Safe, Ethical & Sustainable: A Mantra for All Seasons?



John McDermid provides some guiding principles on how to achieve and manage the safety of complex systems whose failure causes and consequences go beyond the concerns of traditional safety engineering. He sketches some new models for safety engineering and proposes the adoption of the mantra "safe, ethical and sustainable" to not only focus the attention of the community on the key issues, but also to influence politicians and policies.

There are many principles and "laws" relevant to systems and safety engineering. At a time when the community is grappling with systems of unprecedented complexity I am reminded of Mencken's Law:

"For every complex problem, there is a solution that is clear, simple, and wrong."

There is also the famous [1]:

"All models are wrong, some are useful."

And two further (related) quotes, the first from Humpty Dumpty:

"When I use a word, it means just what I choose it to mean — neither more nor less."

"Words in their primary or immediate signification stand for nothing but the ideas in the mind of him that uses them."

The latter, less well-known perhaps, is from John Locke [2] and may have been the source of Lewis Carroll's ideas for Humpty Dumpty's scornful remark to Alice [3].

Why do I quote these? I want to suggest some new models for how we think about complex problems, with a particular emphasis on safety. These models will be wrong in some facet or some situation, but they will be useful if they help focus thought in a constructive way. And all I have is words. My aim is to convey the "ideas in the mind …" – but perhaps if they are repeated often enough (like a mantra) they will start to have significance in other minds too. But I start with the usage of a word which I sort of hate.

Elegance in Systems Engineering

Systems engineering has long been presented as a multi-disciplinary, or trans-disciplinary approach to solving complex, multi-faceted problems. To me 'elegance' means being 'graceful or stylish' but the systems engineering community has hijacked it to mean [4]:

- Efficacy – how well does it achieve the desired outcomes?
- Efficiency – how economical is it in use of resources both to develop and to operate it?
- Robustness – how well does the system perform in unanticipated circumstances?
- Minimising unintended consequences – how well does the system do in reducing unwanted and unanticipated consequences?

The aim – akin to mine here – is to give some guiding principles to help achieve focus on the key issues when trying to solve complex problems. I accept that some systems can be elegant, but I think the complexities of modern systems, the difficulties of dealing with brownfield systems design, the interconnectedness and interdependencies of systems, and the fact that efficacious and robust solutions can be downright ugly, makes me think this is a somewhat naïve characterisation. The four principles seem sensible, but I sort of hate the label 'elegant'. Overall, I view this as a good example as it indicates the kind of models I want to create – guiding principles that are useful, although wrong in some sense.

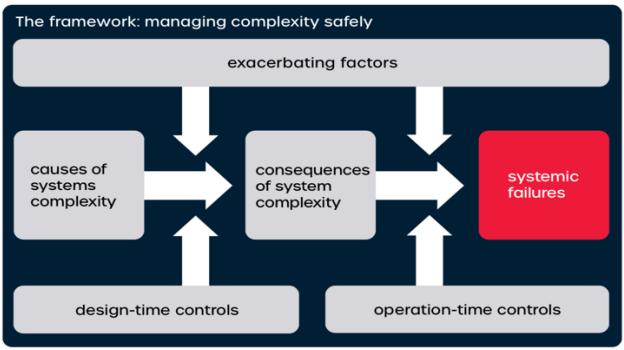
Complex Systems

I have used the term 'complex systems' several times. What do I mean by this? It is common to make a distinction between complicated and complex systems, saying that complex systems exhibit 'emergent properties', that is, properties of the whole that are not simply properties of the individual system components (and their inter-relationships). Some argue that this definition is unscientific – but science doesn't have the answers to everything – moreover, according to Aristotle [5]:

"In the case of all things which have several parts and in which the totality is not, as it were, a mere heap, but the whole is something besides the parts."

I hope I may be excused for relying on this unscientific definition, since the implied uncertainties of 'emergence' indicate one of the key challenges we need to face in safety engineering. We can't assume that the system development will produce results that safely manage all these uncertainties from 'day one', thus we also need to consider the overall governance of the systems in deployment, to identify unintended consequences, to learn from them and provide feedback to improve the system. To me this is another reason that the notion of 'elegance' in systems is somewhat naïve – it implicitly assumes that the system can be designed to achieve its intended use 'full stop'. The reality is, however, that we need to produce systems that are 'good enough' (safe enough) that we can deploy them and then manage them to achieve acceptable safety through life, recognising that what is acceptable may change over time, as technology or society's expectations evolve.

A study on Safety of Complex Systems for the Royal Academy of Engineering [6] introduced a model of how complexity, as opposed to mechanistic failures, contribute to systemic failures, see the figure. It shows the role of both design-time and operation-time controls for reducing the likelihood of systemic failures, or for mitigating the consequences. The exacerbating factors are those issues that can have adverse impacts, particularly on the controls.



The study also introduced a three-layer model for governance of such systems spanning:

- Governance – cross-jurisdictional incentives and requirements for organisations to adhere to best practice through regulations, standards, soft law, etc.
- Management – risk control and trade-offs in an organisation, management of supply chain dynamics and the sustainment of long-term knowledge
- Task & Technical – the behaviour of the technological elements of the system, the users and other stakeholders, in their context of use

The task & technical level is the traditional province of safety engineering (and it seems the main focus of the work on elegance in systems engineering), but it is insufficient to address and manage the problems of complex systems. The report presents some examples of using the framework, showing the issues in the governance, management and task & technical levels, for each of the elements in the model – causes of complexity, exacerbating factors, etc. For example, in the case of the Uber Tempe accident [7] we can see:

- Exacerbating factor (management layer) – casualisation of labour, using untrained safety drivers (gig economy)
- Causes (task & technical) – mentally unstimulating but critical tasks of system supervision

This work provides two key aspects of the broader models – ways of thinking about systemic failures arising from complexity, and the need to consider causes and controls at the levels of management and governance, as well as those within the more traditional scope of safety engineering. It also highlights the need to consider safety management through life; of course, this is not new, but treating safety management as a continuum rather than having a discontinuity as the system enters service is at least a new emphasis.

Benefits and Harms

Safety mainly focuses on harms to people, i.e. death and injury, and ways of reducing the attendant risk. Of course, benefits are considered, albeit in a simplistic, or limited way. For example, a cost benefit analysis can be used to support a claim that a risk has been reduced As Low As Reasonably Practicable (ALARP). This however is a narrow application focused on risk reduction and doesn't consider the wider societal advantages or benefits of having the system in the first place.

Thus, the proposed shift in our models of safety is to consider benefits as well as harms in all aspects of the design and analysis of systems. In terms of models, this requires definition of a trade-space of benefits and harms that should be considered in designing and analysing systems. This should be much broader than the conventional focus of safety engineering on harm to individuals and should include consideration of society and the environment. A challenge is the need to trade-off between incommensurable factors; this is not easily resolved but doing pairwise comparisons between designs that are close in the trade-space is one possible tactic.

Individual, Societal and Environmental

Although broadening models, it is still desirable to take a human-centric approach, and thus I use well-being as an 'umbrella' [8]. In this content, it is possible to divide benefits and harms into three categories: individual, societal and environmental, with examples of potential benefits and harms shown in the table below. These are intended to be illustrative, and the factors to be taken into consideration would need to be identified for a specific system.

Although broadening [safety] models, it is still desirable to take a human-centric approach

	Benefits	**Harms**
Individual	Personal autonomy Health	Physical injury Mental illness
Societal	Safe working environment Equitable access to resources	Social exclusion Inequitable risk distribution
Environmental	Biological diversity Clean water	Warming of the atmosphere Plastic build-up in the oceans

Many of these benefits and harms have an ethical dimension. For example, if introducing autonomous vehicles net reduces the level of fatal accidents on the roads but does not give any decrease in the number of fatalities for cyclists, then this would be a case of inequitable risk distribution. To be fair to the systems' engineers, this could be seen as an interpretation of what they meant by 'efficacy' and 'minimising unintended consequences'.

Many of these benefits and harms have an ethical dimension

The shift in terms of models supporting safety engineering should be clear – but, maybe now, beginning to seem like 'scope creep' on a global scale!

Toujours L'Attaque Surface

A large proportion of modern engineered systems utilise computer-based control, and many have high degrees of interconnectivity. Communication between systems and between systems and the infrastructure, can help in terms of efficacy and efficiency, but it opens up possibilities of cyber-attacks. Napoleon was famous for saying 'toujours l'attaque' (always attack).

I am advised by people who run industrial and public infrastructure that this idea is alive and well in the "hacker community", and that their systems are under continuous attack. Highly connected systems present a large attack surface, and poor security controls in one part of the overall system may enable access to more critical parts.



As an illustration, the figure shows a teenage boy and the TV remote control that he modified and then used to move points under a tram in Lodz, Poland, causing a derailment.

A further example is the 2001 attack on the Maroochydore sewage plant in Queensland, Australia which released about 1 million litres of sewage [9]. These incidents show the need to consider the interaction of security and safety and there are now many proposals for combined approaches to security and safety analysis, including some focused on early-stage design [10].

Perhaps controversially, I am of the view that the *risks* from cybersecurity are relatively low. The likelihood of attacks on many systems is high, but the probability of success is quite low and the more complex the system, the less likely the attacks are to succeed; in the Maroochydore case, the attacker was a disgruntled former employee, i.e. had inside knowledge. Thus, the models need to include cyber-security, to address means to reduce the potential attack surface and to provide effective security controls. This will reflect legitimate societal concern, but the focus should be on security in the role of its contribution to safety [10].

Safe, Ethical and Sustainable

If we were to draw together the models I have hinted at above, and also provide the supporting detail, e.g. in terms of analysis methods, then it would be clear that they are very complex; indeed, there are over 100 elements in the Safety of Complex Systems framework alone and the other models are multi-dimensional too.

It is not easy to produce a good summary – or mantra – but I propose "safe, ethical and sustainable". Safe – the primary focus continues to be on individual health and safety. Ethical – as new systems exhibit increasing levels of autonomy, moving decision-making from humans to machines, there are many issues including the potential for unfair distribution of risk or, unjustifiably, holding someone liable for outcomes which are beyond their control. Sustainable – due to the importance of sustainability in itself, and the human and societal effects of environmental damage; for example, global warming is already a major source of individual harm [8]. This aligns with the focus on individual, societal and environmental benefits and harms.

It is not easy to produce a good summary – or mantra – but I propose "safe, ethical and sustainable"

Perhaps it is better to view this as a question – repeatedly asking if a system being designed or used is 'safe, ethical and sustainable' won't immediately suggest all the details of the models I have alluded to, but it is a prompt and a route into those models.

Conclusions

I believe that safety engineering is at a crossroads. It needs to adapt to the complexities of current and emerging systems and to societal and environmental issues such as the impact of global warming. Some might argue that this is too big a change in role for the community, and I would view it as a target to strive for, not as an immediate objective. However, there is one immediate objective which I believe the community needs to adopt.

Bolt's "A Man for All Seasons" [11], focuses on the struggle between Henry VIII and Sir Thomas More, the chancellor, over issues of religion, power, and conscience. Whilst religion is outside our concerns here, being the 'conscience of power' and drawing to the attention of politicians not only the harms that complex systems can bring, but also their benefits, is something to which the community can, and I believe should, contribute.

I am pleased to see the Royal Academy of Engineering taking a lead in respect of the role engineering can play in reducing (the impact of) global warming. I see a similar model for safety engineers. If we consistently and persistently use the phrase 'safe, ethical and sustainable – treating it as a mantra (for all seasons) – this might begin to resonate with those in power and thus enable the safety engineering community to shape the future in a positive way, something we have perhaps not been good enough at in the past.

References

[1] G. Box, "Robustness in the Strategy of Scientific Model Building", 1978.
[2] J. Locke, "An Essay Concerning Human Understanding" (1689). Various reprints.
[3] L. Carroll, "Through the Looking Glass" (1871). Various reprints.
[4] M. D. Griffin, "How Do We Fix System Engineering?", 61st International Astronautical Congress, Prague, Czech Republic, 27 September – 1 October 2010, pp. 1-9.
[5] Aristotle, "Metaphysics", (See: W.D Ross et al. Aristotle's Metaphysics. Oxford University Press, 1925, pp 8-10.)
[6] "Safer Complex Systems: An Initial Framework", https://www.raeng.org.uk/publications/reports/safer-complex-systems (2020), accessed November 2021.
[7] National Transportation Safety Board. "Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian Tempe, Arizona March 18, 2018". (Published 2019).
[8] J.A. McDermid, Z Porter, Y Jia, "Consumerism, Contradictions, Counterfactuals: Shaping the Evolution of Safety Engineering", Safer Systems: The Next 30 Years, Proceedings of the 30th Safety-Critical Systems Symposium (SSS'22), 8th -10th February 2022, SCSC-170 (Published 2022).
[9] Maroochy sewage spill: https://www.risidata.com/Database/Detail/maroochy-shire-sewage-spill, accessed November 2021.
[10] J.A. McDermid, F. Asplund, R.Oates, J. Roberts, "Rapid Integration of CPS Security and Safety", IEEE Embedded Systems Letters (2018).
[11] R. Bolt, "A Man for All Seasons", Vintage (1990).

Image attribution
Lead image: ID 163537157 © Savagerus | Dreamstime.com. Lodz derailment: credit policja.pl

John McDermid Professor of Software Engineering and Director of the Assuring Autonomy International Programme (AAIP) at the University of York

John McDermid has worked on safety of computer-controlled systems for about 40 years and now leads the AAIP, focusing on the safety of robotics and autonomous systems, including those using machine learning. He has acted as an advisor to industry and government internationally and contributed to the development of standards. He is currently engaged on work on the safety and ethics of autonomous vehicles. He has supervised about 40 PhD students and published around 450 papers. He is a Fellow of the Royal Academy of Engineering and was awarded an OBE in 2010.