BIAS IN REAL-WORLD ACCIDENT CAUSE-FINDING

Kathryn Woodcock
Industrial Engineering, University of Toronto

Regardless of the actual causal processes of particular accidents, it is the causes identified by the analyst which determine what responses arise, and how safety is managed in industry. However, some beliefs have been expressed that cause-finding might incorporate predictable types of bias, notably derived from particular motivations as well as common cognitive phenomena. The two major models used to understand the cause-finding process are the attribution model and the diagnosis model. Although experimentally convenient, the attribution model is sufficiently different from the real world cause-finding task that its validity may be questioned, thus these findings need further verification. Others have characterized accident cause-finding as a diagnostic exercise, which introduces another dimension, consisting of time, search sequence, and stopping criteria. Although diagnosis does correspond better with the task of cause-finding, little is known about the real-world accident causal search process. Some debiasing principles exist but in order to make effective use of these, we need to know more about the task and what biases occur, and when, where, and how they intrude.

INTRODUCTION

Further refinement of theories of industrial accident causation may actually confer little further benefit on the problems in workplace, because it is not the actual abstract truth about the causation of accidents which determines safety management, but the conclusions drawn about the cause by those responsible for the workplace. More important, factors not deemed to be causes will often not lead to preventive measures, and will leave open possibilities for future mishaps. Any systematic biases introduced through this process threaten to undermine the safety management strategies that are developed on the basis of accident analysis.

The problem is not without irony. Ergonomists advocate the design of error-tolerant systems because they presume the potential for human fallibility. In contrast, it seems that this fallibility is presumed absent from managerial or specialist levels in organizations. An operator error would typically evoke ergonomic commiseration with inadequate interface design. On the other hand, recognition that subjectivity had played any part in accident cause-finding would probably until recently have been considered simply a failure to apply recommended models with sufficient skill or diligence. However, it has finally been recognized that actual accident analysis entails subjectivity: “Workplace participants are seldom totally objective, unbiased processors of information.” (DeJoy, 1992, p. 165).
Superficially, biases are seen to be self-serving. For example, the injured person in an industrial accident wishes to avoid being wrongly accused—or rightly accused—of disregarding instructions, carelessness, or substance abuse. Witnesses wish to ensure no culpability is attached to their errors or omissions, or may be biased to protect the honour of the injured colleague in expectation of a reciprocal favour. A supervisor wishes to avoid implications of inadequate training or supervision, and senior management wishes to avoid any findings that incur costs for the company. These arguments have such strong face-validity and have been asserted by so many writers that they have been subjected to little empirical challenge.

In the domain of the practitioners, cause-finding is thought to be susceptible to bias by reporter, witnesses, and the analyst. Initial reports are misleading or incorrect, and require follow up. For example, Hammer (1989) cautioned the safety engineer to avoid prejudgments that would lead to error in assimilating data, and to avoid relying on uncorroborated statements because victim and witness reports could be unreliable. He also cautioned that management would be motivated to document accidents in a biased way, and that even the safety specialist would have difficulty being unbiased, since the hazards in the environment should have been removed by his/her prior work. Practitioners are advised to prevent these sorts of bias by diligence and purity of motive, using systematic methods of analysis, where possible, and engaging the participation of others who can contribute a neutral influence on the investigation. However, as Fischhoff has observed (cited in Fischhoff et al, 1981), simply assuming that practitioners actually use prescribed methods is risky. Furthermore, he continued, good method does not guarantee good outcome nor does good outcome indicate a good method was followed.

BIAS

‘Bias’ need be nothing more than a ‘causal asymmetry’; it has been noted in that people tend to accept responsibility for success and deny responsibility for failure. However, the interpretation of these findings has not been unanimous. They have been explained variously as an altered perception motivated by a desire to maintain or enhance self-esteem, a cognitive error caused by differences in salience between the actor and observers, or potentially an error not of perception or cognition but of description, in the sense of impression management. This asymmetry has been found so frequently that it has been called the “fundamental attribution error”: a pronounced tendency for naive observers to attribute behaviours to a disposition of the person, even when the attributer is aware that behaviour is entirely constrained by the situation.

The fundamental attribution error phenomenon has also been questioned, on the grounds that it does not always occur (hence is hardly ‘fundamental’) and cannot be called an ‘error’ in the absence of criteria for accuracy (Harvey, Town & Yarkin, 1981). The latter authors defined bias as “a subjectively-based tendency to prefer a given cognition over its possible alternatives” whereas an error would constitute holding to a hypothesis in the face of contrary facts. Suggesting that a bias in preference of internal attribution is generally an error disregards the fact that internal explanations may often be correct, and external explanations incorrect. Commentators on observed causal asymmetries have not distinguished error from bias. A consistent explanatory pattern is a bias. Failing to identify the factors contributing to an accident is an error. Depending on the actual factors of the case, a bias—even a strong one—may not be an error. However, presumably, if an
explanation is preferred it will be chosen more frequently than merited, and probably not overlooked on the occasions when it is true. Nevertheless, determination of which causal diagnosis is ‘right’ and which is ‘erroneous’ is a precarious venture, and looking for errors in analysts’ judgment is an imprudent research approach which merely pits the accident analyst against the researcher.

It might be argued that a bias is an economical method of problem solving, and this may be true—even desirable—as long as the preferred explanation is true often enough, the costs of selecting it wrongly are not excessive, and the costs of undertaking a more effortful and less biased method of explaining the problem are high enough to make this economy important. Indeed, the practice of shortcuts around intermediate problem solving steps is often called expertise, not bias. The ‘preferred cognitions’ are preferred because experience has shown them valid often enough.

In most situations applying expertise, shortcuts can do no worse than lead into a dead end. The critical characteristic is that the system provides some kind of feedback that indicates that it has reached a dead end—the television still doesn’t turn on, the patient’s symptoms don’t subside, the reactor temperature continues to increase—and the expert attends to this feedback, taking a new approach if necessary. In contrast, the causes of industrial accidents generally don’t produce their adverse outcomes continuously until remedied. The system typically resumes apparently normal operation, regardless of remedy. If the right cause isn’t remedied, the causes would be expected to result in a recurrence with the low frequency of a ‘rare event’. The system’s response is complex, delayed, ambiguous, and stochastic. Therefore determining whether the hypothesized explanation is correct can’t rely entirely on system response to the remedy.

The problem more closely resembles the domain of business auditing. The state of either system is difficult to discern, and even more so is the relationship of the system state to its various components and attributes. Auditors have evolved audit procedures to balance the impossibility of learning everything about any such system with the need to learn enough to form an audit opinion. In this case, rather than expertise representing the ability to shortcut directly to the correct explanation as confirmed by the system response, expertise is represented by the ability to generate appropriate multiple hypotheses and the minimum necessary amount of data collection with the maximum diagnosticity to choose the most correct hypothesis among them. In marked contrast to the lack of comparable research into industrial accident cause-finding, one finds a rich body of research into the auditor’s information processing (from Libby & Lewis, 1977 and earlier, to Hogarth, 1993).

DECONSTRUCTING ACCIDENT ATTRIBUTION RESEARCH

In an effort to investigate the occurrence of biases in real-world accident cause-finding, a limited amount of experimental research has been undertaken (DeJoy, 1985, 1987, 1992; LaCroix & DeJoy, 1989). This research has used an ‘attribution’ paradigm. Attribution is the inference from a set of circumstances of the causes of behaviour (Kelley, 1973). Its goals are explanation of the action and prediction of future actions (Shaver, 1983). Attribution begins with the observation of an action (including indirect observation, through receipt of a written description). The intention of the actor is judged, then an attribution is made to some person or environmental factor.

Kelley’s attribution theory proposed that naive causal explanations for behaviours or feeling-states incorporate intuitive statistical judgments resembling computation of
covariance of the effect with the potential causal explanations. The covariance model predicts that characteristics of the event influencing attribution will comprise whether the worker has worked safely on other tasks (distinctiveness), whether the worker has worked safely on the same task before (consistency), and whether other workers have performed the same task without incident (consensus). Where a worker has generally worked safely in the past, and the same job has resulted in injury to other workers, an injury will focus causality on the job and not the worker.

**Overextension of attribution paradigm**

Attribution is well-suited to such holistic and non-interactive judgments as attributing causes for behaviours and states. The paradigm of attribution was extended to events affecting a person. If an attribution could be made when someone dropped a concrete block, then an attribution could be made when a concrete block fell on someone. It is a simple matter to present a description of that situation and request inferences about effort (carefulness), task difficulty (risk), luck, and ability.

At some point in advancing this line of inquiry, the event becomes sufficiently complex that there is more than one person involved: not only is there a victim, on whose head the block is dropped, but there is the dropper, or perpetrator. There is ambiguity as to which person the observer will identify with when there is both a perpetrator and a victim in a particular event (Vidmar & Crinklaw, 1974). It could be argued that the paradigm begins to break down at that point. However, the attribution paradigm has been stretched to cover accidents, which often involve multiple actors and multiple environmental factors, and even multiple events.

A lack of clarity as to what constitutes an ‘accident’ may be part of the problem. Mitchell & Wood (1980) had applied the attribution paradigm to a study of supervisory attributions for nursing patient care ‘accidents’ using the covariance model. In this case, the attribution paradigm is a good fit. The focal person in the stimulus events was clearly the nurse, with the patient more or less assuming the role of a product and the accidents in question resembling a quality control flaw more than an occupational accident.

It appears that the attribution paradigm has been extended incrementally from a clearly fitting situation to the point where the fit has slipped. The task of accident cause-finding differs from attribution in two major ways. First, accidents differ from the sort of ‘actions or feeling-states’ that best fit the paradigm. The nature of accidents is that they often involve more than one focal person. In addition to potentially having separate victim and perpetrator, there are myriad actors in different roles involved with management, supervision, design, and peripheral activities. An accident involves a complex sequence of events, from the penultimate action back to some design decision or indeed an act of God (Rasmussen, 1986). Secondly, the process of cause-finding is not like a jury trial, in which a presented data set is evaluated integratively in deliberate or intuitive conformity with covariance principles.

**Doesn’t accident cause-finding involve information search as well as data evaluation?**

Although analysis of variance is a tidy analogy for seeking the abnormality that resulted in the effect, statistical theories of inference and hypothesis testing such as analysis of variance are applied after data have been collected. These tools depend on the assumption that the search process, identification of the source and means of acquisition of the data have
all been adequate. Using the analogy of analysis of variance for attribution implies that information acquisition is of negligible importance or that there are no biases or other shortcomings at that level, when this is clearly implausible in a real world task (Gigerenzer & Murray, 1986.)

The principles of covariance were proposed to apply to naive attribution. Although an attributional response may occur on receipt of the initial information, it is improbable that this would be the final causal explanation of people acting in a professional role. Even if initial attributions made by actual practitioners reflect the suggested patterns of bias, this may have no bearing at all on the final cause-of-record. On the other hand, it may have, if the bias was also reflected in limited information search, such as only seeking information to confirm and elaborate on that initial impression.

A belief in the existence of non-iterative cause-finding can only be justified in two cases. The first of these is the case where a ‘script’ is evoked by the pattern observed. Scripts are mindless formula that describe what happens in a given situation, removing all causal connections in the process (Kass, 1992). It is unnecessary, for example, to think about why a waiter provides a menu in a restaurant; this is simply part of the restaurant script. Scripts could be retrieved in a real world situation if some common recurring accident scenarios have been stored over many years of experience. The causes of events such as slipping while walking in a hall or needlestick injuries (puncture usually on the hands while recapping or discarding used hypodermic syringes) are commonly given on accident reports as “just an accident”. These outcomes have apparently become part of a script that suppresses causal reasoning in those cases. It may not be a good thing to dismiss these events as “just an accident” but it may be reasonable to presume that this happens.

The second instance of non-iterative cause-finding is created by an experimental task that is unrealistic. It has been remarked that people are willing to make predictions despite lack of predictive value of their indicator for the outcome being predicted, as though they believe that an experimenter would not ask them to make a prediction if the provided data were not sufficient (Tversky & Kahneman, 1974). Hence paper-and-pencil attribution experiments could successfully obtain judgments, even judgments that suggest a pattern of bias. However, the same situation in the real world would not be solved in that non-iterative way, for two reasons. First, people may prefer to consider multiple options. When provided with plenty of time to reach their causal explanation, subjects with a low need for closure appeared to consider additional factors and/or reconsider the initial factors, thus avoiding the pattern of overattribution (to internal explanations) bias exhibited by high-closure subjects (Webster, 1993). Secondly, the cause-finding task and professional role entail more complex evaluation. As Hogarth (1993) argued, to understand decision making, understanding the (real) task is more important than understanding the people.

For real-world accidents, the person doing the cause-finding participates in determining what information will be obtained, and often in obtaining it, as well as evaluating it. It would be reasonable to expect some harmony between the kind of biases which intrude at the search strategy stage and at the interpretation stage. In attribution experiments, all subjects in the same treatment condition consider the same data set. For example, one group of subjects may receive a data set designed to represent evidence of a stable external cause (such as a hazard in the work). The attributions made by subjects are analyzed to determine whether this elicits excess attributions to other causes, such as bad luck or lack of effort. This method effectively superimposes someone else’s search
Bias in real-world accident cause-finding

process on the subject. Thus, experimenters may be providing subjects with data those subjects would never be in a position to evaluate in the real world, due to searching elsewhere for evidence.

Flawed search for information and inadequate revision of initial inaccurate hypotheses are no less likely to be the source of a poor cause-finding than a bias in the assessment of a data set. Without knowledge for how search affects bias, the only available debiasing remedies are to “be careful not to be biased,” and “when in doubt, avoid internal explanations because those have the reputation of being biased.” While sensitizing people to known biases is one debiasing option, one must be careful to avoid actively creating a bias (Fischhoff, 1982). In view of the concerns just expressed, the past research does not justify taking the risk of affirmatively creating a bias as far as real-world accident cause-finding is concerned.

DIAGNOSIS

When bias is found in experiments that omit the search stage of cause-finding, it is easy to presume that the biases are either motivational or based on cognitive fallibility. Introducing the factor of search strategy expands the number of possible explanations and creates questions of how search strategy affects bias, and in fact whether bias can be mitigated by more extended consideration, as Webster’s (1993) findings suggest. An alternative paradigm that has been applied to accident cause-finding is that of diagnosis (Rasmussen, 1993).

A diagnostic model recognizes that first the analyst must seek and obtain information on potential factors. Then the factors are evaluated, and a causal chain is assembled. Rasmussen (1986) described a fault as being something that is the cause of deviation from the standard; found on the causal path backward from this effect; something which is accepted as a familiar and therefore reasonable explanation; and for which a cure is known. Rasmussen noted that human error is both a familiar explanation and readily ‘remediable’ by instruction or reprimand, therefore particularly likely to be found. This is consistent also with the ‘bias’ found in many attribution experiments. Rasmussen’s explanation seems to support the idea that it is not (only) the interpretation of the information which is biased, but the search (and termination of the search) for possible causes.

Several of the roles described by Rasmussen (1993) could be adopted by a safety specialist for the cause-finding task, notably the ‘analyst’, the ‘repair man’, and the ‘attorney’. Depending on the role adopted and the stop rule associated with the role, a particular situation could lead to different information sets being evaluated, and different conclusions. The analyst seeks an explanation which is plausible and familiar, which for any given set of circumstances could be explained as a component fault, an operator error, or a design oversight. The attorney seeks to allocate responsibility, and will seek to identify the person who made an error while in control of his acts. The metaphorical ‘repair man’ is responsible for restoring the normal system functioning, hence the search for explanations will be contained within the set of remedies available. Similarly, Wong & Weiner (1981) suggested that the process of “attributional search” (and where it led) could reflect whether the goal of explanation was adaptive or defensive, which would be determined by the task and role of the person doing the explaining.

Studies of students explaining academic performance (Wong & Weiner, 1981) and auditors (Lauer & Peacock, 1992) show that the sequence of questions reveals
characteristics of the diagnostic task, such as hypothesis generation and testing, and information seeking priorities reflecting diagnostic objectives. Both studies reinforced the value of monitoring the process rather than simply the outcome of the explanation process. Almost nothing is known about the actual real world accident cause finding task and practices.

Existing knowledge about heuristics and biases in decision-making is pertinent to cause-finding. Heuristics assist in retrieving knowledge structures and stored causal explanations (schemata). Cause-finding may be influenced by schemata reflecting beliefs about accident causation, such as the careless-accident schema, or the accident-as-indication-of-undiscovered-hazards schema, or a situation-specific schema. Different analysts bring different frames of reference to the analysis and explanations will focus on what the analyst considers unusual or surprising about the event (Kahneman & Miller, 1986; Rasmussen, 1993). The ways in which cognitive heuristics apply to the complex cause finding task, and ways prior research suggests they might bias it, are too numerous to tabulate here. Research experiments need to be structured so as to allow those cognitive processes to occur in a naturalistic setting, providing for ambiguity and irrelevant information as well as informational cues (Fiedler, 1982). This will reveal how the actual task is structured as well as how information is processed.

Finally, although a search may appear to be ‘biased’, and erroneously overlooking causal factors, it must be remembered that the safety specialist functions in a context of bounded rationality within an industrial organization. The organizational dynamics known to them (and unknown to a naive researcher) may establish criteria for solutions that appear to be unrelated to the causes of the accident. This means that research into cause-finding biases must investigate real practitioners and attempt to define the bounds of their rationality. It is hardly original to question the validity of using undergraduate students instead of ‘real people’. If the research is applied, say, to naive attribution, there may be no harm in this practice. However, where the research is supposed to reflect practitioners’ real world judgments, it becomes a role-playing exercise. Role-playing only reveals the role-players’ beliefs about the role-incumbents.

IMPLICATIONS

The intent of this paper was not to reinterpret the findings of prior studies and assert that the biases they have suggested do not exist. Rather it questioned the research paradigm, and advocates that cause-finding research must incorporate the search process and the bounded rationality of real practitioners. Although this reorientation could find that bias is mitigated through search, this is improbable. The conventional wisdom in the real world agrees that bias exists. As well, available information about diagnostic search suggests a number of opportunities for the intrusion of biases, both through search strategy and choice of stop rule. However, our knowledge of diagnosis also suggests that expertise entails individual differences particularly in schemata and the knowledge base of expertise. There may be no universal bias, but rather different processes and types of bias depending on the degree of expertise. Above all, in order to assist the performance of the task we need to understand the task, not just abstract fallibility of humans in generic decisions.

With knowledge of the diagnostic search process for real-world accident cause-finding and the degree of individual practitioners’ variability, we would be in a better
position to provide decision aids for debiasing cause finding. If searches are too narrow, reminding the analyst of additional fault tree branches might be helpful. If searches tend to focus on confirming favoured explanations, we might develop some means to encourage the recruitment of negative evidence. If hasty closure is imposed by task pressures, automated tools to speed up processing—or data to justify more time—may reduce this biasing influence. If the bounded rationality is the chief factor, the industrial community needs to be alerted to this in order to resolve to change it. If searches tend to rapidly close onto unproven remedies, we might provide some base-rate data or encourage and assist monitoring of its effectiveness to counteract overconfidence and the planning fallacy. If searches appear to be simplified just to accommodate information overload, we might offer decision-aiding devices to record and integrate user-gathered information, or even consider expert systems. All of these decision aids sound helpful, but if provided where not necessary could merely get in the way, introduce new biases (as the practitioner searches for the hidden message behind why they were provided), reduce the appetite for future ‘help’ by communicating to practitioners that others don’t understand them, while allowing us to convince ourselves we have done something helpful without actually being helpful. Choosing among the options require more insight into what practitioners currently experience in real-world accident cause-finding.

REFERENCES

DeJOY, D.M. 1985 Attributional processes and hazard control management in industry. *Journal of Safety Research* 16:61-71


LIBBY, R. & LEWIS, B.L. 1977 Human information processing research in accounting: the state of the art. *Accounting, Organizations and Society* 2(3):245-268


