Designing IT to make healthcare safer

Professor Harold Thimbleby
www.harold.thimbleby.net

“Commercial air travel didn’t get safer by exhorting pilots to please not crash. It got safer by designing planes and air travel systems that support pilots and others to succeed in a very, very complex environment. We can do that in healthcare, too.” — Don Berwick

“Computing’s central challenge, “How not to make a mess of it,” has not been met. On the contrary, most of our systems are much more complicated than can be considered healthy, and are too messy and chaotic to be used in comfort and confidence.” — Edsger Dijkstra

1 Introduction

If preventable error in hospitals was a disease, it would be a big killer—and recent evidence suggests that preventable error is the third biggest killer, after heart disease and cancer [9]. Worse, the death rates for error are likely to be an under-estimate; for example, if somebody is in hospital because of cancer, if an error occurs their death is unlikely to be recorded as “preventable error” when it is far easier to say the disease took its inevitable course. It is tempting, but wrong, to blame hospital staff for errors [1]. Healthcare has rising costs, increasing expectations, and we are getting older and more obese, increasingly suffering from chronic diseases like diabetes. However we might try to interpret the evidence, it is clear that healthcare is in crisis.

One would imagine that computers (IT) would be part of the way forward. IT, new technologies, and the “paperless NHS” are all frequently pushed as obvious ways forward. In fact, while computers make every industry more efficient, they don’t help healthcare [5]. Healthcare is now widely recognised as turning into an IT problem [8].

In contrast, we know drug side-effects are expected and unavoidable, so we would be skeptical if the latest “wonder drug” could be all that it was trying to promise. We rely on a rigorous process of trials and review before drugs are approved for general use [4], and we expect a balance between the benefits of taking a drug and suffering from its side-effects. The review process protects us. Yet there is no similar review or assessment process for healthcare IT systems.

So-called “user error” should be considered the unavoidable side-effect of IT; therefore, IT systems should be better regulated to avoid or manage side-effects. To do so, we need principles and strategies to improve IT safety, which we will discuss in this lecture.

If we can work out how to improve computers, there will be enormous benefits. Unlike calls to improve training or other human processes, if we can improve IT, then everybody benefits automatically.

2 Some example problems

2.1 Lisa Norris

At the Beatson Oncology Centre in Glasgow, software was upgraded and the implications were not reviewed. The original forms for performing calculations continued to be used, and as a result a patient,
Lisa Norris, was overdosed. Sadly, although immediately surviving the overdose she died. The report [10] was published just after her death, and says:

“Changing to the new Varis 7 introduced a specific feature that if selected by the treatment planner, changed the nature of the data in the Eclipse treatment Plan Report relative to that in similar reports prior to the May 2005 upgrade . . . the outcome was that the figure entered on the planning form for one of the critical treatment delivery parameters was significantly higher than the figure that should have been used. . . . the error was not identified in the checking process . . . the setting used for each of the first 19 treatments [of Lisa Norris] was therefore too high . . . ” [10, x6–10, p.ii]

“It should be noted that at no point in the investigation was it deemed necessary to discuss the incident with the suppliers of the equipment [Varis 7, Eclipse and RTChart] since there was no suggestion that these products contributed to the error.” [10, x2.7, p.2]

This appears to be saying that whatever a computer system does, it is not to be blamed for error provided it did not malfunction: the revised Varis 7 had a feature that contributed to an error, but the feature was selected by the operator. Indeed, the report dismisses examining the design of the Varis 7 (or why an active piece of medical equipment needs a software upgrade) and instead concentrates on the management, supervision and competence of the operator who made “the critical error” [10, x10.4, p.43]. It appears that nobody evaluated the design of the new Varis 7 [10, x6.21, p.24], nor the effect of the changes to its design, despite an internal memorandum some months earlier querying unclear control of purchased software [10, x6.22, p.24].

2.2 Olivia Saldana

In 2001 the radiographer Olivia Saldana was involved with the treatment of 21 patients who died from radiation overdoses. Treatment involves placing metal blocks to protect sensitive parts of the patient’s body, and then to calculate the correct radiation dose given that the blocks restrict the treatment aperture. Saldana drew the shapes of the blocks on the computer screen, but the computer system did not perform the correct calculation. The computer should have detected its inability to perform a correct calculation [3]; it was a bug of which Saldana was unaware. She was imprisoned for manslaughter, even though the manufacturer, Multidata Systems, was already aware of the bug in 1992.

2.3 Silently throwing away user keystrokes

Handheld calculators are used throughout healthcare, but rather than show you a drug dose calculation, let’s do something much simpler and easier to understand. I live in Wales, and I am interested in what proportion of the world’s population is welsh. I therefore use a calculator to find out 3,063,500 divided by 6,973,738,433, which is the population of Wales divided by the population of the world (being numbers I got off the web so they must be right). Remember that you use a calculator because you do not know what the right answer is. Here, I obtain the following results (ignoring least significant digits):

<table>
<thead>
<tr>
<th>Calculator</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Casio HS-8V</td>
<td>0.04…</td>
</tr>
<tr>
<td>Apple iPhone portrait</td>
<td>0.004…</td>
</tr>
<tr>
<td>Apple iPhone landscape</td>
<td>0.0004…</td>
</tr>
</tbody>
</table>

If you didn’t know what the right answer is, you still don’t! These are all market-leading products, yet none of these calculators reports an error — only the last is correct. Whatever is going on inside the Apple iPhone, it could clearly report an error since it provides two different answers even if it doesn’t know is right!

The first electronic calculators appeared in the 1960s. We are no longer constrained by technology, and we’ve had some fifty years to get their designs right; it is hard to understand why calculators used in healthcare are not dependable.

2.4 Kimberly Hiatt and Kaia Zautner

Kimberly Hiatt made an out-by-ten calculation error for a dose of CaCl (calcium chloride). The patient, Kaia Zautner, a baby, died, though it is not obvious (because the Kaia was already very ill) whether the overdose contributed to the death. Hiatt reported the error and was escorted from the hospital; she subsequently committed suicide — she was the “second victim” of the incident [24]. Notably, after her death the Nursing Commission terminated their investigation into the incident, so, we will never know exactly how the error occurred [22].

How did the miscalculation occur? It is possible that Hiatt made a simple keying slip on a calculator,
such as pressing a decimal point twice, resulting in an incorrect number and, if so, the calculator could have but would not have reported it, or perhaps the pharmacy computer had printed an over-complex drug label that was too easy to misread? It is possible some of the equipment (calculator, infusion pump, etc) simply had a bug and, despite being used correctly, gave the wrong results.

2.5 Zimed Syringe Driver
The Zimed AD Syringe Driver is, according to its web site [25], “a ground-breaking new infusion device that takes usability and patient safety to new levels. . . . Simple to operate for professionals, patients, parents, or helpers.”

In our team’s paper [2] we showed that this device permits use error that it ignores, potentially allowing very large numerical errors. One cause for concern is over-run errors: a nurse entering a number such as 0.1 will move the cursor right to enter the least significant digits of the intended number, but an over-run (e.g., an excess of move-right key presses) will silently move the cursor to the most significant digit. Thus an attempt to enter 0.1 could accidentally enter 1000.0 by just one excess keystroke with no warning to the user.

2.6 Denise Melanson
Denise Melanson had regular infusions of chemotherapy. She was given an overdose and her death was followed by a Root Cause Analysis that has been made public by the Institute of Safe Medication Practices [7]. Elsewhere we have criticised the difficult calculation aspects of the infusion [19,23] and shown how the relevant design problems might have been avoided, but here we highlight two issues raised by a short two hour experiment performed as part of the Root Cause Analysis:

- Three out of six nurses participating in the experiment entered incorrect data on the Abbot AIM Plus infusion pump; all were confused by some aspect of its design. Three did not choose the correct mL option (the infusion pump displays the mL per hour option as mL) — one nurse selected mg/mL instead. Some were confused with the infusion pump using a single key for both decimal point and an arrow key.
- The drug bag label was confusing, causing information overload — it had over 20 numbers printed on it. It also displayed 28.8, the fatal dose, incorrectly causing confirmation bias (the correct dose was 1.2 mL per hour, but both nurses omitted to divide by 24 hours in a day; they both obtained 28.8, and the bag’s presentation of this number helped mislead them).

One wonders why the manufacturers of the infusion pump and the programmers of the pharmacy label printing system did not perform similar experiments to help them redesign and improve their products. Indeed, evaluation is required by the appropriate international standards, such as ISO 62366 Medical devices—Application of usability engineering to medical devices, which require manufacturers to identify hazards, perform experiments with users in realistic scenarios, and to have an iterative design plan.

Drugs with such bad side effects that are so easily demonstrated (in just a single two hour experiment!) would never be approved.

3 From problems to solutions
When you get cash from a cash machine, you put your card in, key in your PIN code, grab your cash—and walk away leaving your card behind. So instead of that disaster, cash machines force you to take your card before giving you any cash. You went there to get cash, and you won’t go away until you have it—so you automatically pick up your card. This is a simple story of redesigning a system to avoid a common error. Notice that the design has eliminated a type of error without having to retrain anybody. Similar thinking used to redesign car petrol caps, so you don’t leave them behind in petrol stations.
Figure 1: Kahneman’s two cognitive systems [11] illustrate how conscious thinking (System 2) can only know about the world through what it perceives using System 1. System 1 is fast and effortless, like a reflex; whereas System 2 is conscious, “logical” and powerful, but requires conscious effort. Many errors occur when System 2 delegates decisions to System 1, e.g., in so called attribute substitution [11]. Poorly designed healthcare systems do not work with System 1 to warn System 2 of the need for thinking.

In the NHS, the UK’s largest employer, any solution to problems that relies on retraining is not going to be very successful. Instead, we should redesign the IT. If IT is designed properly, then automatically things work more safely — let’s call this the Thimbleby Principle (as we shall see this is a much more powerful idea than the Japanese “zero defect” manufacturing concept of poka-yoke). The question is, how do we do that?

The common thread in the problems listed above is that errors go unnoticed, and then, because they are unnoticed, they are unmanaged and then lead to harm. If a user does not notice an error, they cannot think about it and avoid its consequences (except by chance). The broader consequences of this truism are explored in detail by Kahneman [11], and illustrated in figure 1. Kahneman’s two systems are “in the head” and are used to help understand how we think.

Figure 2 presents a more memorable model. We have two “inner personalities,” who we will call Homer Simpson and Spock because they closely resemble how these bits of us behave. The point is that Spock is inside our heads and cannot know anything about the world but for Homer’s interpretation of it. Unfortunately Homer often misleads Spock, and often Spock delegates important decisions to Homer. Well-known human factors issues like “tunnel vision” can be explained with this model.

Most IT systems try to talk to Spock — after all, they are designed to be “logical.” Instead we argue that — particularly for handling errors — they should talk to Homer. For example, in the “throwing away keystrokes” problem (section 2.3 above) the problem was that Spock was supposed to know things were going wrong, but Homer didn’t notice (why would he?) so Spock carries on — and ends up making an error. If instead the IT system bleeped and interrupted Homer, then Spock would have to wake up and intervene, correcting the error. (Our research has shown this is extremely effective [2, 14, 16, 17, 20, 21, 23].)

Take the UK plans to move the NHS to being paperless. This sounds like an obvious advantage, as patients have piles of paper records that get lost and mixed up, and are often in the wrong place when they are needed. Instead, “if only they were fully computerised” (some say) there would be no paper and no problems. Wrong [18]. Computer screens are smaller than what is needed to show all of a patient’s records, so Homer cannot see all the records. If Spock is unwary (which of course happens frequently, for instance at the end of a long shift), Homer will not realise that what they can’t see needs scrolling into view, and Spock won’t even notice that they have not seen something. Homer didn’t see it, so it wasn’t there.

Obviously Spock is supposed to be highly-trained, and therefore will remember to scroll the screen so that he reviews all the patient’s pertinent information. Unfortunately, Spock will sometimes make mistakes and not realise it because Homer isn’t clever enough to help him. For example, if Spock is tired, he will leave difficult decisions to Homer (attribute substitution), and if Homer can’t see something “it isn’t there.”
Spock then has his Rumsfeld moment: there are unknown unknowns, and he doesn’t even know it — and the patient suffers.

As before, if the IT system works with Homer, we can avoid the problem. If there is important information that is out of sight, the computer screen has to have a flashing icon (which attracts Homer’s attention, possibly a beep as well) that announces that there is more information to read — Homer sees this and re-

Figure 2: A more memorable version of figure 1! Homer (System 1) is quick and effortless, if not completely reliable; Spock (System 2) is conscious, “logical” and powerful, but requires effort. In a very real way, Spock is “inside” Homer, and knows nothing about the world except what Homer tells him. In particular, however clever Spock might be, if Homer misleads him, he cannot know the truth about the world and won’t be able to think correctly about what to do.

Figure 3: Example EU product performance labels: tyre label (left) and energy efficiency label for white goods, like freezers (right). The A–G scale has A being best.

...minds Spock to scroll. It does not take much to get Homer to jog Spock into thinking, and then Spock’s training will click in. It is worth saying that this should all work well unless Spock is totally fatigued or suffering from tunnel vision or other human factors issues because of work pressure, etc — in which case we have to rely on teamwork coming to the rescue. Hopefully there are other people nearby not subject to tunnel vision, and those people’s Homers will register the importance of reading more of the screen.

4 Making informed purchasing decisions

The Kahneman model explains why bad IT systems are purchased. Homer gets excited about the latest, sexiest technology, and Spock is struggling to work out what is best, so attribute substitution comes into play. Homer convinces Spock what is best—let’s call this the Homer Reflex. This is why we all rush out to buy the latest mobile phone or iPad, despite Spock worrying that it won’t be compatible with everything we have on our existing system! Spock literally cannot start thinking when Homer is so excited. Another trick is to provide discounts to get Homer excited — if it’s the same thing but cheaper it must be better; or, provide a big discount on consumables (which Homer likes) rather than explain something safer (which is hard work for Spock to grasp).

Makary gives a brilliant discussion of the misleading temptations of hospital robotics [13]. What is perfect for consumerism may be a disaster when it comes to hospital procurement [18]. The latest shiny
gizmo (especially if it has wifi . . . ) is not necessarily the best or the safest. Somehow we have to change things so that Spock gets engaged with the serious purchasing decision.

The EU found the same problem with consumers buying car tyres. Homer chose cheap or pretty tyres; Spock never got a look in to think about safety or fuel consumption.

European Union (EU) legislation now requires car tyres to show their stopping distance, noise level and fuel efficiency at the point of sale. An example is shown in figure 3. The legislation follows on from similar successful schemes to show energy efficiency in white goods. Customers want to buy better products, so now when they buy a tyre, they — their Homer’s — can see these safety factors and then factor them into their decision making.

The EU has not specified how to make tyres better, they have just required the quality to be visible. The manufacturers, under competition, work out for themselves how to make their products more attractive to consumers. Indeed, thanks to energy efficiency labelling, product efficiency has improved, in some cases so much so that the EU has extended the scales to A*, A**, A***. The point is, normal commercial activity now leads to better products.

By analogy with tyres medical devices should show safety ratings — ones that are critical and that can be assessed objectively — and start with these to stimulate market pressure to make improvements.

It is easy to measure tyre stopping distances and we agree that stopping several meters before an obstacle instead of hitting it is desirable. There is no similar consensus in healthcare. Therefore any measurement process has to be combined with a process to improve, even create, the measurements and methodologies themselves. (And we need to fund more research.)

It is interesting to read critiques of pharmaceutical development [4] and realise that at least in pharmaceuticals there is a consensus that scientific methods should be used, even if the science actually done has some shortcomings. In healthcare devices there isn’t even any awareness that the things need evaluating. Homer is too excited — in fact, Kahneman has a term for this “the illusion of skill” meaning that people think they are skilled at deciding that IT and computer-based things are good, but this is illusory; Homer, not Spock, is making the decisions.

In the UK, about 2,000 people die on the roads annually, and cars (and tyres) are subject to legally required pre-market safety checks, and to annual safety checks, as well as random road-side checks. The safety checks are a combination of visual inspection, rigorous tests, and using advanced equipment (e.g., to measure exhaust emission quality). After a road accident, proof of roadworthiness is usually required, as it is obvious that a faulty car may cause an accident because it is faulty. In the UK, there are an estimated 42,000–80,000 per year (based on [9], simply scaling by UK/USA populations), which is 20 to 40 times higher, yet there is no legislation comparable to the many requirements imposed on road transport.
5 Conclusions

Today's enthusiasm for IT recalls the original enthusiasm for X-rays (figure 4) — Clarence Dally, an early adopter, suffered radiation damage and later died from cancer only a few years after Röntgen’s first publication [15]. It is now obvious that X-rays carry risks and have to be used carefully.

Today’s healthcare IT is badly designed; the culture blames users for errors, thus removing the need to closely examine design. Moreover, often manufacturers require users to sign “hold blameless” contracts [12], which means the user is the only person left to blame. Mortality rates in hospitals can double when computerised patient record systems are introduced [6]: computers are not the unqualified “X-ray” blessings they are often promoted as being.

A safety labelling scheme would raise awareness of the issues and stimulate competition for safer devices and systems. It could be done voluntarily, with no regulatory burden on manufacturers. By keeping rating labels on devices for their lifetime, patients would also gain increased awareness of the issues.

We need to improve. Kimberley Hiatt and many other people might still be alive if their calculators, infusion pumps, robots, linear accelerators, and so forth had been safer.

Acknowledgements

Partly funded by EPSRC Grant [EP/G059063/1]. With many thanks to Ross Koppel. Full background for this lecture can be found in [17].

References

[22] Various authors. runningahospital.blogspot.co.uk/2011/05/i-wish-we-were-less-patient.html; josephineensign.wordpress.com/2011/04/24/to-err-is-human-medical-errors-and-theconsequences-for-nurses. I wish we were less patient; To Err is Human: Medical Errors and the Consequences for Nurses, 2013.