

What the Death Star Can Tell Us About System Safety

Guy Walker¹(✉), Paul Salmon², and Neville A. Stanton³

¹ Heriot-Watt University, Edinburgh, UK
G.H.Walker@hw.ac.uk

² University of the Sunshine Coast Accident Research Centre,
Queensland, UK
psalmon@usc.edu.au

³ University of Southampton,
Southampton, UK
N.Stanton@soton.ac.uk

Abstract. Resilience engineering requires that organizations review their own systems to proactively identify weaknesses. Imagine, then, having to identify a critical flaw in a highly complex planetoid sized orbital battle station, under extreme time pressure, and with no clear idea at the outset where the vulnerability will lie? This was the challenge faced by the Rebel Alliance in the film *Star Wars*. One of the belligerents, the Imperial Empire, considered it highly unlikely a weakness would be found even if the other belligerent were in possession of a full technical readout of the Station. How could it be done? The first option presented in this paper is to employ traditional error identification methods. The findings show the limitations of this component-based approach because it did not predict the actual vulnerability exploited. The second option is to use a systems-based method to model the Death Star's functional constraints and affordances. This method did detect the film ending, and several others. It also provides a compelling narrative around the use of reductionist methods for systems problems, and some wider implications for method selection in more earth-bound settings.

Keywords: Resilience · Scale · Variety · Predictive efficiency

1 Introduction

This paper is about resilience, systems-thinking, and fundamental questions about how to match methods to problems. It is also about the Death Star, a fictional battle station from the 1977 film *Star Wars*, which premiered (twice) at the nearby Chinese Theatre in Hollywood. The late seventies also had a wider cultural resonance. It is widely held to be the point in history when the modernist age, and its dominant scientific paradigm of determinism and reductionism, began giving way to a post-modernist, probabilistic, systems-based paradigm [1]. The story depicted in *Star Wars* contains some interesting parallels with the challenges faced by safety scientists today. It remains an open question whether or not our ability to proactively identify all potential weaknesses is fully formed, or whether our methods always match the types of problems which give

rise to brittle, accident prone systems. “Don’t be too proud of this technological terror you’ve constructed” runs one of the famous lines of dialogue, and a similar warning could be levelled at many of today’s systems. In an act of hubris around which the film’s storyline hinges, the Imperial Empire dismissed the idea that a weakness could be found and exploited, even if the Rebel Alliance, the other belligerent in the conflict, were in possession of a full technical readout of the Station. Let alone one being found in the four day time-frame between the plans being received and the Death Star arriving in orbit over the planet Yavin ready to destroy it. How could this feat of system’s analysis be achieved? Was it merely a cinematic sleight of hand or could it actually be done? More specifically, could it be done with the methods commonly used in the late-1970’s and would the methods of today fare any better? These are important questions for safety scientists because if the answer is no, then we need to think more carefully about our methods and when and where to apply them.

1.1 A Long Time Ago, in a Galaxy Far, Far Away...

Star Wars Episode IV, the first in an original trilogy of films, was released into the US domestic market in May 1977, had a second launch in August of the same year, and a general release in UK cinemas from March 1978. Among other films, such as *Jaws*, *Close Encounters* and *Superman*, it heralded the dawn of the Blockbuster genre [2]. It helped end the interregnum left by the widescreen Roadshow era of films common in the 1960’s, and stem (albeit temporarily) the precipitous decline in cinema audiences through the 1980’s [2, 3]. The center piece of the original Star Wars film was a planetoid sized orbital battle station referred to as the ‘Death Star’, around which much of the storyline centers (Fig. 1).

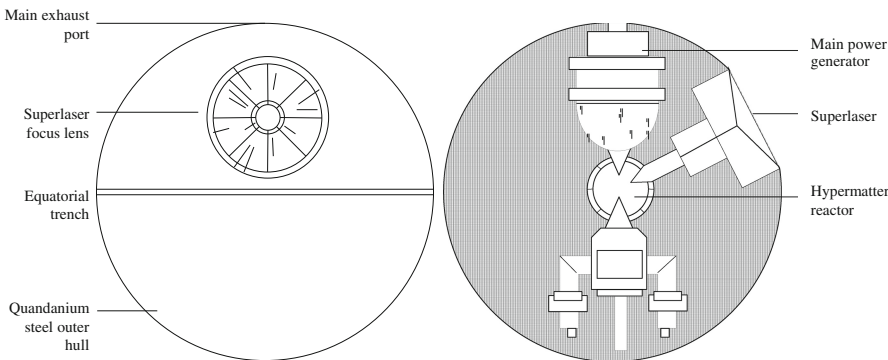


Fig. 1. High level schematic of the DS1 Orbital Battle station

The Mark I version of the battle station was a 100 km diameter spheroid constructed from Quandanium steel, a metal alloy mined from asteroids. The internal superstructure was devoted to a large Sienar Fleet Systems SFS-CR27200 Hypermatter Reactor, and its ancillary systems, which supplied all the power needs for propulsion,

life-support, defensive and offensive weapons systems. Principle among the latter was the superlaser, possessing enough power to destroy entire planets. The large conical focusing nexus, positioned on the northern hemisphere, is a dominant feature of the battle station's visual identity. With the core devoted largely to energy and offensive purposes, accommodation and living space resided on the surface protected by turbo-laser towers and a magnetic shield system.

Docking bays, hangars, ion drive arrays and thermal exhaust ports were located in a trench which ran continuously around the station's equator [4].

2 Describing the Death Star

2.1 Formal Rationality

To be emblematic of an ultimate technological terror weapon circa 1977 the principles of formal rationality needed to find particularly vivid expression. Formal rationality is a prominent part of an 'implicit theory' that has guided modern organizational design since the industrial revolution, and is one of many universal themes picked up by the film [5, 6]. A formally rational organization [7] is labelled a bureaucracy and the Death Star is its ultimate expression. Rationalizing organizations like this exhibit a tendency towards hierarchies and the maximization of the following attributes:

- Efficiency: The Death Star is "...the most efficient structure for handling large numbers of tasks...no other structure could handle the massive quantity of work as efficiently" [8],
- Predictability: "Outsiders who receive the services the [Death Star] dispense know with a high degree of confidence what they will receive and when they will receive it" [8],
- Quantification: "The performance of the incumbents of positions within the [Death Star] is reduced to a series of quantifiable tasks...handling less than that number is unsatisfactory; handling more is viewed as excellence" [8]
- Control: the Death Star "may be seen as one huge nonhuman technology. Its nearly automatic functioning may be seen as an effort to replace human judgment with the dictates of rules, regulations and structures" [8].

Organizations like the Death Star, designed along bureaucratic lines, can be seen as a way of imposing control theoretic behavior on a large scale, and in so doing, trying to make inputs, processes, outputs, even humans (and in this case other life-forms too), behave efficiently, predictably, quantifiably and with maximum control.

2.2 Scale

Another defining feature of the Death Star is its size. Why was it so big? A Work Domain Analysis (WDA [9]) reveals the physical need to accommodate an enormous reactor core and super laser in order to serve the Functional Purposes of (a) presenting the galaxy with a powerful symbol, (b) subjugating worlds, (c) enabling the galaxy to be ruled unchallenged and d), enacting the doctrine of 'ruling through fear' [4]. An

analytical side effect of the Death Star's enormous size are the changes that happen to its behaviors when viewed at different scales [10]. This behavior changes in specific ways depending on how an organization is designed. The relationship between the different behaviors observable at different scales is termed a complexity profile and defined as "the amount of information necessary to describe a system as a function of the level of detail provided" [10]. In the case of the Death Star its primary behaviors are visible at very large scales indeed. The reason for this is that entities and actors within the Death Star behave in highly coordinated ways. This arises as a direct result of the hierarchical, bureaucratic, formally rational way in which it is organized. As we zoom in on the Death Star, decreasing our scale of observation to that of small groups of actors, that high level of coordination gives rise to a distinctive property. Although the effects of bureaucratic organizations can be viewed from a large scale of observation the Death Star's fine scale behaviors are not always especially complex. Consider for a moment the serried ranks of storm troopers marching in time, or the squadrons of TIE Fighters in strict formation. The Death Star, therefore, is an example of a complex organization, in terms of its control structures, rules, myriad procedures, patterns of vertical communication and, of course, its technology, which nonetheless only really permits agents to undertake comparatively simple tasks [11]. This reflects "a fundamental principle of complex, and not so complex, systems: when parts of a system are acting together, the fine scale complexity is small" [10].

2.3 Variety

The second point about bureaucracies is that the large scale behaviors they are capable of emitting cannot be more complex than the actors at the top of the structure. The Emperor is the supreme overlord of the Death Star, and while undoubtedly a complex individual even his complexity is ultimately limited [10]. Bureaucracies like the Death Star, therefore, are large scale but also low variety. Variety is a cybernetic concept referring to the total number of states a system can adopt [12]. In discussions of scale it is often variety that is being used as shorthand for complexity. Less complex is equated with only limited numbers of behaviors (low variety), while more complex is equated with large numbers of behaviors (high variety). Ashby's Law of Requisite Variety tells us that:

"in active regulation only variety can destroy variety. It leads to the somewhat counterintuitive observation that the regulator must have a sufficiently large variety of actions in order to ensure a sufficiently small variety of outcomes in the essential variables [...]. This principle has important implications for practical situations: since the variety of perturbations a system can potentially be confronted with is unlimited, we should always try to maximize its internal variety (or diversity), so as to be optimally prepared for any foreseeable or unforeseeable contingency." [13].

In other words, a resilient system must have a sufficient repertoire of responses, and the agility to use them, such that it can track the dynamics of its environment. If not, it becomes vulnerable. In Star Wars Episode IV variety literally did destroy variety. The Death Star had insufficient internal variety or diversity to be able to counter the threat posed by the Rebel Alliance. Its hierarchical organization was exceptional at amplifying

the scale of the individual agent at the top of the hierarchy (i.e. the Emperor) but it was ultimately “not able to provide a system with larger complexity than that of its parts” [10]. In stark contrast, the Rebel Alliance’s attack was more complex than the person(s) at the top of its organizational structure. The focus here was on self-synchronizing teams, effects based operations, compatible awareness (via The Force), all of which created the conditions for greater freedom of action, that is to say more variety.

Naturally, a degree of caricaturing and stereotyping has been necessary to draw out these fundamental differences between the two belligerents, so in order to temper any over-enthusiasm for one approach or another it is important, firstly, to state that scale versus complexity is a trade-off. There are situations which require sheer scale, and others that require high organizational variety. Secondly, organizations often need both at the same time. Aspects of complexity (and the response to it) vary as a function of time and of behavior; they are not orthogonal. As such, it relies on organizations trying to match their ‘approach’ to the extant ‘problem’. There are occasions where coordination and scale are needed, but likewise, occasions when coordination and scale are redundant. Similarly, there are occasions when a non-deterministic, systemic approach will lack coordination and scale, and other occasions where it matches the target problem’s variety perfectly, or even, as in the case of the Death Star, exceeds it.

The methodological challenge is to become better at choosing which approach to match to which problem. Redundancy in the scale of an analysis, or not enough of it, is inefficient and can potentially lead to inaccurate analysis outcomes. The core challenge for analyzing complex issues like system resilience is that, quite often, methods are applied with little deep appreciation of the tacit assumptions being made in doing so. The remainder of this paper will use the Death Star as an exemplar of the perils and pitfalls of failing to match analysis methods to important properties of the target problem being analyzed.

3 Analyzing the Death Star

3.1 Predictive Efficiency

Did the Battle of Yavin, in which the strongly asymmetric force of Rebel Alliance fighters managed to destroy the Death Star, exhibit the property of emergence? Emergence describes behavior that is not deducible from its low level properties, behavior that does not adhere (at any reasonable or tractable scale of analysis) to the logic of causal determinism. It can be defined as follows:

“Emergence is the phenomenon wherein complex, interesting high-level function is produced as a result of combining simple low-level mechanisms in simple ways” [14].

The combination of simple low-level mechanisms in simple ways to give rise to complex, interesting, high-level function describes the Rebel Alliance’s attack on the Death Star well. To quote the classic sociotechnical literature, “though [the Rebel Alliance’s] equipment was simple, their tasks were multiple”, the agent in this organization “...had craft pride and artisan independence” [15]. The Rebel Alliance’s attack on the Death Star is an example of a simple organization (and equipment) ‘doing’

complex ‘things’. We see, therefore, strong evidence of emergent behaviors arising from simple low level tasks and equipment which means, methodologically speaking, that we need to predict the emergent behavior rather than its detailed component level antecedents. In other words, a focus on the simple low-level mechanisms means that the interactions between them, which give rise to the systemic behavior that is really the required unit of analysis, is lost. The question is how to tell whether a component or systems focus is needed? Whether the problem being confronted is best described with deterministic component methods, or systems methods, or a combination of both. A useful concept is Relative Predictive Efficiency (RPE [16]) expressed as follows:

$$\text{RPE} = E/C$$

E is ‘excess entropy’ or the extent to which a system can be adequately modelled. Here there is a comparison to be made between the system behaviors predicted by a model compared to those behaviors actually observed. Any disparity between ‘expected’ and ‘observed’, and in what quantity, represents some equivalent of ‘excess entropy’, or ‘E’ in the formula. For example, the simple organization (like the Rebel Alliance) which is able to do complex (unexpected) things not predicted by an analysis based purely on component parts, would measure highly on the parameter E. C, on the other hand, is ‘statistical complexity’. It is a measure of the size and/or complexity of the system’s model at any given scale of observation. This can be measured in a number of ways [17]. For example, it is possible to consider the number of ‘build symbols’ in the model (i.e. the number of goals/sub-goals in a Hierarchical Task Analysis (HTA) which produce the overall goal), the sophistication of the model (i.e. the number of logical operators used in HTA plans), or the connectivity (the maximum number of links that can be removed before a task analysis splits into two), and so on. RPE is a simple concept but it can help the analyst decide what approach to take: strict reductionism (and a focus on component antecedents of system behavior) or systemic (and a focus on the system’s emergent behavior itself). Emergence exists, and therefore systemic methods become more appropriate, “if the higher level description of the system has a higher predictive efficiency than the lower level” [16]. RPE, therefore, provides insight into the type of analysis suited to a particular complex system, and in this respect could lead to considerable savings in analysis time and effort for a corresponding increase in predictive efficiency. This can be demonstrated by replicating the analysis of the Death Star’s technical plans alluded to in the film using two different approaches.

3.2 Method

To illustrate the practical methodological issues around scale, variety and predictive efficiency two very distinct approaches to analyzing the Death Star’s resilience were trialed. The detailed technical findings are to be presented in a future paper but a high-level summary can be provided here in order to demonstrate the key concepts being discussed. The first method is a reflection of the deterministic world-view then dominant at the time of the original film’s release in 1977. For the purposes of demonstration a

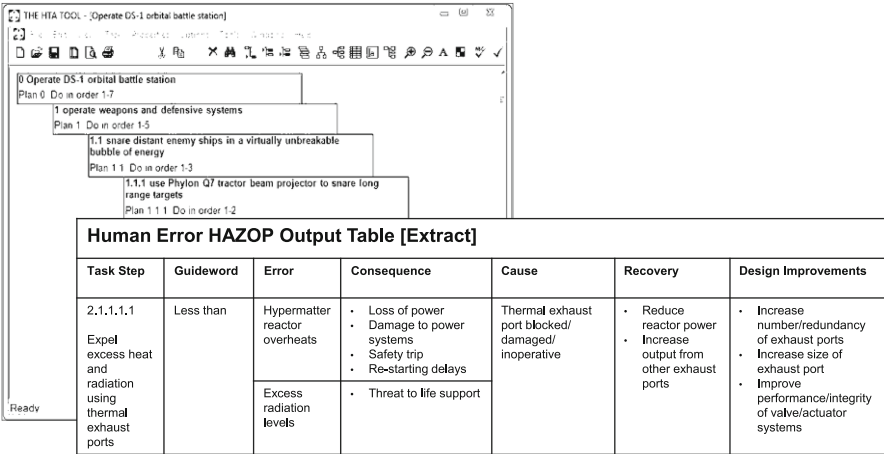


Fig. 2. HTA of the DS-1 Orbital Battle Station

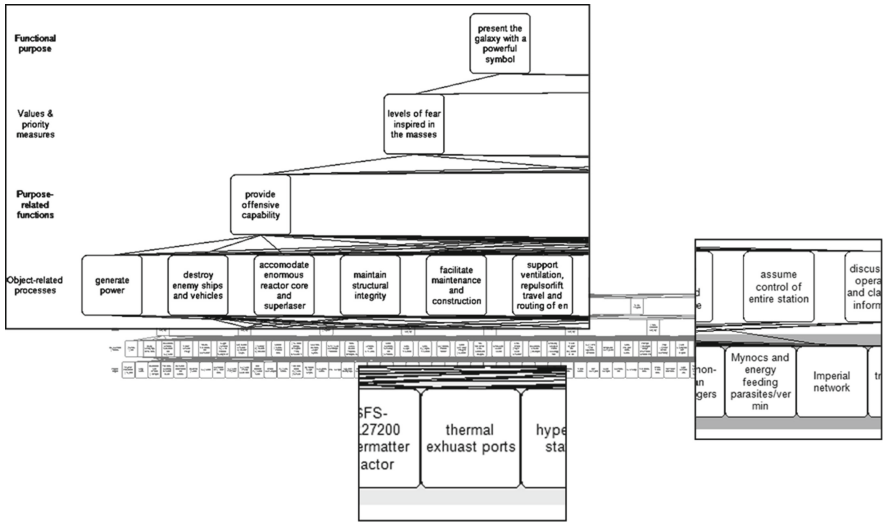


Fig. 3. Work Domain Analysis (WDA) of the DS-1 Orbital Battle Station

HTA of the Death Star was created (see Fig. 2) and used to drive a Human Error (HE) HAZOP analysis [19].

The second method is a reflection of the systems world-view increasingly dominant at the present time. For the purposes of this demonstration a WDA [9] of the Death Star was created and systematically interrogated for key constraints and affordances (see Fig. 3). Both analyses were driven from technical data contained in [4] in preference to relying on Bothan spies (ethical approval was not forthcoming given the deaths described in the film).

3.3 Outcomes

Having obtained a full technical readout of the Station it was then possible to replicate the analysis of it performed in the film using both HE-HAZOP and WDA. Did the methods detect the actual film ending? In the case of the HE HAZOP, no. The component-level analysis revealed a very large quantity of potential system failings and human-error potential (see forthcoming companion paper for more detail). The first fundamental problem was that none of these component risks related strongly to the key risks actually exploited in the film. For example, the infamous Thermal Exhaust Port, down which proton torpedoes were fired leading to the Battle Station's complete destruction, activated very few HE-HAZOP guidewords, certainly far fewer than other system components/agents. The HE-HAZOP was able to predict that the Thermal Exhaust Port might fail or expel too much or too little heat, however, it could not predict the proton torpedo strike or the knock-on effects of this. The countermeasures identified focused on making the exhaust port more reliable, or on monitoring its performance, or, perversely given what we know about the film, make the exhaust port bigger (and therefore easier to hit). The second fundamental problem was the time needed to complete the analysis which, in the current demonstration, exceeded the time available between the Death Star plans being received and the Death Star itself arriving over the planet Yavin ready to destroy the rebel base. In other words, to complete a detailed HTA and HE-HAZOP of the size and complexity needed for the Death Star took longer than four days, meaning the Rebel Alliance would have been destroyed whilst still undertaking the analysis. The third fundamental problem – the most fundamental of all – is that because of all the above Excess Entropy was high and Relative Predictive Efficiency low.

The Work Domain Analysis, on the other hand, did detect the actual film ending. Or at least, there were explicit and uncontroversial affordances between the Thermal Exhaust Port and Object-Related Processes around 'accommodating the enormous reactor core and super-laser', 'generate power' and 'expel excess heat and radiation'. These in turn were linked to key Death Star capabilities such as 'provide offensive and defensive capability' and 'energy and propulsion'. These in turn directly affected all of the higher-level Functional Purposes. Depicting these interdependencies between the Thermal Exhaust Port and other functions, affordances, and purposes would certainly lead an astute rebel planner to the conclusion that degrading these links would have a significant impact on the Death Star's functioning. The method was also relatively quick. Not only was there time to consider the critical vulnerability already known about from the film, but numerous others emerged. There were links which connected 'energy feeding parasites' to the Station's ability to provide life support, meaning that one way of exploiting this vulnerability would be to proactively *create* a space vermin infestation. Principal among the linkages that were explored within the Work Domain Analysis was the relationship between the Imperial [computer] Network and all the other critical functions of the Station. Here was an obvious vulnerability which could be exploited via some form of disabling computer virus: indeed, the astro-droid character R2D2 plugged in to the Imperial Network twice in the film and could conceivably do so to upload a virus. That being said, the first computer virus released into the wild happened in 1983, some years after the film's release, which itself communicates an interesting facet about

the changing shape of risk and what analysts may or may not consider to be viable threats. It is certainly telling, from a safety science perspective, that in the more recent film prequel 'Clone Wars' R2D2 does precisely this. The critical point here is that the constraints and affordances are present and observable in the system, and if analysts are not able to explore them for all conceivable eventualities there exist methods, such as the Strategies Analysis Diagram [20], which will perform this function in an exhaustive and systematic manner. Potentially, every pathway to failure, and indeed success, can be elicited.

4 Conclusions

Was it a cinematic sleight of hand or could a full technical read-out of the Death Star be analyzed in time to discover a critical vulnerability and launch a pre-emptive strike? Using HE-HAZOP, a method popular in industry at the time of the film's release in the late 1970's, the answer is no. This approach took longer than the time available and it did not identify the critical vulnerability exploited in the film. Using Work Domain Analysis, a method in more common currency today (albeit one originally developed in the late 1960's) the answer is yes. Not only did it detect the key relationships linking the thermal exhaust port to the Battle Station's destruction, it did so in a short enough time to act on the results. More than that, it revealed further pathways which could lead to the Death Star's failure, and other potential film endings. What this comparison of methods illustrates is the role of scale, variety and predictive efficiency in making contingent decisions about what methods to apply to what problems. This is an increasingly important question because (a) the paradigm has shifted towards greater use of systems concepts, (b) many research grand challenges occur at the non-linear intersection of people and technology, and (c), every time we use a method we make tacit assumptions about the nature of the problem we are trying to solve. This paper has travelled to a galaxy far, far away to demonstrate that sometimes those assumptions can be at odds with what we are trying to achieve, with potentially disastrous consequences. Considerations of variety, scale and predictive efficiency are tractable means to think afresh about sociotechnical problems and direct our analysis efforts in more cost-effective and expedient ways. May The Force (of this contingent approach to method selection) be with you.

References

1. Berman, M.: *All That Is Solid Melts Into Air: The Experience of Modernity*. Penguin, London (1982)
2. Stringer, J. (ed.): *Movie Blockbusters*. Routledge, London (2003)
3. Haines, R.W.: *The Moviegoing Experience, 1968-2001*. McFarland, Jefferson, NC (2003)
4. Windham, R., Reiff, C., Trevas, C.: *Imperial Death Star DS-1 Orbital Battle Station: Owner's Workshop Manual*. Haynes, Sparkford, UK (2013)
5. Rinzler, J.W.: *The Making of Star Wars*. Ebury, New York (2007)
6. Campbell, J.: *The Hero with a Thousand Faces*, 3rd edn. New World, New York (2008)

7. Weber, M.: The protestant ethic and the spirit of capitalism (1930). e-book available at: http://www.ne.jp/asahi/moriyuki/abukuma/weber/world/ethic/pro_eth_frame.html
8. Ritzer, G.: The McDonaldisation of society. Pine Forge Press, London (1993)
9. Naikar, N.: Work domain analysis: concepts, guidelines, and cases. CRC Press, Boca Raton (2013)
10. Bar-Yam, Y.: Making things work: solving complex problems in a complex world. NESCI: Knowledge Press, Cambridge (2004)
11. Sitter, L.U., Hertog, J.F., Dankbaar, B.: From complex organizations with simple jobs to simple organizations with complex jobs. *Hum. Relat.* **50**(5), 497–536 (1997)
12. Ashby, W.R.: Introduction to Cybernetics. Chapman & Hall, London (1956)
13. Heylighen, F., Joslyn, C.: Cybernetics and Second Order Cybernetics. In: Meyers, R.A. (ed.) *Encyclopedia of Physical Science and Technology* (3rd edn.), vol. 4, pp. 155–170. Academic Press, New York (2001)
14. Chalmers, D. J.: Thoughts on emergence (1990). Available at: <http://consc.net/notes/emergence.html>
15. Trist, E., Bamforth, K.: Some social and psychological consequences of the longwall method of coal getting. *Hum. Relat.* **4**, 3–38 (1951)
16. Crutchfield, J.P.: The calculi of emergence: computation, dynamics and induction. *Phys D* **75**, 11–54 (1994)
17. Hornby, G.S.: Modularity, reuse, and hierarchy: measuring complexity by measuring structure and organization. *Complexity* **13**(2), 50–61 (2007)
18. Halley, J.D., Winkler, D.A.: Classification of emergence and its relation to self-organization. *Complexity* **13**(5), 10–15 (2008)
19. Kletz, T.: *An Engineer's View of Human Error*, 2nd edn. Institution of Chemical Engineers, Rugby, Warwickshire (1991)
20. Cornelissen, M., McClure, R., Salmon, P.M., Stanton, N.A.: Validating the strategies analysis diagram: assessing the reliability and validity of a formative method. *Applied Ergonomics* **45**(6), 1484–1494 (2014)