Turning apples into bananas

How big data undermines safety and what can be done about it

Nippin Anand
PhD

An able seaman has been 'reprimanded' for leaving the gangway hanging out while the vessel was shifting berth.

- Root cause: 'Lack of awareness';
- Corrective action: 'Risk assessment'.

A third officer has been served a warning letter for missing out on monthly checks on a fire extinguisher, resulting in non-conformance during a safety audit.

- Root cause: 'Complacency';
- Corrective action: 'Follow the procedures'.

A crew member trips over an obstruction on deck and hurts himself.

- Root cause: 'Lack of awareness';
- Corrective action: 'Risk assessment'.

A chief officer who submitted a near miss report stating that he came close to a fishing vessel during the coastal passage has been sent on a refresher training course.

- Root cause: 'Lack of planning';
- Corrective action 'Training and supervisory control'.

Go through any reporting system and you will find hundreds of such reports. With intense budget controls and an over-zealous commitment to safety, many organisations are turning towards software solutions, colloquially speaking 'Big Data', to measure and manage the state of safety.

Looking deeper

Listening to the other side of these reports is disturbing but gives valuable insights. The able seaman points out that moving the vessel alongside the berth with one crew member forward and aft during the night was not unusual. The captain had consciously made a decision not to wake up other crew members in the middle of the night. What was unusual, however, was that a crane swung out and obscured the captain's sight of the gangway from the bridge. When things went wrong, the able seaman was held to blame for not informing the captain that the gangway was hanging out.

The third officer has his own version of the story. As a safety officer he must ensure that each portable fire extinguisher is visually inspected and ticked off every month. With the best will in the world, there are instances when an odd fire extinguisher can be missed out when you have a list of 300 to check. 'But nobody is interested that I check the other 299. One mistake and I got a warning from the captain,' he says. It does not end there. He adds, 'Now every time we have inspection, I am worried, I cannot sleep for many days. Maybe I forgot something, I am going around at night after my watch to make sure I have not missed anything. It's not easy to find another job if I lose my job.' This is a watch officer who is on the bridge for at least eight hours a day on a vessel laden with hydrocarbons.

The chief officer became furious when I probed into the near miss. 'They make me do a three days course, I find it insulting. It is them who should do better planning. I tell them don't load six high containers on the forward hatches especially on the sides, it obstructs visibility from the bridge. But they don't care. And when I reported this, they are telling me I was not careful on my watch.'

The question is, can we ever get insights of this kind from purely data-based reporting systems or are we simply camouflaging them with technology?

Measuring safety

Attempting to measure safety is a key purpose of any reporting tool. Safety is defined as the condition of being protected from harm or injury to an acceptable level. But this definition is not without its problems. To the consumers, stakeholders and society at large (in a legal and media sensitive environment) the notion of an 'acceptable level' of safety can easily become arational (that is, outside the realm of reason) and unrealistic. Take an example of what is regarded as 'acceptable' – 'If it saves even one life, it is worth the effort'. More eloquent examples include 'all accidents are preventable' (AAAP), or the prospect of an accident free future.

On the face of it, there is nothing wrong with such virtuous statements, – but the thinking behind them is based on emotion. Our view of what is acceptable is not based on a calculated decision based on professional judgment. Rather, it is the unfortunate outcome of a society where expert decisions and professional judgment can easily come under attack in the press. For anyone tasked with measuring safety a natural response would be to say that everything is a risk, and nothing can be acceptable. Curiosity and questioning are out; orthodoxy and fear are instilled in people.

Businesses operating in fierce market competition often struggle with such arational responses. At times the costs can become unsustainable, and on other occasions the responses themselves make a mockery of safety. This is how we end up investing heavily in barriers and protective devices and implementing behavioural safety tools despite no evidence of injury or harm – and sometimes even in the face of evidence that the 'safety' measures are counterproductive.

A second and, in my view, bigger problem with measuring safety is that it relentlessly aims for perfection. There is a very specific language that underpins this thinking and quite often it has little or nothing to do with safety. For instance, compliance, conformance, all accidents are preventable, zero tolerance, and the zero-accident vision. A handful of unrealistic goals are established, and the focus turns towards petty mistakes rather than overall progress. It does not matter if you have checked hundreds of fire extinguishers, what matters is one that is left out. Any improvement that does not fit with those narrow set of goals means nothing. As we build sophisticated reporting tools to measure safety, we need to be aware that the end purpose can become trapped in arational thinking and unrealistic goals.

Big Data or Big Brother?

What about the actual reporting and analysis of data? As copious data is poured into software models, we are tempted to believe that this will help us predict what lies ahead, alleviate uncertainty and improve business performance. Technology will comb through petabytes of data and provide insights into individual behaviour. Software systems are already becoming capable of identifying patterns of human behaviour and correlating them with individuals and groups of workers. HR software tools can gather extensive data to assist organisations with hiring decisions and predicting employees' performance at work.

But imagine how spurious correlations can easily go wrong when it comes to measuring safety. Once a few 'unsafe' individuals are identified this can lead to detailed monitoring of those individuals, subsequently finding even more problems in their behaviour. The accident-prone workers will remain logged in the system for a long time. In the wake of an accident (or an unsatisfactory inspection outcome) the first thing would be to recall the history of the workers and take them to task. This may exacerbate the problem of accountability and blame.

Algorithmic injustice

At the organisational level, reporting accidents, defects or any shortcomings has long been considered a sign of failure in many safety critical industries. So why make so much fuss now that we are assigning the analysis to computers?

Data scientist Cathy O'Neil, author of the book *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, tackles this issue through a series of case studies. O'Neil argues that while humans learn to adapt quickly when things don't seem to work, computers can get fixated on erroneous correlations that become difficult to break down. When this happens the scale of damage becomes inconceivable.

One such example was recounted by a ship manager writing on LinkedIn. One of their vessels developed a small oil leak in the main engine while transiting a busy shipping lane. The issue was reported, the vessel anchored, changed the part in question, and was back underway within an hour, problem solved. Later, the company saw that the vessel had been downgraded by a fully automated rating system from five stars to three. It took months to return to five star status, during which period several chartering opportunities were lost. There was no recourse, and the entry could not be changed, as the system was fully automated.

This is a telling example of algorithmic injustice in the digital era.

Turning apples into bananas

Data manipulation can affect the outcome in other ways. One operator states, 'Our client takes the risks of dropped objects very seriously, so we scan through our incident reports to check for terms such as 'dropped objects' and 'deck' to ensure we do not have issues there.' What if over a period of time the computers are trained to carefully ignore those buzzwords and catch-phrases that affect their performance indicators and market situation? You are not making those incidents disappear – but you are making them appear as something they are not. How?

Imagine a database full of pre-determined phrases like 'complacency'; 'procedures not followed'; 'lack of planning'; 'lack of situational awareness,' etc. Now get rid of 'dropped objects' from the database when they become a problem, and assign any 'dropped object' incidents to one of these categories instead. The same search engines that were once designed to identify problems could become a weapon to conceal them. Over a period of time computers will learn to figure out our affinities and aversions – and the 'dropped objects' issues could disappear without us intending them to do so.

Back in 2017 there was a CNN commercial showing a photograph of an apple. The commercial stated: 'This is an apple. Some people might try to tell you that it's a banana. They might scream 'Banana, banana, banana' over and over and over again. They might put BANANA in all caps. You might even start to believe that this is a banana. But it's not. This is an apple.' Data manipulation is very good at turning apples into bananas, even unintentionally.

Data without context

There is a further problem with the phrases discussed above. Rich and vivid human stories are stripped of their context and simplified into standard phrases of 'human error'. The thinking behind all this is that for every accident or non-compliance there is a cause, and generally a bad one. Bad causes precede bad consequences and those bad causes can be traced back if we searched far enough (hence the term root cause analysis). This is what Erik Hollnagel refers to as the 'causality credo'.

But causation is flawed. Causes can be imaginary and fabricated to serve certain purposes, as in the case of the chief officer who was enrolled for a training course. Similarly, the quest for the cause will typically end with the last man or woman in the chain, as in the case of the able seaman who 'forgot to secure the gangway'.

And then there are instances when causes are mistaken for consequences and vice versa. Consider the crewmember who tripped on deck. Was it *because* he did not pay enough attention to the obstruction on deck? Or could it be that a bad design *caused* the crew to stumble? The former puts the blame on the worker, while the latter aims to find ways to design out the obstruction. Could algorithms ever expose the flaws of causation and get to the human stories behind the reports? That depends on the data we chose to collect (and ignore) and the questions that we ask of that data.

Rethinking safety in the digital era

Data analysts would tell us that software tools can work with massive volumes of data to 'automatically discover trends and patterns' or find 'non-obvious causal relationships' in the data. This should mean that the entire process, from data collection, to reporting and analysis, is exploratory and driven by curiosity and science. This cannot be true at least for safety where the purpose itself is ill-defined, convoluted (to serve multiple interests of which protecting people from harm and injury is just one purpose) and driven by a pre-determined set of narrow goals.

The examples cited here highlight that there is more to safety data than just observing patterns of individual behaviour or actions. As the educational psychologist Jerome Bruner pointed out, 'It is practically impossible to understand a thought, an act, a move of any sort from the situation in which it occurs.' We can think of safety as an individual's problem or we could think of it as the capacity of our people to succeed despite some very common patterns mirrored across organisations and work-sites – poor technological designs, poorly written instructions and procedures, conflicting demands, resource constraints.

If we can combine big data with deeper stories, we can rebuild trust in reporting, demystify the flaws of causation and build richer understanding and analysis. Technology offers us two choices (not necessarily binary but co-existing). We could continue with the same old concepts of safety in the hope that by using software tools we may get better results, or we could genuinely transform safety into a business and performance tool.

Disclaimer: The views expressed by the author may not represent the views of the organisation that the author represents.