Open Research Online

The Open University's repository of research publications
and other research outputs

Leading indicators of operational risk on the railway: A novel use for underutilised data recordings

Journal Item

For guidance on citations see FAQs.

Version: Version of Record

oro.open.ac.uk

Leading indicators of operational risk on the railway: A novel use for underutilised data recordings

Guy Walker *, Ailsa Strathie

Institute for Infrastructure and Environment, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom

A B S T R A C T

Flight Data Monitoring (FDM) is the process by which data from on-board recorders, or so-called 'black boxes', is analysed after every journey to detect subtle trends which, if allowed to continue, would lead to an accident. An opportunity has been identified to advance the state of the art in FDM processes by coupling recorder data to established Human Factors methodologies so that issues arising from the strategically important human/machine-system interface can be better understood and diagnosed. The research has also identified a significantly underused source of recorder-data within the railway industry. Taking this data, the paper demonstrates how key areas of driver performance can be quantified using a simple behavioural cluster detection method coupled to sensitivity and response bias metrics. Faced with a class of operational accident that is increasingly human-centred, an underused source of data, and methods that can join it to established human performance concepts, the potential for detecting risks in advance of an accident are significant. This paper sets out to describe and demonstrate this potential.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Data recording

Data recording is the act of automatically logging information on system parameters over time. Data recording has become increasingly ubiquitous in rail transport operations. Entire national train fleets are now required to carry recorders which continuously extract data on how individual trains are being driven, at increasing rates, and across an increasing range of parameters. The outflow of data is extensive and growing yet comparatively neglected. What could it be used for? In this paper we argue it could be used to tackle the most important strategic risk issues currently faced by rail operators and authorities worldwide.

1.2. Brief history

The act of automatically recording data on system parameters over time is referred to as 'data logging' or 'data recording'. In the aviation industry the generic term data logging falls under the specific heading of Flight Data Recording, which itself comprises several individual procedures and devices. The most prominent of these is what is termed colloquially as the 'black box', which represents the combination of a Flight Data Recorder

(FDR) and a Cockpit Voice Recorder (CVR). Other systems under the heading of Flight Data Monitoring include various Aircraft Condition Monitoring Systems (ACMS), such as engine health monitoring (e.g. the Rolls Royce EHM programme) and the wide range of parameters available from modern avionics (e.g. ARINC 573) via so-called 'Quick Access Recorders' (QARs).

Data recording can trace its origins back to the allied fields of metrology, instrumentation, telemetry, predictive maintenance and condition monitoring. The Wright Brother's 1903 'Wright Flyer 1', one of the world's first powered aircraft, was equipped with "*instruments to record air velocity, engine revolutions and time while in the air*" (Ford, 2012) and herein lie the very early antecedents for the sophisticated Flight Data Recording and Monitoring that exist today. The rail sector, however, can lay claim to even earlier and more sophisticated examples of instrumentation. Stephenson's Rocket (1829), for example, had instrumentation for boiler pressure and water level, and in 1838 the Great Western Railway in the UK constructed the first 'dynamometer car', using equipment designed by Charles Babbage to integrate various readings into an accurate representation of train performance.

The use of data logging as a tool in safety science is a post-war development. It evolved amid a wider context that included a marked increase in post-war air travel, the development of new jet airliners, and accidents in which passenger aircraft 'crashed without trace' leaving investigators perplexed as to the root cause. Most notable among these were the De Havilland Comet Crashes of

* Corresponding author.

1953 and 4. During the subsequent investigations it was noted that "*anything which provides a record of flight conditions, pilot reactions, etc. for the few moments preceding the crash is of inestimable value*" (Warren, 1954). The prototype 'Flight Memory Unit' (as the black box was then referred) was manufactured from early magnetic audio recording technologies and a primitive crash survivable enclosure. The device could superimpose signals from some of the aircraft's primary controls onto approximately 30 feet of metal wire at a rate of approximately eight signals per second. The device was configured so that the metal wire looped continuously, storing four hours of voice and data, continually over-writing itself.

In 1958 the UK Air Registration Board became aware of the Flight Memory Unit and due to the national importance of the jet aviation industry and the potential safety barrier that the Comet crashes represented to continued foreign sales, the concept was considered important enough to warrant further development. A British clock making company called S. Davall and Sons were able to acquire production rights and develop the first commercial 'black box', or Davall Type 1050 "Red Egg", as it was then called. Improvements now enabled readings to be captured at a rate of 24 per second, greater accuracy in the data collected from aircraft instruments and controls, and the flexibility to record voice, data or both. To do this, up to 40 miles of stainless steel wire was needed as a recording medium. An unexplained air crash in Queeensland in 1960 led to the mandatory fitment of cockpit voice recorders like these in Australia. Regulations also appeared in the United States as early as 1958, and legislation also followed in 1960 (Morcom, 1970). In Britain, changes were made to the Air Navigation Order as early as 1960 although a lengthy period of consultation and evaluation ensued, meaning that it did not become mandatory to carry a flight data recorder until 1965. With legislation imminent, however, the supply and fitment of recorders was well underway prior to this. Indeed, the first crash investigation to make substantial use of the data provided by an FDR occurred in 1965 when a BEA Vanguard fitted with a Davell Type 1050 'red egg' crashed in poor weather at London's Heathrow airport.

Early data recorders were relatively stand-alone devices. The recorder carried its own sensors and, apart from an electrical supply, operated relatively independently of the host aircraft (Campbell, 2007). Calibration proved to be a problem, with the actual state of the aircraft systems not necessarily being identical to those indicated by sensors in the recorder, or even sometimes the same as those displayed to the pilots on their cockpit displays. This 'system architecture' was to change with the advent of avionics. Avionics is the collective term given to aircraft electrical systems. The Boeing 787 and Airbus A380 represent the current state of the art and an expression of what is sometimes referred to as the 'electronic aircraft'. Here, air, mechanical and hydraulically operated systems are replaced by electrical systems, all of which reside on a communications network that can be interrogated by various aircraft systems, including flight data recorders. Rather than a stand-alone device, data recorders are now part of a comprehensive data acquisition architecture that relies on the integration of data from myriad sources via a Flight Data Acquisition Unit (FDAU), common communications protocols (ARINC 573, 717 and 767), and the use of quick access recorders as well as crash survivable 'black boxes'. Modern flight data recorders are solid state devices with the ability to continuously record over 2000 separate parameters for in excess of 30 days. The separation between a crash survivable data and voice recorder, mandated by law and used for accident investigation, and a Quick Access Recorder (QAR), not mandated but used for operational and safety purposes by airlines and regulators, occurred in the 1970s. It arose from a growing recognition that easy access to flight data, both routine and abnormal, was of value.

While the aviation sector has a long history with on-vehicle data recording devices for the purposes of safety and crash investigation, these are a much more recent innovation in the rail sector. Experience in the UK is quite typical. Here, fitment has only been mandatory since 2002 but has been the subject of discussion within the industry for many years (Uff, 2000), indeed, a Railway Group Standard (GO/OTS203) was issued in 1993 in recognition of the fact that the technology existed and was beginning to be fitted in isolated cases. The situation the industry faced was one in which costs (in terms of installation and operation) of fitting data recorders were estimated at £13,000 per unit, with savings due to investigations and repairs estimated at only £3200. In simple terms, this required investment of £75 million and would need to prevent at least two equivalent fatalities each year to show positive financial benefits (Uff, 2000, p. 177). On this basis, widespread fitment of data recorders could not be justified. Privatisation of the rail industry in 1994 and a number of coincident high profile crashes (Southall in 1997 and Ladbroke Grove in 1999) served to accelerate the adoption of data recorders. In the Southall Inquiry report it is noted that "*In my view, the cost-benefit figures produced and the conclusions that they suggest amply demonstrate the shortcomings of CBA* [Cost Benefit Analysis] *as a decision-making tool* [. . .] *I believe that the general fitting of data recorders is long overdue and that this view is shared by the great majority of the industry.*" (Uff, 2000, p. 178). By 2002, Railway Group Standard GM/RT 2472 made data recorders mandatory in all new UK trains from 07 December 2003 onwards, and required that existing trains be fitted with them by 31 December 2005.

Modern trains share some conceptual similarities with aircraft in that they too make extensive use of electrical actuation (the brakes are 'electro-pneumatic' for example), rely on communications between disparate devices and systems, data buses (i.e. the Train Data Bus) and various forms of standardised communications protocol. In other words, they possess a roughly equivalent form of 'avionics' and a data bus (or 'buses') through which an on-board recorder can acquire information. There is not the same degree of conformity as in comparable avionics systems. Critical differences between the rail and aviation data acquisition architectures are, firstly, that the functions of a 'Flight Data Acquisition Unit' (FDAU) are incorporated within the On Train Data Recorder (OTDR) device itself. Likewise, so are some of the functions of a Quick Access Recorder (QAR), and as a result the data must be downloaded manually via serial cable, USB or other memory device. At the present time there is not a standard 'data frame' for OTDR data, with each device manufacturer using a proprietary version and associated analysis software. At the present time the emphasis is on individual data download and analysis for the purposes of driver training and assessment (as per the Southall Inquiry recommendations) or else incident investigation, rather than large scale data storage and industry wide analysis of 'normal' operations. Some modern rolling stock is able to wirelessly download diagnostic information for the purposes of condition monitoring but at the time of writing this is the exception rather than the rule.

1.3. Pushing the envelope

Regardless of measure, whether it takes into account exposure by distance or time, the risk to the travelling public and workforce of using and operating the railway is exceedingly low. In Europe the probability of a fatality is approximately 0.57 per billion miles (Evans, 2011), or two fatalities per 100 million person travel hours (EU, 2003). This figure arises despite the fact that exposure in time and distance have increased dramatically in some countries. In the UK, for example, between 1995 and 2012 the risk exposure by passenger distance rose by 25 billion kilometres or 58% (DfT, 2011). At the same time estimated mean fatal train accidents per

billion train kilometres has fallen by approximately 9.1% annually (Evans, 2011). Risk exposure is accompanied by an increase in the overall intensity of operations. The UK railway system currently supports 1.3 billion passenger journeys (ORR, 2012) with 16% more trains timetabled in 2013 compared to 1995, most of which are running at higher passenger occupancy levels. This equates to 296.2 million train miles travelled in 2010/11 (ORR, 2012), all of which have at one point or another been recorded on an OTDR device.

The UK rail network is the fifth busiest in the world behind China (1.86bn journeys), Germany (1.95bn), India (8.03bn) and Japan (22.67bn). Relative to population size the UK has the third highest rail usage in the world at approximately 21 journeys per head of population. The network is currently loaded with 1.5 billion passenger journeys per year, which compares to an historical peak of 1.43 billion which occurred in 1946. At this time the network comprised approximately 30,400 route kilometres compared to a 2013 network of 15,777 route kilometres. To support even greater numbers of journeys to those achieved in 1946 on a network 48% smaller means, quite simply, that more trains are using less track at higher speeds, or in other words, the system is being constantly pushed "back to the edge of the performance envelope" (Woods and Cook, 2002, p. 141). The 'performance envelope' is defined by Rasmussen (1997) as a set of specific boundaries within which transport systems reside. There is a boundary of economic failure: these are the financial constraints on a system that influence behaviour towards greater cost and operational efficiencies. Then there is a boundary of unacceptable workload: these are the pressures experienced by people and equipment in the system as they try to meet economic and financial objectives. The boundary of economic failure creates a pressure towards greater efficiency, which works in opposition to a similar pressure against excessive workload. Because transport systems involve human as well as technical elements, and because humans are able to adapt situations to suit their own needs and preferences, these pressures introduce variations in behaviour (Qureshi, 2007; Clegg, 2000). "*Over a period of time, [human] adaptive behaviour causes [the system] to cross the boundary of safe work regulations and leads to a systematic migration toward the boundary of functionally acceptable behaviour. This may lead to an accident if control is lost at the boundary*" (Qureshi, 2007, p. 6). Systems that exist at these 'limits of controllability' represent a significant challenge for the human operators within them (Lupu et al., 2013).

1.4. The black box paradox

There are three key paradoxes inherent in this wider picture. Firstly, because so few major rail accidents occur there is no longer enough data to construct reliable forward looking estimates (e.g. Evans, 2011). When safety performance data reaches the level of that achieved in the rail sector it instead starts to become characterised by unpredictable periodicities, cycles or discrete events. This is becoming evident in EU rail safety data, with one large scale rail accident occurring on average every six years (EU, 2003). In other words, the safety data is 'levelling off' with a persistent class of human/machine-systems accident now elevated to the status of a key strategic risk (RSSB, 2009).

Secondly, "*there is widespread concern within the industry that the background indicators – rather than the headline grabbing ones – have remained worryingly stable*" (Wolmar, 2012). An example of this is UK data on Signals Passed At Danger (SPAD) incidents. In the period since the introduction of a countermeasure called the Train Protection and Warning System (TPWS), after which there were initial improvements, there has been comparatively little variation in the overall SPAD rate. For example, the rate for Quarter 4 2012 is the same as Quarter 3, and indeed the same

(or very nearly the same) as on seven previous reporting periods since 2005 (e.g. ORR, 2012).

Thirdly, and most paradoxically, is that the opportunities to use black box data for their original purpose (i.e. post-accident analysis) are diminishing at the same time as the technical capabilities of data recorders are increasing. What this means is that enormous quantities of non-accident data is being collected day in and day out, but not currently used.

1.5. Flight (rail) data monitoring

This paper is premised on best-practice safety science approaches developed within the aviation domain, specifically a process called Flight Data Monitoring (FDM). This is a "*a systematic method of accessing, analysing, and acting upon information obtained from digital flight data records of routine flight operations to improve safety. It is the pro-active and timely use of flight data to identify and address operational risks before they can lead to incidents and accidents.*" (CAA, 2003). FDM is mandatory for operators of aeroplanes of a certified take-off mass in excess of 27,000 kg. In effect, it is a way of using data collected from routine operations to detect trends which, if allowed to continue, might eventually lead to an accident. Changes are made to address issues, and the changes themselves are monitored for their possible effects on other parts of the system. The traditional approach to FDM is focused on exceedance or event detection. Events are defined as: "*deviations from flight manual limits, standard operating procedures and good airmanship*" (CAP 739, p16). Computer software is used to automatically scan FDR data for instances of these deviations, and a set of core events that cover the main areas of interest are quite standard across operators. Event detection is commonly based on simple statistical techniques and automatic algorithms that detect different phases of flight and events therein. FDM is a highly successful way of making use of black-box data in a pro-active manner, but it too is challenged by the emerging class of human-systems problem that is the focus of this paper. Indeed, while having systems that automatically detect events, it is still incumbent on so-called FDM analysts to manually interpret the lower-order 'trace plots' that data recorders produce in order to derive meaning from them. As such, there is considerable value in being able to robustly transform these trace plots into higher order representations, to detect psychologically meaningful patterns therein, and to automatically derive human performance metrics that can help to assess risk.

1.6. Human Factors leading indicators

Leading indicators are measurable precursors to major events such as an accident (Reiman and Pietikainen, 2012). The indication of a precursor 'leads', or comes before, the actual event itself. Lagging indicators are the opposite. These are so called 'loss metrics' that can only become apparent after an event (Rogers et al., 2009). Leading indicators are said to be 'proactive' because they enable steps to be taken to avoid seriously adverse consequences. Lagging indicators are said to be 'reactive' in that a seriously adverse event needs to occur before it can be learnt from (Hinze et al., 2013). For this reason, leading indicators are also sometimes referred to as 'positive performance indicators' and lagging indicators as 'negative performance indicators'. The concept of leading and lagging indicators originally derives from the field of economics and the need to understand 'business cycles' (i.e. growth, recession, investment, divestiture, etc.) and to predict when one phase of a 'cyclical process' such as this will change to another (Mitchell and Burns, 1938). The terms have been appropriated more recently by the safety science community (e.g. Grabowski et al., 2007; Hinze et al., 2013; Reiman and

Pietikainen, 2012), particularly in view of developments in Safety Management Systems (SMS) since the 1990s. Leading indicators, in a Safety Management context, can be defined as "*something that provides information that helps the user respond to changing circumstances and take actions to achieve desired outcomes or avoid unwanted outcomes*" (Step Change, 2003, p. 3).

The basic 'research problem' can be stated thus: despite considerable improvements in safety performance in both rail and aviation sectors, a persistent class of accident/near accident continues to occur. This class of accident/near accident resides at the interface of people and machine-systems. What is required is a means to detect the presence/emergence of such problems before they manifest themselves as a serious operational incident. This paper describes how black box data can be coupled to existing Human Factors methods to provide leading indicators of trends residing at this interface. Specifically, it examines OTDR data on driver responses to an in-cab warning to reveal the types of errors that may be more likely to arise if the discovered trends continue.

2. Method

2.1. Data file and parameters

The OTDR data file is a continuous download for a single traction unit. The recording started at 05:34:57 on the 6th July 2012 and ceased at 21:36.32 on the 11th July 2012. This is a period of 136 h, 1 min and 35 s during which the train made 107 journeys and travelled 1638 miles. The raw data takes the form of a Comma Separated (CSV) file containing a data matrix 191,021 time samples (rows) deep by 72 parameters (columns) wide: a total of 13,753,512 data points. The mean sampling rate is 2.56 s. The logger itself scans the parameters for changes at a rate of 20 mS but, in the present system, to economise on memory requirements data are only logged when one of the 72 parameters changes. The OTDR device itself was a UK Railway Group Standards compliant Arrowvale unit which recorded 72 parameters, 25 of which are in addition to those mandated. In terms of data classification four of the parameters; time, distance and two speed signals derived from a driven and non-driven axle, are continuous ratio data. The remaining 68 are nominal/binomial (i.e. on or off).

2.2. Rolling stock

The sample of OTDR data was obtained from a Class 153 'super sprinter', unit number 153,306. This is a single-unit diesel powered railcar built between 1987 and 1988. Class 153s are 23.2 m in length and have an un-laden weight of 41.2 tons. They seat 72 passengers, comprise a riveted aluminium body shell affixed to a steel underframe, and are equipped with four electrically powered single-leaf Bode doors. The prime-mover is an under-slung turbocharged 6 cylinder Cummins NT855 diesel engine producing 285 bhp. A Cummins-Voith T211r hydraulic transmission drives both axles of the leading BT38 bogie via a Gmeinder final drive. The Unit's maximum operating speed is 75 mph. It is fitted with electro-pneumatic clasp brakes, with cast iron brake pads acting directly on the tread of the wheel(s) via compressed air actuation. Air suspension is provided for additional passenger comfort and refinement. Tight-lock compact BSI auto-couplers mean that Class 153's can work flexibly in unison with several other DMU classes. The present unit worked solo for the duration of the data collection period.

2.3. Journeys and routes

Data collection took place on the Great Eastern (Route 7) and West Anglia (Route 5) regions of the UK's rail network. The strategic 'backbone' of the Great Eastern region is the Great Eastern Main Line (GMEL) originating from London Liverpool Street and travelling North East to Norwich. There are numerous branch lines attached to the GMEL providing services to commuter areas and freight hubs. Data for the present analysis was derived from three journeys between the towns of Ipswich and Felixstowe, a distance of 25 km, over which there are 14 AWS sites.

2.4. Automatic Warning System

The purpose of the Automatic Warning System (AWS) is described thus (McLeod, Walker & Moray, 2005):

"*AWS serves two functions. The first function is to provide an audible alert to direct the driver's attention to an imminent event (such as a signal or a sign). The second function, linked to the first, is to provide an on-going visual reminder to the driver about the last warning. [AWS] is there to help provide advance notice about the nature of the route ahead, and thus communicate to the driver the need to slow down or stop*" (p.4).

AWS alerts and reminders are triggered by an electro-magnetic device placed between the tracks approximately 200 yards prior to the signal, sign or other event to which it refers (although this distance can vary according to local circumstances). Sensors underneath the train detect the presence of magnetic fields and activate AWS accordingly. AWS has three system states:

State 1 – no additional action required

If the event to which an AWS activation is caused requires no action by the driver (such as a signal showing a green aspect) then a bell or simulated chime (at 1200 Hz) will sound briefly. The visual display will also remain inactivated. The driver behaviour in this case is to proceed; there is no requirement to cancel AWS using the cancellation button nor is there any specific need to enact driving behaviours in addition to those that are current or planned.

State 2 – attention is required towards some imminent signal, sign or event

If the event to which an AWS activation is caused requires (or potentially requires) the driver to slow down or stop then a steady alarm or horn sound (at 800 Hz) will sound. The visual display will activate (turning from an all-black display to yellow and black, known as the 'sunflower').

State 3 – acknowledge (and continue to be reminded of) the imminent or previous signal, sign or event

The act of cancelling AWS (by pressing a button) stops the horn sound, and the sunflower continues to be displayed. Failure to cancel AWS within approximately 2 s leads to an immediate emergency brake application which cannot be cancelled (Railway Safety, 2001). This level of braking may cause some discomfort to passengers and the event will be logged on the OTMR equipment. Repeated failure to cancel the AWS within the time allowed is likely to lead to an investigation followed by some remedial training and coaching of the driver.

AWS is a legacy system that was originally conceived as a means to prevent Signals Passed At Danger (SPADs). Several major accidents saw the use of AWS, and the number of events it now refers to, being extended. AWS now provides warnings in six circumstances:

1. (Certain types of) permanent speed restriction,
2. all temporary speed restrictions,

3. (some) level crossings,
4. SPAD indicators,
5. cancelling boards,
6. and other locations (such as unsuppressed track magnets, and depot test magnets).

Unfortunately, the simple two state warning (bell/horn) and reminder (black/yellow) is unable to discriminate between these six different events. There are approximately 29,000 AWS sites around the UK railway network, which equates to a mean of 1.6 AWS indications received in the train cab every 1.6 route miles, or 2.7 activations (either a bell or horn sound) every minute when travelling at 100 mph. Many warning indications require no action from the driver, simply a press of the cancellation button. Many other warnings occur in situations when the correct behaviour at that time is to accelerate the train (it is moving at slow speed or departing from a station, for example). The task of the driver, therefore, is not a simple one of hearing the warning and pressing a button. It requires them to interpret the source of the warning and the context in which it is occurring, and then decide on the correct course of action. The confusion that this could cause for drivers has been cited in several accident inquiry reports (e.g. Cullen, 2001).

3. Results and discussion

3.1. Behavioural clusters

Process charting techniques are used to represent complex real-world activity in an easy to read graphical format using standardised symbols and layout. The process chart methodology has a long application history, with early examples dating from the 1920s (e.g. Gilbreth and Gilbreth, 1921). It has been used extensively in military and high hazard domains as a way of understanding the interaction between people and systems, particularly in terms of identifying human error potential. The method has been used in both rail and aviation settings before. In this application, process charts offer a novel way of converting raw 'trace plots' derived from data recorders into an alternative representation, one that makes it easier to:

• Discern how larger journey phases break down into smaller component activities.
• The order and timing that component activities occur.
• Who is performing what activity.
• The presence of distinct activity clusters.

The 72 parameters extracted from the data recorder were classified into:

Symbol	Description
◇	Operator decision (e.g. proceed on basis of received information?)
▢	Operator action (e.g. move control)
▷	Information transmitted (e.g. to another part of the system via a communications medium)
○	Information received (e.g. from system interface or other agent/actor)
▣	Automatic action (e.g. an action performed autonomously by the system)

Once classified the process chart itself was constructed. This involved creating a timeline and columns for each 'agent' in the system. In the case of the railway example six such agents/columns have been used (Fig. 1). As different recorder channels become active, the corresponding process chart symbol is inserted into the relevant column at the correct point on the timeline. The sequence of activities and their dependence on each other defines when these symbols are linked. Thus an activity/symbol that occurs *after* another activity/symbol becomes linked 'vertically'. Activities that are performed *concurrently* are linked 'horizontally'. Fig. 1 shows the 'normative' structure of operations associated with receiving and cancelling an Automatic Warning System alert.

Having the ability to detect behavioural clusters grants the opportunity to assess whether such structures are typical or atypical. Indeed, whether they are one of a number of different behavioural responses within a wider repertoire, and whether one cluster of behaviour is implicated in risk outcomes more than another, and under what circumstances. From the collected data three different 'clusters' were detected. The first cluster is the normative 'perceive-decide-act' sequence. Here the infrastructure on the track triggers an in-cab warning horn. The driver perceives (hears) this, has enough time to classify it (0.89 s) and respond by pressing the cancellation button. The second cluster is the 'predictive cancel' sequence. In this case the infrastructure on track triggers the in-cab warning horn but the driver responds so quickly that it is not possible to have perceived, classified and responded to the warning. Instead, the driver has seen the track infrastructure and has anticipated the in-cab warning and timed their response to coincide with it starting. The third cluster is the 'multiple predictive cancel' sequence. As in cluster two, the driver can see the track infrastructure ahead and is pressing the cancellation button numerous times before hearing the in-cab warning horn, and several times after the warning has sounded and been cancelled (see Fig. 2).

3.2. Response bias

Based on the analysis above it is clear, firstly, that cancelling an AWS warning is not merely a perceptual one of hearing and seeing the different system alerts and indications, it is also cognitive: driver's not only have to discriminate a 'stimulus' from within a 'noisy' environment, but correctly classify it and respond. Secondly, there are different strategies that drivers employ to perform this apparently simple task. Signal Detection Theory (SDT) helps to untangle these different aspects by separating out a person's sensitivity to stimuli (how easy it is to detect something) and their response bias (their preference for responding one way or another to the stimuli). SDT helps us to understand why a particular 'stimulus', which might be very loud, visible or unambiguous, is not always responded to in the ways we expect (or vice versa). Signal Detection Theory classifies human responses to stimuli in the environment in four ways. The responses that drivers made within each of these categories are shown in Table 1.

The ability to accurately detect stimuli in the environment and correctly classify them is the desired outcome. Under the Signal Detection Theory (SDT) paradigm this requires a high number of hits *and* a low number of false alarms. For example, if the reset button was pressed in response to ANY warning indication this will ensure a 100% Hit rate but will also increase the rate of False Alarms. Accuracy in this case is low. If, on the other hand, the driver is trying to do the opposite, to avoid False Alarms and instead maximise Correct Rejections, they would not respond to ambiguous 'signals', they would instead 'play it safe'. This would increase the number of Correct Rejections but it would also increase the number of 'Misses'. Accuracy in this case is also low. Signal Detection Theory enables us to separate sensitivity (d') from decision bias (C). Sensitivity is a measure of accuracy and tells us how easy it is to distinguish a particular environmental stimuli
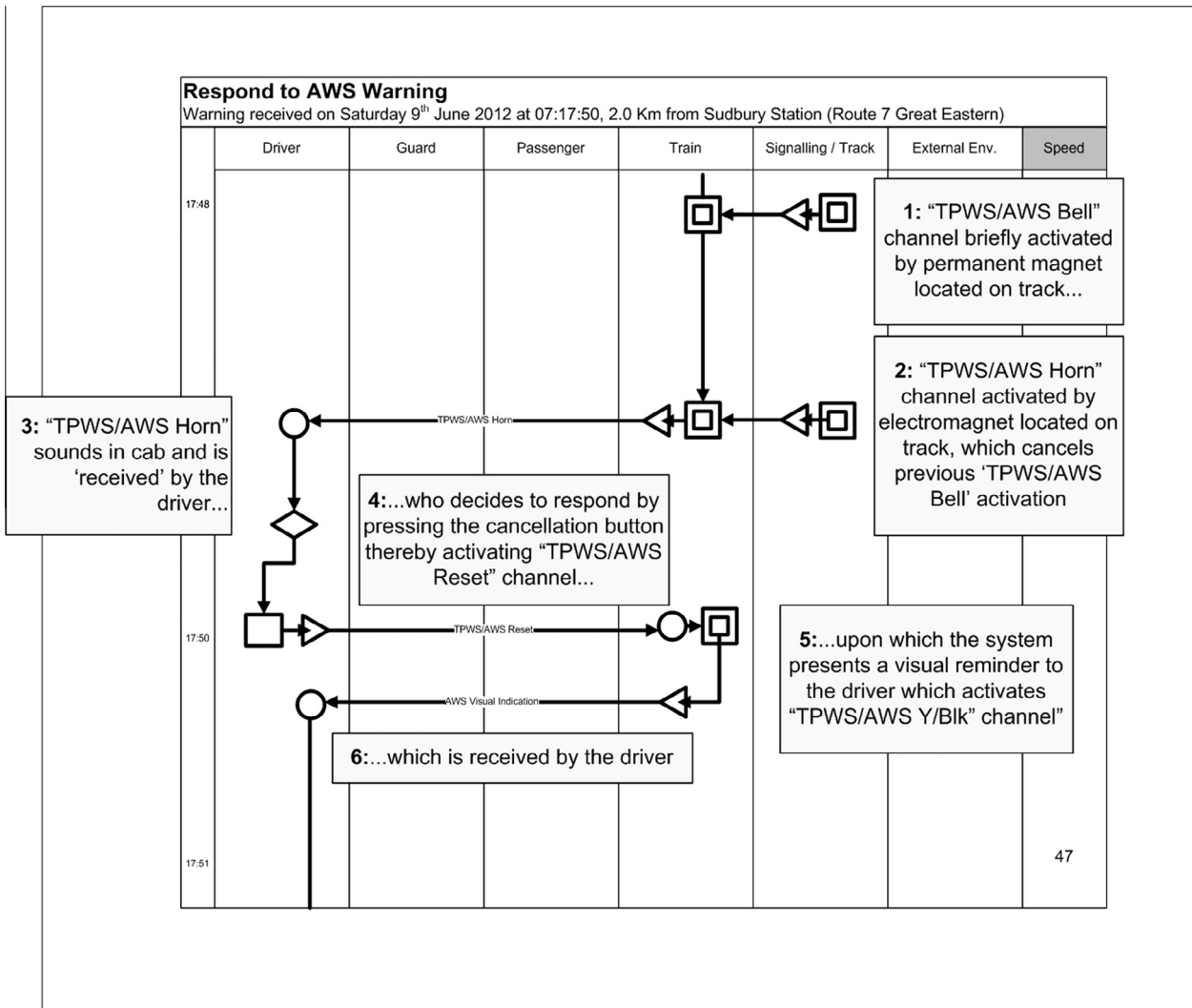
Fig. 1. Annotated process chart showing the type and sequence of operations required to correctly respond to an AWS warning.
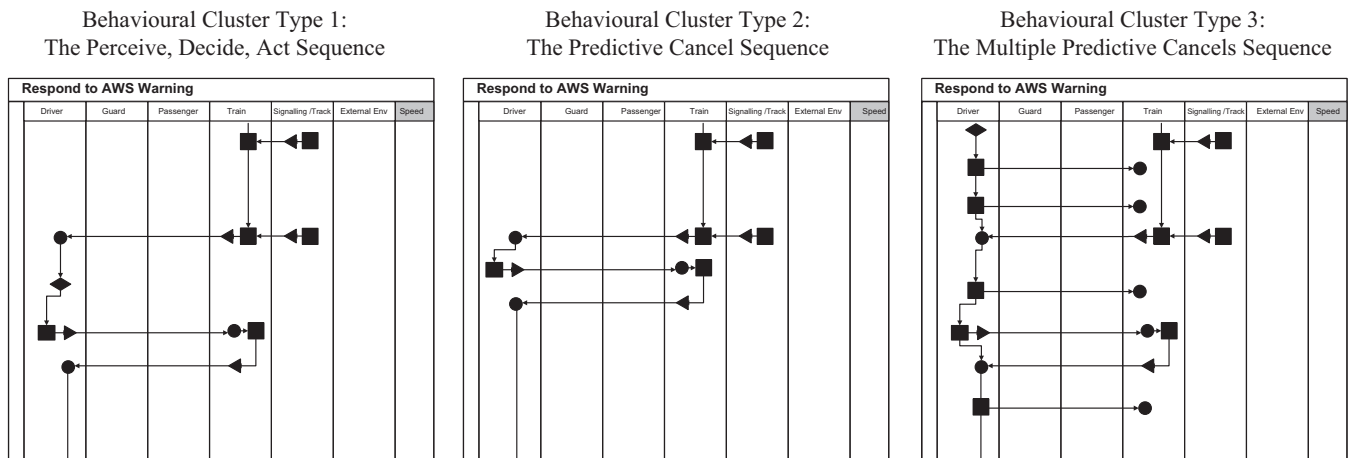

Fig. 2. Three clusters of behaviour associated with cancelling an Automatic Warning System (AWS) alert were detected.

(e.g. an in cab warning). Decision bias tells us whether one response is more probable than another.

Human responses to certain stimuli vary because of incentives embodied in the environment. In train driving, for example, there is a strong incentive in normal operations to prevent unnecessary applications of the emergency brake because an AWS warning was missed. This is because an unwanted brake application cannot be cancelled and is highly inconvenient. On the other hand, the consequences of False Alarms are low because nothing happens when the cancellation button is pressed in error. Because of this it would be expected that train drivers might adopt a 'liberal' decision bias and be willing to prioritise False Alarms (redundant

Table 1
Driver responses to the Automatic Warning System (AWS) can be characterised according to the Signal Detection Theory (SDT) paradigm as follows.

Status of AWS/TPWS system	Categorisation	Number	%
AWS horn followed by reset	Hit	21	30
AWS horn followed by no response	Miss	0	0
AWS bell followed by reset or no activation followed by reset	False alarm	22	31
AWS bell followed by no response	Correct rejection	28	39
Totals		71	100

Table 2
Driver's sensitivity to the Automatic Warning System (AWS) alerts.

Journey	Hits	Misses	Correct rejections	False alarms	d-Prime
1	5	0	11	0	4.65
2	9	0	6	7	2.30
3	7	0	11	15	2.13
				Mean	3.03

Table 3
Results of decision bias (c).

Journey	Hits	Misses	Correct rejections	False alarms	Criterion
1	5	0	11	0	0
2	9	0	6	7	−1.21
3	7	0	11	15	−1.26
				Mean	−1.24

presses of the cancellation button) over Correct Rejections (not pressing the cancel button unnecessarily). Added to this is the discriminability of the 'stimulus' people are responding to, a stimulus that occurs in a 'noisy' real-world environment. By noise we refer to other stimuli, competing demands and distractions in the environment as well as the background noise inherent in human perceptual and cognitive processes. Decision bias and sensitivity interact with these 'noisy' transport environments to make some future responses more likely than others, in ways that are not

always immediately apparent. For example, highly visible warnings that are apparently 'missed', or control actions that are at odds with the situation.

3.2.1. Sensitivity

Sensitivity to a stimulus is given by the metric d-prime, which was calculated as follows:

$$d' = z(FA) - z(H)$$

where $z(H)$ is the number of Hits expressed as a z-value subtracted from the same Z-transformed False Alarm rate. The results obtained are shown in Table 2:

The d-prime figure measures the strength of the stimulus, which in this case is an AWS warning. A value of 3.03 indicates that drivers are highly sensitive to it: in this situation it is unambiguous and easy to discriminate from the wider background of noise, distractions, other contextual factors, etc. Expressed more formally, the responses drivers' are providing when an AWS warning is overlain on top of the 'contextual noise' is 3.03 standard deviations 'different' from the responses they give when the signal is absent (and only the 'contextual noise' is present). Sensitivity provides an important leading indicator concerning the discriminability of information needed for driver's to develop accurate situational awareness. The same 'stimuli' may yield different levels of sensitivity depending on external/contextual factors. A warning that was not expected, ambiguous, not fully understood or masked may lower sensitivity.

3.2.2. Decision criterion

Decision bias/criterion is given by the metric c, which was calculated as follows:

$$c = -\frac{Z(H) + Z(FA)}{2}$$

The results obtained are shown in Table 3:

Decision Bias is independent of sensitivity and relates not to the discriminability of the 'signal' but to the payoffs involved in making one response in favour of another. Thus, regardless of how easy it is to discriminate a stimulus a counter intuitive
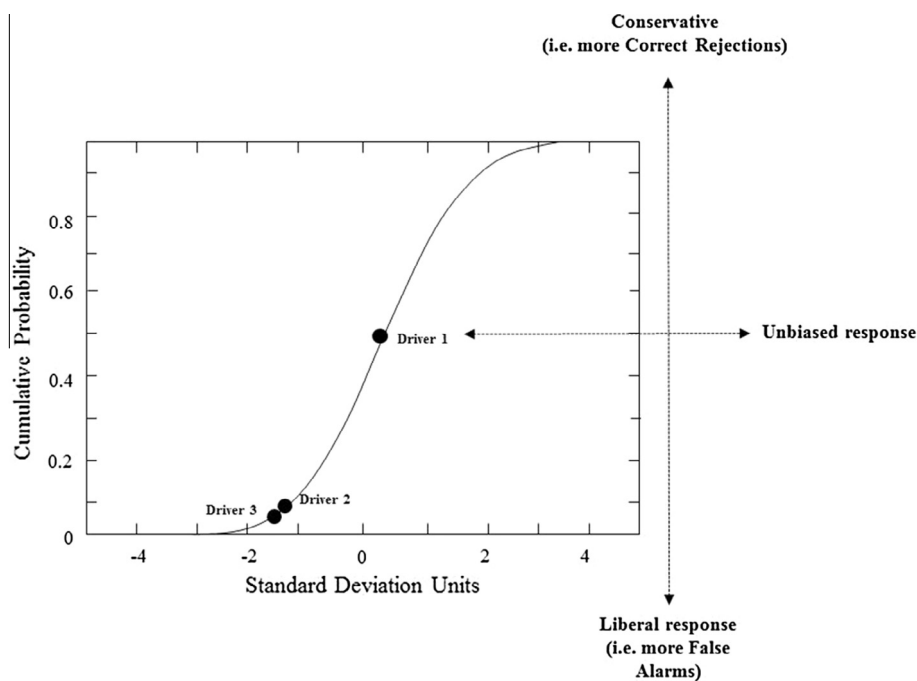

Fig. 3. Cumulative probability plot showing the extent and direction of response biases exhibited by the drivers.

response may still be favoured. This is because the consequences of False Alarms, Misses and Correct Rejections vary with the context. Psychological research shows that decision bias is more unstable and situationally dependent than sensitivity and, therefore, a potentially valuable Leading Indicator.

The mean decision bias value across the three sampled drivers was $c = -1.24$ which indicates a liberal bias. Driver's make more responses that indicate the AWS signal is present than it is absent. In other words, they are prioritising False Alarms over Correct Rejections which, in turn, provides a clue as to the sorts of error that may be more likely to occur in future (i.e. warnings that are cancelled incorrectly). Assuming that drivers' 'internal responses' to the AWS warning are normally distributed (as per Signal Detection Theory) it is possible to plot individual driver decision bias' into the chart below which provides an important diagnostic tool in defining risky psychological/decision making states:

According to Fig. 3, Driver 1 shows no systematic bias in their responses to the AWS warning. They respond correctly to the AWS warning on every occasion and his/her False Alarm rate is zero. Drivers 2 and 3 are different. They are exhibiting a strong 'liberal response bias' meaning that they are much more inclined to exhibit 'false alarm' responses (and behavioural clusters 2 and 3). With the ability to detect these changes in decision bias comes the possibility to analyse (a) the extent to which different biases interact with accident/incident rates (i.e. is a liberal bias of this magnitude associated with particular types of risk) and (b) how the context influences human decision making (and therefore how that context can be modified to 'un-bias' human responses).

4. Conclusions

This paper has described how the outputs from transport data recorders can be coupled to existing Human Factors methods to provide advanced indication of strategically important Human Factors risks. The 'black box paradox' is that the opportunities to use these devices for their original post-accident purpose are diminishing at the same time as their technical capability and data richness are increasing. In addition, the types of risks are changing, with progress in technical areas of reliability and performance exposing issues around human–machine system interaction. To continue to ensure safety in the face of increased risk exposure and operational intensity, better use of this data needs to be made. This paper demonstrates how black box data from the rail sector can be turned into useful 'information' in the form of Human Factors leading indicators of risks associated with the use of an in-cab warning device. Sensitivity provides a measure of how much useful information there is in the environment and the extent to which drivers can discriminate it from the background of contextual noise. Warnings, stimuli and so forth may, in an engineering sense, appear to be unambiguous, yet they may be considerably less so cognitively. Sensitivity provides a measure of this which can, in turn, be associated with changing risk. Decision bias reveals the likelihood that one type of driver response will be favoured and how this interacts with risk. In a wider application it would be possible to examine decision bias in a systematic way looking at differences between drivers and between particular routes. This could provide insight into driving styles and indicate whether particular aspects of a route result in a shift in decision bias. For example, a specific AWS signal on a particular route may result in a high level of predictive pressing (high false alarms) relative to most others, identifying this as a more risky section of journey. Relationships such as these would need to be established based on large-scale

future research but even on a smaller sample of data the method was able to detect potentially important differences between drivers, with some adopting a much more liberal response bias than others. The principle, however, is a much more important one. We have demonstrated that Human Factors methods like these can accept recorder data as an input, are amenable to the kind of software implementation that would be required in a full-scale application, and point the way towards Leading Indicators of strategic importance in safety science more generally.

References

CAA, 2003. CAP 739: Flight Data Monitoring. A Guide to Good Practice. Civil Aviation Authority, UK.

Campbell, N., 2007. The evolution of flight data analysis. In: ASASI 2007: Proceedings of Australian Society of Air Safety Investigators conference, Civic Square, Australia. <http://asasi.org/papers/2007/The_Evolution_of_Flight_Data_Analysis_Neil_Campbell.pdf>.

Clegg, C.W., 2000. Sociotechnical principles for system design. Appl. Ergonomics 31, 463–477.

Cullen, W.D., 2001. The Ladbroke Grove Rail Inquiry: Parts 1 and 2 Report. HSE Books, London, UK.

Department for Transport (DfT), 2011. Department for transport statistics. Retrieved 12 July 2012. <http://www.dft.gov.uk/statistics>.

European Transport Safety Council, 2003. Transport Safety Performance in the EU: A Statistical Overview. Brussels: European Transport Safety Council. Retrieved July 17, 2012, <http://www.etsc.be/documents/statoverv.pdf>.

Evans, A.W., 2011. Fatal train accidents on Europe's railways: 1980–2009. Accident Anal. Prevention 43 (1), 391–401.

Ford, M., 2012. FDM Technology and Data Recovery. Presented at FDM Short Course. Cranfield University, UK.

Gilbreth, F.B., Gilbreth, L.M., 1921. Industrial Efficiency. American Society of Mechanical Engineers, New York.

Grabowski, M., Ayyalasomayajula, P., Merrick, J., Harrald, J.R., Roberts, K., 2007. Leading indicators of safety in virtual organizations. Saf. Sci. 45 (10), 1013–1043.

Hinze, J., Thurman, S., Wehle, A., 2013. Leading indicators of construction safety performance. Saf. Sci. 51 (1), 23–28.

Lupu, M.F., Sun, M., Xia, R., Mao, Z.-H., 2013. Rate of information transmission in human manual control of an unstable system. IEEE Trans. Human–Machine Syst. 43 (2), 259–263.

McLeod, R.W., Walker, G.H., Moray, N., 2005. Analysing and modelling train driver performance. Appl. Ergonomics 36, 671–680.

Mitchell, W.C., Burns, A.F., 1938. Statistical indicators of cyclical revivals, NBER Bulletin 69, New York; reprinted as Chapter 6. In: Moore, G.H. (Ed.), Business Cycle Indicators. Princeton University Press, Princeton, NJ, p. 961.

Morcom, A.R., 1970. Flight data recording systems: a brief survey of the past developments, current status and future trends in flight recording for accident investigation and operational purposes. Aircraft Eng. Aerospace Technol. 42 (7), 12–16.

Office of Rail Regulation (ORR), 2012. National Rail Trends Yearbook, 2010–2011. ORR, London.

Qureshi, Z.H., 2007. A review of accident modelling approaches for complex socio-technical systems. Proceedings of the Twelfth Australian Workshop on Safety Critical Systems and Software and Safety-Related Programmable Systems, vol. 86. Australian Computer Society, Adelaide, Australia, pp. 47–60.

Rail Safety & Standards Board (RSSB), 2009. The Railway Strategic Safety Plan 2009–2014. Rail Safety and Standards Board, London.

Railway Safety, 2001. Automatic Warning System (AWS) GE/RT8035. Railway Safety, London.

Rasmussen, J., 1997. Risk management in a dynamic society: a modelling problem. Saf. Sci. 27, 183–213.

Reiman, T., Pietikainen, E., 2012. Leading indicators of system safety – monitoring and driving the organizational safety potential. Saf. Sci. 50 (10), 1993–2000.

Rogers, A., Evans, R., Wright, M., 2009. Leading indicators for assessing reduction in risk of long latency diseases. Greenstreet Berman Limited for Health and Safety Executive.

Step Change, 2003. Leading Performance Indicators: Guidance for Effective Use. Step Change in Safety, Aberdeen.

Uff, 2000. The southall rail accident inquiry report. Health and Safety Commission, London.

Warren, D.R., 1954. A device for assisting investigation into aircraft accidents. ARL-MECH-ENG-TECH-MEMO-142.

Wolmar, 2012. Rail Magazine, 697: Safety is the Hidden Success Story. <http://www.christianwolmar.co.uk/2012/05/rail-697-safety-is-the-hidden-success-story>.

Woods, D.D., Cook, R.I., 2002. Nine steps to move forward from error. Cognition, Technol., Work 4 (2), 137–144.