National Aeronautics and
Space Administration



Managing Failure by Design In Sociotechnical Systems

Jon Holbrook, PhD

National Aeronautics and Space Administration

May 8, 2024

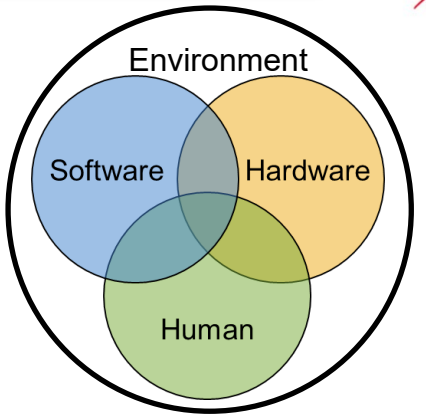
IFS Briefing

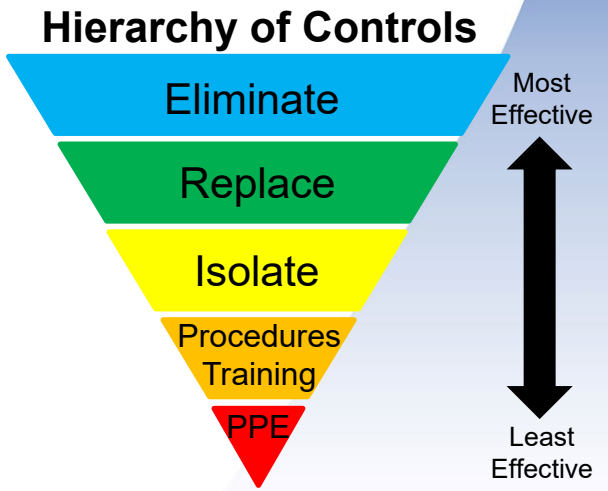
Bottom line up front

- Failure to acknowledge how human cognitive systems work – and how they evolved to work – leads to misconceptions and oversimplifications about human performance
  - Miscalibrations of what humans contribute to system performance
  - Misinterpretations about what humans are “bad at” or “good at”
- Successful long-term solutions for designing socio-technical systems require knowledge of how people think and work
  - This is true whether designing systems to replace humans, work with humans, or rely on humans to intervene
  - Experience “being human” is an insufficient qualification for being an expert on how people think

All complex sociotechnical system fail

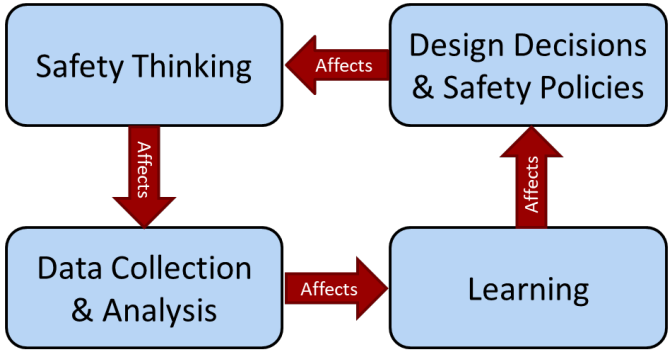
- All system components are potential sources of failure
- Failure prevention through design
  - Design to achieve a state in which conditions leading to an undesired outcome are controlled
- Designing for failure prevention can reduce failures but doesn't make systems failure-proof
- Designs must also enable preparing for and recovering from both expected and unexpected failures





Data-driven failure management

- In today's aerospace industry data talks…
  - … and it talks a lot about errors and failures!
  - … and exceedances
  - … and deviations
  - … and violations



- Even a "safety event" refers only to a *loss* of safety
  - An undesired outcome
  - Increased risk of an undesired outcome
  - Conditions that could result in an undesired outcome (i.e., close call)

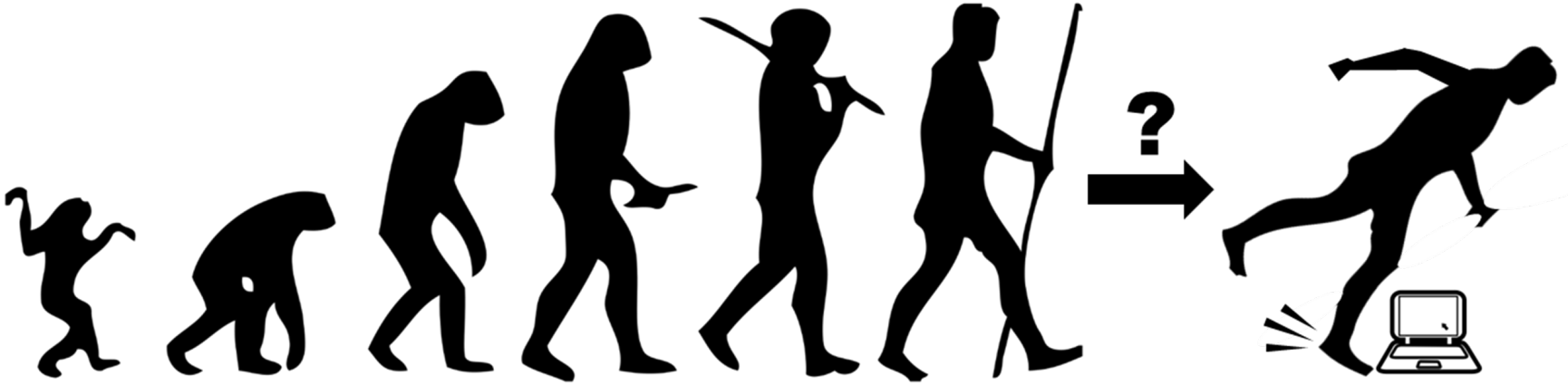
Absence of evidence ≠ evidence of absence

- When we characterize task performance only in terms of errors and failures, we ignore the vast majority of human impacts on the system
- When design or policy decisions are based only on failure data, they are based on a very small sample of systematically non-representative data
  - Sampling bias: Results based on non-random samples can be erroneously attributed to the phenomenon under study rather than to the method of sampling
- When the only available human performance data are about failure, then data-driven designs only consider that humans fail

To err is human

- Claybrook & Kildare (*Science*, 2018)
  - "… humans are inherently bad at monitoring semi-autonomous systems and are readily distracted."
  - *Note*: The cited reference (Cunningham & Regan, 2015) doesn't actually say that…
- NASA APPEL "Cognitive Bias in Engineering Decision-Making" course description:
  - "… will help participants understand how decision-making can be compromised at all levels in an organization through reliance on heuristics…"
- "To fast forward to the safest possible operational state … network operators will be interested in the path that realizes full machine autonomy as quickly as possible." (Uber, 2016)

Did millions of years of natural selection go wrong?



H. habilis

H. erectus

H. heidelbergensis

H. sapiens

H. sapiens sapiens

H. sapiens errorum?

2.5 million years ago

Today

Image credit (left image): Jose-Manuel Benito, released into the public domain.
Available at: https://commons.wikimedia.org/wiki/File:Human_evolution_scheme.png.

The mind is a successful adaptive system

- Evolutionary adaptation is the adjustment of organisms to their environment to improved their chances of survival in that environment
- Natural selection produces computational systems that solve adaptive problems reliably, quickly, and efficiently, but those systems are functionally specialized (Cosmides & Tooby, 2013)

Environments and tools that are poorly aligned with what our cognitive systems evolved to do limit human capabilities for resilient performance

Illusions of memory *(Howe, Garner, & Patel, 2013)*

- "False memories" are an inevitable consequence of a powerful, reconstructive memory system
- What determines whether they have a positive or negative consequence depends solely upon how they are later put to use

Associative problem solving was faster and more accurate when critical information was falsely recalled than when it wasn't

*"The person who **misremembers** seeing a predator while foraging for food might be more cautious upon their return to gather more food than the person who **accurately remembers** that only signs (e.g., feces, scent) of the predator had been present on an earlier visit."*

Judgments under uncertainty *(Cosmides & Tooby, 1996)*

- Literature on cognitive biases has been interpreted to suggest that human inductive reasoning mechanisms do not embody a calculus for computing probability (i.e., people are "bad at" intuitive statistics)
- Evolutionary psychology, however, predicts that humans should be able to successfully represent probabilities as frequencies

The probability of a single event is intrinsically unobservable

There's a 20% probability that today's hunt in the north canyon will be successful.	vs	We were successful 2 out of the last 10 times we hunted in the north canyon.

Judgments under uncertainty

(Cosmides & Tooby, 1996)

- Medical diagnosis problem (famous for eliciting "base-rate neglect" cognitive bias) presented either as a probability problem or as a frequency problem

Probability Condition

A test to detect a disease (disease X), whose prevalence is 1/1000, has a false positive rate of 5%. Assuming that you know nothing about the person's symptoms or signs, what is the chance that a person found to have a positive result *actually* has the disease? _____%

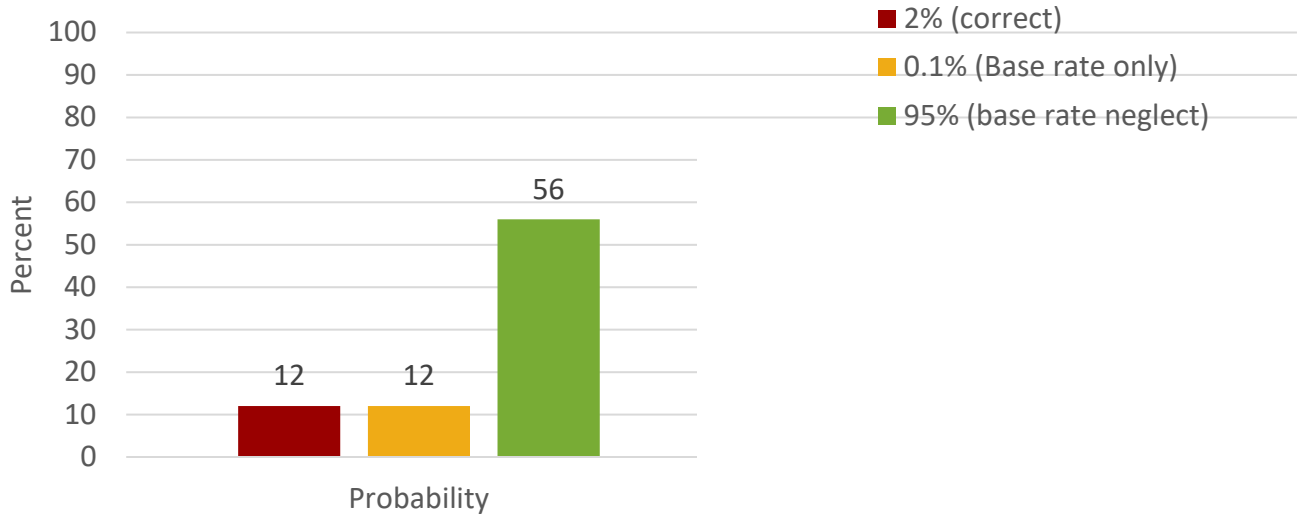
Frequency Condition

1 out of every 1000 Americans has disease X. Every time the test to detect disease X is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, out of every 1000 people who are perfectly healthy, 50 of them test positive for the disease. From a random sample of 1000 Americans, on average, how many people who test positive for the disease will *actually* have the disease? _____ out of _____.

Judgments under uncertainty

(Cosmides & Tooby, 1996)

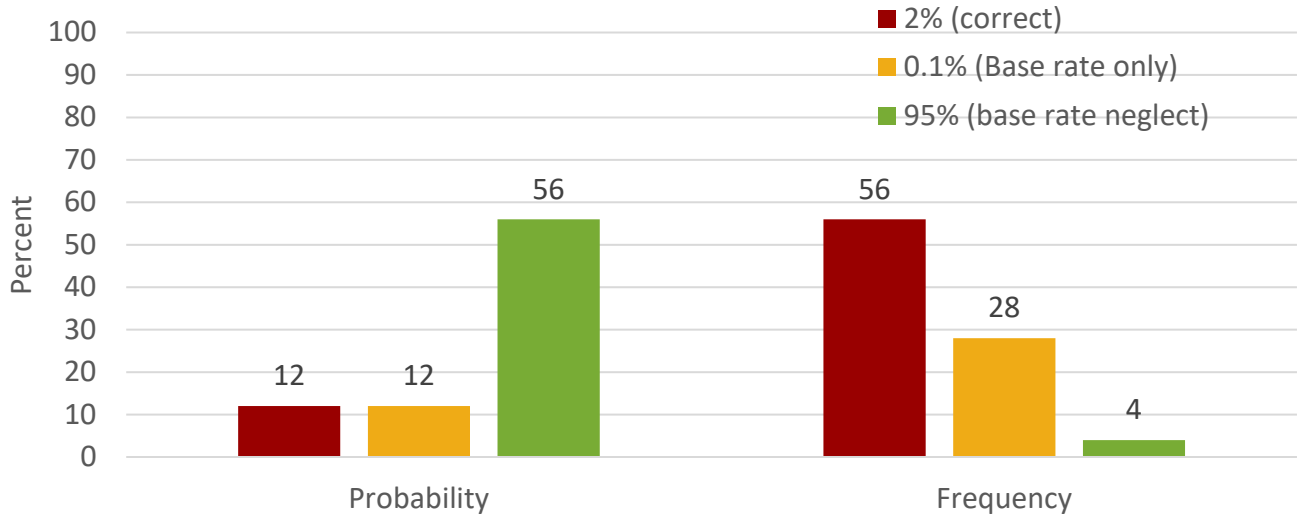
- Correct = ~ 2% (or ~ 1 out of 50)
- Base rate conservative (considering *only* the base rate) = ~ 0.1% (or ~ 1 out of 1000)
- Base rate neglect (ignoring the base rate) = ~ 95% (or 950 out of 1000)



Judgments under uncertainty

(Cosmides & Tooby, 1996)

- Correct = ~ 2% (or ~ 1 out of 50)
- Base rate conservative (considering *only* the base rate) = ~ 0.1% (or ~ 1 out of 1000)
- Base rate neglect (ignoring the base rate) = ~ 95% (or 950 out of 1000)



A *simple rewording of the problem can elicit correct Bayesian reasoning and eliminate base-rate neglect for a majority of subjects*

A real-world base-rate fallacy example

Probability Condition

Human error contributes to 80% of aviation accidents[1]

Humans cause most of the accidents, so:
Remove them
Replace them
Isolate the system from them

(1) Weigmann & Shappell, 2003

A real-world base-rate fallacy example

Probability Condition

Human error contributes to 80% of aviation accidents[1]

Humans cause most of the accidents, so:
Remove them
Replace them
Isolate the system from them

Frequency Condition

- From 2007- 2016, there were 244,000,000 commercial aviation departures, of which 388 resulted in accidents[2]
- Human error was implicated in ~310 out of the 388 accidents (i.e., 80%)
- Pilots intervened to manage aircraft malfunctions on ~48,800,000 of the 243,999,612 non-accident flights (i.e., 20%)[3]

(1) Weigmann & Shappell, 2003; (2) Boeing, 2017; (3) PARC/CAST, 2013

A real-world base-rate fallacy example

Probability Condition

Human error contributes to 80% of aviation accidents[1]

Humans cause most of the accidents, so:
Remove them
Replace them
Isolate the system from them

Frequency Condition

- From 2007- 2016, there were 244,000,000 commercial aviation departures, of which 388 resulted in accidents[2]
- Human error was implicated in ~310 out of the 388 accidents (i.e., 80%)
- Pilots intervened to manage aircraft malfunctions on ~48,800,000 of the 243,999,612 non-accident flights (i.e., 20%)[3]

Pilots intervene to keep flights safe from aircraft malfunctions over 157,000 times for every 1 time that a human error is implicated in an accident.
Does removing, replacing, or isolating them still sound like the safest idea?

(1) Weigmann & Shappell, 2003; (2) Boeing, 2017; (3) PARC/CAST, 2013

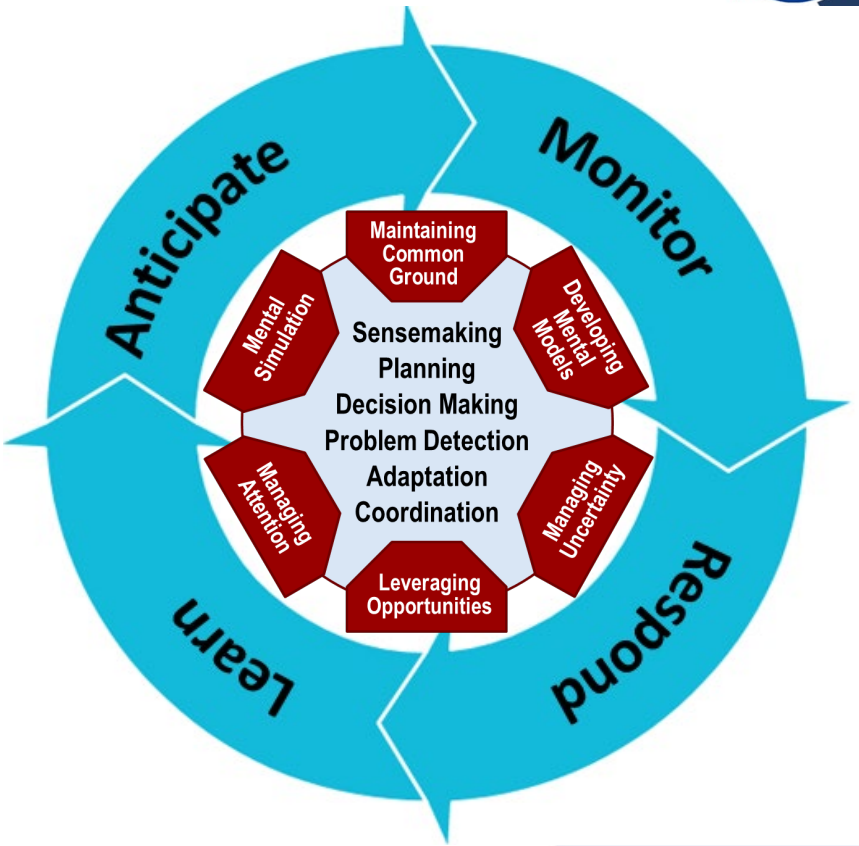
Humans cause safety

- Analysis of 20 “major incidents” by PARC/CAST Flight Deck Automation Working Group (2013)
  - “These major incidents represent cases where pilots successfully mitigated the risk of the malfunction.”
  - Equipment malfunctions identified as a threat in 11 of 20 major incidents
  - No checklist or defined procedure to address equipment malfunctions in 5 of these 11 incidents (i.e., failure not anticipated by designers)

Reasoning beyond procedures is essential for many recovery actions

Human potential for resilient performance

- Resilient performance[1] comprises the capabilities to
  - **Anticipate** – consider what might happen in the future
  - **Monitor** – know what is happening and where to look for change
  - **Respond** – know what to do and have the opportunity to do it
  - **Learn** – know what has happened and why
- Resilient performance enables macrocognitive functions[2] (i.e., "goals")



Outer circle adapted from Hollnagel (2015)[1] ; Inner circle adapted from Klein et al. (2003)[2]

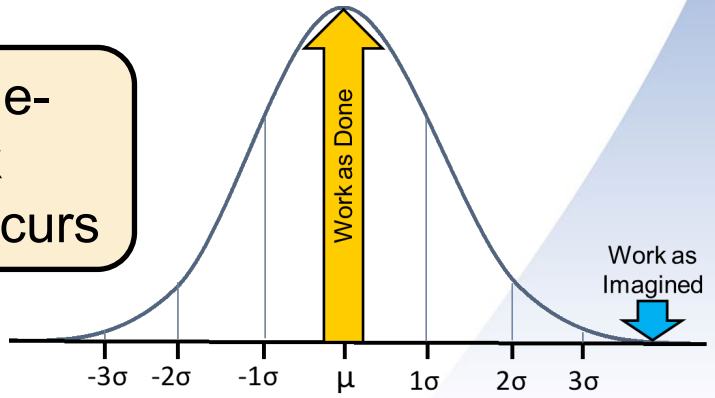
Consequences of designing out failure factors

- Designs intended to protect the system from error-prone humans can inadvertently limit the capability to intervene and adapt
- For example…
  - Inadequacies in training can limit the ability to **anticipate**
  - Creating impoverished environments when automating tasks can limit the ability to **monitor**
  - Simplified interfaces and inability to over-ride or take control from automation can limit the ability to **respond**
  - Designs that restrict feedback can inhibit the ability to **learn**

Designing for “nominal” performance

- ***Nominal*** describes performance according to plan or design
- Because the human’s role in preventing failures is rarely acknowledged, it is often assumed that “nominal” performance is free of failure and free of human intervention
- Even in ultra-safe, well-studied systems, failures and the need to deal with them are *routine*
  - Airline pilots intervene to deal with system malfunctions on 1 in 5 “normal” flights
  - These systems are all carefully designed, rigorously tested, and fully certified

When “nominal” is characterized as a blue-sky, “everything works” condition, we risk designing for a state that rarely if ever occurs



We don’t think like we think we think

Just for fun, try Googling “the most complicated thing in the universe”… ☺

- Why everyone isn’t an expert on how human’s think just because they’re human
  - People do not have reliable introspective access to their own cognition (Nisbett & Wilson, 1977)
  - It doesn’t *feel* like our introspections are unreliable, so people overestimate their ability to do so (Dunning & Kruger, 1999)
  - We employ scientific methods to understand how people think
  - We employ human performance experts to understand the science

Managing failure by design: A path forward

There are engineering solutions to designing for failure management, and they are *integrated* solutions

- Include expertise in human performance on your design team
- Recognize system failures as both routine and unavoidable – account for that in design and in testing
- Measure what happens, not just what fails
- Don't simply design out failure factors without considering the possible reduction of preparation and recovery factors
- Question dogmatic assumptions about what people are "bad at" or "good at"
- Support capability for resilient performance through system designs that align with what our cognitive systems evolved to do

Thank you