What can the Post Office 'Horizon' scandal teach about Artificial Intelligence deployment?

Chris Baber[1], Brandon King[2], Paul Salmon[2] and Yihao Jiang[1]

[1]University of Birmingham, UK, [2]University of Sunshine Coast, Australia

SUMMARY

In this paper we use the Post Office Horizon scandal. Broadly, accounting errors arising from a computer system were falsely attributed to Subpostmasters who were subsequently accused of theft, fraud, and false accounting. Over 700 prosecutions have been successfully appealed and convictions quashed. Coupled with severe technical shortcomings, there was confirmation bias in the decisions of Post Office Limited to prosecute Subpostmasters and an assumption in UK Law that computers do not make mistakes (unless proven otherwise). From a review of evidence and newspaper reports of specific cases in the Post Office Horizon scandal, we construct and analyse Accimaps. We argue that a common problem across these cases is how different sorts of 'black box' of the Horizon system meant that it lacked transparency for all stakeholders.

KEYWORDS

Post Office Horizon; Accimaps; Socio-Technical Systems; Transparency

Introduction

In the UK, a public enquiry into problems arising from the 'Horizon' computer system has been widely reported. These problems were brought to the public's attention via a BBC television programme 'Mr. Bates vs the Post Office' in 2023. In the UK, all Post Office branches are franchised by Sub post-masters (SPM) from Post Office Limited (POL). Most SPMs also run other businesses from their Post Office, e.g., a general store or newsagents etc. This means that Post Office branches might have two sets of tills and two sets of accounts, only one of which is returnable to POL. Further, the franchise agreement states that SPMs are responsible for any shortfalls in returns to POL. This means that if the return is incorrect, the SPM will need to make up the missing money (there does not seem to be a requirement to handle over-payment).

In the Horizon scandal, over 730 SPMs were prosecuted by POL between 2000 and 2014. Many of these were accused of theft, which could lead to a prison sentence, and were offered a plea deal of 'false accounting', a lesser charge, and repay the 'missing' money. Most accepted this deal. It is worth noting that POL does not have any special powers under UK law, but (like any other citizen) can bring its own private prosecutions. In contrast to prosecutions brought by the Police, private prosecutions do not require scrutiny by the Crown Prosecution Service[1]. Where POL differed from other citizens it the resources that it could apply to this process and the fact that it employs its own investigators. External auditing of accounts has not received much attention to date. In many instances, SPMs agreed to settle out of Court and pay the outstanding balance (and in the cases that have been reported, SPMs had been paying smaller shortfalls from their own money prior to being brought to Court).

[1] https://www.gov.uk/government/consultations/oversight-and-regulation-of-private-prosecutors-in-the-criminal-justice-system/oversight-and-regulation-of-private-prosecutors-in-the-criminal-justice-system-consultation

When computers were first considered in relation to crime (in the late 1970s) computer evidence was treated as 'hearsay' and, hence, inadmissible. In section 69 of the 1984 Police and Criminal Evidence Act (PACE), there was a requirement for the prosecution to prove that the computer was working correctly before evidence could be admitted. Checking all computers related to a crime was burdensome, so section 69 of PACE was repealed (under section 60 of the Youth and Criminal Evidence Act in 2000). Thus, the presumption in English law is that a computer is 'working properly'. It is the burden of the person challenging evidence to prove that the computer was not working properly. In most of the *Horizon* cases, Courts assumed that the computer was working properly, and the Defence was not given access to documents from POL or Fujitsu that could allow this assumption to be challenged, and (except in a couple of cases) the Court would not accept a claim that the computer had malfunctioned.

In his instructions to the jury in the case of Seema Misra in 2010, the judge asked, "Do you accept the prosecution case that there is ample evidence before you to establish that Horizon is a tried and tested system in use at thousands of post offices for several years, fundamentally robust and reliable?" Misra pleaded guilty to six counts of false accounting and the jury found her guilty of stealing £74,000 (Brooks and Wallis, n.d.). This verdict was given despite the judge's observation that, "There is no direct evidence of her taking any money...There is no evidence of her accumulating cash anywhere else or spending large sums of money or paying off debts, no evidence about her bank accounts at all. Nothing incriminating was found when her home was searched."

From this brief introduction, it should be apparent that we are considering a Sociotechnical System in which technology (*Horizon*) is embedded in an organizational culture (in POL) that was, at best, suspicious of the accounting of SPMs. The Horizon scandal was compounded by a legal system that presumes computer technology is reliable and by denials by POL that there were problems with *Horizon*. As early as 2003, in the trial of Julie Wolstenholme, the judge accepted expert witness testimony that *Horizon* was faulty. In response, POL stated "It is denied that the said computer system was unfit for its purpose and it is averred that the same worked adequately."[2]

Perrow (1990) suggested that any reasonably complex system is always in a near accident state. Such states are kept in check because safeguards are in place to limit potential risks, or because people can monitor and intervene to control the system, or because regulations exist to penalize the outcomes of critical states. Sometimes all three fail and the Horizon scandal is an example of this. In terms of system failures, one might expect the 'system' under consideration to be sufficiently transparent for stakeholders to understand and manage (Woods, 1996). An initial step in an Ergonomics analysis of the Horizon scandal would be to ask (a) what defined the boundaries of the system under consideration, and (b) was the system transparent to its operators and stakeholders.

In this paper, we consider the boundaries of the system through the lens of Accimaps, and transparency in terms of 'black boxes'. In computing, a 'black box' is a process that is opaque to its users, typically because of its complexity, and the user will be expected to trust the output of this process. The black box problem is increasingly pertinent with the application of Artificial Intelligence, AI (von Eschenbach, 2021) and we believe that the problems relating to a non-AI computers in a Sociotechnical System can illustrate the challenges that AI might pose.

A brief overview of the UK Post Office Horizon Computer System

International Computers Limited (ICL) was awarded a contract from UK Government in 1996 for a computer system to support a Benefits Payment Card for payment of unemployment and other

[2] https://www.computerweekly.com/news/366546032/Post-Office-tried-to-convince-independent-IT-witness-that-he-was-wrong-about-Horizon [accessed 12th February 2025]

benefits by Post Office branches. This project, *Pathway*, was discontinued due to "greater than expected complexity", with the House of Commons public accounts committee calling it "one of the biggest IT failures in the public sector". However, in 1999 ICL signed a contract to automate Post Office operations by the Pathway team and this resulted in *Horizon*. Fujitsu had bought 80% of shares in ICL by 1990 and by 1998 was the sole shareholder. *Horizon* proved to be a lucrative contract for Fujitsu. It is estimated that Fujitsu has received around £2.5 billion[3] for Horizon operations. In addition to a fixed payment for delivering the system (and upgrading it in 2021, 2023 and 2024), Fujitsu receives payment for each of the 2 million+ transactions processed each day.

Horizon integrated systems and services from a variety of providers, e.g., Oracle, Escher, ICL, ATOS, Computacenter, using a variety of programming languages, e.g., VisualBasic, C, and C++, and databases built in Oracle. Thus, *Horizon* was an amalgamation of sub-systems, combined using a variety of languages and built on the already discredited *Pathway* system. One might anticipate that, without appropriate testing, integration would be problematic. To make matters worse, the development team for the project was small and not familiar with the nuances of the various systems. A member of the development team, David McDonnell told the 2024 public inquiry "of eight [people] in the development team, two were very good, another two were mediocre but we could work with them, and then there were probably three or four who just weren't up to it and weren't capable of producing professional code".[4]

For any database transaction, one can apply four criteria: Atomicity (statements are treated as single units to be executed); Consistency (changes to the database must be predictable); Isolation (transactions from different users of the database need to be kept separate); Durability (changes to data will be saved, even if the system fails). The examples considered below demonstrate failures of atomicity (because the transactions were duplicated); consistency (the examples suggest an inability to synchronize the various parts of the system, which led to mismatches); isolation (the examples suggest that operations had the potential to interfere with each other); durability (some entries (and errors) could not be un-done). A detailed discussion of these issues can be found online.[5] As noted previously, *Horizon* was an amalgamation of systems that were integrated together and combined with the poorly built *Pathway* system. A key issue with *Horizon* was how it managed data. While there was an Oracle database, much of the data management used an XML structure that wrote messages to a message store and there was no agreed catalogue or dictionary of the message, so no attempt to ensure their consistency (Wallis, 2021).

Horizon included an electronic point of sales (EPOS) terminal and accounting transaction management system for Data Reconciliation between the Post Office branch and POL baock-office computers. Connection between branches and back-office was via a telephone line (until the introduction of Horizon Online). This would inevitably cause delays in transmission of data, particularly in branches that were in remote locations. As an accounting system, *Horizon* ran according to double entry book-keeping. When the SPM was recording a transaction, this involved two double entry baskets: Transfer Out from one Stock Unit and Transfer In to another Stock Unit. When a Transfer was made, then an indicator of outstanding transfers should be updated. But, this indicator was not displayed on all EPOS terminals so the SPM would not know if the Transfer had been made, so might attempt to repeat the Transfer. Balancing of the accounts assumed that there were no outstanding transfers (that is, it requires Transfer Out to equal Transfer In). However, if (as happened in the Callendar Square 'bug', mentioned below) there were 2 Transfers In for a single

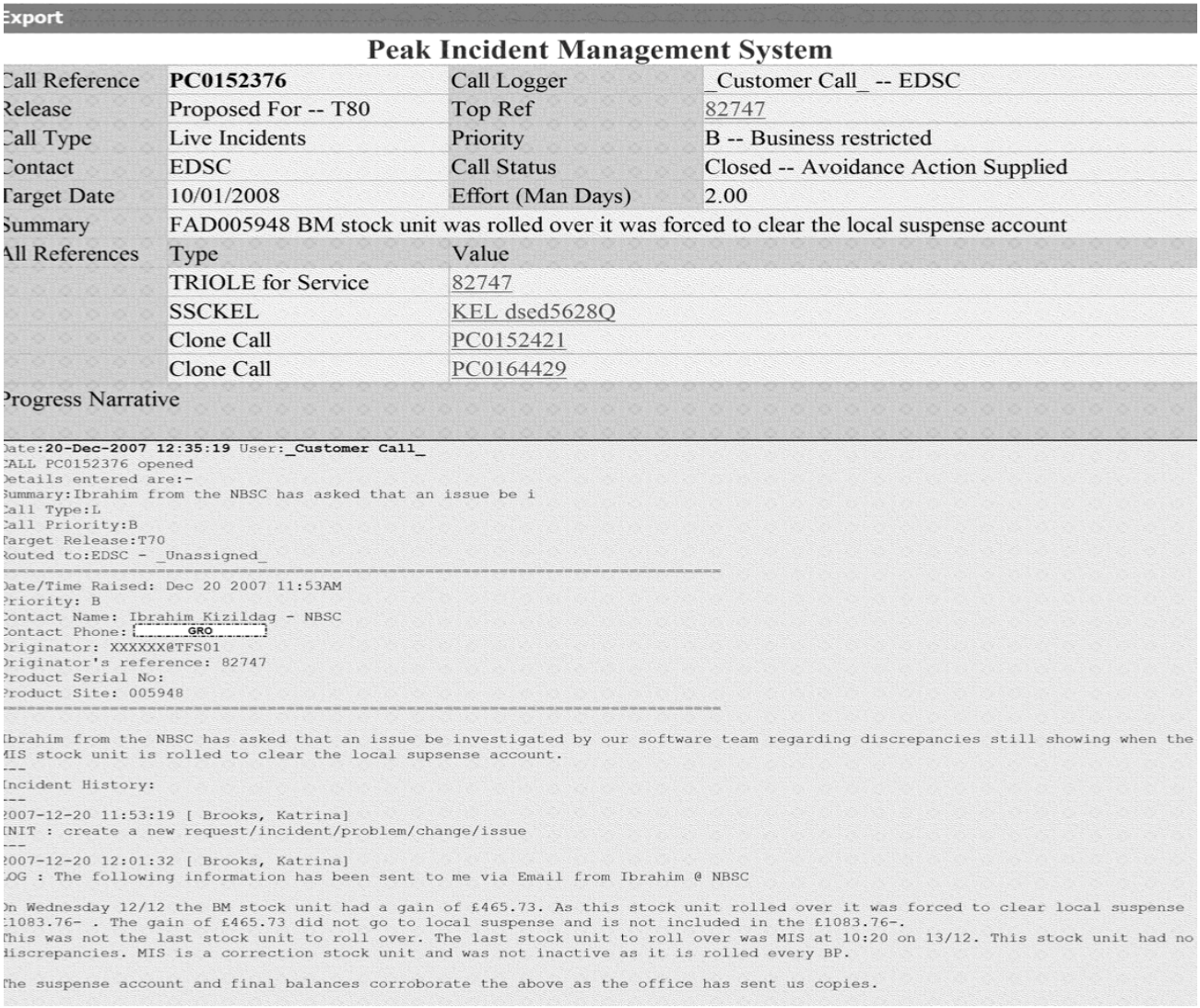
[3] https://www.computerweekly.com/news/366586814/Post-Office-Horizon-replacement-project-labelled-unachievable-as-taxpayer-bill-reaches-1bn

[4] https://www.theguardian.com/uk-news/2024/jan/09/how-the-post-offices-horizon-system-failed-a-technical-breakdown [accessed 31st January 2025]

[5] https://evidencecritical.systems/2021/07/15/what-went-wrong-with-horizon.html [accessed 30th January 2025]

Transfer Out, then this affects the cash held in the branch and is defined as a loss. According to Wallis (2021) "The cash account wasn't an account in the traditional sense of the word. It was a program which crawled through every transaction on each Horizon terminal in each branch at the end of the day's trading. It then came up with a figure which should correspond exactly with the amount of physical cash on the premises. That figure was then automatically uploaded to the Post Office's central servers overnight. It is a relatively simple task to describe, but not necessarily to execute. Given its central importance to the financial integrity of the Horizon system, it had to be bullet-proof. It wasn't. The code was not good enough." (p.12)

From the early deployment, the development team maintained a log of 'bugs' (although a full list of these has not been disclosed). Fujitsu engineers maintained a Known Error Log (KEL) and the PEAK incident management system. An example of a PEAK record is shown in figure 1.



Export

Peak Incident Management System

Call Reference	**PC0152376**	Call Logger	_Customer Call_ -- EDSC
Release	Proposed For -- T80	Top Ref	82747
Call Type	Live Incidents	Priority	B -- Business restricted
Contact	EDSC	Call Status	Closed -- Avoidance Action Supplied
Target Date	10/01/2008	Effort (Man Days)	2.00
Summary	FAD005948 BM stock unit was rolled over it was forced to clear the local suspense account		
All References	Type	Value	
	TRIOLE for Service	82747	
	SSCKEL	KEL dsed5628Q	
	Clone Call	PC0152421	
	Clone Call	PC0164429	

Progress Narrative

Date:**20-Dec-2007 12:35:19** User:**_Customer Call_**
CALL PC0152376 opened
Details entered are:-
Summary:Ibrahim from the NBSC has asked that an issue be i
Call Type:L
Call Priority:B
Target Release:T70
Routed to:EDSC - _Unassigned_

Date/Time Raised: Dec 20 2007 11:53AM
Priority: B
Contact Name: Ibrahim Kizildag - NBSC
Contact Phone: [redacted]
Originator: XXXXXX@TFS01
Originator's reference: 82747
Product Serial No:
Product Site: 005948

Ibrahim from the NBSC has asked that an issue be investigated by our software team regarding discrepancies still showing when the MIS stock unit is rolled to clear the local supsense account.

Incident History:

2007-12-20 11:53:19 [Brooks, Katrina]
INIT : create a new request/incident/problem/change/issue

2007-12-20 12:01:32 [Brooks, Katrina]
LOG : The following information has been sent to me via Email from Ibrahim @ NBSC

On Wednesday 12/12 the BM stock unit had a gain of £465.73. As this stock unit rolled over it was forced to clear local suspense £1083.76- . The gain of £465.73 did not go to local suspense and is not included in the £1083.76-.
This was not the last stock unit to roll over. The last stock unit to roll over was MIS at 10:20 on 13/12. This stock unit had no discrepancies. MIS is a correction stock unit and was not inactive as it is rolled every BP.

The suspense account and final balances corroborate the above as the office has sent us copies.

Figure 1: extract from PEAK PCO152376

Detica[6] identified four areas of risk (in terms of operational practice) in POL operations: "…non-conformance to Post Office policy and processes by branches, with an institutionalised acceptance that errors, workarounds and non-conformance exists; Complexity and fragmentation of information systems which hamper efforts both to gain an insight into branch behaviour and root

[6] https://www.jfsa.org.uk/uploads/5/4/3/1/54312921/document_25_-_detica_netreveal_fraud_analysis_011013_1.pdf [accessed 6th February 2025]

causes; Ineffective process, policy and working practice in the central operational teams to gather information, prioritise and act in a co-ordinated manner; Technology available to central operational teams are not fit for purpose; analysis of large data sets is performed on an ad-hoc basis of data subsets copied into Excel and tasking of teams is initiated and managed through email."

The first point is of interest to this paper because it highlights the complexities in defining a single root cause of the 'technical' issues relating to *Horizon.* The Detica report provides the example of selling non-Post Office products through a Post Office till. As noted previously, many Post Office branches are run in conjunction with other shops. While the policy might be that only Post Office products (stamps, postage etc.) will be put through the *Horizon* EPOS terminal, it was common for SPMs to also record transactions of items from their other shop (e.g., greetings cards) through the EPOS and to record these as 'Postage other'. This transaction would then be reversed at the end of the day (or the accounting period). As the Detica report notes, "Instances of non-conformance also generate operational noise which hides deliberate attempts to defraud it. This has resulted in a large number of false positives when looking for fraud, and inhibits Post Office's ability to detect fraud early, resulting in larger losses." Thus, dealing with 'deliberate fraud' was a challenge for POL. Given the prevalence of multiple businesses being from Post Office branches, POL might have assumed that SPMs were using money to cross-subsidise these other businesses in Post Office branches. This could have led to the belief that the scale of this 'theft' was revealed by *Horizon.* Consequently, there may have been a degree of confirmation bias behind POL prosecutions of SPMs.

Technical Failings of Horizon

Descriptions of some technical failures are provided in the Technical Appendix to 'Alan Bates and others and the Post Office Limited'. POL argued that there were no technical problems with *Horizon.* At first, 'bugs' were named after the Post Office branch in which they were reported. This naming convention suggests that any problem is due to the branch rather than the computer system. This might also reflect Fujitsu's assumption that the 'bugs' were due to human error rather than technical failings. As Graham Ward, Fujitsu security case manager, commented in an email to staff at Fujitsu about a report on Noel Thomas (who was jailed for £48k alleged false accounting) "Given the allegations made by the postmasters, I'm sure you'll agree that it's very much in ourselves and Fujitsu's interest to challenge the allegations and provide evidence that the system is not to blame for the losses provided."[7]

Callendar Square and Dalmellington bugs

The Dalmellington bug occurred when a user entered a transaction of £8000 on the EPOS terminal. The screen froze, so the user hit the 'enter' key again – duplicating the entry. With second and third presses of the enter key, the transaction accumulated to £32,000. As noted above, the SPM was liable for the discrepancy of £24,000, which was paid from personal funds. The incident was raised in PEAK PCO126042 (15/09/05) which also noted an instance recorded in 2000. Related issues were noted in PEAKS PC0056922, PC0075 92, PCO083101, PCO193012, PCO103864. KEL JSimpkins338Q, May 2002 seems similar and describes the root cause as a Riposte error "Timeout occurred waiting for lock (0xC1090003)". Riposte was supplied by Escher and, in 2006, the problem was fixed by Escher, although there is a Peak (PC0193012 - referenced by JSimpkins338Q - which was raised in 2010 and again a reboot fixed the issue, and these PEAKs closed in 2010.

[7] https://www.computerweekly.com/news/366589716/Metropolitan-Police-set-to-investigate-one-of-its-own-staff-in-Post-Office-probe#:~:text=In%20the%20email%20to%20Fujitsu,his%20edit%20of%20the%20statement [accessed 2nd February 2025]

Callendar Square is reported in the following PEAKS: Peak PCO126042 raised 15/09/2005; Peak PCO126376 raised 21/09/2005; PC0056922; PC0075 92; PCO083101; PCO193012. It is reported in KEL JBallantyne5245K, KEL JSimpkins338Q (from 2002). The root cause was a time-out bug in Riposte that prevented EPOS terminal from writing messages.

Accimaps

We constructed Accimaps of individual cases using source materials from the public enquiry and court cases, supplementing these with newspaper and magazine articles, journal papers, and opinion pieces of websites. Originally developed by Rasmussen (1997), Accimaps visualize the interplay of factors that contribute to an incident at levels of Socio-Technical System, i.e., wider society, legal and regulatory, organizational, technical, individual. These are a popular approach to accident analysis (Salmon et al., 2023). We are interested in how technical failings of *Horizon* (which are not always well documented) contribute to the scandal.

Accimap of a known technical failure: Girobank reconciling

In PEAK PC0044232 (KEL_MWright531p) there was a £505.72 discrepancy. There seem to be five complicating factors: (i.) this involved a giro (which is like a cheque, in this instance one issued by the UK benefits agency) which is cashed by the SPM for a customer after the last collection of giros from the branch (every day Girobank would send a van around the post offices to collect any giros that had been cashed, take these back to Girobank to process, and then Girobank would pay the post office back the money that had been paid out for each giro); (ii.) because the SPM knew that the last collection had been missed, he cancelled the giro transaction after entering it in *Horizon*, intending it to be recorded the following day after the giro had been collected; (iii.) Girobank receive the giro but also notification from Horizon that the transaction had been cancelled, so their system issues an error notice; (iv.) the SPM should (according to company policy) reject the error notice so that the amount of cash can be recorded in Horizon correctly; (v.), the SPM did not know this until the end of the Trading Report period (which was every week - not every day).

Figure 2 illustrates how the Horizon system (in addition to technical problems) operated in a wider technical system in which the timing of events and interconnection of different processes created a complicated set of relations. In order to deal with these complications, heuristics were applied and these included counter-intuitive tasks, such as cancelling and then correcting the Girobank error notice, which needed to be performed at a different time to the original transaction (in this case, several days after the transaction).

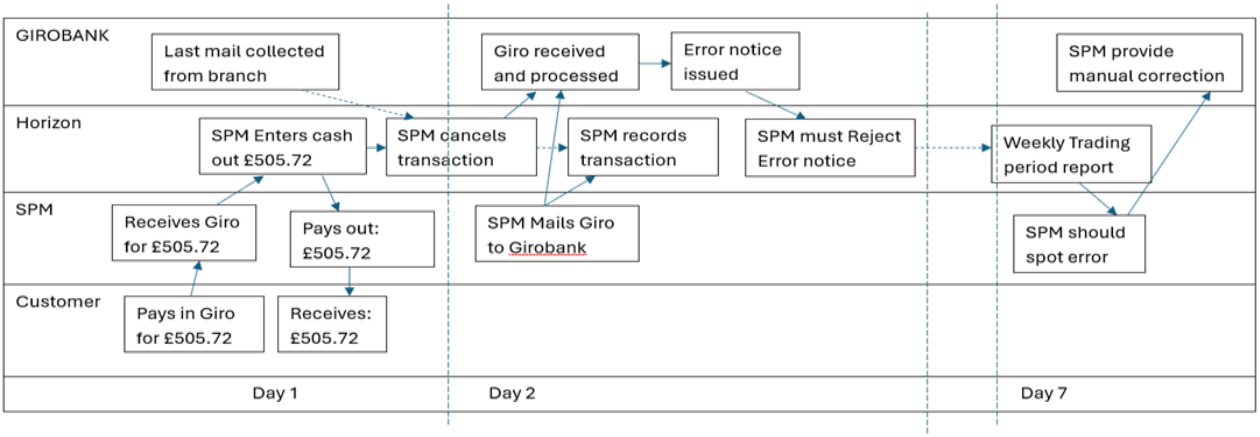


Figure 2: Accimap of Girobank discrepancy issue

Accimap of wider systemic failure: The case of Jo Hamilton

Jo Hamilton was featured in the BBC 'Mr. Bates vs the Post Office' and her case is sufficiently well reported to allow us to construct the Accimap (a section of this is shown in figure 3). She became SPM in 2003. *Horizon* showed shortfalls in the accounts. Assuming that these were the result of problems that arose from her mistakes in using the system, she paid the shortfalls from her own money. From 2003 to 2006 the errors accumulated and she re-mortgaged her house to continue making payments. In 2006, the discrepancy was over £10,000 and she was suspended. When Hamilton was suspended, she was not allowed back into the Post Office branch. This meant that she and her defence struggled could not access the branch computer. It is plausible to assume that, even if such access was granted, the records of transactions might not have been stored in a reliable manner. Consequently, the prosecution case relied on the assumption of a correctly working computer. In 2008 she was taken to Court and charged with the theft of £36,644.89. To avoid a prison sentence, she agreed to a plea bargain in which she pleaded guilty to 14 counts of false accounting. She had to remortgage her house again to pay the outstanding amount plus £1000 towards prosecution costs. She was also sentenced to a community order.

The initial problems in Jo Hamilton's case appear similar to those of the Dalmellington 'bug' (although there is a lack of technical detail both in Hamilton's accounts of the incidents and the Post Office's prosecution statements). Hamilton sought assistance (through hundreds of calls) from the help-desks offered by Post Office Limited. There were at least 3 levels over which the help desks operated:

Level 1: HSH is Fujitsu Horizon Helpdesk – call handlers reply to SPM questions using POL operating procedures and gives advice from scripts. So, these are unlikely to help in solving a problem. There was also a National Business Support Centre which was POL helpdesk which responds to security problems, complaints about Horizon, other operating problems.

Level 2: SMC is Fujitsu Horizon unit that monitor the 'event storms' from counter (which I believe are the Electronic Point of Sales (EPOS) terminals in branches). These can access Known Error Logs (KELs) to find previous fixes. There was also (in the earliest deployment) the ICL Pathway System Support Centre.

Level 3: Fujitsu System Support Centre dealt with 'new' problems and logged these in a system called PINICL (which was later replaced by a system called PEAK).

The likely flow would be for the SPM to call a level 1 helpdesk and receive a standard answer, such as turn the system off and on again which appeared to solve some of the issues associated with bugs such as Dalmellington. This help was provided from a script and would not directly address specific problems. If the problem was not solved, then this might escalate to level 2 where the solution to the problem could be retrieved from a KEL. If this did not work then the problem might escalate to level 3 where, if there was not a KEL, would be recorded as a PEAK.

Conclusions

We propose lessons that can be learned from *Horizon* are applicable to the deployment of Artificial Intelligence (despite *Horizon* using poorly designed technology). The lessons relate to Sociotechnical systems and can be understood in terms 'black boxes' at different levels.

- For sub-postmasters *Horizon* was opaque and produced confusing and inconsistent output. This was compounded by poor user interface design, limited training, and unpredictable software. This resulted in a lack of transparency that made it impossible for the SPM to determine the source of a problem in the computer. In a few instances, the SPM kept a parallel record of transactions which could be used to challenge the *Horizon* output.

- In English Law, the assumption is that any computer is working correctly at the time of any incident, unless there is evidence to the contrary. This presumption means that the burden is on the defence to prove failure of a computer. Lack of access to computer logs for defence lawyers meant that *Horizon* was a black box for them. The manually kept records (noted above) were seldom sufficient to challenge the computer records in Court.

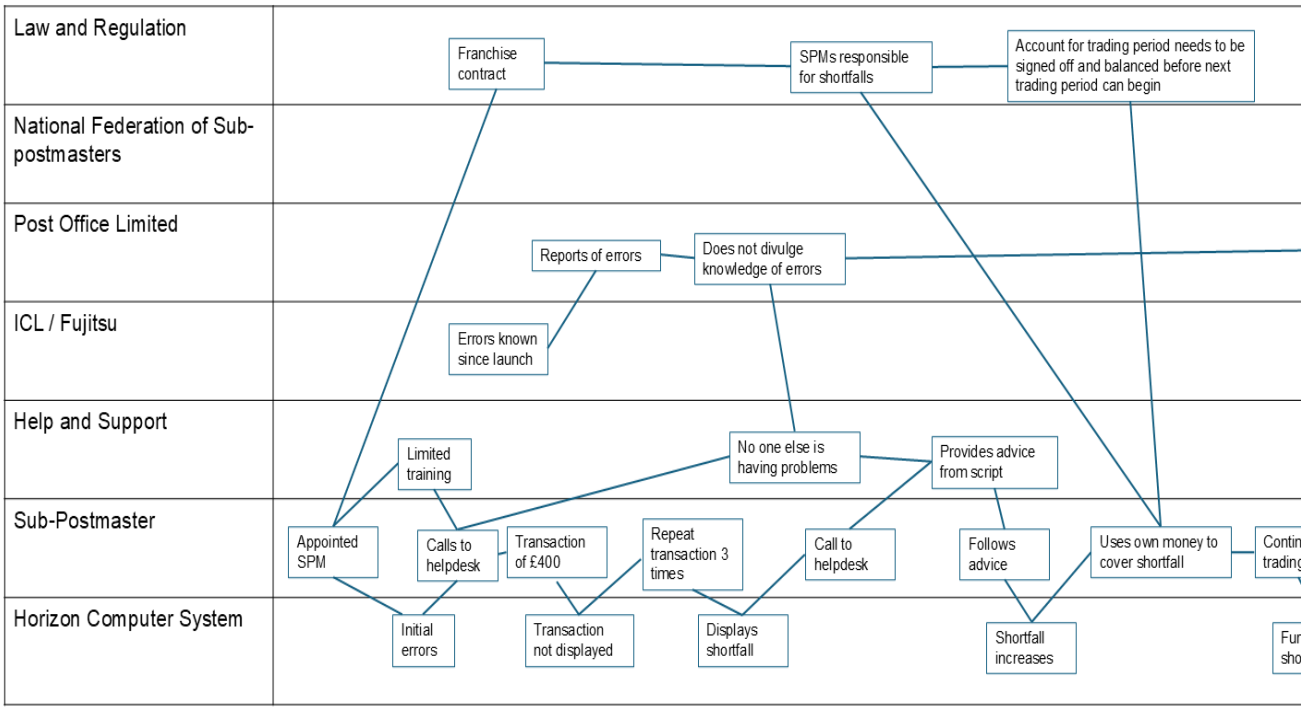


Figure 3: Extract of Accimap summarising Jo Hamilton's incidents

- For system developers a combination of limited competence, integration of incomplete systems, and limited oversight of failures, meant that *Horizon* was a black box for the team who developed and deployed it. Fujitsu ran a team that attempted to correct errors through remote access to Post Office branch computers although, as late at 2015, POL told a House of Commons inquiry that "There is no functionality in Horizon for either a branch, Post Office or Fujitsu to edit, manipulate or remove transaction data once it has been recorded in a branch's accounts."
- For Post Office Limited there seemed to be a lack of knowledge or interest in the operation of the computer system and how it might not be performing correctly. This suggests a black box perspective of technology taken by management and their advisors. More pernicious is the implication that viewing the *Horizon* system as a black allowed the confirmation bias of assumed guilt of SPMs.

We propose that multi-level, multi-stakeholder black boxes are a critical issue in sociotechnical systems, and that the transparency issues they create play a causal role in adverse events. This work is critical as it points to the need for improved transparency in AI technologies, both in terms of the AI itself, the human-machine interface, and associated procedures, operating rules, and regulations.

References

Brooks, R. and Wallis, N., Justice lost in the post: how the post office wrecked the lives of its own workers, Private Eye, https://www.private-eye.co.uk/special-reports/justice-lost-in-the-post [accessed 30th January 2025]

Woods D. D. (1996). Decomposing automation: Apparent simplicity, real complexity. In Parasuraman R., Mouloua M. (Eds.), *Automation and human performance: Theory and applications* (pp. 3–17). Erlbaum.

von Eschenbach, W.J. Transparency and the Black Box Problem: Why We Do Not Trust AI. *Philos. Technol.* 34, 1607–1622 (2021)

Perrow, C., 1984, *Normal Accidents: living with high-risk technologies,* New York: Basic Books

Rasmussen, J., 1997, Risk management in a dynamic society: a modelling problem, Safety Science, 27, 183-213

Salmon, P.M., Hulme, A., Walker, G.H., Waterson, P., Berber, E. and Stanton, N.A., 2020. The big picture on accident causation: A review, synthesis and meta-analysis of AcciMap studies. *Safety science*, *126*, p.104650.

Wallis, N. (2021) *The Great Post Office Scandal*, Bath Publishing