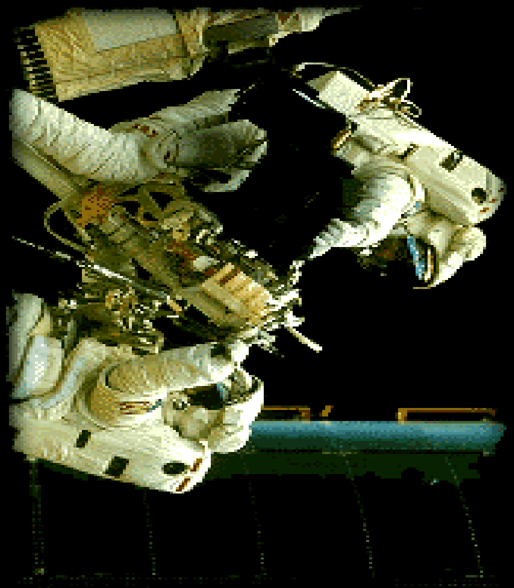
Assessment of Uncertainty:

Models, Computation, and Decisions

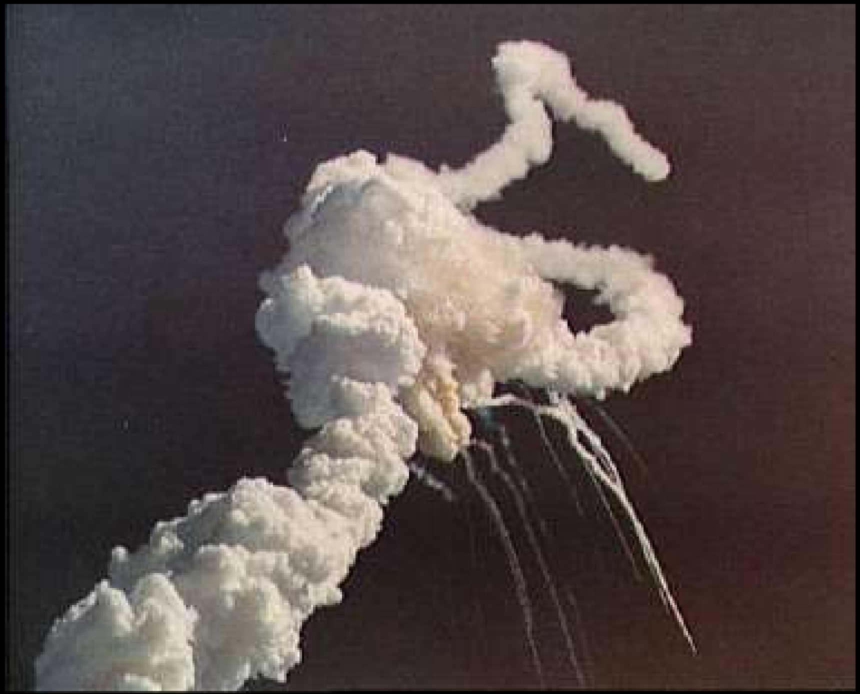
Martin A. Wortman
Department of Industrial & Systems Engineering
Texas A&M University

What's at stake?



The design and operation of high-consequence engineered systems carries considerable responsibilities. We are each concerned that these responsibilities are appropriately addressed.

“Design and operation” are all about making risk encumbered decisions. Often these decisions have outcomes that can be tragically expensive.



Recent advances in communication and computational technology have greatly expanded the role of computer-based models, in supporting design and operations.

The fidelity of analytical models and computer code supporting design and operation decisions must be congruent with the consequences of decision outcomes.

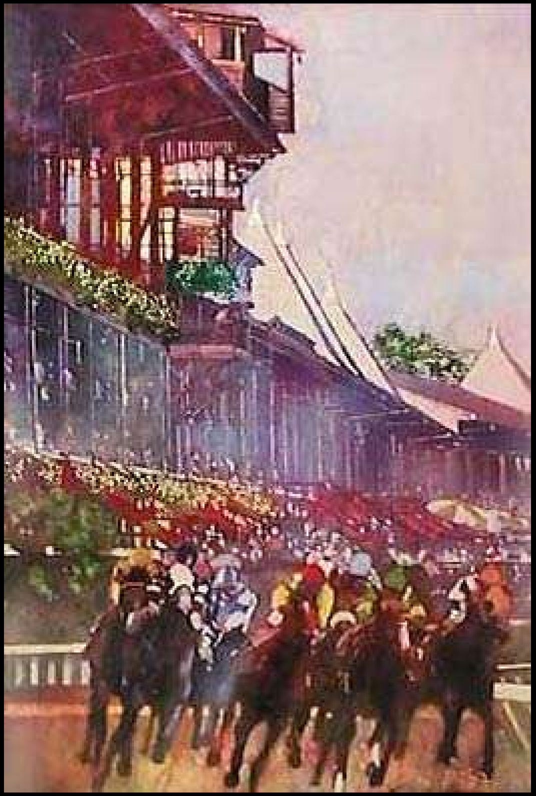
Computational Models and Uncertainty

Engineers employ analytical models with the singular purpose of decision support. Typically, analytical models are grounded in physics (empirical science). That is, we employ mathematics (logic) to model observations (physics) with the purpose of inferring the (future) value of decision alternatives.

Computer codes implement analytical models – they are a model of an analytical model. Hence, codes are a model of a model of the physics.

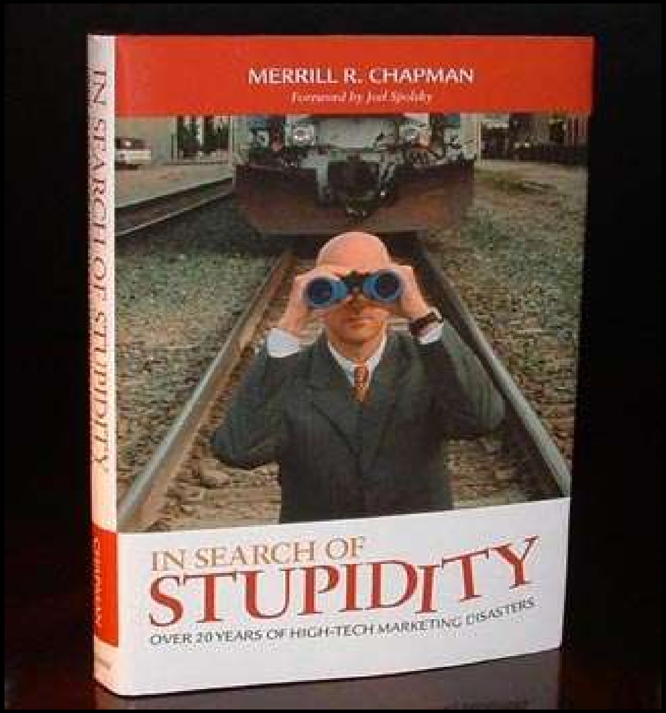
All models are fraught with uncertainty; hence, the future is uncertain.

Conclusion: Engineers are gamblers. They place audacious bets, supporting their judgement with layered and contingent models having an unquantifiable fidelity with reality.



Verification & Validation

Verification concerns modeling; validation concerns decisions.



Validation is the process of ensuring that a conclusion is correctly derived from it's primises.

Horst the killer bee was led to meet his destiny...

Verification is an ongoing activity focused on ensuring that one model agrees with another model (*e.g.*, code produces results that agree with an analytical model – *bug elemination*).

Verification

Verification can never be assurred – it is not an observable state or condition. Verification is NOT similar finding needles in a haystack.

Ed was unlucky enough to find the needle in the haystack!

There is no possibility of a general stopping rule for the verification process.

We can only claim that models agree (within measure of uncertainty) at specific epochs of observation.

It is important to recognize that models under comparison are not always independent of the observation process. For example, we fix software bugs when we find them; yet, we may introduce new bugs when attempting to remove old ones.

Because of the verification stopping time issue, statistical estimators measuring agreement among models are not available through ergodic theorems.

Verification

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, P)$, $n \in \mathbb{R}$ be a filtered probability space on which an observation process $\{Z_n; n \in \mathbb{R}\}$ is measurable. Let the random variable T^* be the observation for which models under comparison *disagree* for the last time (*i.e.*, one model is verified with respect to the other). T^* is *not* a stopping time for this filtered space.

Validity and Decisions

- A decison is the act of selecting the from among a set of alternatives.
- Decisions leads to an irrevocable committment of resources.
- Decisions necessarily have consequences (outcomes) that are uncertain.
- Decision making and gambling are, for all practical purposes, synonomous.
- Validity = chose the *best bet*.

Alternatives

- Let $\mathbb{A} = [a, b] \subset \mathbb{R}$ contain the point 0. $\mathbb{A}$ is set of possible rewards, with 0 the *status quo*.
- Each decision alternative α has a corresponding reward distribution function $F_\alpha(\cdot) \in D(\mathbb{A})$.
- So which reward distribution do you like best?

Normative wagering.

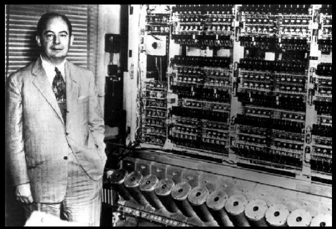
Axiom 1 *(Weak Ordering) There is a preference relation* $\succcurlyeq$ *among the elements of* $D(\mathbb{A})$ *that is complete and transitive.*

Axiom 2 *(Continuity). For every set* $F \in D(\mathbb{A})$ *the sets* $\{G \in D(\mathbb{A}) : G \succcurlyeq F\}$ *and* $\{G \in D(\mathbb{A}) : F \succcurlyeq G\}$ *are closed in the topology of weak convergence.*

Axiom 3 *(Independence) For all* $F, G, H \in D(\mathbb{A})$ *and all* $\lambda \in [0, 1]$, $F \succcurlyeq G$ *implies* $\lambda F + (1 - \lambda)H \succcurlyeq \lambda G + (1 - \lambda)H$.

A Separation Theorem

Theorem 1 *(Expected Utility Theorem.) Let* $\succcurlyeq$ *be a binary relation on* $D(\mathbb{A})$*. There exists a continuous function* $u : \mathbb{A} \to \mathbb{R}$ *(unique up to affi ne transformations) such that* $F \mapsto \int_X u(x)dF(x)$ *represents* $\succcurlyeq$ *if and only if Axioms 1, 2, and 3 are satisfi ed.*



This separation theorem is an ordinal result!

Comment

- When the three axioms are satisfied, utility will separate preferences in the space of alternatives.
- One searches among the alternatives for the one having the highest expected utility.
- Observation: Complete specification of reward distributions is not required to separate certain preferences using the expected utility theorem.

Decomposition of $D(\mathbb{A})$

- $F_i \in D(\mathbb{A})$, $i = 1, 2, ...N$ is the reward distribution of the i^{th} alternative
- $u : \mathbb{A} \to \mathbb{R}$ is the utility function.
- $\mathcal{A}_i \subset D(\mathbb{A})$ is such that $F_i \in \mathcal{A}_i$; here, $\mathcal{A}_i$ is a family of distribution functions that contains the distribution of alternative i.

Calculus of Variation

- $u_i^{min} = min_{F \in \mathcal{A}_i}\{\int u(x)dF(x)\}$ and $u_i^{max} = max_{F \in \mathcal{A}_i}\{\int u(x)dF(x)\}$.
- For each decison alternative i, there is an interval $I_i = [u_i^{min}, u_i^{max}]$ such that $\int u(x)dF_i(x) \in I_i$.
- $G = (V, E)$ is the interval graph of the set $\{I_i; i \in \mathcal{A}\}$
- For $i^* = \arg\max_i\{u_i^{min}\}$, I_{i^*} is the interval having the greatest lower bound.

Lemma 1 *With $G = (V, E)$ and $\{I_i; i \in \mathcal{A}\}$ defined as before, $v_{i*} \in V$ belongs to a maximal clique and is incident to no other vertices.*

Corollary 1 *1. Any decision alternative with distribution belonging to the set $\mathcal{A}_j$ with $u_j^{max} < u_{i*}^{min}$, is less preferable than any alternative with distribution in $\mathcal{A}_{i*}$.*

2. Any decision alternative k for which $u_k^{max} \geq u_{i*}^{min}$ is indistinguishable from the most preferred alternative.

The Physics

- Let $\{\underline{Y}_t; t \in \mathbb{R}\}$ be a stochastic analytical model.
- Let $\{T_n; n \in \mathbb{N}\}$ be a sampling process.
- $\{\underline{X}_n; n \in \mathbb{N}\}$ is the discretized model, with $\underline{X}_n = \underline{Y}_{T_n}$
- $\underline{X} = \{\underline{X}_0, \ldots, \underline{X}_K\}$ finite stochastic model supporting computation
- $F(\underline{x}) = P\{\underline{X} \leq \underline{x}\}$ is the distribution of $\underline{X}$.
- $r : \mathbb{R}^{K+1} \times \mathbb{R}^N \to \mathbb{R}$ is reward.

Utility & Computation

The marriage of physics with the expected utility theorem is not so easily made.

Testing the validity of a decision and its supporting models and methodologies is not three foot putt.

$$E[u(r)] = E[(u \circ r)(\underline{X})] = \int_{\mathbb{R}^{\mathbb{N}}} (u \circ r)(\underline{x}) dF(\underline{x})$$

Induced Sets of Distributions

- Orderings of event probabilities, conditional probabilities, and independence between events introduces constraints that can be written in the canonical form $g_i(F(\underline{b}_{i_1}, \ldots, F(\underline{b}_{i_k})) \leq 0$,
- where $g_i : \mathbb{R}^k \rightarrow \mathbb{R}$ is functional representation of event probability constraints in terms of distribution $F_{\underline{X}}(\underline{x})$ on the random variables $\underline{X}$.
- b_{i_j} is a vector of constants $\forall j$.

Induced Sets of Distributions

- Ordering constraints induce a set of distribution functions $S \in D(\mathbb{A})$ such that $S = \{F \in D(\mathbb{A}) : g_i(F(\underline{b}_{i_1}, \ldots, F(\underline{b}_{i_k})) \leq 0, i = 1, \ldots, n\}$
- The expected utility of reward associated with this alternative is bounded and lies within a compact interval having end points u^{min} and u^{max}.

Non—linear Programming

Utilitiy intervals over sets of distributions are obtained by solving (large) non–linear programming formualtions.

$$u^{min} = \min_{F \in S} \int_{\mathbb{R}^{\mathbb{N}}} (u \circ r)(\underline{x}) dF(\underline{x})$$

and

$$u^{max} = \max_{F \in S} \int_{\mathbb{R}^{\mathbb{N}}} (u \circ r)(\underline{x}) dF(\underline{x}),$$