



The Mind Is a Successful Adaptive System

Jon Holbrook, PhD

National Aeronautics and Space Administration

March 27, 2024

NESC HF TDT F2F

Humans, we have a problem



- NASA is attempting and supporting some of the hardest things humans have ever tried to do

However...

- There appear to be prevalent misconceptions and misunderstandings about what people do and how people think

Designing systems with an inadequate understanding of what humans do and how our cognitive systems evolved to work can introduce unrecognized and unknown risks

Humans, we have a problem



➤ Evidence that there's a problem

- Sampling bias in the human performance data we collect and analyze
- Limited recognition of humans as the primary source of system resilience
- Dogmatic and widely accepted views on what humans “are bad at”
- System designs that limit humans’ capabilities to adapt and intervene
- Inadequate resourcing of humans in system design
- Everyone seems to think they’re an expert on how the brain works



The value of focusing on human error

- Measuring errors is a common practice in the study of cognition
 - Thresholds, illusions, cognitive biases, limits of memory and attention, etc.
- Such errors are systematic and can therefore be used to study the logic by which our cognitive systems work (Gilchrist, 2006)
- The literature on human performance is rich with findings of cognitive failures and methods to identify, label, and measure them



The value of focusing on human error

- Measuring errors is also a common practice in the applied field of Human Factors
 - “The goal of human factors is to reduce human error, increase productivity, and enhance safety and comfort...” (Human Factors & Ergonomic Society website)
 - Concerns for human error were a major influence in establishing the field of human factors (Helander, 1997)
 - Human error has since become increasingly emphasized in product and system design and in the operations of various organizations (Sharit, 2012)

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)

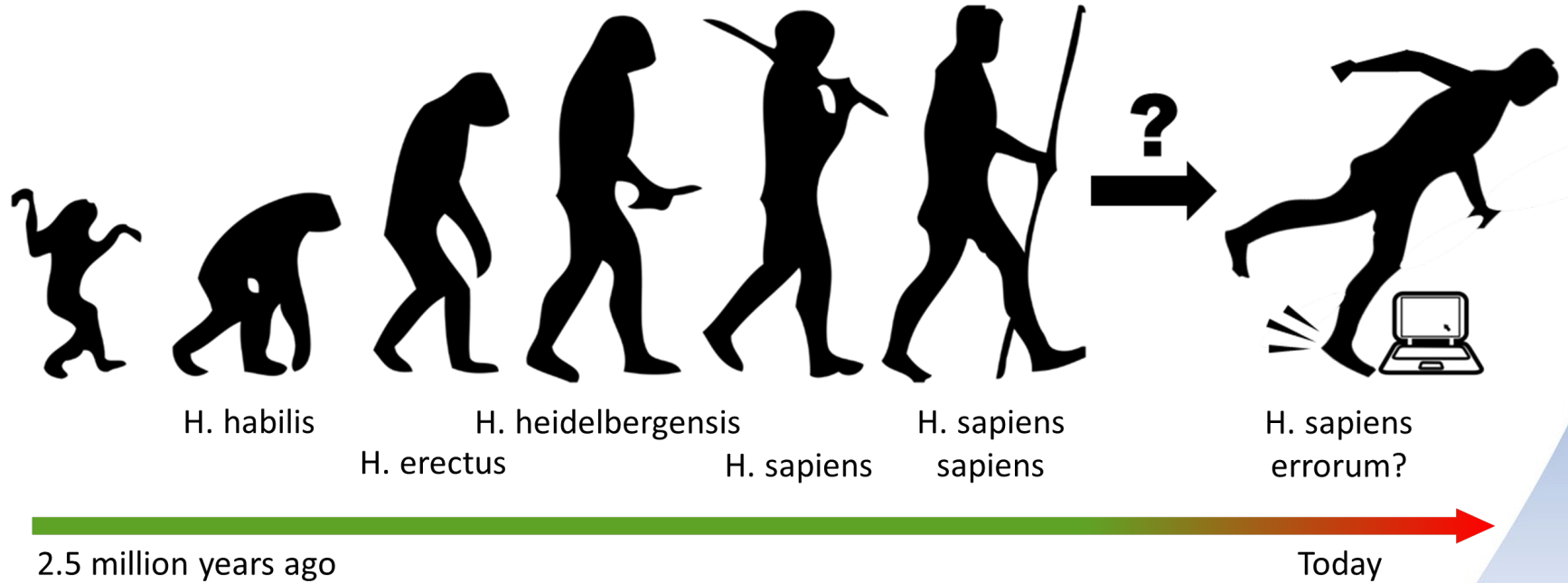
Absence of evidence $\neq$ evidence of absence



- 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

Removing or restricting the only demonstrated reliable source of failure preparation and recovery without first understanding or acknowledging the capability being removed or restricted introduces unknown risks

Did millions of years of natural selection go wrong?



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

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

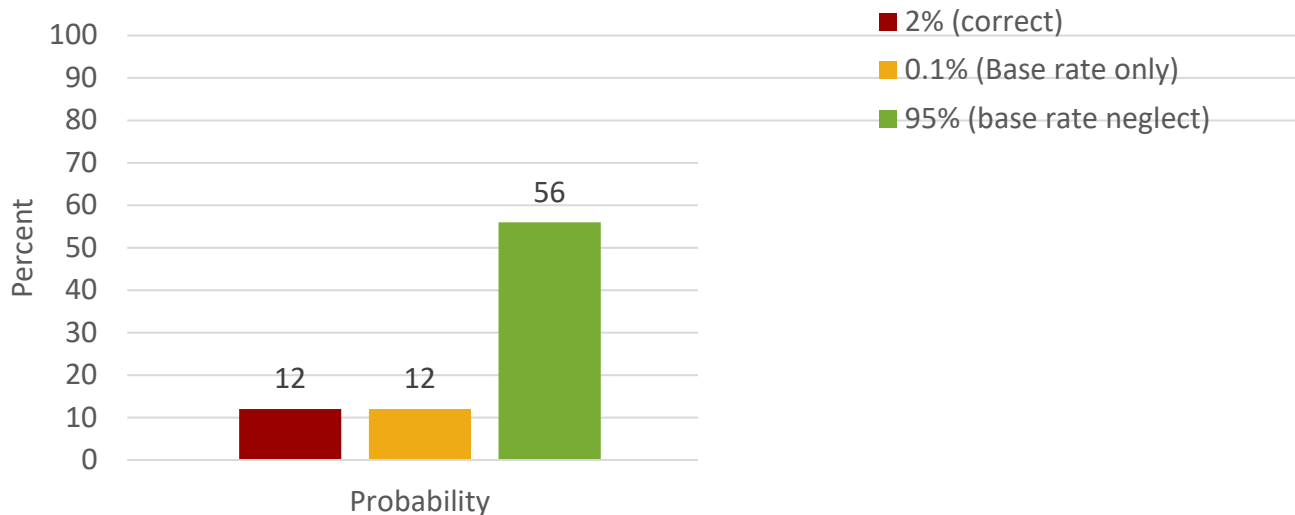
1 out of every 1000 Americans has disease X. Every time the test to detect disease X is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, out of every 1000 people who 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 = $\sim 2\%$ (or ~ 1 out of 50)
- Base rate conservative (considering *only* the base rate) = $\sim 0.1\%$ (or ~ 1 out of 1000)
- Base rate neglect (ignoring the base rate) = $\sim 95\%$ (or 950 out of 1000)

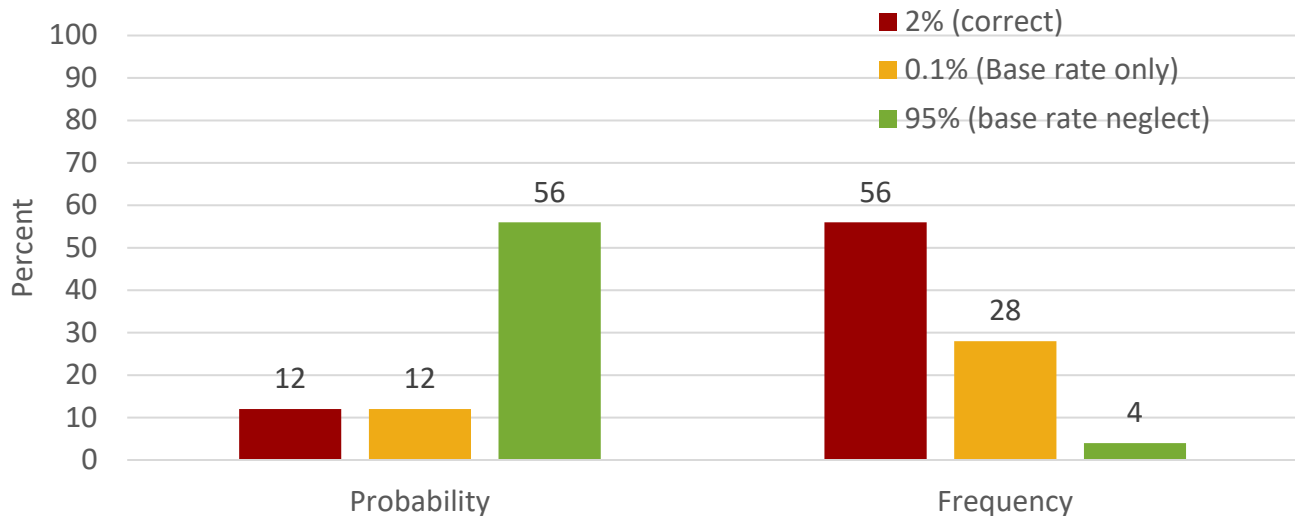


Judgments under uncertainty

(Cosmides & Tooby, 1996)



- Correct = $\sim 2\%$ (or ~ 1 out of 50)
- Base rate conservative (considering *only* the base rate) = $\sim 0.1\%$ (or ~ 1 out of 1000)
- Base rate neglect (ignoring the base rate) = $\sim 95\%$ (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 accidents¹



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 accidents¹



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 accidents²
- 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 accidents¹



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 accidents²
- 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?

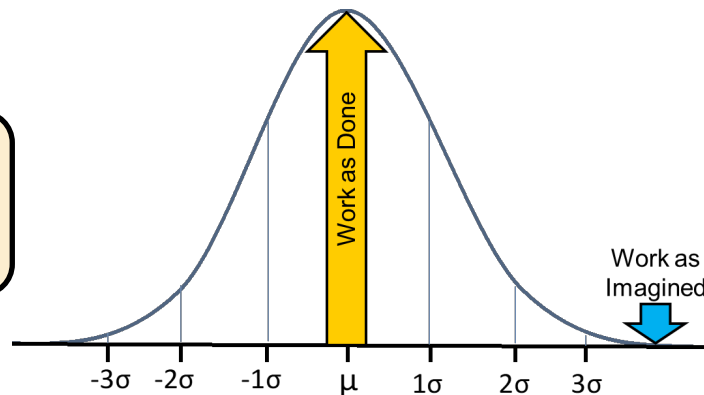
(1) Weigmann & Shappell, 2003; (2) Boeing, 2017; (3) PARC/CAST, 2013

Designing for “Nominal” Performance



- **Nominal** describes performance according to plan or design
- 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

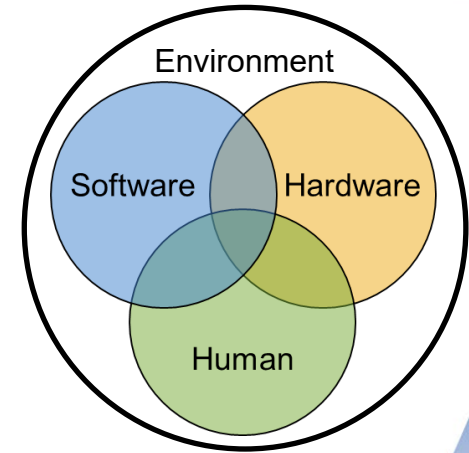
What are we planning and designing *for*?



System failures are “predictably unavoidable”*



- Designing for failure prevention can reduce failures but doesn't make systems failure-proof
 - All system components are potential sources of failure
- Designs must also enable preparing for and recovering from both expected and unexpected failures
 - Humans are the primary source of failure preparation and recovery



* van der Schaaf & Kanse (1999)

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...
 - Creating impoverished environments when automating tasks can limit ability to **notice or interpret what's happening**
 - Simplified interfaces and inability to over-ride or take control from automation can limit the ability to **respond**
 - Inadequacies in training can limit the ability to **know what to do now and to accurately predict what is going to happen later**
 - Designs that restrict feedback can inhibit **learning**

Work as done, not work as imagined



- Skilled operators succeed through failure recovery far more often than they fail through error
- Reasoning beyond procedures is essential for many recovery actions
- When do we need to equip people to do more than follow rules?
 - 14 CFR 121.557(a) : “In an emergency situation that requires immediate decision and action, the pilot in command may take any action that he considers necessary under the circumstances. In such a case he may deviate from prescribed operations procedures and methods, weather minimums, and this chapter, to the extent required in the interests of safety.”

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

Failure to adequately resource humans in design



- There are rarely (if ever) enough resources to go around in system design
 - Under-resourcing brittle system components guarantees failure
 - We get away with under-resourcing humans because they ARE resilient, and adjust their functioning to compensate for being under-resourced
 - But there are limits to this that are unrecognized and unacknowledged
- Often justified by claims that optimizing or enabling human task performance is a nice-to-have not a need-to-have (i.e., can't lead to loss of mission or loss of crew).

How we can help



- Change how we, as a community, talk about human factors & human performance – change the conversation that humans are the safety problem
- Recognize system failures as both routine and unavoidable – account for that in our study designs
- 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”
- Support capability for resilient performance through system designs that align with what our cognitive systems evolved to do



Thank you

jon.holbrook@nasa.gov