CITY
UNIVERSITY OF LONDON
EST 1894

How to reason about Risk,

given Inevitable Doubt on Arguments for High Dependability

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

[1]Centre for Software Reliability – City St George's, University of London, U.K.
[2]Adelard , NCC Group, London, U.K.

Background: for engineered systems with potential for unintended serious harm

- sensibly, regulations demand

  before allowing large scale operation,
  demonstration that harm from operation is unlikely enough

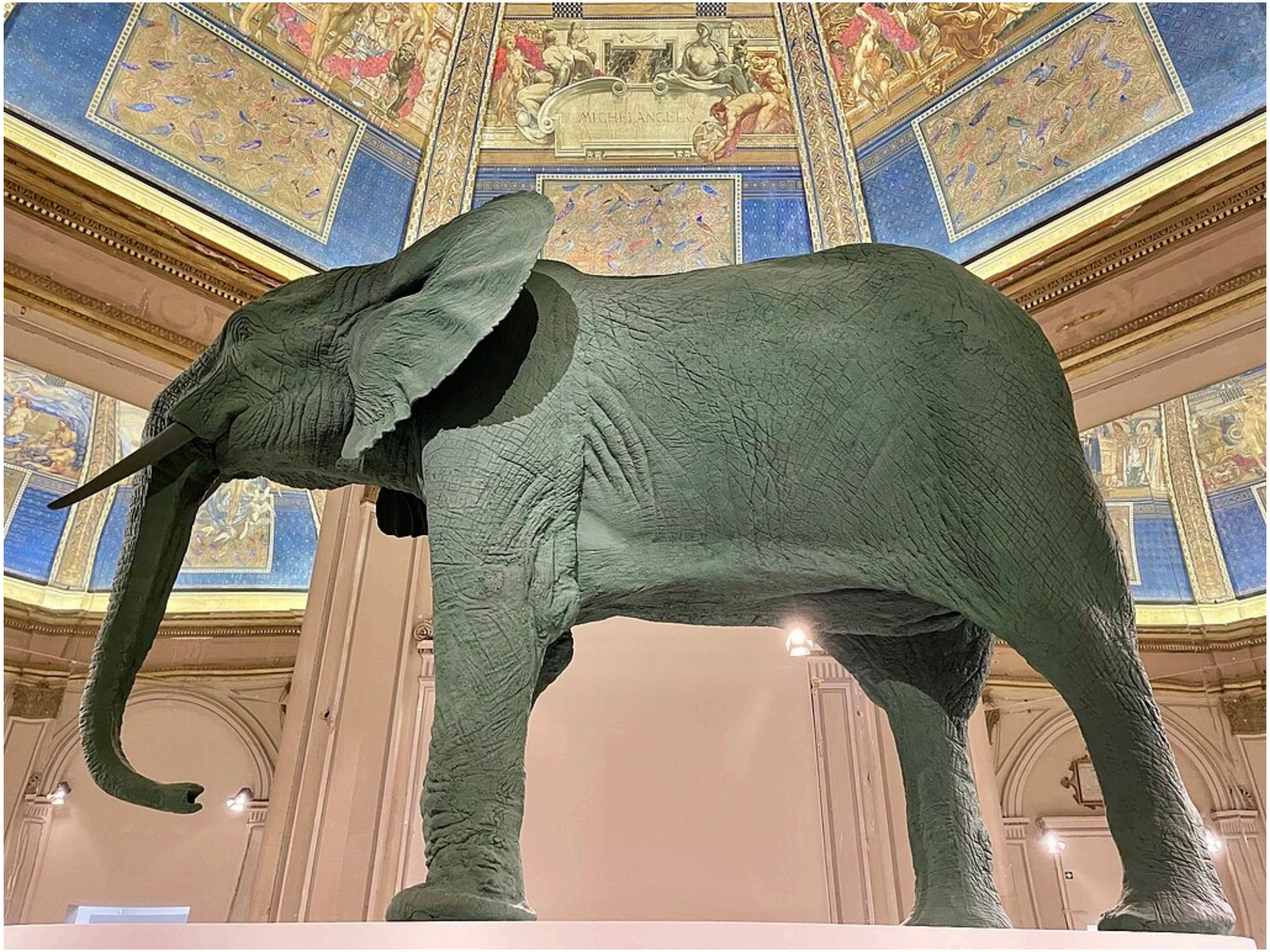
- serious effort is spent on this demonstration

- indeed we have remarkably safe operation in many areas
- although the safety levels required are ***hard*** to demonstrate in advance

  e.g. $\leq \mathbf{10^{-9}}$ per flight hour probability of *catastrophic failure conditions*

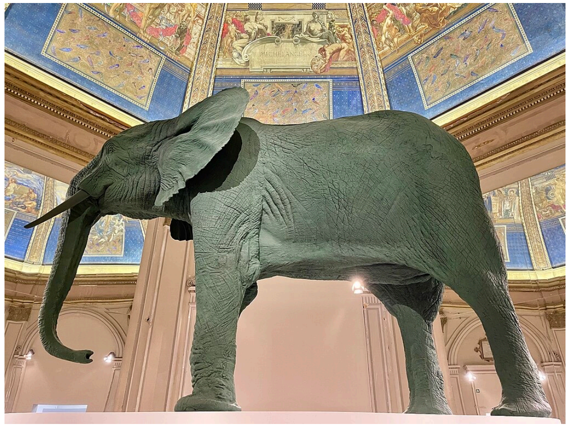
All this should give everyone peace of mind...

But ...



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

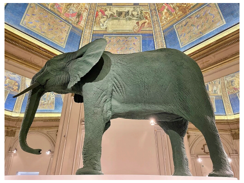
The elephant in the room... epistemic uncertainty



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 *(e.g. "airworthiness directives")*
  - *or* disasters happen (think Boeing 737 MAX)
  - e.g. in airliners, nuclear reactors, ... a fraction of systems have proved not to be as safe as required and "demonstrated"

the advertised risk figures may be badly wrong



as often pointed out by sociologists, antinuclear protesters, ... and more quietly among specialists

However, usually

- the new system type is gradually deployed
- seeing safe, surprise-free operation rightly reassures us about safety
- surprisingly, this process is not part of formal certification / authorisation processes
- ... how can then regulators, insurers, users take the right decisions?

... how safe ***should*** we trust a newly approved system to be?

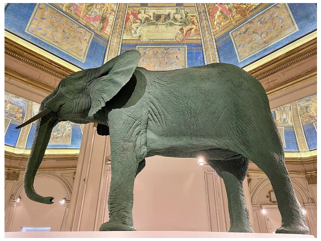
Simple scenario: we have a *good* argument showing that a system is safe enough...

Suppose e.g. for a new aircraft type

- **proved** probability of mishap per flight **$\leq 10^{-6}$**
  if the argument is correct
- but if it's wrong, this probability is **unknown** – might be 1!
- assume **90% confidence** that it is correct

what should the airline / regulator / insurer / passenger think of risk?

The upper bound on probability of mishap in the first flight **is**

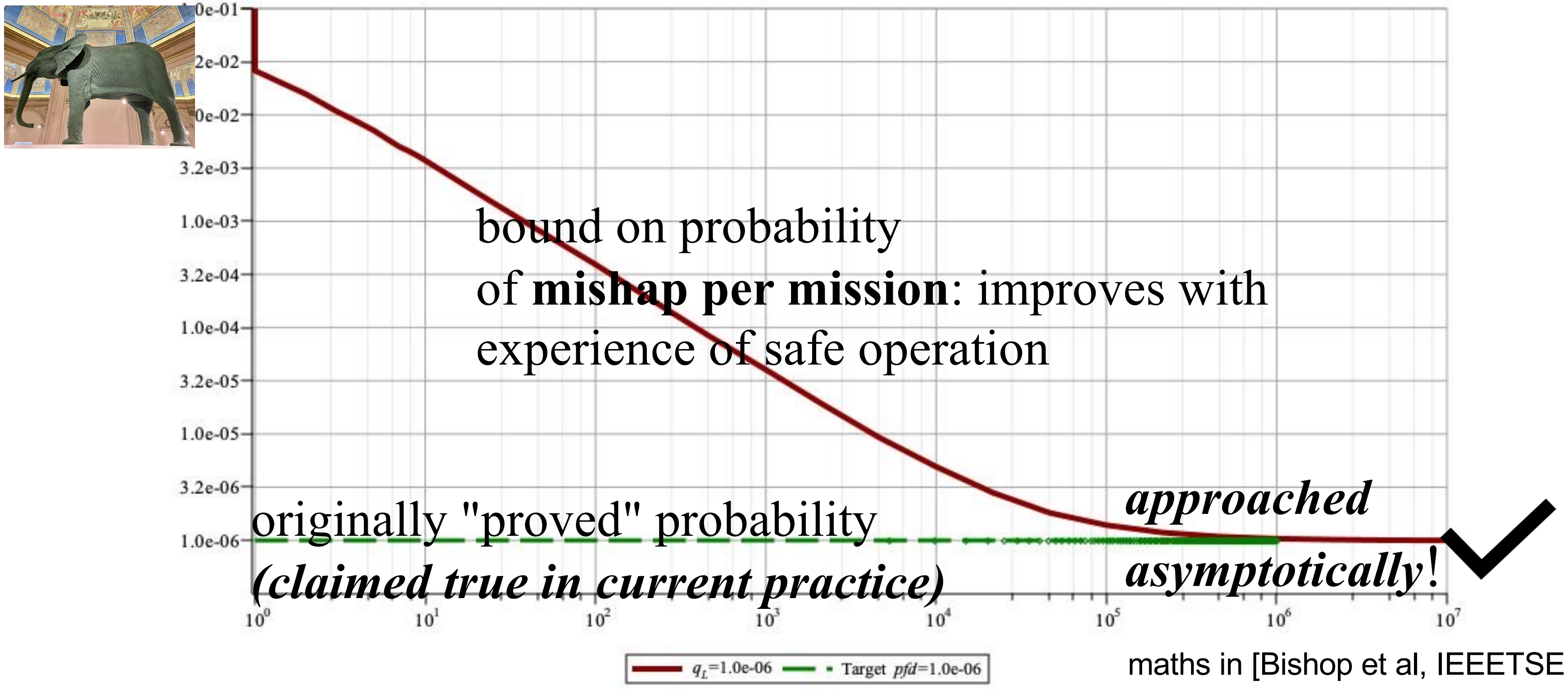


$$0.9 \times 10^{-6} + (1-0.9) \times 1 \quad , \quad \text{i.e.} \quad \sim 0.1$$

a lot more than the advertised 10^{-6} !

The good news: as we see more and more safe operation, we can show...

- how much less likely this system is to be in the unlucky 10%
- that even if it does, lack of mishaps so far proves they cannot be ***very*** likely. Thus:



bound on probability of **mishap per mission**: improves with experience of safe operation

originally "proved" probability ***(claimed true in current practice)***

approached asymptotically! ✓

maths in [Bishop et al, IEEETSE 2011]

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

So, is the decision process for acceptance broken ? What is to be done?

- **acknowledge** inevitable doubt and the attendant **risk**
- study history: learn roughly *how much* we should doubt *proved* safety claims, for each kind of system and of claim
- exploit good practices (e.g. strict monitoring in operation) to support ***rational*** growth in confidence

- improve safety arguments
  - include "backup" sub-arguments (more modest claims with higher confidence)
  - improve confidence in main claim? (hard! Any low-hanging fruits?)
  - change claims? E.g. overall fleet risk (Bishop *et al* 2022)
  - exploit more historical evidence about risk parameters

- make the improved theory actually help the process: learn from psychology/sociology of decision under uncertainty

Thank you for your attention..

Questions, comments, resonance with situation in **your** area?

Interest in case study projects?

Do Email us:
{P.Bishop, A.A.Povyakalo, L.Strigini}@city.ac.uk

Theorems, extensions, references: Arxiv article "scheduled to be announced at Thu, 19 Sep 2024 00 GMT"

Some background:

Bishop, P., Povyakalo, A. & Strigini, L. (2022). Bootstrapping confidence in future safety based on past safe operation. ISSRE 2022, ISSN 1071-9458 doi: 10.1109/ISSRE55969.2022.00020, https://openaccess.city.ac.uk/28641/

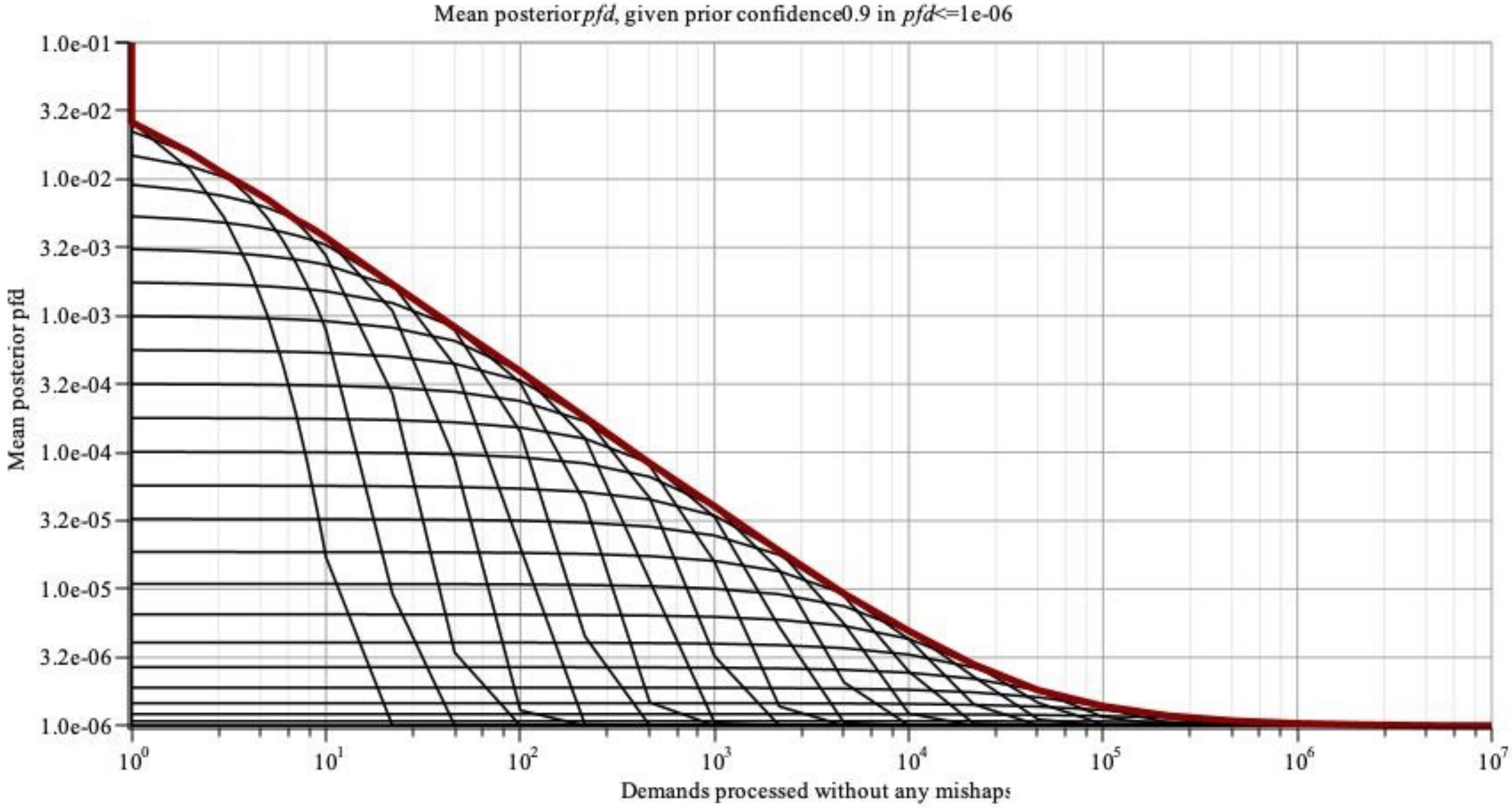
Bishop, P. G., Bloomfield, R. E., Littlewood, B. , Povyakalo, A. A. & Wright, D. (2011). Toward a Formalism for Conservative Claims about the Dependability of Software-Based Systems. IEEE Transactions on Software Engineering, 37(5), pp. 708-717. doi: 10.1109/TSE.2010.67 , https://openaccess.city.ac.uk/id/eprint/1070/

Littlewood, B. & Strigini, L. (1993). Validation of Ultrahigh Dependability for Software-Based Systems. Communications of the ACM (CACM), 36(11), pp. 69-80. doi: 10.1145/163359.163373 https://openaccess.city.ac.uk/id/eprint/1251/

Additional slides

How did we draw that curve of worst-case *pfd*?

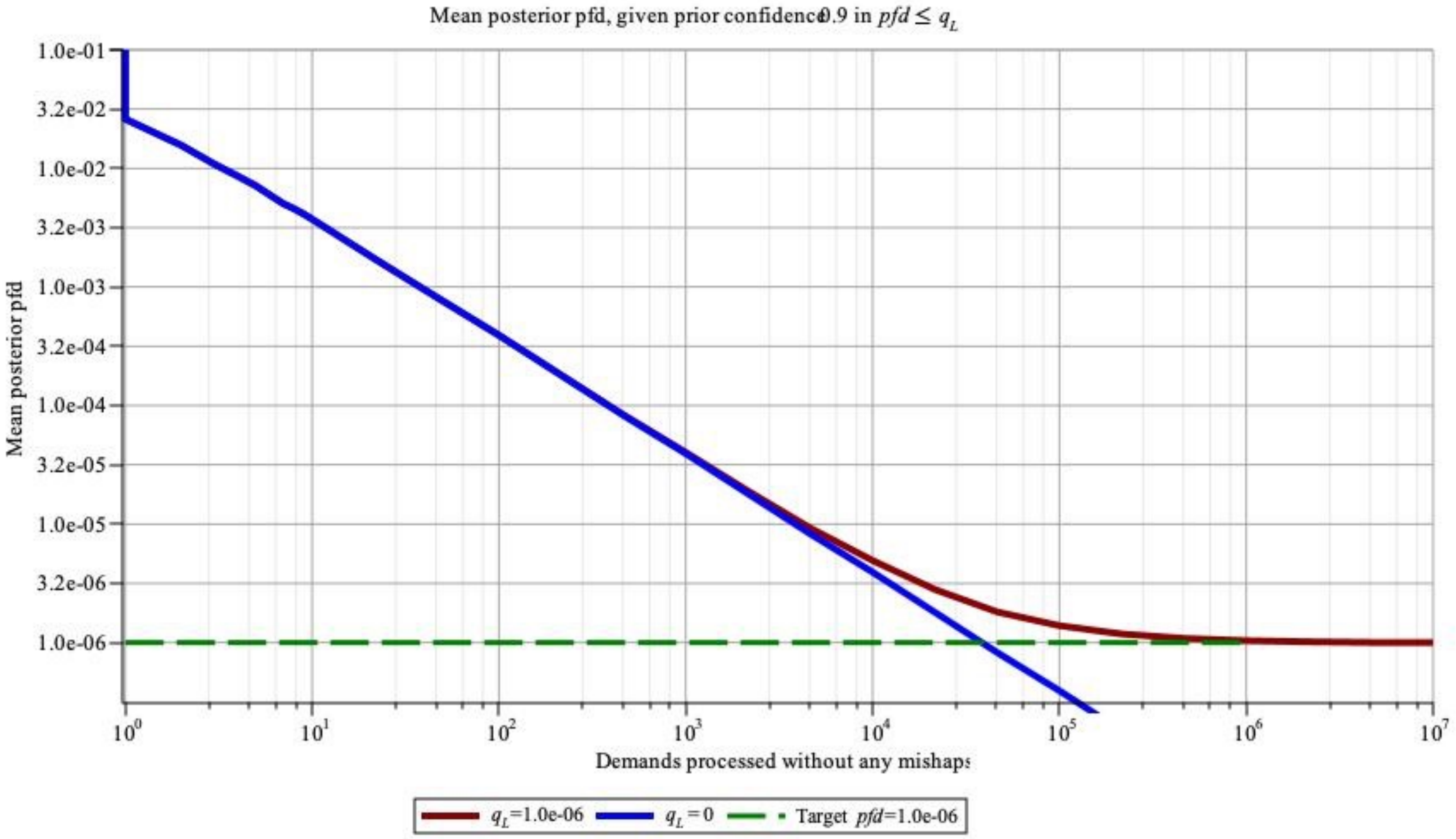
"conservative Bayesian inference"



- which distribution is "worst-case" changes with increasing amounts of past successful operation
- so the evolving worst-case prediction is given by the envelope above

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

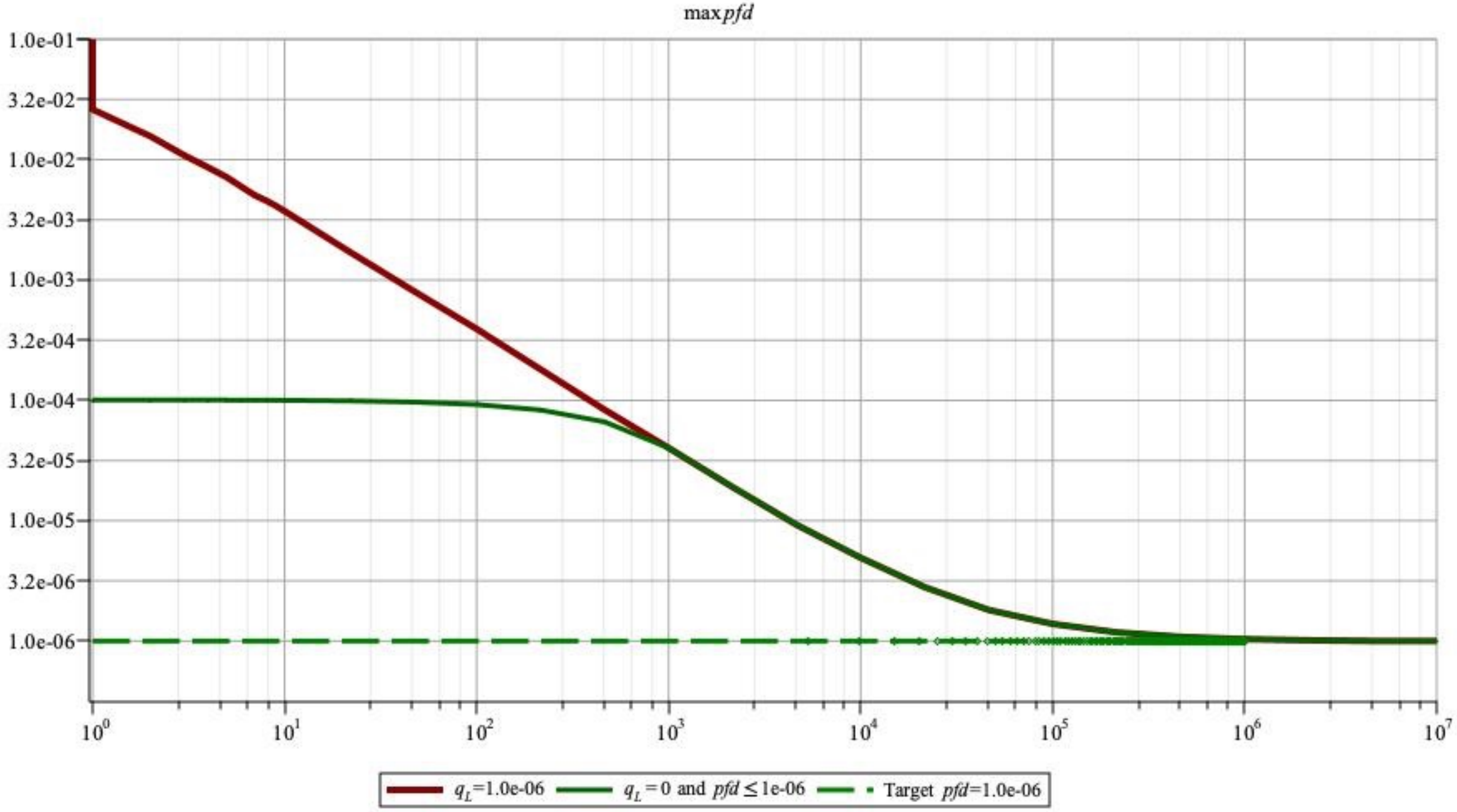
your curve will asymptotically approach that lower q_L



it helps – but 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



lower initial risk

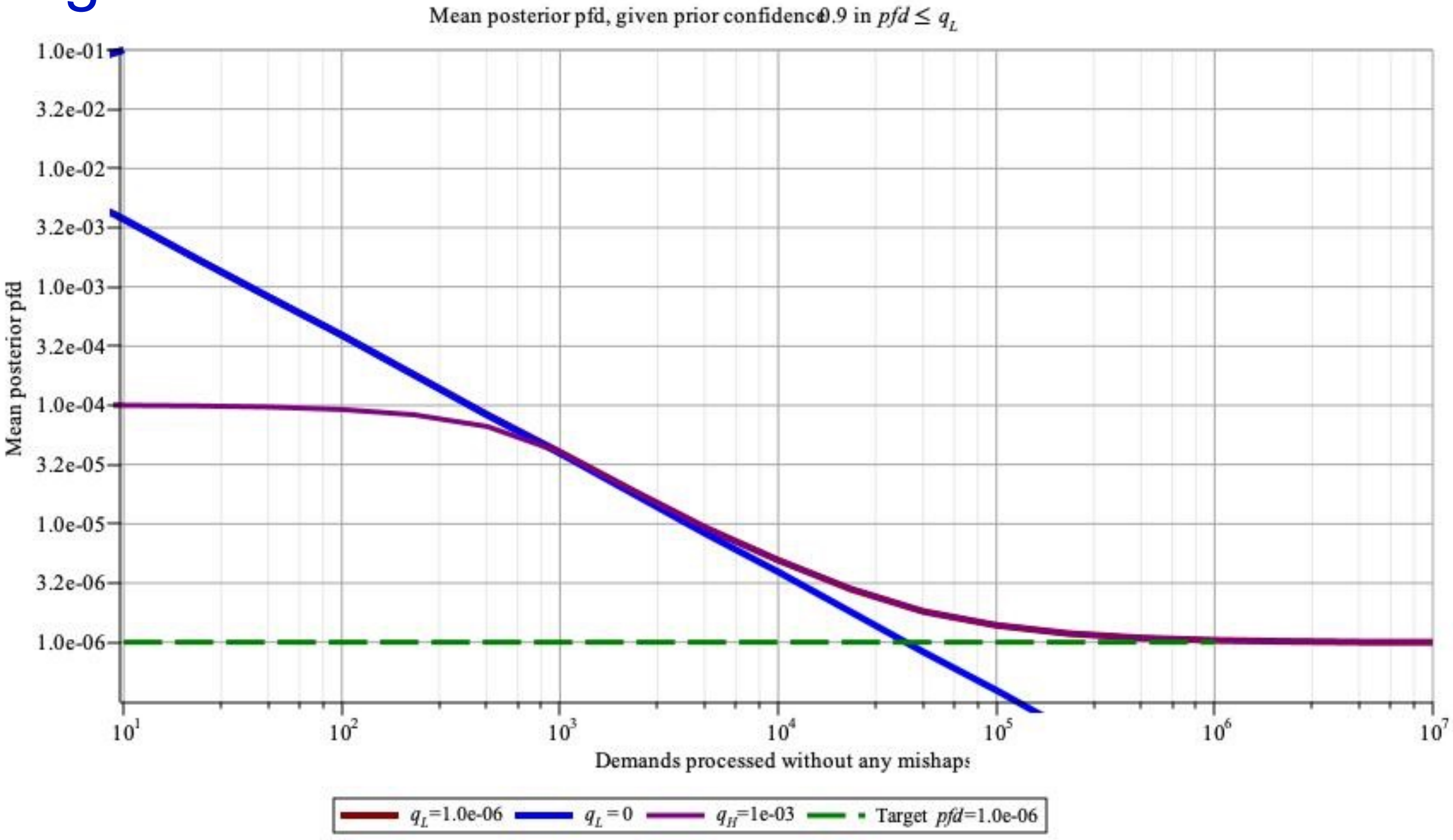
same asymptote

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

Combine both...?

It helps.

Still long time to reach desired risk level



We can do better: multiple backup arguments, each claiming less but with more confidence

by studying the actual evidence about the specific system

Why the current fiction that a verified claim is true?

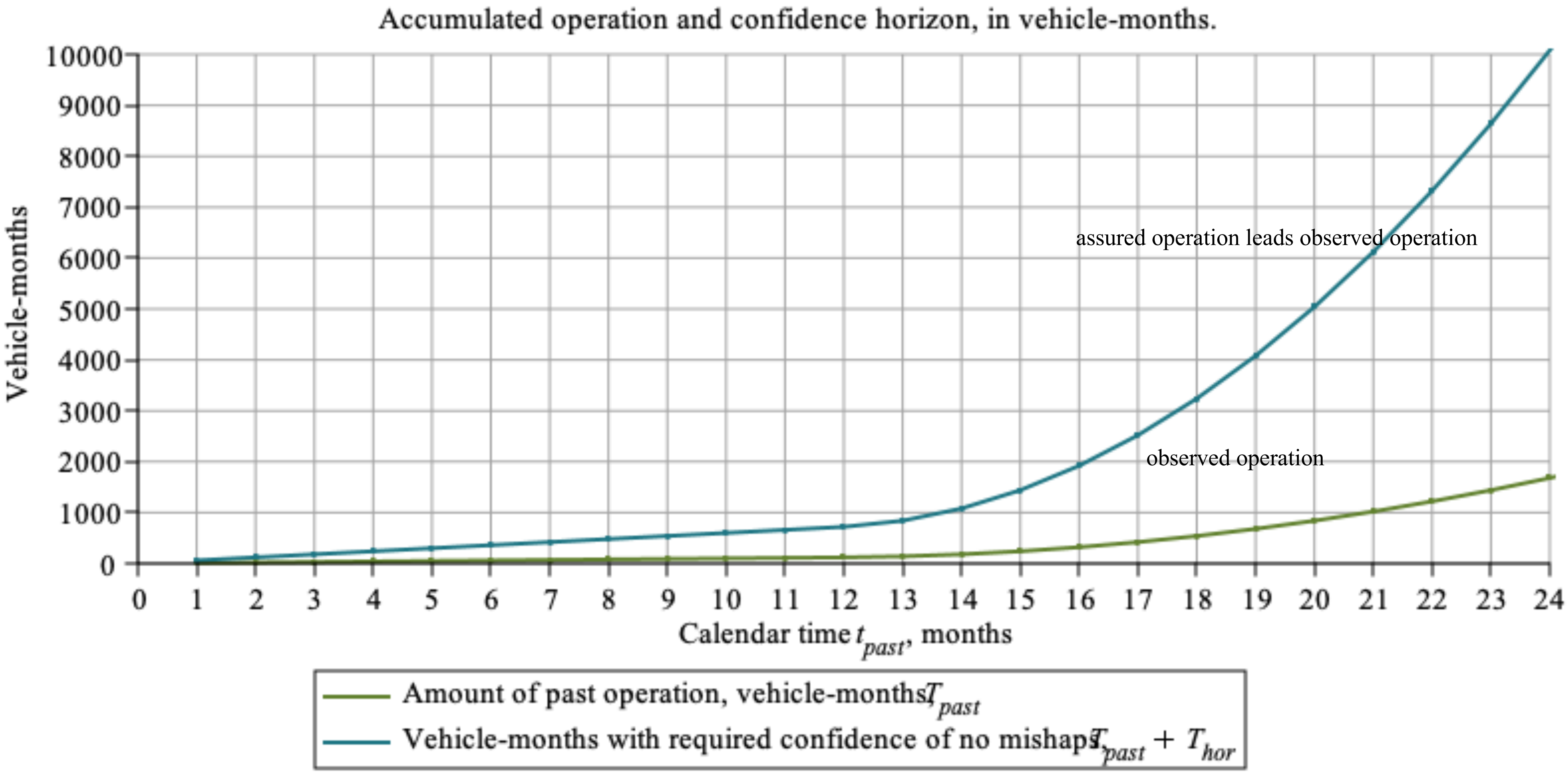
- simpler
- inevitably, commercial/political pressures
  - who feels like defending "gambling with people's lives"?
- but importantly also:
  - human minds treat "epistemic uncertainty" differently from "aleatory uncertainty"
    - + people may accept that "safe" means "low probability of accidents" rather than "no accidents"
    - + but are uneasy accepting uncertainty ***about that probability***
  - treating the latter uncertainty by probability goes against the grain
    - + for many lay people and experts 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?
    - + but certainly the myth favours other forms

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]


Accumulated operation and confidence horizon, in vehicle-months.
Vehicle-months
Calendar time t_{past}, months
assured operation leads observed operation
observed operation
Amount of past operation, vehicle-months T_{past}
Vehicle-months with required confidence of no mishaps $T_{past} + T_{hor}$