Science and Public Policy, volume 21, number 4, August 1994, pages 233-248, Beech Tree Publishing, 10 Watford Close, Guildford, Surrey GU1 2EP, England.

Computer safety

Computer-related accidental death: an empirical exploration

Donald MacKenzie

Despite widespread interest in computer system failures, there have been few systematic, empirical studies of computer-related accidents. 'Risks' reports in the Association for Computing Machinery's Software Engineering Notes provide a basis for investigating computer-related accidental deaths. The total number of such deaths, world-wide, up until the end of 1992 is estimated to be 1,100 ± 1,000. Physical causes (chiefly electromagnetic interference) appear to be implicated in up to 4% of the deaths for which data were available, while 3% involved software error, and about 92% failures in human-computer interaction.

Donald MacKenzie holds a personal chair in Sociology at the University of Edinburgh. His contact address is Department of Sociology, University of Edinburgh, 18 Buccleuch Place, Edinburgh EH8 9LN, Scotland. This work was supported financially by the Science and Engineering Research Council (grant J58619), the Economic and Social Research Council (grants WA35250006 and R000234031) and the Joint Committee of the above two Research Councils (grant H74452).

The author is grateful for bibliographic help, data, ideas and pointers received from Robin Bloomfield, Nick Curley, Bob Lloyd, Peter Mellor, Peter Nicolaisen, Gene Rochlin, Scott Sagan, and, especially, Moyra Forrest and Rosi Edwards. He owes a broader debt to Peter Neumann and the many contributors to 'Risks' reports in *Software Engineering Notes*.

JUST HOW SAFE, or how dangerous, are the computer systems on which lives depend? How many lives have been lost through failures of such systems? What are the causes of such accidents?

Although there is a large literature on computer system safety, it contains little in the way of systematic, empirical answers to these questions. Published discussions often begin by highlighting a handful of dangerous failures, but typically make no attempt to place these in the context of any wider record.

There is, it is true, widespread awareness of the potential dangers of computer systems, and considerable research work and substantial sums of money are being devoted to technical means for making computer systems safer. This effort to find a solution is entirely necessary and desirable. Its chances of success might, however, be enhanced by detailed investigation of the problem.

The aim of this article is to indicate what might be involved in an empirical investigation of fatal accidents involving computer systems. The article's contribution to our knowledge of these accidents is at best modest. It is based on patently incomplete data sources, rendering its quantitative conclusions dubious. There are difficulties with its central category of 'computer-related accidental death'. There are both conceptual and empirical problems in its attempt to categorise the causes of such deaths.

Nevertheless, I hope that, precisely by virtue of these inadequacies, this paper will spark further work on this topic. One of its conclusions — that there is a pressing need for public agencies to begin

 0302-3427/94/040233-16 US$08.00

systematic, cross-sectoral data collection in this area — indeed seems to follow irresistibly from the very inadequacies of the existing record. Others of the article's conclusions, such as that computer-related fatalities have, to date, seldom been caused by technical design error alone, seem reasonably robust, despite the deficiencies in the data drawn on here.

Definition of terms

What is meant by 'computer-related accidental death'? Each of the four words in this phrase requires some justification or elaboration, beginning with the last.

Death

There are three reasons for focusing on accidents involving death, rather than all computer-related injury. First, the latter would be too broad a category for any sensible analysis. It would, for example, be necessary to include the large numbers of cases of ill health resulting from computer-terminal use, of which cases of upper limb disease (or 'repetitive strain injury') are perhaps the most prominent.

Second, the only available source of international, cross-sectoral data (described below) is indirectly dependent on press reports. Deaths are, to put it crudely, more newsworthy than non-fatal injuries, and so there is a far better chance of obtaining reasonable coverage of deaths than of injuries.

Third, accidental deaths often trigger formal enquiries. These provide useful information on the causes of computer-related fatal accidents, information that is absent in most cases of non-fatal injury.

To allow a reasonable period for reports of such deaths to enter the public domain, the cut-off point of this analysis is the end of December 1992. As far as possible, however, I have attempted to encompass all earlier cases of computer-related accidental death, world-wide.

Accidental

Some computer systems are meant to kill people. Since my interest is in unintended and erroneous behaviour in computer systems, it would not be appropriate to include in the analysis deaths caused by military computer systems when these function as intended.

A more difficult issue is deaths of civilian bystanders caused by computer-controlled offensive military systems whose primary targets are opposing military forces. Such deaths have clearly been substantial in number, from the Vietnam War, in which such systems first found major use, to the Gulf War and its aftermath.

In one sense, these are accidental deaths: the designers and operators of such systems would, ideally, prefer them not to take place. On the other hand, a certain level of 'collateral' civilian death is typically an anticipated and tacitly accepted feature of some kinds of military operations. Furthermore, it is extremely difficult to obtain reliable data on such incidents. I have, therefore, reluctantly decided to exclude such deaths from my analysis.

However, I have sought to include in the data set some deaths in military operations that are the result of system failures that are in some more clear-cut sense accidental in nature (rather than 'by-products' of normal system operation). The analysis includes deaths resulting from computer-related failures of defensive military systems and from computerrelated accidental crashes of military aircraft. It also includes the 1983 shooting down of a Korean airliner by Soviet air defences (where the accidental element is the navigational error that led the plane to stray into Soviet air space), and the 1988 downing of an Iranian airliner by the USS *Vincennes* (where the accidental element is the misidentification of this plane as an attacking military aircraft).

Computer

I have deliberately taken a broad view of what constitutes a 'computer', including in my definition any programmable electronic device or system, and not only those incorporating a full general-purpose digital computer. An industrial robot (so long as it is both electronic and programmable), a computer numerically-controlled machine tool, and a programmable cardiac pacemaker would, for example, all fall under my definition of systems which incorporate a 'computer'.

Nevertheless, some problems remain. The first generation industrial robots installed in the 1960s typically had pneumatic and electromechanical, rather than electronic, control systems (Dhillon, 1991, page 38). They would therefore fall outside my definition, but in reports of cases of robot-related death it is often unclear whether this kind of robot, or a more sophisticated electronic device, was involved.

It is too narrow to include as computer-related only cases of 'technical' failure of a computer system: also incorporated are cases where there has been a breakdown or error in human interaction with the system

Related

These definitional problems are, however, negligible compared to the difficulty of saying when a given accidental death is computer-*related*. The mere presence of a computer, even one playing a safety-critical role, in a system which suffers an accident is not sufficient for any reasonable categorisation of a death as computer-related. Rather, the presence of the computer must be causally important to the accident.

On the other hand, it would be too narrow to class an accident as computer-related only when a computer system problem was its *sole* cause. Major accidents often, perhaps usually, have multiple causes (Perrow, 1984; Oster *et al*, 1992, page 25).

Furthermore, it would, in my opinion, also be too narrow to include only cases of 'technical' failure of a computer system. I have also incorporated cases in which no technical failure is evident, but there has been a breakdown or error in human interaction with the system. Of course, such accidents can be, and frequently are, attributed to 'human error'. Yet system design often contributes to human error: for example, when the user interface of a computer system increases the probability of certain kinds of mistake, or when the safe functioning of a system requires its human operators to perform perfectly in tasks which are known to be error-prone (Norman, 1990).

Also included in my definition of 'computer-related' are accidents in which false confidence in computer systems, or specific misunderstandings of them, seem to have been a major factor in leading operators to adopt, or persist in, courses of action which they otherwise would have avoided or abandoned.

These considerations mean, however, that there is inevitably a degree of judgement involved in the categorisation. Just when does the role of a computer system in the sequence of events leading to an accidental death become important enough to justify calling it a 'computer-related' death? While seeking to exclude cases in which the computer system's role was minor, I have also tried to avoid being over-stringent, on the grounds that it is easier for a critical reader to exclude a case as not sufficiently computer-related than to scrutinise for possible inclusion all the conceivable 'marginal' cases.

This kind of (obviously contestable) judgement is not the only dilemma. The widely publicised failure late in 1992 of the new computerised dispatch system at the London Ambulance Service indicates another problem. There is no doubt that considerable suffering and some degree of physical harm to patients resulted from this failure. Patients also unquestionably died in London on the crucial days of 26 and 27 October and 4 November.

Yet there are matters of delicate medical judgement involved in assessing whether the lives of those who died might have been saved had ambulances reached them earlier. The coroners involved seem to have taken the view that they would not have been saved. Therefore, the London Ambulance Service case has to be excluded from my list of computer-related deaths. I should note, however, that, were the case to be included, the findings of the enquiry into this incident, which highlight the interaction of technical and organisational failings, would reinforce, rather than undermine, the conclusions below (see London Ambulance Service 1993), while the number of deaths involved is not such as to alter greatly the quantitative totals.

Similarly, to take a case that is included in the data set, many cancer patients died after receiving underdoses in computerised radiotherapy at the North Staffordshire Royal Infirmary between 1982 and 1991, but there are clearly difficult clinical judgements to be made as to which of those deaths are attributable to the underdosing. No figure more precise than "tens ... rather than hundreds" has been given (Milhill, 1993).

Another, pervasive, problem is that there is frequently sharp disagreement over the causes of an accident. On the outcome of such disagreement may hinge issues of civil liability and sometimes even criminal culpability (such as homicide or manslaughter charges). Unless he/she has the resources to mount his/her own investigation, the best the researcher can do is to turn to the most authoritative available source: an official enquiry, or, in some cases, an independent report.

In practice, however, it is often wise to be sceptical of even these sources. For example, Martyn Thomas, a leading commentator on matters of computer system safety, suggests that "the probability of the pilot being blamed for [an air] crash is more than twice as high if the pilot died in the crash" (quoted in *Software Engineering Notes*, April 1992, page 30).

In a substantial number of cases, furthermore, I have been not able to find either the report of an official enquiry or that of a thorough independent investigation. In these cases, I have erred on the side of inclusion, at least so long as there seemed to me to be a not wholly implausible case for their computer-relatedness.

Unlike many official enquiries, research such as this does not seek to allocate blame, and I have felt it better to include cases that *may* be computer-related, rather than to exclude them because computer-relatedness cannot be proven. Critical readers may, however, wish to excise from the totals those incidents for which I have described data quality as "poor" or "very poor", as well as drawing on the bibliographic materials cited here, to form their own opinion of the degree of computer-relatedness of the better-documented cases.

A more particular problem concerns what this

data set suggests are the two most important 'technical' causes of computer-related accidental death: electromagnetic interference and software error. A broken part will often survive even a catastrophic accident, such as an air crash, sufficiently well for investigators to be able to determine its causal role in the sequence of events. Typically, neither electromagnetic interference nor software error leave physical traces of this kind. Their role can often only be inferred from experiments seeking to reproduce the conditions leading to an accident. While this can on occasion be done convincingly, it is sometimes far from easy, and the suspicion therefore remains that these causes are under-reported.

Method

My primary source of cases was the remarkable compilation of reports of computer-related accidents and other failures that has, as a result of the efforts of computer scientist Peter Neumann, accumulated over the years in the pages of the Association for Computing Machinery's newsletter *Software Engineering Notes*, established in 1976. To begin with, these reports were a sporadic feature of Neumann's "Letter from the Editor".

In the early 1980s, however, the volume of such reports grew sharply, and in August 1985, an online electronic news network, called RISKS Forum, was set up, moderated by Neumann, with many contributors. This Forum (accessible on Internet) has become the basis of a section on "Risks to the Public" in each ordinary issue of *Software Engineering Notes*.

Although the resultant record has deficiencies from the point of view of systematic analysis — these are discussed below — this material forms a unique and invaluable data source. There is no doubt that its very existence has been a spur to a great deal of the research work relevant to computer safety. Inspection of existing articles dealing with the topic makes clear how important *Software Engineering Notes* and the RISKS forum have been in publicising accidents involving computers: see, for example, the influential lecture by Thomas (1988).

The method I used to gather cases was very simple. I examined each issue of *Software Engineering Notes* carefully for instances of apparent computer-related accidental death. These were cross-checked against the helpful indexes regularly produced by Peter Neumann in case one should be missed in the sheer volume of material. Wherever possible, I then sought either the report of an official enquiry into, or an independent investigation of, the particular incident described. At the very least, an attempt was made always to check the original published source, wherever this was quoted.

'Over-reporting' and 'under-reporting' of computer-related deaths are potential problems: 'under-reporting' is the more intractable as there is no straightforward way of investigating its extent

Apart from the general issues raised in the previous section, there are clearly two potential problems in this use of *Software Engineering Notes*: the 'over-reporting' and 'under-reporting' there of computer-related accidental deaths. 'Over-reporting' is more common than might be imagined. Computer professionals have shown commendable zeal in searching for and publicising cases of computer-system failure. (There is, indeed, an interesting puzzle for the sociology of the professions in the contrast between this attitude and what seems to be the typically less zealous attitude of other professionals, for example doctors or lawyers, in uncovering and publicising errors by their colleagues.)

Reasonably often, incidents originally reported in the journal as involving computer-related accidental death subsequently turn out not to have been computer-related. The newsletter has itself often included corrections, and in other cases my own research suggested computer involvement to be negligible. These instances are excluded.

In other cases, no reliable source of information could be found on which to base such a judgement. As noted above, most of these are included in the data set, with warnings as to the poverty of information on them. A handful of cases which appeared *prima facie* merely apocryphal were, however, excluded; the number of deaths involved is small, so the effect on the overall pattern of the data is not great.

Unfortunately, under-reporting is far more of an intractable problem than over-reporting. *Software Engineering Notes* makes no pretence to be comprehensive in its coverage. Neumann, for example, is careful to title his indexes "Illustrative risks to the public". The cases reported in the RISKS forum and *Software Engineering Notes* are typically culled from press coverage: only a minority come from the reporter's personal experience (and these are almost always the less serious incidents, not those involving death).

Furthermore, there is an enormous preponderance of English-language newspapers and journals amongst the sources quoted. At best, therefore, only those computer-related fatal accidents that find their way into the English-language press appear to be covered.

In the absence of any comparable alternative

Table 1. Cases of possible computer-related accidental death (to end of 1992)

Date(s)	Number of deaths	Location	Nature of incident	Probable main cause(s)	Main reference(s)	Data quality
Physical causes						
?	1	USA	Accidental reprogramming of cardiac pacemaker	Interference from therapeutic microwaves	Dennett (1979)	Poor
?	1	USA	Accidental reprogramming of cardiac pacemaker	Interference from anti-theft device	*SEN*, 10(2), 6; *SEN*, 11(1), 9	Poor
1982	20	South Atlantic	Sinking of HMS *Sheffield* following failure to intercept Argentinean Exocet missile	Interference from satellite radio transmission	*Daily Mirror*, 15/5/86, 1; *Hansard*, 9/6/86	Fair
1982	1	USA	Car accident	Fire may have caused failure of anti-skid braking system	*San Francisco Chronicle*, 5/2/86, 12	Very poor
1986	2	Libya	Crash of US F111 during attack on Tripoli	Possible electromagnetic interference	*SEN*, 14(2), 22	Very poor
1982-87	22	?	Crashes of US military helicopter	Possible electromagnetic interference; denied by makers, US army	*AW&ST*, 16/11/87, 27-28	Contro-versial
1988	1	UK	Operator killed by computer-controlled boring machine	Machine restarted unexpectedly due to faulty capacitor	Edwards (nd)	Good
Software error						
1986	2	USA	Overdoses from radiation therapy machine	Error in relationship between data entry routine and treatment monitor task	Leveson & Turner (1992)	Very good
1991	28	Saudi Arabia	Failure to intercept Iraqi Scud missile	Omitted call to time- conversion subroutine; delayed arrival of corrected software	GAO (1992);Skeel (1992)	Good
Human-computer interaction problems						
Medical						
1982-91	"in the tens"	UK	Underdosing by radiation therapy machine	Correction factor for reduced source-skin distance in isocentric therapy applied twice (already present in software).	West Midlands Regional Health Authority (1992); North Staffordshire Health Authority (1993)	Good
Military						
1987	37	Persian Gulf	Failure to intercept attack on USS *Stark* by Iraqi Exocet missile	Alleged lack of combat-readiness; possible defective friend/foe identification or switching off of audible warning	Sharp (1987), Committee on Armed Services (1987); Adam (1987); Vlahos (1988)	Fair
1988	290	Persian Gulf	Shooting down of Iran Air airliner by USS *Vincennes*	Stress; need for rapid decision; weapon system human interface not optimal for situation	Fogarty (1988)	Good
Air						
1979	257	Antarctica	Crash of airliner on sightseeing trip	Communication failure re resetting of navigation system; continuation of flight in dangerous visual conditions.	Mahon (1981)	Fair, but aspects contro-versial
1983	269	USSR	Shooting down of Korean Air Lines airliner following navigational error	Autopilot connected to compass rather than inertial navigation system	*AW&ST*, 21/6/93, 17	Fair
1988	4	UK	Collision of 2 RAF Tornado aircraft	Use of identical navigational cassettes by different aircraft	*Sunday Times*, 11/3/90, A9	Fair
1989	12	Brazil	Crash of airliner after running out of fuel	Incorrect input to navigation system (?)	*SEN* 15(1), 18	Contro-versial

(continued)

Table 1. *(continued)*

Date(s)	Number of deaths	Location	Nature of incident	Probable main cause(s)	Main reference(s)	Data quality
1992	87	France	Crash of airliner into mountain during night-time approach	Vertical speed mode may have been selected instead of flight path angle; limited cross-checking between crew; possible distraction; no ground proximity warning system	Sparaco (1994)	Fair
Robot-related						
1978-87	10	Japan	Workers struck during repair, maintenance, installation or adjustment of robots	Workers entered envelope of powered-up robots; in some cases, deficiencies in training and absence of fences	Nagamachi (1988)	Fair
1984	1	USA	Heart failure after being pinned by robot	Worker entered envelope of powered-up robot	Sanderson *et al* (1986)	Fair
Involving other automated plant						
1979	1	USA	Worker struck by automated vehicle in computerised storage facility	Absence of audible warning; inadequate training; production pressure	Fuller (1984)	Good
1983-88	13	France	Accidents to operators/ installers/repairers of automated plant	Insufficient individual detail given in source	Vautrim and DeiSvaldi (1989)	Fair, but too aggre-gated for current purpose
1988	1	UK	Maintenance electrician killed by unexpected movement of automatic hoist	Disconnection of proximity switch, sent signal to controller; machine not isolated	Edwards (nd)	Good
1989	1	UK	Setter/operator killed by palletiser	Machine cycled when boxes interrupting photoelectric beam removed; transfer table not isolated	Edwards (nd)	Good
1991	1	UK	Maintenance fitter killed by hold-down arm of feed unit to log saw	Fitter's body interrupted beam of process sensor; machine not isolated	Edwards (nd)	Good
1991	1	UK	Maintenance fitter killed in automatic brick-making plant	Fitter inside guarding enclosure observing cause of misalignment of bricks	Edwards (nd)	Good
?	3	Nether-lands	Explosion at chemical plant	Typing error caused wrong chemical to be added to reactor	*SEN*, 18(2), 7	Fair
Insufficient data						
1986	1	USA	Overdose of pain-relieving drugs	Error in medical expert system(?)	Forester and Morrison (1990)	Very poor
1989	1	USA	Failure of school-crossing pedestrian signals	Breakdown in radio communications link to computer(?)	Emery (1989)	Poor
1990	1	USA	Collision of automated guided vehicle and crane	Unclear	*SEN*, 16(1), 10	Very poor
1990	1?	USA	Delay in ambulance despatch	Logging program not installed(?) Unclear whether death result of delay	*SEN*, 16(1), 10	Poor
c 1983	1	West Germany	Woman killed daughter after erroneous medical diagnosis	'Computer error'	*SEN*, 10(3), 8	Very poor
c 1984	1	China	Electrocution	Unclear	*SEN*, 10(1), 8	Very poor
c 1989	1	USSR	Electrocution	Unclear	*SEN*, 14(5), 7	Very poor
?	2?	?	Sudden unintended car acceleration	Unclear	*SEN*, 12(1), 8-9; *Business Week*, 29/5/89, 19	Poor; contro-versial

Acronyms: *SEN* is the Association for Computing Machinery, *Software Engineering Notes*
AW&ST is *Aviation Week and Space Technology*
GAO is the General Accounting Office

source, however, there is no straightforward way of investigating the extent of under-reporting. The impression I formed was that coverage of 'catastrophic' accidents, such as crashes of large passenger aircraft, is good. These will always be reported in the press, extensive enquiries will typically ensue, and the subscribers to RISKS seem carefully to scrutinise reports of such accidents and enquiries for any suggestion of computer involvement.

It seemed likely, however, that coverage of less catastrophic accidents, such as industrial accidents involving robots or other forms of computer-controlled automated plant, would be poorer. These will typically involve only a single death; they take place on the premises of an employer who may have no wish to see them widely publicised; and they may be regarded by the media as too 'routine' to be worth extensive coverage. Accordingly, I investigated these separately through contacts in the robot industry and the UK Health and Safety Executive.

It turns out that *Software Engineering Notes*'s coverage of fatal accidents involving robots is reasonable: indeed, there seems to have been a degree of over-reporting rather than under-reporting. This good coverage probably arises because robot accidents have been regarded by the media as a newsworthy topic. On the other hand, even the small amount of systematic data I have found on fatal industrial accidents involving more general types of automated plant makes it clear that this kind of accident is greatly under-reported in *Software Engineering Notes*. I would indeed hypothesise that this is the most important gap in the data recorded below.

Overall data

There are around 1,100 computer-related accidental deaths in the overall data-set generated by the above methods: to be precise, 1075 plus the 'tens' of the North Staffordshire radiation therapy incident (see table). The data's limitations, discussed above, mean that these figures are far from definitive. Despite extensive literature searches, data on several of the incidents remain extremely poor.

Those inclined to attribute accidents to human error alone would probably deny that many of the 'human-computer interaction' cases are properly to be described as computer-related. It might also be argued that some of the deaths (for example, those resulting from failure to intercept a Scud missile and from the Soviet downing of the Korean airliner) should not be classed as accidental. There are, furthermore, a variety of particular problems in the diagnosis of other incidents (some of which are discussed below) which might lead a critic to exclude them too.

Only a small minority of incidents — perhaps only the Therac-25 radiation therapy incidents — seem entirely immune from one or other of these exclusionary strategies, although to force the total much below 100 would require what seem to me to be bizarre definitions, such as a refusal to accept the North Staffordshire deaths as computer-related.

In other words, more stringent criteria of what is to count as a computer-related accidental death could reduce the overall total well below 1,100. On the other hand, the fact that the mechanisms by which a death reaches *Software Engineering Notes* are far from comprehensive means that there is almost certainly a substantial degree of under-reporting in this data set.

In particular, there must have been more fatal industrial accidents involving computer-controlled automated plant than the 22 cases recorded here. Systematic data were available to me only for Britain and France, and for limited periods of time. Comprehensive coverage of other advanced industrial nations would increase the overall total considerably.

Furthermore, the relatively small number of instances from outside the English-speaking world (particularly from the former Soviet bloc) is suspicious. Reliance on computers is more pervasive in western industrial nations than in the former Soviet bloc and third world, but probably not to the extent that the geographic distribution of the accidents recorded here might suggest.

Any attempt to correct for this under-reporting is obviously problematic. It seems to me unlikely, however, that any plausible correction could boost the total by much more than around a further 1,000. For that to happen would require that one or more catastrophic computer-related accidents, involving at least several hundred deaths, has gone unrecorded. That is possible, but, given the number and diligence of Neumann's correspondents, unlikely.

Therefore, the findings of this analysis on the total number of computer-related accidental deaths, world-wide, to the end of 1992, can be expressed, in conventional format, as 1,100 ± 1,000. The relatively large error band appropriately conveys the twin problems inherent in this exercise: that more stringent definition would reduce the total considerably, while correction for under-reporting could plausibly just about double it.

Aside from the total number of deaths, the other most salient aspect of this data set is the causes of the incidents it contains. I have divided the accidents into three rough-and-ready categories, according to the nature of their dominant computer-related cause: physical failure of a computer system or physical disturbance of its correct functioning; software error; or problems in human-computer interaction. While space and inadequate data prohibit description of every individual incident, some discussion of the type of accident to be found in each category may be of interest.

Physical causes

Up to 48 deaths fall into this category. Apart from one resulting from a capacitor failure, and another (dubious) case in which a safety-critical computer system may have failed fatally because of fire, deaths involving physical causes have all been the result of electromagnetic interference, when a programmable system is reprogrammed or its normal operation otherwise impeded by stray radio signals or other electromagnetic emissions.

There are two reported deaths resulting from the accidental reprogramming in this way of cardiac pacemakers. Several military system accidents have also been alleged to have been caused by electromagnetic interference, although (perhaps because of the particular difficulty noted above of diagnosing electromagnetic interference retrospectively) these cases are almost all controversial. In only one of them has electromagnetic interference been stated officially to be the cause: the failure of HMS *Sheffield*'s defensive systems to intercept an attacking Argentinean Exocet missile during the Falklands War.

At the time of the attack, the *Sheffield* was in urgent radio communication, through its satellite communications transmitter, with another vessel in the British task force. Interference from this transmitter prevented the *Sheffield* from picking up warning signals on its electronic support measures equipment until too late to intercept the Exocet attack. Published reports leave unclear what precise aspect of the equipment was interfered with (although the distinction is difficult for a modern system of this kind, it clearly could be the radar rather than an information processing aspect), but there seems to me to be sufficient indication here of possible 'computer-relatedness' to merit the inclusion of this case in the data set.

Software error

Much of the discussion of the risks of safety-critical computing has focused on software error, and the data set contains two incidents (involving a total of 30 deaths) which are clearly of this kind. Two deaths resulted from overdoses from a computer-controlled radiation therapy machine known as the Therac-25. (A third patient also died from complications related to a Therac-25 overdose, but he was already suffering from a terminal form of cancer; the autopsy on a fourth overdosed patient revealed her cause of death to have been the cancer from which she suffered rather than radiation over-exposure).

Two deaths resulted from overdoses from a radiation therapy machine when a software error shifted the mode of operation from x-ray to electron, while leaving the intensity at the high current required for x-ray therapy

The Therac-25 has two therapeutic modes: electron mode, used for treating tumour sites on or near the surface of the body; and x-ray mode, used for treating deeper tumour sites. The latter involves placing in the path of the electron beam a tungsten target (to produce the x-rays) and also what is called a 'beam flattener' (to ensure a uniform treatment field). Because the beam flattener greatly reduces the intensity of the beam, x-ray therapy requires around 100 times more electron-beam current than electron-mode therapy. If the stronger current were used without the target and beam flattener being in place, then the patient would receive a massive overdose.

Unfortunately, a software error (described in detail in Leveson and Turner, 1992) meant that there was a particular form of data entry on the Therac-25 which caused precisely this to happen, because it shifted the mode from x-ray to electron, while leaving the intensity at the current required for x-ray therapy. The data that appeared on the system's display did not reveal that this had taken place, and the fatal error was diagnosed only with some difficulty. Investigation also revealed a further dangerous software error, although this seems not to have been implicated in the two deaths included here (Leveson and Turner, 1992).

A software error also caused the failure of the Patriot air defence system at Dhahran during the 1991 Gulf War which led to the deaths of 28 American soldiers in an Iraqi Scud missile attack, the largest single Allied loss in the campaign. When tracking a target, sophisticated modern radar systems, such as that used for Patriot, process not the entire reflected radar beam, but only a portion of it known as the 'range gate'. An algorithm embedded in the system software shifts the range gate according to the velocity of the object being tracked and the time and location of its last detection. An error in the implementation of the range gate algorithm was the cause of the failure to attempt to intercept the attacking Scud (General Accounting Office, 1992).

Patriot's internal clock keeps time as an integer number of tenths of seconds, and that number is stored as a binary integer in the Patriot computer's registers, each of which can store 24 binary digits, or 'bits'. For use in the range-gate algorithm, this integer number of tenths of a second is converted into a 48-bit floating-point[1] number of seconds, a conversion that requires multiplication of the integer by the 24-bit binary representation of one tenth. The binary representation of one tenth is

non-terminating, and so a tiny rounding error arises when it is truncated to 24 bits, which, if uncorrected, causes the resultant floating-point representations of time to be reduced by 0.0001% from their true values (Skeel, 1992).

Patriot was originally designed to intercept relatively slow targets such as aircraft. Amongst the modifications made to give it the capacity to intercept much faster ballistic missiles was a software upgrade increasing the accuracy of the conversion of clock time to a binary floating-point number. Unfortunately, at one place in the upgrade a necessary call to the subroutine was accidentally omitted, causing a discrepancy between the floating-point representations of time used in different places in the range-gate algorithm. The result was an error that was insignificant if the system was used for only a small amount of time, but which steadily increased until the system was 'rebooted' (which resets time to zero).

The problem was detected prior to the Dhahran incident. A message was send to Patriot users warning them that "very long run times could cause a shift in the range gate, resulting in the target being offset" (General Accounting Office, 1992, page 9). A software modification correcting the error was dispatched to users over a week before the incident. However, the matter was reportedly treated as not one of extreme urgency because Army officials "presumed that the users [of Patriot] would not continuously run the batteries for such extended periods of time that the Patriot would fail to track targets" (General Accounting Office, 1992, page 9); rebooting takes only 60 to 90 seconds.

Unfortunately, on the night of 25 February, Alpha Battery at Dhahran had been in uninterrupted operation for over 100 hours, a period sufficient for the error to cause loss of tracking of a target moving as fast as a Scud. As a result, no defensive missiles were launched against the fatal Scud attack.[2] The following day the modified software arrived.

Human-computer interaction

In this category there were 988 plus 'tens' of deaths. The accidents centring on these problems are typically 'messier' in research terms than those which have clear-cut 'technical' issues at their core. Precisely because they result from problems in interaction, blame can be a contentious matter. System designers can see the fault as lying with operators. These operators, in their turn, sometimes make allegations of defective technical functioning of the system, often allegations for which no decisive evidence can be found, but which cannot be ruled out *a priori*.

These blame-seeking disputes cloud over what is typically the key point. Many safety-critical systems involving computers rely for their safe functioning on the correctness of the behaviour of both their technical and their human components. Just as failure of technical components is typically regarded as a predicable contingency (and guarded against by duplication or triplication of key parts, for example), so human failure should be expected and, as far as possible, allowed for.

Medical

For the sake of convenience, I have divided the cases of human-computer interaction problems into five broad categories: medical, military, air, robot-related, and those involving other automated plant. The medical case is the most clear-cut. It is the systematic underdosing in isocentric radiotherapy for cancer that took place at the North Staffordshire Royal Infirmary between 1982 and 1991.

In isocentric therapy the system's focal distance is set at the centre of a tumour and the machine is rotated so that the tumour is 'hit' from several different angles. In calculating the required intensity of radiation, it is necessary to allow for the fact that the distance between the source of the beam and the skin of the patient will be less than the 100 cm standard in forms of radiotherapy for which each beam is directed not at the tumour but at the point in the skin overlying it. If not, the patient will be overdosed.

Prior to computerisation, this correction was always calculated and entered manually. Unfortunately, this practice continued at the North Staffordshire hospital after a computerised treatment plan for isocentric radiotherapy was introduced in 1982, because it was not realised that the correction was already being made by the system software. The error was not detected until a new computer planning system was installed in 1991. The result was the underdosing by various amounts of around 1,000 patients.

Subsequent investigation (North Staffordshire Health Authority, 1993) suggests that 492 patients may have suffered an adverse effect from underdosage, of whom 401 had died by mid 1993. However, radiation therapy for cancer has a far from total success rate even when conducted perfectly, and so many of these patients would have died in any case. As noted above, the clinical verdict was that the deaths resulting from the error were likely to be "in the tens rather the hundreds" (Milhill, 1993).

Military

The two military cases are much less clear-cut in their causes, and their interpretation has been controversial. While patrolling the Persian Gulf in 1987 during the Iran-Iraq war, the US frigate *Stark* was struck by two Exocet missiles fired by an Iraqi

When an Iraqi Exocet missile hit a US frigate the computer system could have defined the Exocet as 'friendly' rather than 'hostile', or it may have produced a warning which was not noticed by the operator who had switched off the audible alarm

aircraft. Like HMS *Sheffield*, the *Stark* was equipped with computerised systems designed to detect and intercept such an attack.

The subsequent US Navy investigation focused mainly on an alleged lack of combat-readiness of the *Stark* (Sharp, 1987; Vlahos, 1988); it should be noted, however, that the United States was at war with neither party to the conflict, and indeed was widely seen as a *de facto* supporter of Iraq. More particularly, it remains puzzling that while the *Stark*'s electronic warfare system detected the Iraqi Mirage fighter, its crew appear not to have received a warning from the system about the incoming missiles.

Both of the main candidate explanations of this would lead to the classification of the incident as computer-related. One possibility is that the system may have detected the missiles but had been programmed to define the French-made Exocet as 'friendly' rather than 'hostile' (this suggestion was also made in attempts to explain why the *Sheffield* failed to intercept the Exocet attack on it, but was denied by the UK Ministry of Defence). The *Stark*'s SLQ-32 electronic warfare system "had Exocet parameters in its software library, but this software might have been flawed or out of date, a problem the Navy has admitted" (Vlahos, 1988, page 65). Another possibility is that the system did produce a warning, but that this was not noticed by its operator. The operator had switched off the audible alarm feature because the system was issuing too many false alarms.

In the shooting down the following year of the Iranian airliner, there is no evidence of any technical malfunction of the sophisticated Aegis computerised combat system on board the *Vincennes*. Data tapes from the system are entirely consistent with what in retrospect we know to have been the true course of events.

It is clear that the crew was operating under considerable stress. The ship was fighting off several fast-manoeuvring Iranian small boats, while having to turn abruptly at full speed to keep its weapons engaged on the targets (it had a fouled gun-mount); such turns cause a vessel such as the *Vincennes* to keel sharply. Furthermore, memories of the surprise airborne attack on the *Stark* were still fresh, and there was little time available in which to check the identification of the radar contact as a hostile Iranian military aircraft.

However, this accident should perhaps not be ascribed simply to human error (although the case for its computer-relatedness arguably remains the most marginal of the major cases in the data set). A key role in the mis-identification of the Iranian airliner as a military threat was played by the perception of it as descending towards the *Vincennes*, when in fact it was (and was correctly being analysed by the Aegis system as) rising away from it.

Stress undoubtedly played a major role in this misperception. However, the US Navy's report on the incident suggested that "it is important to note, that altitude cannot be displayed on the LSD [Aegis large screen display] in real-time". After the investigation of the incident, the Chairman of the Joint Chiefs of Staff recommended

> "that a means for displaying altitude information on a contact such as 'ascending' or 'descending' on the LSD should ... be examined ... [and] that some additional human engineering be done on the display screens of AEGIS" (Crowe, 1988, page 8).

More generally, it is noteworthy that it was the highly computerised *Vincennes* which misidentified the radar contact, while its technologically more primitive sister ship, the USS *Sides*, correctly identified the Iranian aircraft as no threat (Rochlin, 1991). A possible reason for this is discussed in the conclusion.

Air

The air incidents are also instances for which there is no evidence of technical malfunction, but where problems arose in human interaction with an automated system. The most recent of them has been the focus of intense scrutiny because it involved the first of the new generation of highly computerised 'fly-by-wire' aircraft, the Airbus A320,[3] one of which crashed in mountainous terrain after an over-rapid night-time descent in bad weather to Strasbourg-Entzheim airport.

The possibility that there had been a technical failure of the A320's Flight Control Unit computer system was not ruled out by the crash investigators, but was judged a "low probability" (Sparaco, 1994, page 30). Instead, their central hypothesis is that the two-man A320 crew (who both died in the accident) may have intended to instruct the flight control system to descend at the gentle angle of 3.3 degrees, but by mistake instructed it to descend at the extremely rapid rate of 3,300 feet per minute. Although a letter designation on the Flight Control Unit screen and distinct symbols on the primary flight displays indicate which mode has been selected, both modes were represented by similar two-digit numbers (the interface has subsequently

been redesigned so that the vertical speed mode is now represented by a four-digit number).

Analysis of the cockpit voice recorder suggests that "there was limited verbal communication, coordination and crosschecking between the two pilots" (Sparaco, 1994, page 31), whose attention may have been distracted from their speed of descent by a last minute air traffic control instruction to change runways and terminal guidance systems. The carrier operating the particular aircraft in question had declined to install automated ground proximity warning systems in its A320 fleet, at least in part because it believed such systems to give too many false alarms in the type of operation it conducted, so no warning of imminent impact was received by the crew.

The cases involving air navigational errors are in a broad sense similar. Modern long-range civil air transports and nearly all modern military aircraft are equipped with automatic navigation systems, most commonly inertial systems (which are self-contained, not reliant on external radio signals). Inertial navigators are now extremely reliable technically, perhaps to such an extent that undue reliance is placed on their output, with other sources of navigational data not always checked, and flights sometimes continued under what might otherwise be seen as overly dangerous conditions.

Yet such automated systems do have vulnerabilities. Inertial navigation systems need to be fed data on initial latitude and longitude prior to take-off. In civilian airliners inertial navigators are typically triplicated, to allow the isolation of individual errors. However, some configurations contain an override that allows data to be entered simultaneously to all three systems instead of individually to each. Furthermore, if the inertial system is to 'fly' the plane (via an autopilot), details of the requisite course must also be entered (typically in the form of the latitude and longitude of a set of way-points, and often as a pre-prepared tape cassette) and the correct 'connection' must be made between the inertial system and the autopilot.

The most notorious of the resulting incidents is the 1983 episode when a Korean Airlines airliner strayed into Soviet airspace and was shot down. The reasons for it over-flying Soviet territory attracted much speculation, and some lurid conspiracy theories, at the time. Data tapes from the airliner released recently by Russia, however, seem to point to a simple, undetected mistake: the aircraft's autopilot was connected to its compass rather than to the inertial navigation system. The aircraft therefore followed a constant magnetic heading throughout its flight rather than the intended flight plan.

While robot-related accidents have attracted considerable interest, there has been much less attention to fatal industrial accidents involving other kinds of automated plant, although the latter appear likely to be considerably more numerous

Robot-related

The robot-related deaths in the data set seem to manifest a common pattern, one also to be seen in non-fatal robot-related accidents, on which, unusually, considerable data is available. The key risk posed by robotic systems, in contrast to more conventional industrial machinery, is that the movements of the latter are typically repetitive and predictable, with danger points being obvious, while robot motion on the other hand is much less predictable (see, for example, Altamuro, 1983). A robot may suddenly start after a period of inactivity while internal processing is going on; the direction of movement of a robot 'arm' may suddenly change; all points in a robot's 'work envelope' (the three dimensional space which it can reach) are potentially hazardous.

Deaths and other serious accidents involving robots are nearly always because a worker is present within the envelope of a powered-up robot. Often, the worker is struck from behind and is pushed into another machine or against a fixed obstacle.

Workers are typically instructed not to enter the envelopes of powered-up robots, so it is tempting to ascribe all such accidents to 'human error' alone. But to do this would be to miss several points. First, the human error involved is an entirely foreseeable one, and so one that should be anticipated in system design. Second, in some early installations no barriers were present to inhibit workers from entering the envelope, and training was sometimes inadequate.

Third, there is little reason to think that workers enter robot envelopes gratuitously. They may, for example, be cleaning or attending to some small snag in the robot installation. It may be that there are pressures in the situation, such as to maintain productivity, that encourage workers to do this without switching off the power supply.

Fourth, some fatal accidents have occurred when a worker did indeed switch off power to the robot, but it was switched back on inadvertently by himself or another worker while he was within the robot envelope. Installation design could, at least to some extent, guard against this happening.[4]

Other automated industrial plant

While robot-related accidents have attracted considerable interest, there has been much less atten-

tion to fatal industrial accidents involving other kinds of automated plant, although the latter appear likely to be considerably more numerous. A particularly dangerous situation (as occurred in three of the five UK fatalities identified by Edwards) arises when workers enter or reach into the automated plant when it has stopped but is still powered-up, and so can be restarted by sensors, control system faults or signals from other locations (Edwards, no date, page 7).

As in the robot case, accidents of this type should not be disregarded as mere gratuitous and unpredictable 'human error'. The two systematic studies which I have been able to locate (Vautrin and Dei-Svaldi, 1989 and Edwards, no date) both suggest that accidents with automated plant typically involve inadequate system designs which make some necessary work activities — such as finding and rectifying faults, adjusting workpieces and, especially, clearing blockages — dangerous.

Sometimes there is deficient guarding or defects in isolation systems. Other dangers arise from having a process 'stop' device which halts the machine, but does not isolate it: the resultant accidents are far from unpredictable. More generally, accidents involving unsafe systems of work typically point to organisational, rather than individual, failings. For example, the maintenance electrician killed in Britain in 1988 by unexpected movement of an automatic hoist was reportedly "expected to maintain a system which had been supplied without an interlocked enclosure, and without any form of operating or maintenance manual" (Edwards, no date, page 26).

Conclusions

How safe are computers?

The data presented here are clearly insufficient for any quantitative measure of levels of risk associated with computer systems. For that to be possible, we would need to know not just numbers of accidental deaths, but also levels of 'exposure': total usage of computerised radiation therapy machines; total passenger-miles or hours flown in fly-by-wire aircraft or planes reliant on inertial navigators; total hours of work spent in proximity to industrial robots or close to automated plant; and so on. I do not possess this data, nor am I sure that the aggregate result of such an exercise would be meaningful: the risks involved in such different activities are scarcely commensurable.

Furthermore, even the crudest quantitative assessment of the benefits and dangers of computerisation would also require data on the risks of analogous activities conducted without the aid of computers. In limited spheres such as radiotherapy and (perhaps) civil aviation the comparison might be an interesting research exercise,[5] but often it is impossible. For example, effective defence against ballistic missiles without the aid of computers is hard to imagine; there is therefore simply no comparator case.

So I can answer the question of the overall safety of computer systems in only the crudest sense: the prevalence of computer-related accidents as a cause of death. In that sense, a total of no more than around 2,000 deaths so far, world-wide, is modest. For example, in 1992 alone, there were 4,274 deaths in the UK in road traffic accidents (Smithers, 1993). By comparison, computer-related accident has, up until now, not been a major cause of death.

Nevertheless, there are no grounds here for complacency. In the context of activities with a generally excellent safety record, such as scheduled air transport, even a small number of major accidents becomes most worrying. In addition, deaths are sometimes only the visible tip of what can be a much larger 'iceberg' of serious injuries, minor injuries and 'near misses'.

This is, for example, clearly the case for accidents involving robots and other forms of automated plant. Edwards's data set (Edwards, no date), for example, contains 14 major injuries, and 40 minor ones, for each fatality.[6] These multipliers would probably be smaller in other sectors, notably air travel,[7] but there have clearly been a substantial number of computer-related injuries to add to the total of fatalities. Furthermore, even a cursory reading of 'risks' reports in *Software Engineering Notes* leaves one convinced that the number of 'near misses' is likely to be considerable.

We are dealing here with a relatively new problem, for which the record of the past is unlikely to be a good guide to the future, since the incidence of computerisation, its complexity and its safety-criticality seem to be increasing rapidly (see, for example, Rushby, 1993, pages 127-128).

True, an unequivocal trend in time in the data set cannot be established: the numbers of deaths are dominated too much by the three incidents in 1979, 1983 and 1988 in each of which over 200 people were killed. It is, however, striking that there is no well-documented case of a computer-related accidental death before 1978. Of course, that may to some degree be an artefact of the reporting system: 'risks' reports in *Software Engineering Notes* were only beginning then. But attention to the problem of computer safety goes back at least to the late 1960s (Peláez 1988), and so it seems unlikely that there can be large numbers of deaths prior to 1978 that have gone unrecorded in the literature.

Need for systematic data collection

The attempt to conduct an exercise such as this, quickly reveals the need for systematic data collection about computer-related accidents. There are

occasional pieces of excellent scientific detective work, such as Skeel's (1992) uncovering of the precise role of rounding error in the Dhahran incident, a role not fully evident even in the otherwise useful report by the General Accounting Office (1992). There is one superb detailed case-study: Leveson and Turner's (1992) investigation of the Therac-25 accidents. There are also 'islands' of systematic data on particular sectors, such as Edwards' (no date) noteworthy study of accidents on computer-controlled plant in Britain.

The RISKS Forum and *Software Engineering Notes,* however, remain the only cross-sectoral, international database. Remarkable and commendable efforts though they are, they are no substitute for properly-resourced, official, systematic data collection.

A large part of the problem is the diversity of regulatory regimes which cover safety-critical computing. By and large, computer use is covered by the regulatory apparatus for its sector of application, apparatus which normally will predate digital computer use in that sector, and which will naturally be influenced strongly by the history and specific features of the sector.

Yet there is a strong argument that the introduction of digital computers, or of programmable electronic devices more generally, introduces relatively novel hazards which have common features across sectors. Software-controlled systems tend to be logically complex, so operators may find it difficult to generate adequate 'mental models' of them. Their complexity also increases "the danger of their harbouring potentially risky design faults", and

> "the largely discrete nature of their behaviour ... means that concepts such as 'stress', 'failure region', 'safety factor', which are basic to conventional risk management have little meaning" (Randell, 1989, page 21).

Digital systems are characterised by the

> "discontinuity of effects as a function of cause. There is an unusual amplification of the effects of small changes. Change of a single bit of information (whether in a program or data) can have devastating effects" (Neumann, 1988, page 3).

Installing programmable systems in duplicate or triplicate offers only limited protection, since software or hardware design errors can be expected to produce 'common-mode failures' which manifest themselves in each system simultaneously. Even installing different systems may be less of a protection against common-mode failure than might be imagined, because in some cases the individual programs produced by separate programmers can still contain "equivalent logical errors" (Brilliant, Knight and Leveson, 1990, page 238).

There is evidence that the existence of independent data-gathering systems makes systems safer, especially when data is on 'incidents' as well as 'accidents', on a no-fault and confidential basis, and results are well publicised

If this is correct (and amongst the cases presented here some of these phenomena can be found[8]), the risks associated with computer systems can be expected to have generic, technology-specific features as well as sector-specific, application-specific ones. It could thus be that a great deal of important information is being lost through the partial and predominantly intra-sectoral nature of current information gathering.

Nor is this simply a matter of the need for an empirical basis for research. There is evidence from other areas that the existence of independent data-gathering systems in itself makes systems safer, especially when data is collected on 'incidents' as well as on actual 'accidents', and when the data-gathering is on a no-fault and confidential basis (to reduce to a minimum the motivations to under-report), and when results are well publicised to relevant audiences. The incident-reporting system in civil air transport is a good example (Perrow, 1984).

The British Computer Society has recently called for a system of registration of safety-related computer systems with mandatory fault reporting. Such a system would be an important contribution to improving the safety of such systems as well as a valuable basis for research (BCS Safety Critical Systems Task Force, 1993).

The technical and the human

Computer-related accidental deaths caused *solely* by technical design flaws are rare. The fatalities in the data set resulting from human-computer interaction problems greatly outnumber those from either physical causes or software errors. True, some of the 'interaction' cases may perhaps mask software error or hardware faults, but, on the other hand, one of the cases of software error, and some of those of physical causes, also have 'interaction' aspects.

For instance, the cause of the Dhahran deaths was not just the omitted call to the time conversion subroutine: assumptions about how the system would be operated in practice, and delays in the arrival of the corrected software, were also crucial. Leveson and Turner (1992, pages 37-38) argue

that even in the Therac-25 deaths — whose cause was perhaps the closest in the well-documented cases in this data set to a 'pure' technical fault — software error "was only one contributing factor". They argue that organisational matters, such as what they regard as inadequacies in the procedure for reporting and acting upon incidents, were also important.

Indeed, multi-causality may be the rule rather than the exception. More computer-related accidental deaths seem to be caused by interactions of technical and cognitive/organisational factors than by technical factors alone: computer-related accidents may thus often best be understood as *system* accidents (Perrow, 1984). In the absence, in many cases, of the depth of understanding now available of the Therac-25 and Dhahran deaths, or of the systematic coverage of Edwards' (no date) study of industrial accidents, this hypothesis cannot be verified conclusively, but such data as are available make it plausible.

There is, however, another worrying category of accident: that of unimpaired technical operation of the system, as far as we can tell, and yet disastrous human interaction with it. Contrasting the *Vincennes*'s erroneous identification of its radar contact and the *Sides*'s correct one, Rochlin (1991) argues that intensive computerisation can result in a changed relationship of human beings to technology, and his argument has wider implications than just for the analysis of this particular incident.

In a traditional naval vessel or aircraft, human beings play a central role in processing the information flowing into the vehicle. By contrast, as computerisation becomes more intense, highly automated systems become increasingly primary. Ultimate human control — such as a human decision to activate the firing mode of an automated weapon system — is currently retained in most such systems.[9] But the human beings responsible for these systems may have lost the intangible cognitive benefits that result from themselves having constantly to integrate and make sense of the data flowing in.

In such a situation, danger can come both from stress and from routine. Under stress, and pressed for time, the human beings in charge of automated military systems cannot be expected always to ask whether the situation they face is one "the elaborate control system in which they were embedded, and for which they were responsible" was designed to meet. We should not be surprised if sometimes they act out "the scenario compatible with the threat the system was designed to combat" (Rochlin, 1991, page 119). Nor should we be surprised if, after hundreds or thousands of hours' experience of flawless functioning of automated equipment such as inertial navigators, pilots or other operators start to trust that equipment too much, and, for example, fail to check other information available to them.

To make computer systems safer, we need to address not merely their technical aspects, but also the cognitive and organisational aspects of their 'real world' operation. The psychologist and organisational analyst have to be involved in this effort along with the computer scientist.

If this does not happen, there is a risk that purely technical efforts to make computer systems safer may fail. Not only are they addressing only part of the problem, but they may conceivably even increase risks, through their effect on beliefs about computer systems. There is a danger of what several contributors to *Software Engineering Notes* have called the 'Titanic effect': the safer a system is believed to be, the more catastrophic the accidents to which it is subject.

Self-negating prophecies

Although this article has focused on the risks of computerisation, it is of course necessary to bear in mind the latter's very considerable benefits. The application of computer systems clearly offers considerable economic advantages. In some applications, it may also be beneficial environmentally, for example in reducing aircraft fuel consumption. There are, furthermore, already examples of programmable electronic systems whose safety records, in extensive practical use, are impressive (see, for instance, Rushby, 1993).

In many contexts, computer use can enhance human safety, for example in automating the most dangerous parts of industrial processes or in providing warning of potentially dangerous situations. Wisely used, relatively simple forms of automation, such as ground-proximity warning systems on aircraft, can potentially save many lives: the dominant form of death in scheduled air travel is now the CFIT (controlled flight into terrain) accident, when a technically unimpaired aircraft nevertheless crashes (Nordwall, 1993).[10]

There is thus every reason for optimism: with good research, careful regulation and intelligent application, the computer's risk-benefit account can be kept positive. However, it is also worth noting that the relatively modest number so far of computer-related accidental deaths, particularly the small number caused by software error, is in one sense puzzling. For, while computer systems appear empirically to be reasonably safe, there are, as noted above, grounds for regarding them as inherently dangerous:

> "A few years ago, David Benson, Professor of Computer Science at Washington State University, issued a challenge by way of several electronic bulletin board systems. He asked for an example of a real-time system that functioned adequately when used for the first time by people other than its developers for a purpose other than testing. Only one

> candidate for this honor was proposed, but even that candidate was controversial ... As a rule software systems do not work well until they have been used, and have failed repeatedly, in real applications." (Parnas *et al*, 1990, page 636).

The reason for this apparent paradox (an error-ridden technology that nevertheless has a reasonably good safety record in practice) is almost certainly conservatism in design: "restraint ... in introducing [computers] into safety-critical control loops" (Leveson, 1992, page 1), and "defence-in-depth" — hardware interlocks, back-up systems, and containment devices which reduce the impact of computer failure.

If this is correct, we have an interesting case of a self-negating prophecy. I have already noted one side of this: that to the extent that we (operators and users) start to believe the computer to be safe (completely reliable, utterly trustworthy in its output, and so on), we may make it dangerous. Here is the prophecy's other side: that up until now we (in this case, system designers rather than users) have generally believed the computer to be dangerous, and therefore have fashioned systems so that it is in practice relatively safe.

Those who work in this field, therefore, have a narrow path to tread. They must do the necessary research to make computer systems safer, and ensure that the results of this research are well implemented, bearing in mind that much of the problem is not technical, but cognitive and organisational. At the same time, they must do nothing to encourage complacency or over-confidence in the safety of computer systems. To make computer systems safer, while simultaneously keeping alive the belief that they are dangerous: that is the paradoxical challenge faced by the field of computer system safety.

Notes

1. Floating-point representation is the analogue of the normal 'scientific' representation of, for example, 1,245,000, for example, as 1.245 times 10^6. There is a fascinating 'politics' of binary floating-point representation (MacKenzie, 1993), which, however, need not detain us here.
2. The main uncertainty in this case is whether a successful interception would have taken place had defensive missiles been launched: the US Army claims only a 70% success rate in interceptions of Scud attacks on Saudi Arabia, and critics (such as Postol, 1991-92) have alleged that the true success rate was substantially lower than that. However, a near miss that did not destroy or disable the Scud warhead might nevertheless have deflected it.
3. There were two other A320 crashes prior to the end of 1992: at Habsheim in Alsace in 1988 (3 deaths) and near Bangalore in India in 1990 (92 deaths). Their interpretation, especially that of the former, has been a matter of dispute (Mellor (1994) argues that both should be seen as computer-related, primarily in an 'interaction' sense). In particular, it was suggested in the report of the official enquiry into the Bangalore crash that the aircraft's pilots may have had undue confidence in the capacity of an automated protection facility ('alpha-floor', triggered by the angle between the aircraft's pitch axis and the air flow) rapidly to bring their aircraft back to a safe condition. Generally, though, the case for the computer-relatedness of these accidents seems to me to be weaker than in the Strasbourg crash, and therefore only the latter is included in the data set.
4. The official investigation of the sole reported US robot-related fatality speculates about a further possible factor: that workers may perceive robots as "something more than machines" and may "personalise" them. Thus the worker in this case had nicknamed his robot 'Robby'. They suggest that "this personalisation may cause the worker's mind to be more focused on the 'teamwork' with the robot rather than upon relative safety issues" (Sanderson *et al*, 1986, page 20).
5. Although comparing like with like is clearly difficult. Computers are now so central to modern long-range civil aviation that the comparison would in effect be of two different epochs in the history of flight.
6. Major injury includes, for example, "amputation of a joint of a finger, fracture of any bone in the skull, spine, neck, arm or leg (but not in the hand or foot) and a penetrating or burn injury to an eye". Minor injury is one which is not (in this sense) major but which causes "a person to be incapable of doing his normal work for more than 3 consecutive days" (Edwards, no date, page 3).
7. The overall ratio of non-fatal serious injuries in air travel to fatalities seems typically to be less than one (Oster *et al*, 1992, page 23).
8. For example, the North Staffordshire radiation therapy incident involved an incorrect mental model of a computerised system. In the Therac-25 case an error hidden amongst the logical complexity of even only modestly large software manifested itself not in gradual deterioration of performance but in a sudden and fatal switch in mode of operation. In the Dhahran incident a tiny cause (an uncorrected rounding error of a ten-thousandth of a percent) led to system failure.
9. In defence against ballistic missiles, firing is generally automatic because of the extremely limited decision time available.
10. Although the general point about over-confidence applies here too: wise use would involve measures to make sure that pilots do not rely exclusively on ground-proximity warning systems to avoid CFIT accidents!

References

J A Adam (1987), "USS Stark: what really happened?", *IEEE Spectrum*, September, pages 26-29.

V M Altamuro (1983), "Working safely with the iron collar worker", *National Safety News*, July, pages 38-40.

BCS Safety Critical Systems Task Force (1993), *Policy Statement on Safety-Related Computer Systems* (British Computer Society, London, 8 June).

S S Brilliant, J C Knight and N G Leveson (1990), "Analysis of faults in an N-version software experiment", *IEEE Transactions on Software Engineering*, 16, pages 238-247.

Committee on Armed Services (1987), *Report on the Staff Investigation into the Iraqi Attack on the USS Stark* (Committee on Armed Services, House of Representatives, Washington, DC).

W J Crowe (1988), endorsement added to Fogarty (1988).

J T Dennett (1979), "When toasters sing and brakes fail", *Science 80*, 1(1), November/December, page 84.

B S Dhillon (1991), *Robot Reliability and Safety* (Springer, New York).

R Edwards (no date), "Accidents on computer controlled manufacturing plant and automated systems in Great Britain 1987-91", study for Commission of the European Community, DG5, typescript, Health and Safety Executive, Birmingham.

E Emery (1989), "Child's death spurs safety inquiries", *Colorado Springs Gazette Telegraph*, 11 January, pages A1 and A3.

W M Fogarty (1988), *Formal Investigation into the Circumstances Surrounding the Downing of Iran Air Flight 655 on 3 July 1988* (Office of the Chairman, Joint Chiefs of Staff, Washington, DC).

T Forester and P Morrison (1990), "Computer unreliability and social vulnerability", *Futures*, June, pages 462-474.

J G Fuller (1984), "Death by robot", *Omni*, March, pages 45-46 and 97-102.

General Accounting Office (1992), *Patriot Missile Defense: Software Problem Led to System Failure at Dhahran, Saudi Arabia*, Report GAO/IMTEC-92-26 (General Accounting Office, Washington, DC).

N G Leveson (1991), "Software safety in embedded computer systems", *Communications of the Association for Computing Machinery*, 34(2), pages 34-46.

N G Leveson (1992), "High-pressure steam engines and computer software", talk to International Conference on Software Engineering, Melbourne.

N G Leveson and C S Turner (1992), *An Investigation of the Therac-25 Accidents*, Technical Report 92-108 (Information and Computer Science Department, University of California at Irvine). Also published in *Computer*, 26(1993), pages 18-41.

London Ambulance Service (1993), *Report of the Inquiry into the London Ambulance Service* (South West Thames Regional Health Authority, London).

D MacKenzie (1993), "Negotiating arithmetic, constructing proof: the sociology of mathematics and information technology", *Social Studies of Science*, 23, pages 37-65.

P T Mahon (1981), *Report of the Royal Commission to Inquire into the Crash on Mount Erebus, Antartica* (Hasselberg, Wellington, New Zealand).

P Mellor (1994), "CAD: computer aided disaster!", typescript, Centre for Software Reliability, City University, London.

C Milhill (1993), "Hospital error killed dozens of patients", *The Guardian*, 30 September, page 3.

M Nagamachi (1988), "Ten fatal accidents due to robots in Japan", in W Karwoski, H R Parsaei and M R Wilhelm (editors), *Ergonomics of Hybrid Automated Systems I* (Amsterdam, Elsevier) pages 391-396.

P Neumann (1988), "Letter from the editor: are risks in computer systems different from those in other technologies", *Software Engineering Notes*, 13(2), pages 2-4.

B D Nordwall (1993), "GPWS to improve regional safety", *Aviation Week and Space Technology*, 26 April, pages 53-54.

D A Norman (1990), "Commentary: human error and the design of computer systems", *Communications of the Association for Computing Machinery*, 33, pages 4-7.

North Staffordshire Health Authority (1993), *Report of the Independent Clinical Assessment commissioned by the North Staffordshire Health Authority on the Effects of the Radiation Incident at the North Staffordshire Royal Infirmary between 1982 and 1991* (UK).

C V Oster Jr, J S Strong and C K Zorn (1992), *Why Airplanes Crash: Aviation Safety in a Changing World* (Oxford University Press, New York).

D L Parnas, A J van Schouwen and S P Kwan (1990), "Evaluation of safety-critical software", *Communications of the Association for Computing Machinery*, 33(6), pages 636-648.

E Peláez (1988), *A Gift from Pandora's Box: the Software Crisis*, PhD thesis (University of Edinburgh).

C Perrow (1984), *Normal Accidents: Living with High-Risk Technologies* (Basic Books, New York).

T A Postol (1991-92), "Lessons of the Gulf War experience with Patriot", *International Security*, 16(3), pages 119-171.

B Randell (1989), "Technology doesn't have to be bad", *Software Engineering Notes*, 14(6), page 21.

G I Rochlin (1991), "Iran Air Flight 655 and the USS Vincennes: complex, large-scale military systems and the failure of control", in T R La Porte (editor), *Social Responses to Large Technical Systems: Control or Anticipation* (Kluwer, Dordrecht) pages 99-125.

J Rushby (1993), *Formal Methods and the Certification of Critical Systems*, Computer Science Laboratory Technical Report 93-07 (SRI International, Menlo Park, California)

L M Sanderson, J W Collins and J D McGlothlin (1986), "Robot-related fatality involving a U.S. manufacturing plant employee: case report and recommendations", *Journal of Occupational Accidents*, 8, pages 13-23.

G Sharp (1987), *Formal Investigation into the Circumstances surrounding the Attack on the USS Stark (FFG31) on 17 May 1987* (Commander, Cruiser-Destroyer Group Two, Miami, Florida).

R Skeel (1992), "Roundoff error and the Patriot missile", *SIAM* [Society for Industrial and Applied Mathematics] *News*, 25(4), July, page 11.

R Smithers (1993), "Road deaths at new low", *The Guardian*, 26 March, page 3.

P Sparaco (1994), "Human factors cited in French A320 crash", *Aviation Week and Space Technology*, 3 January, pages 30-31.

M Thomas (1988), "Should we trust computers?", British Computer Society/Unisys Annual Lecture, Royal Society of London, 4 July.

J P Vautrin and D Dei-Svaldi (1989), "Accidents du travail sur sites automatisés: Évaluation d'une Prévention Technique",*Cahiers de Notes Documentaires*, 136, pages 445-453.

M Vlahos (1988), "The Stark Report", *Proceedings of the US Naval Institute*, May, pages 64-67.

West Midlands Regional Health Authority, (1992), *Report of the Independent Inquiry into the Conduct of Isocentric Radiotherapy at the North Staffordshire Royal Infirmary between 1982 and 1991* (UK).