Safety Science 92 (2017) 190–204

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

Probative blindness and false assurance about safety

Andrew John Rae BEng PhD (University of Queensland) FHEA [a,*], Rob D. Alexander BSc (Keele) PhD (York) [b]

[a] School of Humanities, Languages and Social Science, Griffith University, 170 Kessels Road, Nathan, Qld 4111, Australia
[b] Department of Computer Science, University of York, United Kingdom

ARTICLE INFO

Article history:
Received 26 January 2016
Received in revised form 24 August 2016
Accepted 7 October 2016

Keywords:
Probative blindness
Assurance
Compliance
Qualitative analysis
Accident analysis

ABSTRACT

Safety activities may provide assurance of safety even where such assurance is unwarranted. This phenomenon – which we will call "probative blindness" – is evident both in hindsight analysis of accidents and in the daily practice of safety work. The purpose of this paper is to describe the phenomenon of probative blindness. We achieve this by distinguishing probative blindness from other phenomena, identifying historical instances of probative blindness, and discussing characteristics and causes associated with these instances. The end product is an explanation of the features of probative blindness suitable for investigating the probative value of current safety activities, and ultimately for reducing the occurrence of probative blindness.

1. Introduction to probative blindness

Not all "safety activities" have a positive effect on safety. Some activities neither reduce the risk of harm associated with a system, nor provide more accurate understanding of that risk. These activities have no safety value. Worse, if these activities are believed to be effective, they result in false assurance – unjustified confidence that safety goals have been met.

Safety activities perform three main roles, where any given activity can fill more than one role. "Ensurance" is the direct improvement of the safety of systems or operations. An example of an ensurance-focused activity is changing a design by adding diverse means of performing a safety function. "Assessment" improves knowledge about safety. Quantitative Risk Assessment is the stereotypical assessment technique – it does not directly make a system safer, but is supposed to inform ensurance effort (Apostolakis, 2004) and thus indirectly improve safety. "Assurance" is the demonstration of safety, often directed towards increasing the confidence of stakeholders not directly involved in ensurance and assessment. As with assessment, assurance does not directly make a system safer, but efforts to demonstrate safety may lead ultimately to safety improvement. For example, attempting to construct a formal proof that a design meets its safety requirements may expose ambiguity in the requirements or bugs in the design.

Assessment and assurance are closely linked. The difference between determining safety and demonstrating safety is subtle, with the terms "assessment" and "assurance" often used interchangeably. Regulatory approaches typically assume that an activity that demonstrates safety would equally reveal danger if such danger was present (Menon et al., 2009). Where this is not the case the activity is an instance of "probative blindness" (Rae et al., 2014b).

Any safety technique, applied in the wrong way or under the wrong circumstances, can exhibit probative blindness. Hence, probative blindness is a property of activities; it applies to particular executions of a technique by particular people at a particular time.

An activity is defined as exhibiting probative blindness if it provides stakeholders with subjective confidence in safety disproportionate to the knowledge it provides about real problems.

There are often multiple opportunities to identify and mitigate hazards, so isolated instances of probative blindness are not necessarily catastrophic. When an organisation is prone to probative blindness, however, its beliefs about safety may drift away from reality even as great effort is expended on safety activities. This is why probative blindness is of such concern – it involves substantial wasted effort, and it can actively hide problems. Probatively blind activities can engage skilled people in enthusiastically doing things that increase the risk of harm. This is a particularly galling misuse of good engineers with good intentions. A better understanding of the phenomenon is necessary if we are to build organisations that can select, apply and interpret safety activity to align beliefs about safety with safety reality.

* Corresponding author.
E-mail addresses: d.rae@griffith.edu.au (A. John Rae), rob.alexander@york.ac.uk (R.D. Alexander).

http://dx.doi.org/10.1016/j.ssci.2016.10.005
0925-7535/

In order to learn about probative blindness, we need a research approach matched to the current maturity of our understanding. Moving ahead too quickly – developing sophisticated theoretical models of loosely defined phenomena – can be unhelpful. Before a phenomenon such as probative blindness can be theorised, it must first be distinguished and explored (von Krogh et al., 2012).

This paper presents a case study series designed to characterise historical instances of probative blindness. The case studies show how probative blindness can be distinguished from other phenomena, and provide an initial characterisation of the manifestations and causes of probative blindness.

The paper argues that probative blindness is a distinct and recognisable phenomenon, illustrates the features by which probative blindness can be recognised, and suggests how the causes of probative blindness can be investigated further.

2. Distinguishing probative blindness as a phenomenon

2.1. Belief-shifts have a central role in accident theory

Organisational accident theory suggests that accident prevention hinges on early recognition that a dangerous situation is developing. In other words, there needs to be a shift from believing that the situation is "safe" to believing that the situation is "unsafe". The reasons behind this lack of belief-shift become a central theme of the accident narratives. For example, Weick (1993) described the deaths of thirteen fire jumpers in the Mann Gulch fire in terms of their understanding about how dangerous the fire was. Early impressions that it was a fire that could be extinguished by the next morning were reinforced by the actions of their team leaders. When (too late) they realised that they were in imminent danger, team co-ordination and trust collapsed. In the wake of the Hertfordshire Oil Storage Depot (Buncefield) explosion, the Health and Safety Executive criticised the operators for "not understanding the potential impact of a vapour cloud explosion" (Board, 2006). The implication of this criticism was that the actual beliefs about danger differed from the "correct" beliefs about danger. The Royal Commission into the West Gate Bridge collapse, in discussing a particularly dangerous feature of the construction method, suggested "Neither contractor ... appears to have appreciated this need for great care" (Barber, 1971). A dangerous situation had developed without a corresponding change in the perception of risk.

Whilst individual accident reports will often make claims about things that "could have been known" or "should have been known", there will always be a mismatch between what appears obvious in hindsight and what was actually knowable with foresight (Fischhoff, 2003). Incorrect beliefs that appear unreasonable to investigators probably were rational to those with no knowledge of what was to come. Attempts to provide a general theory of accidents, summarised in Table 1, try to reconstruct this rationality. In particular, they offer explanations for how and why beliefs do not shift to match the real safety of the system (which would have allowed operators or designers to prevent the accident).

Turner (1976) describes the pre-accident period as "disaster incubation". During disaster incubation the organisation does not shift its beliefs about safety despite mounting evidence of problems. Turner's explanation for this problem is a form of bounded rationality, where organisations are unable to pay attention to signals of danger. These signals are important and obvious in hindsight, but before the accident appear as insignificant – even as distractions from more salient concerns.

Subsequent researchers have upheld Turner's characterisation of the problem as a failure to shift beliefs, but have offered alternate explanations for how beliefs are formed, evolve, and are challenged within organisations.

Keyser and Woods (1990) describe the problem of "fixation errors". A fixation error involves a preliminary assessment of a situation that is rational given the information available at the time. This early assessment is not revised as new information becomes available, or even as the situation itself changes. Keyser and Woods provide the example of an operator ignoring alarms because they "know" that the alarms are inconsistent with the "actual" state of the system.

Vaughan (1997), explaining why the space shuttle program did not react to increasing evidence of danger, introduced the concept of "normalisation of deviance". Once a particular warning signal has become absorbed into routine operations, further occurrences of similar signals have no particular salience. Instead of suggesting that the state of affairs is unsafe, they are part of a pattern of information associated with a normal, presumed safe situation.

"Normal Accidents", written by Perrow (1999) in the wake of the Three Mile Island accident, suggests that the complexity of interaction between human and technical systems can render the current state of the combined human-technical system incomprehensible. People form flawed mental models of the system, and then interpret new information (which could have vital safety insights) to fit those models; they are unlikely to quickly update the models themselves in the midst of an emerging dangerous situation.

Kewell (2007) points to the role of reputation as a two-way "cloaking device", both concealing risk from outsiders and preventing insider awareness of danger. Strong existing beliefs, constructed through a process of public relations, institutionalisation and mystification, are resilient to new information, particularly when the source of that information is less socially powerful.

"High Reliability Organisations" (La Porte, 1996) suggests that organisations are safest when they focus on "evidence that contradicts" and eschew hierarchical authority in factor of operational knowledge.

All of these theories make the counter-factual claim that accidents could be prevented if only organisations were better at updating their beliefs. Failure to do so is explained in terms of properties of the organisations – structures, attitudes, technologies, and reputations – but the theories do not directly examine the events in which beliefs fail to shift.

2.2. Probative blindness is one of several belief-shift phenomena

Probative blindness is not intended to be a new theory of organisational accidents. Instead, it is a clarification of one of the phenomena that must be explained by organisational accident

Table 1
Accident theories involving belief-shift.

Theory	Author(s)	Primary concerns
Disaster incubation	Turner, Pidgeon	Bounded rationality, particularly for leadership attention and decision making
Fixation errors	Keyser & Woods	Situation assessment by operational staff
Normalisation of deviance	Vaughan	Differentiating warning signs from routine events at all levels of the organisation
Normal accidents	Perrow	Situation assessment by operational staff
Reputation	Kewell	Interactions between staff with different levels of authority
High reliability organisations	La Porte, Weick, Rochlin, Roberts	Operational decision making