A critical appraisal of reliability estimation procedures for high dimensions. *Probabilistic Engineering Mechanics*, *19*(4), 463–474. https://doi.org/10.1016/j.probengmech.2004.05.004

Serafino, L. (2013). No Free Lunch Theorem and Bayesian probability theory: Two sides of the same coin. Some implications for black-box optimization and metaheuristics. *ArXiv:1311.6041*.

Singh, K. K., Elhoseny, M., Singh, A., & Elngar, A. A. (Eds.). (2021). *Machine Learning and the Internet of Medical Things in Healthcare* (1st edition). Academic Press.

Tiwari, R., Zalkikar, J., & Huang, L. (2021). *Signal Detection for Medical Scientists: Likelihood Ratio Test-based Methodology* (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9780429259753

Valdebenito, M. A., Pradlwarter, H. J., & Schuëller, G. I. (2010). The role of the design point for calculating failure probabilities in view of dimensionality and structural nonlinearities. *Structural Safety*, *32*(2), 101–111. https://doi.org/10.1016/j.strusafe.2009.08.004

Wolpert, D. H. (2021). What Is Important About the No Free Lunch Theorems? In P. M. Pardalos, V. Rasskazova, & M. N. Vrahatis (Eds.), *Black Box Optimization, Machine Learning, and No-Free Lunch Theorems* (pp. 373–388). Springer International Publishing. https://doi.org/10.1007/978-3-030-66515-9_13

Wolpert, D., & Macready, W. (1996). No Free Lunch Theorems for Search. *Technical Report SFI-TR-95-02-010. Santa Fe Institute. S2CID 12890367*.

Wolpert, D., & Macready, W. (1997). No Free Lunch Theorems for Optimization. *IEEE Transactions on Evolutionary Computation*, *1*, 67–82. https://doi.org/10.1109/4235.585893

Yager, R. R. (1987). On the dempster-shafer framework and new combination rules. *Information Sciences*, *41*(2), 93–137. https://doi.org/10.1016/0020-0255(87)90007-7

Zabinsky, Z. (2009). *Random Search Algorithms*. Wiley. https://doi.org/10.1002/9780470400531.eorms0704

Zhong, C., Wang, M., Dang, C., Ke, W., & Guo, S. (2020). First-order reliability method based on Harris Hawks Optimization for high-dimensional reliability analysis. *Structural and*

Multidisciplinary Optimization, *62*(4), 1951–1968. https://doi.org/10.1007/s00158-020-02587-3