CITY
UNIVERSITY OF LONDON
EST 1894

New systems, inevitable doubt, actual risk

and how to deal with all that better

Peter Bishop[1,2], Andrey Povyakalo[1], Lorenzo Strigini[1]

[1]Centre for Software Reliability - City, University of London, U.K.

[2]Adelard LLP, London, U.K.

CSR Building confidence in a computerised world

www.csr.city.ac.uk

For highly critical computer applications we have...

sensible regimes, demanding

- *before* such a system is allowed into operation
- a demonstration that harm from its operation is unlikely enough

and we have remarkably safe operation in many areas

(e.g. scheduled civilian air transport)

- despite "ultra-high" dependability requirements

  like 10^{-9} probability of catastrophic failure conditions per flight hr

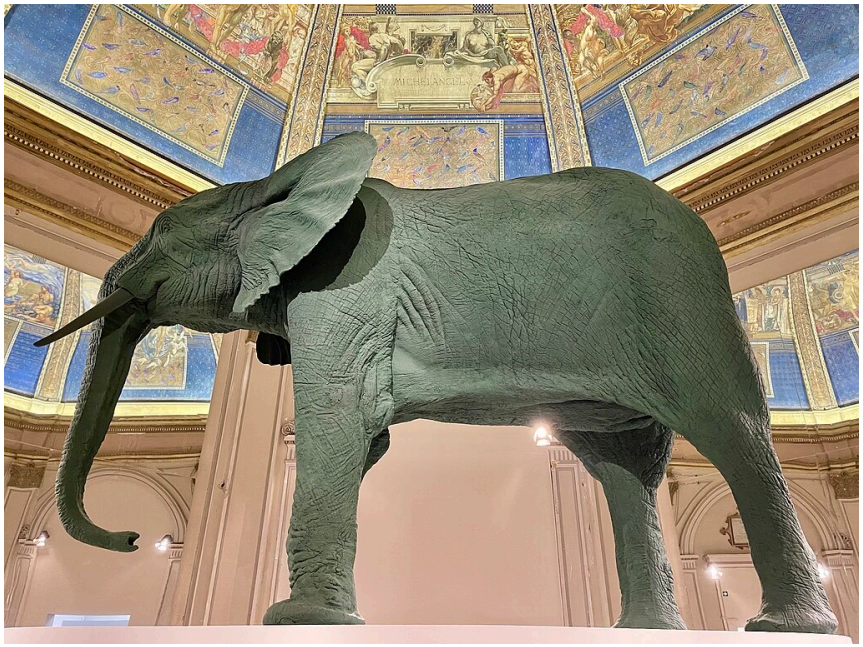
- so when a novel system comes along that requires UHD...

  e.g. "an automated car shall cause death at a rate ≤ 1 in 10^{-10} $mile^{-1}$ "

  ... we *rightly* demand a similarly stringent assurance regime

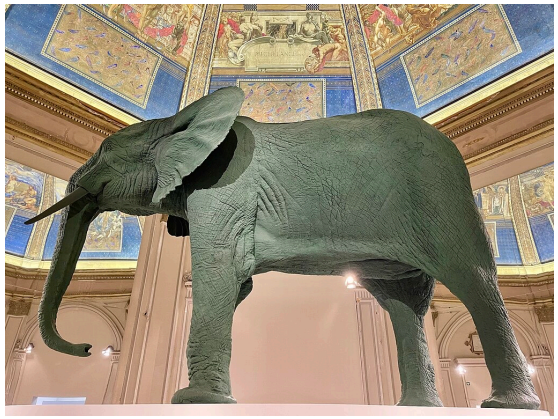
this should buy the public peace of mind... or should it?

There's an elephant in the room...



https://commons.wikimedia.org/wiki/File:The_elephant_in_the_room_at_Arsenale_(52196585578).jpg
license: https://creativecommons.org/licenses/by/2.0/deed.en

The elephant in the room... epistemic uncertainty



https://commons.wikimedia.org/wiki/File:The_elephant_in_the_room_at_Arsenale_(52196585578).jpg
license: https://creativecommons.org/licenses/by/2.0/deed.en

- sometimes that **carefully verified** demonstration of acceptable safety is **wrong**:
  - *in operation after approval,* dangerous flaws are found & fixed *(see airworthiness directives)*
  - *or* disasters happen (think Boeing 737 MAX)
  - e.g. in airliners, nuclear reactors a fraction of new systems have proved not to be ultra-safe
- however, accumulating safe, surprise-free operation under strict monitoring will reassure us about safety of a type

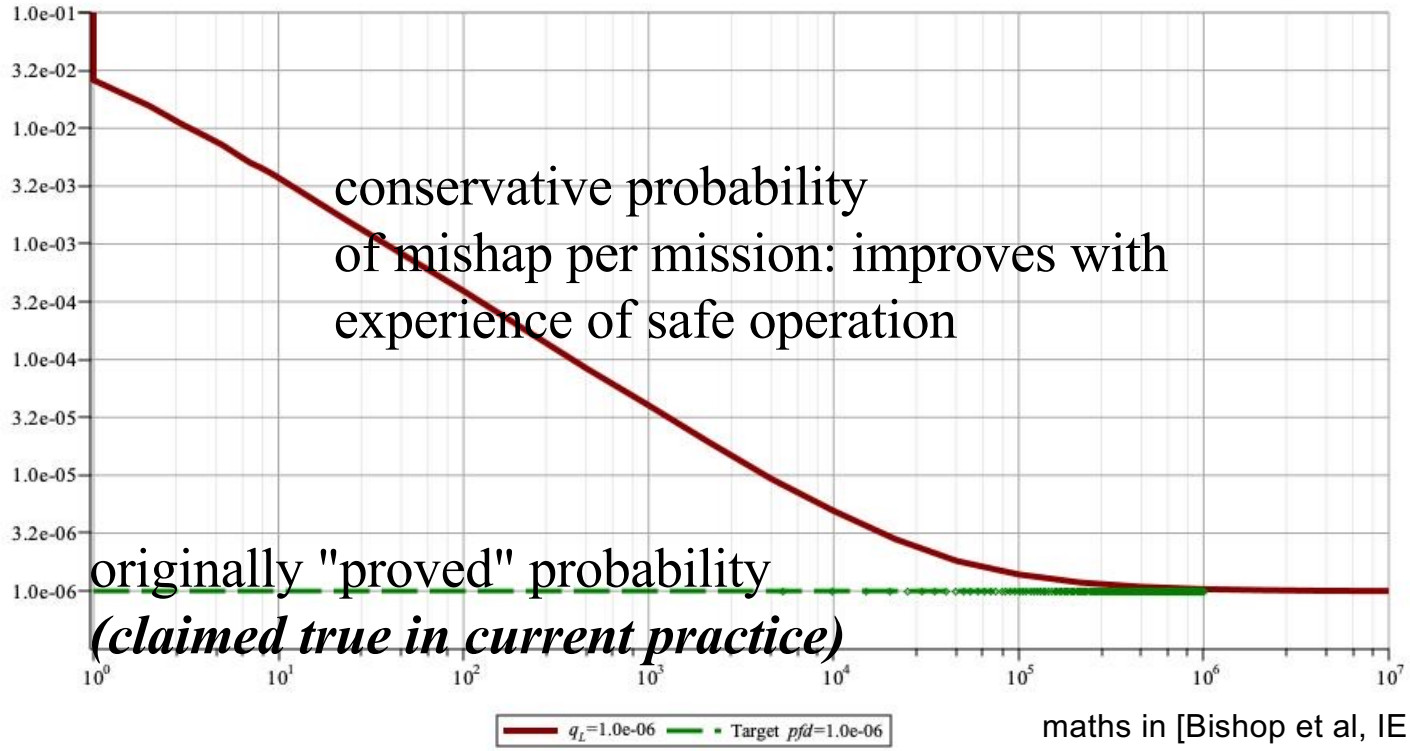
So, given a *good* argument showing that a system is safe enough

Say, it proves that the probability of mishap per mission, *pfd*, is ≤q_L, say ***$q_L \leq 10^{-6}$*** ***if*** the argument is correct

- with probability p_L that the claim is correct of .. say p_L=90%
- what should the airline/regulator/passenger think of risk per flight now?
  - in the range [0, $p_L\ q_L + (1- p_L)\ q_H$]
  - q_H : *pfd* if the argument is wrong – typically unknown
- with the numbers given, when you start operation, real risk per mission is between [~0, ~0.1]
  NOT [0..10^{-6}]

What does a realistic assessment of risk look like?

Observing more and more safe operation you infer that even if this system belongs to the unlucky 10%, it cannot be *very* bad.... thus:



maths in [Bishop et al, IEEETSE 2011]

... you get close to the "demonstrated" low risk in the long run... asymptotically

This more realistic estimate should allow better decisions about licensing, deployment!

What is to be done?

- Given inevitability of doubt, acknowledge the attendant risk
- study history: how much we should doubt our claims, depending on kind of system and claim
- export from mature fields (aviation, nuclear, ...?) not bad theory but good practice: strict monitoring
- control overall fleet risk (our paper, ISSRE 2022)

- to reduce risk that we must live with, improve arguments with
  - not just claims "if argument is correct" but confidence in it
  - "backup" (higher confidence, modest claims) sub-arguments
  - higher confidence (hard! But somewhere low-hanging fruits?)

- to make better theory helpful, use psychology/sociology of risk decisions in the various applications
- (and do the maths: we have been doing that)

Thank you for your attention..

Questions, comments?

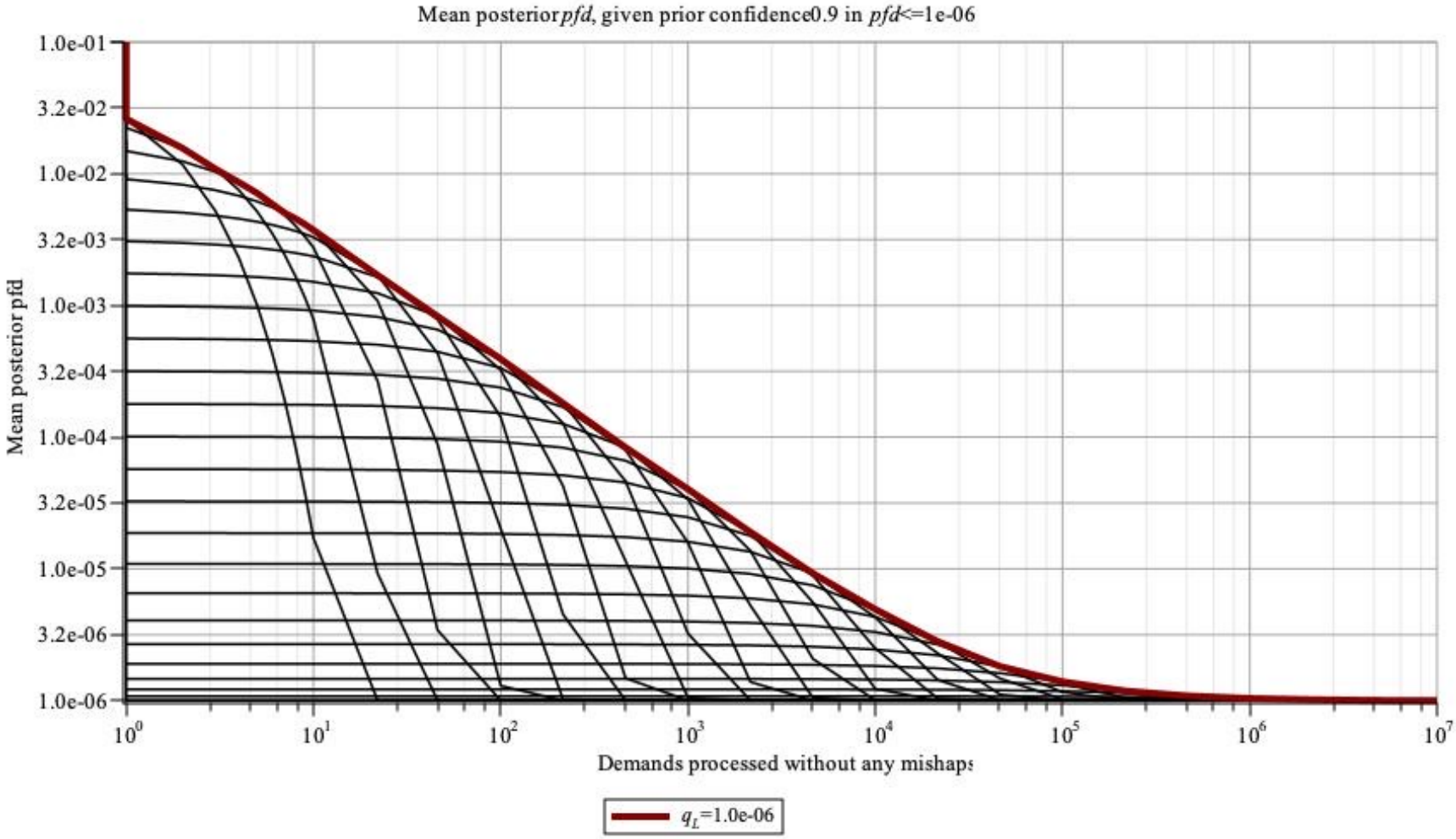
Do Email us

Theorems, extensions and references: a paper will be on Arxiv in a few days, ask us for the URL

SPARES for questions

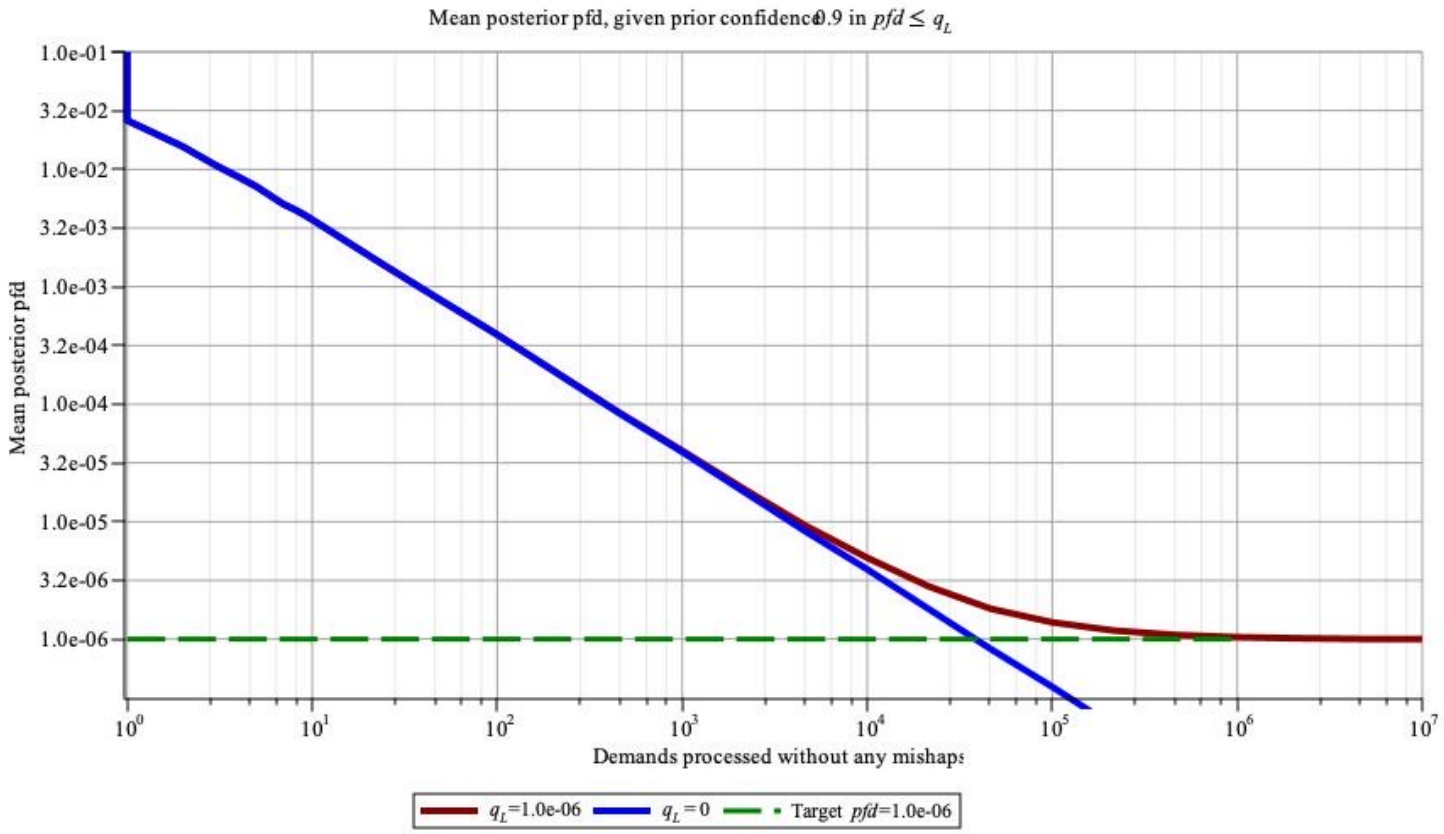
How did we draw that curve of conservative maximum *pfd*?

"conservative Bayesian inference"



Can you improve... by proving a better q_L?

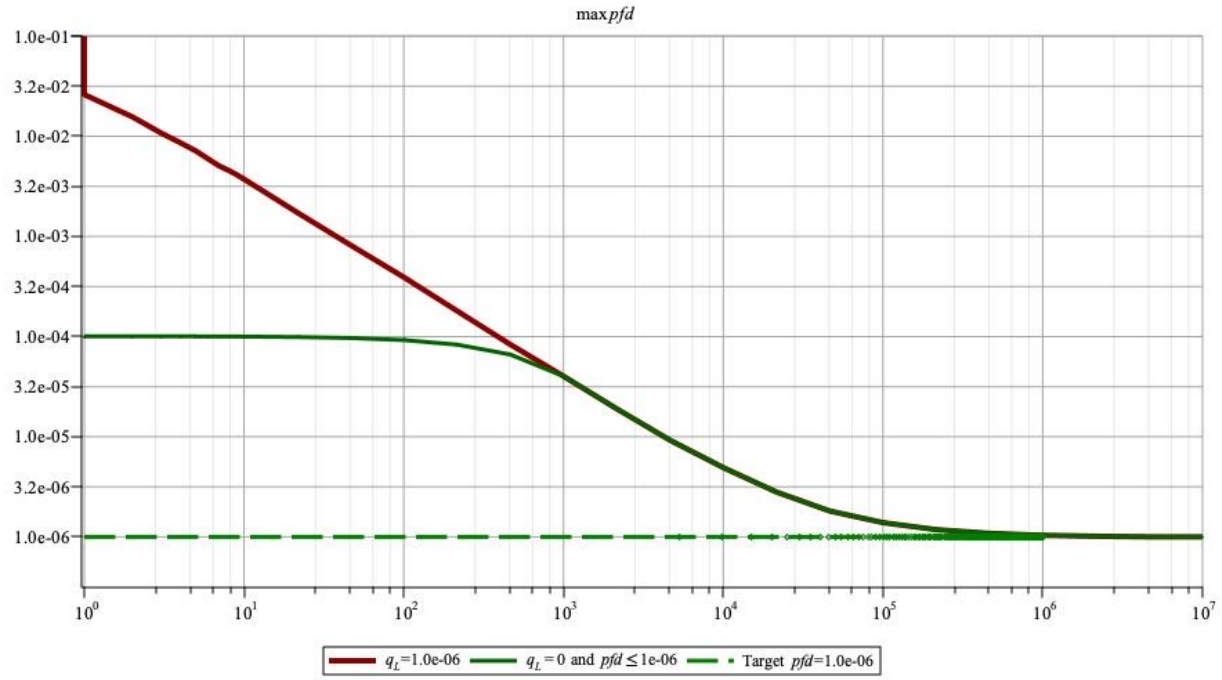
your curve will asymptotically approach that lower q_L



it helps – only in the long run!

How to add "backup" arguments

High prior confidence that if your main argument is wrong, still you know an upper bound on q_H that is <1

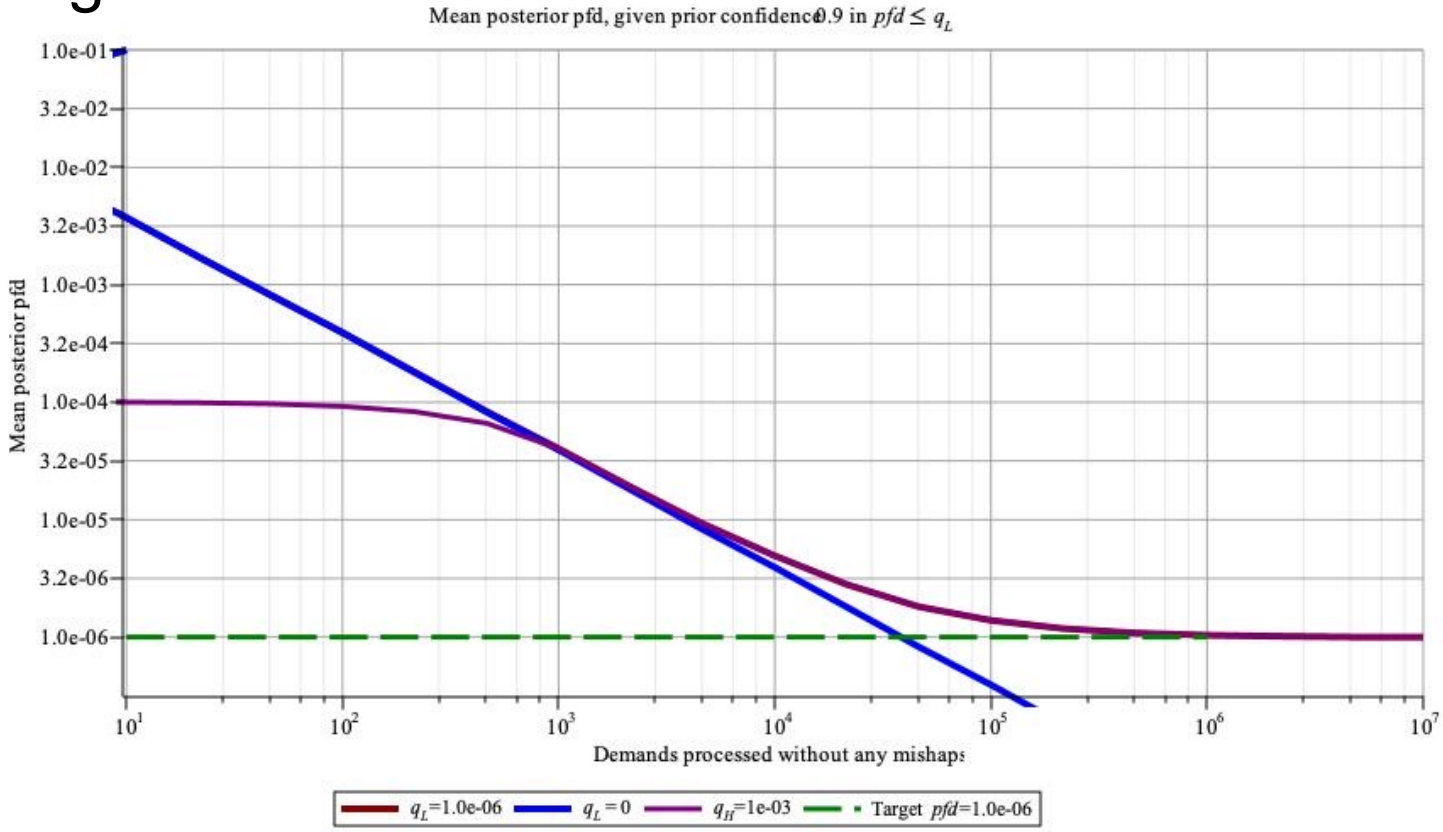


This limits initial risk; after a while, it stops helping

Combine both...?

It helps.

Still long time to reach desired risk level



We could do better: multiple backups arguments, increased confidence in them

by studying the actual evidence about the specific system

Why the current fiction that a verified claim is true?

- simpler
- inevitably, commercial/political pressures
  - who feels like defending "gambling with people's lives"?
- but importantly also:
  - human minds treat "epistemic uncertainty" differently from "aleatory uncertainty"
    + people may accept that "safe" only means "low probability of accidents"
    + but are uneasy accepting uncertainty ***about that probability***
  - treating the latter uncertainty by probability goes against the grain
    + for many experts and lay people alike
    + (despite widespread use of Bayesian approaches to risk)
  - ... despite the distinction being often an illusion

- maybe the current fictitious separation has societal advantages?
    + avoids some forms of corruption of the process?

How do we manage fleet level risk?

Example of "confidence bootstrapping":

incremental deployment contains overall risk of mishap for whole fleet [Bishop et al. . ISSRE 2022]