SAND2009-4216C

The Concept of Independence in Weapon Safety: Foundations and Practical Implementation Guidance

SAND2009-????C

Jeff Brewer

JOWOG-44

July 2009

Sandia National Laboratories

Overview

- Numerical Nuclear Weapon (NW) Safety Requirements
- Assured NW System Safety
- Nuclear Safety Design Principles
- Typical Definitions of Independence
- Independence in NW Safety
- Independence for Human Intent and Other Human Interactions with Weapons
- Conclusions

Requirements are Stringent

Numerical probability requirements typically associated with inadvertent nuclear detonation (IND):

- 1 out of 1 **billion** per weapon lifetime in normal environments
- Less than 1 out of 1 **million** per exposure to abnormal environments

Normal Environments – weapon may be expected to encounter and retain operational capability.

Abnormal Environments – all other possible environments; emphasis is usually on ‘credible’ abnormal environments.

Assured NW System Safety

- Isolation – the predictable separation of detonation-critical elements from compatible energy
- Inoperability – the predictable inability of detonation-critical elements to function
- Incompatibility – the use of energy or information that will not be duplicated inadvertently

Safety Theme – a high-level, concise expression of what will be isolated, inoperable and/or incompatible.

Nuclear Safety Design Principles

Isolation

- Separation
- Barrierization
- Diversion
- Support
- Insulation
- Cancellation

Spatial distance

physical wall, Faraday cage, hermetically sealed volumes, etc.

Lightning arrestor connector, energy spillway, etc.

strong structure

avoid resonance, damping, etc.

Inoperability

- Deploy in pre-disabled state **(i.e. active reversal required for use)**
- Incapacitation
  - **chemical phase change (e.g., melt, freeze, dissolve)**
  - **Mechanical fracture (shatter)**

(Combination)

- Layering
- Redundancy
- Coordination
- Removal
- Defense in Depth

Incompatibility

- Interlocking
- Entropization
- Energy incompatibility
- Information incompatibility

Complicated, multi-part device, etc.

magnetic, electrical, optical, fluidic, etc.

unique signal patterns & corresponding stronglink discriminators

Information independence

- Functional
- Temporal
- Physical Isolation

Weaklink failing irreversibly before a stronglink fails given a threat environment

Multiple barriers of increasing conservatism, safety margins, performance monitoring, etc.

Quantitative Design Goals

Is there a quantitative aspect to the qualitative concept of "assured nuclear weapon system safety"? **Yes!**

If *independent* response of two safety subsystems to all AEs could be achieved at the $1x10^{-3}$ level, then the overall system safety level could be asserted to be achieved at the (1E-3)(1E-3) = 1E-6 level.

The reductionist approach of designing multiple, independent, safety subsystems, while mathematically expedient, requires great confidence in the degree of independence achieved!

Each subsystem, "…must not be subject to chain-of-events coupling between subsystems or common-mode failures in which both subsystems are damaged or bypassed by the same event…each must serve its purpose even if the other subsystems are defeated, damaged, or fail" – P. D'Antonio (1998)

Typical Definitions & Applications of Independence

- Event (A) and event (B) are independent events if $P(A|B) = P(A)$ and $P(B|A) = P(B)$, thus $P(A \cap B) = P(A)P(B)$. Recall that in general, $P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$, given $P(A) \neq 0$, $P(B) \neq 0$.
- Two events are independent if the outcome of one event does not influence the other event; i.e., knowing the outcome of a flip of a fair coin provides no additional insight about whether the next coin toss will reveal a head or tail.
- Beware not to confuse *independent* with *mutually exclusive*
- In the domain of formal experimentation, most common statistical test require independence between events.
  - Independence forms the basis of hypothesis testing
  - To detect dependence between selected/manipulated factors, it is necessary to minimize the effect of sources of dependence which may not be controlled
- Typical examples: fair coin flips, cards from well-shuffled decks, fair die rolls, balls from a well-mixed urn, casino & lottery games

All involve well-defined and fixed boundary conditions or rules—unlike inadvertent nuclear detonation where many uncertainties regarding AEs exist

Independence in the Weapon Safety Context

Independence – design of subsystems to prevent common-mode and common-cause failures such that the failure of one subsystem does not affect the failure of another subsystem.

General approaches for achieving independent safety subsystem designs.

- Spatial separation
- Geometric differences
- Temporal spacing of functions
- Different materials
- Different energy levels
- Different orientations of otherwise similar components
- Different energy types of operation
- Different design teams
- Different chemical phases
- Different logic structures/algorithms/protocols
- Independent verification and vulnerability review teams searching for dependencies between the designs
- …

Sandia National Laboratories

Independence for Human Intent & other Human Interactions

Intent Unique Signal Information (IUSI) – provides unambiguous communication of intent to detonate from the source (one or more humans) to a weapon in a manner highly unlikely to be inadvertently generated.

Implemented as a carefully engineered sequence of bi-valued events transmitted separately to the relevant safety device in the weapon:

A,B,B,B,B,A,A,A,B,A,A,A,B,B,A,A,B,B,B,A,B,A,A,B

— SAND91-1269 by Spray & Cooper (1991)

How do you provide IUSI to a weapon in a way maximizing independence?

- Two people each entering 24 bi-valued events for one of two patterns?
- One person entering 48 events for both patterns?
- Generate a large portion of IUSI from a small amount of information provided by the crew?

Independence for Human Intent & other Human Interactions

1. Only provide IUSI to the “inanimate engineered system” only when operationally required. No storage of IUSI either algorithmically or using memory.
   - Question, “if the IUSI was changed, what parts of the system would need to be modified to accommodate the change?”
2. At least two distinct human actions are required, where *human action* is defined as resulting from human intellect, will and sophisticated motor skills such that the action bypasses multiple barriers in the inanimate engineered system that provide **functional**, **temporal**, and **physical** independence between the safety devices (destination) in the weapon and the IUSI (source).

Concepts for Increasing Independence

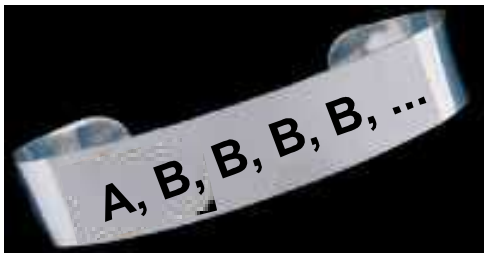
Function, **Time**, **Physical Isolation**,

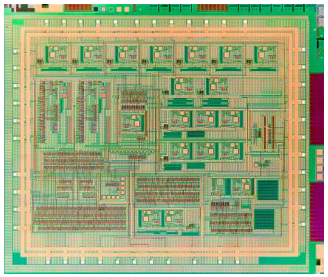
—while not providing mutually exclusive sources of "dependence," they are proposed as helpful concepts in the search for tendencies toward independence (both with respect to "energy" and "information")

- Functional independence – ***minimize*** the functional connection of IUSI to any part of the weapon system
- Temporal independence – ***minimize*** time of exposure of IUSI **patterns** to the functional (inanimate) system, ***maximize*** time separation of **events** in a pattern (always communicated sequentially), and **maximize** time separation of multiple patterns for multiple safety subsystems—within practical limits
- Physical independence – ***maximize*** the physical isolation of IUSI from the weapon system (e.g., spatial separation, size and/or number of barriers)

Concepts for Increasing Independence

- **Functional information independence**





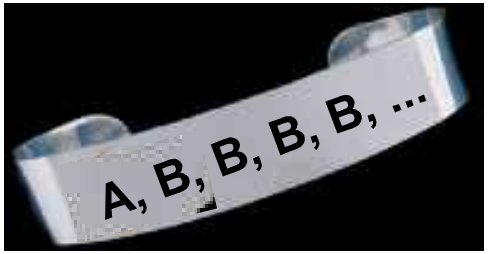
- **Temporal information independence**

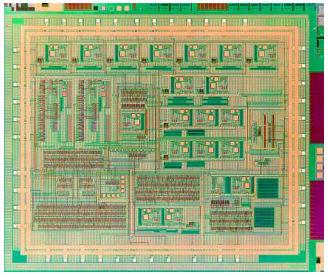
A , **A** , **B** , ...

Safety Subsystem # 1

Safety Subsystem # 2

- **Physical isolation (e.g., spatial separation, size and/or number of barriers)**





Additional Independence Characteristics

Many **bias processes** impact human behavior and some may also be embedded in the behavior of inanimate engineered items , as they arise from application of human knowledge. Understanding of these bias processes is needed to promote and preserve desired independence to protect against IND.

Bias – a systematic tendency or heuristic which limits a comprehensive application of available knowledge, experience, and related data to decisions and/or actions. Biases, tendencies or heuristics of human decision making are not inherently bad; they are methods of mentally taking shortcuts in recognizing a situation, which normally allow people to quickly select the most plausible choices first, followed by the less plausible choices. However, biases or heuristics that tend to work in specific, often "simple" information settings, sometimes lead to severe and systematic errors in other settings (e.g., more complex) such that they hinder proper interpretation of available information and data and lead to inappropriate perceptions, decisions, and actions.

Overestimation of independence between redundant-type events

- Common mode failures
- Social Shirking
- Overcompensation

Combinatorics, probability, statistics and related *critical thinking skills*

Structure of human cognitive abilities

Values, personality, interests, group identity, substantive knowledge, and *critical thinking skills*

Normative Knowledge

insensitivity to sample size	High
means and medians estimated well	High
coefficient of variation is noticed	High
variance largely ignored	High
gambler's fallacy	High
small probabilities overestimated	High
large probabilities underestimated	High
regression to the mean	High
as number of options change, probability assignments change dramatically	Low
overestimate the probability of conjunctive events (series combinations)	High
underestimate the probability of disjunctive events (parallel combinations)	High
overestimate independence between redundant-type events	High

Degree to which education & practice of concepts can mitigate

Easiest to change, given disciplined effort

Availability

anchoring effect	Medium
illusory correlation	Medium
recency	Medium
imaginability	Medium
salience	Medium
retrievability	Medium
representativeness	Medium
explicitness	Medium
framing effect	Medium

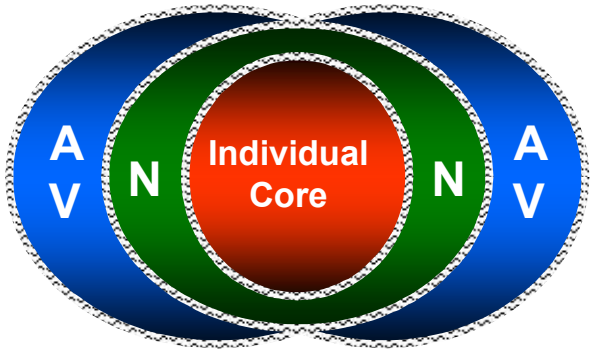
Degree to which knowledge of cognitive processes can mitigate (i.e., knowledge of hierarchical, distributed, parallel processing abilities of the central nervous system—the machinery with which we perceive, learn, remember, & communicate)

More difficult to change, given disciplined effort

Individual Specific

loss aversion†	Low
law of effect†	Low
constantly requiring more	Low
locus of control†	Low
ambiguity aversion†	Medium
confirmation bias†	Medium
hindsight bias†	Medium
false consensus bias†	Medium

Degree to which explicit self-knowledge can impact these



Most difficult to change even with disciplined effort

†Key for "overconfidence" phenomena

Overestimation of Independence Between Redundant-Type Events

- **Common mode failures**—one event causes multiple failures
  - E.g., improper training in searching for defects, faulty test procedures
- **Social Shirking**—phenomenon in which individuals or groups reduce their reliability by assuming that others will "take up the slack."
  - Probabilities of errors for a checker of someone else's work will be much higher than the probabilities of errors for the original performer (i.e., checker usually does not expect to find many errors)
  - E.g., "Trust" of co-workers or subordinates leading to cursory inspections/verification of safety critical work
- **Overcompensation**—results when the addition of extra items (alleged to be redundant) encourages individuals or groups to increase production or engage in riskier behavior.
  - E.g., increasing throughput after adding inspectors, reckless driving in safer cars, "child-proof" medicine bottles leading to increased poisoning

WAYWARD 6

Additional Independence Characteristics Not Discussed Here…

- Exposition of AE context influencing design activities
- Independence between intended delivery environment & other conceivable physical environments (e.g., accidents, test, maint.)
- Discussion & mathematical defense of serial communication of IUSI for promoting independence
- Heuristics for increasing IUSI independence in digital systems
- Extensive description of bias processes and mitigation measures impacting dependencies related to human behavior
- Independence-increasing measures in:

Manufacturing	Testing
Assembly	End of life transportation
Beginning of life transportation	Disassembly
Maintenance	…

Sandia National Laboratories

Conclusions

- Assured Safety, a method for striving to meet stringent numerical safety requirements, is accomplished through application of well-defined nuclear safety design principles
- Proper application of the concept of independence to safety subsystem design is essential for “assured safety”
- Typical definitions of independence were presented
- Selected elements from of a methodology for improving the technical basis supporting independence assumptions were presented—esp. for IUSI & other human interactions with weapons
- This presentation is intended to be the first in a series detailing methods (and their mathematical basis) for applying the concept of independence to improve weapon safety