ORIGINAL ARTICLE

Sidney Dekker · Erik Hollnagel

Human factors and folk models

Abstract This paper presents a discussion of the susceptibility of human factors to the use of folk models. The case of automation-induced complacency is used as a guiding example to illustrate how folk models (1) substitute one label for another rather than decomposing a large construct into more measurable specifics; (2) are immune to falsification and so resist the most important scientific quality check; and (3) easily get overgeneralised to situations they were never meant to speak about. We then discuss the link between models and measurements, where the model constrains what can be measured by describing what is essential performance, and where the model's parameters become the basis for specifying the measurements. We propose that one way forward for human factors is to de-emphasize the focus on inferred and uncertain states of the mind, and shift to characteristics of human performance instead.

Keywords Model · Mind · Cognition · Human factors · Explanation · Falsification

1 Introduction

"Principles taken upon trust, consequences lamely deduced from them, want of coherence in the parts, and of evidence in the whole, these are every where to be met with in the systems of the most eminent philosophers, and seem to have drawn disgrace upon philosophy itself." David Hume (1711–1776). A Treatise of Human Nature (Introduction)

S. Dekker
Department of Mechanical Engineering,
University of Linköping, 58333 Linköping, Sweden

E. Hollnagel (✉)
Department of Computer and Information Science,
University of Linköping,
58333 Linköping, Sweden
E-mail: eriho@ida.liu.se

Although it is difficult to adequately account for complex behaviour, it is a necessary part of the jobs of human factors professionals, be they investigators, designers, or researchers. One way to overcome this difficulty in the applied human sciences has been the introduction of terms and concepts that aim to cover and explain large portions of human behaviour (Sarter and Woods 1997). Examples range from long established labels such as decision-making and workload, to more recent additions such as complacency and situation awareness.

Today, human factors have a sizeable stock of concepts that are used to express insights about the functional characteristics of the human mind – or covert information processes – that underlie complex behaviour. It has become increasingly common in, for example, accident analyses to mistake the labels themselves for deeper insight (Woods 1993), but although it is tempting to do so it is definitely wrong. Therefore "loss of situation awareness", "automation complacency", and "loss of effective crew resource management" can now be found among the causal factors and conclusions in accident reports (National Transportation Safety Board 1994; Aeronautica Civil de Columbia 1996). This usage takes place without further specification of the psychological mechanism that might possibly be responsible for the observed behaviour – much less of how such mechanism could force the sequence of events toward its eventual outcome. The labels refer to concepts that are intuitively meaningful in the sense that everyone associates something with them, so they feel that they understand them. People furthermore tacitly assume that others understand the concepts named by the labels in the same way and that they therefore also implicitly agree on the underlying "mechanisms". The ease by which these labels are used and swapped around as common currency in an industry or scientific community reinforces this practice. If this goes on for long enough, it leads to the syndrome of "The Emperor's New Clothes": people may no longer dare to ask what these labels mean, lest others suspect they are not really initiated in the particulars of their business.

The use and popularity of these labels are evidence of the psychological strategies people turn to when confronted with the daunting complexity of modern technological systems. In scientific terms it can be seen as a tendency to rely on folk models (Hollnagel 1998a; Stich 1985). Such common-sense models are not necessarily incorrect, but compared to articulated models they focus on descriptions rather than explanations and are therefore very hard to prove wrong. This article examines the characteristics that set apart folk models used to explain human performance from other models and that limit their ability effectively to further the growth of human factors knowledge. The arguments address the following three questions:

– Are there characteristics that allow us to separate folk models from other models of human performance?
– Can we distinguish folk models from models that are just immature and still lack a firm empirical foundation?
– Is the field of human factors in any way particularly vulnerable to folk models?

The concrete reason for exploring these questions is the recent problem of human performance created by high levels of automation on commercial aircraft flight decks (Federal Aviation Administration 1996; Sarter and Woods 1997). Various provisional explanations for the observed problems have been proposed, including complacency (Wiener 1988), loss of situation awareness (Endsley 1999) and loss of effective crew resource management (Waldman 1999). Although these explanations all are suitable candidates for an examination of the nature of folk models, we concentrate on one – complacency – and will only touch lightly on the others. Although the substance of our arguments comes from developments in aviation, the issues addressed (such as complacency, automation and indeed the tendency to rely on folk models) affect all human factors in every application area.

2 Characteristics of folk models

2.1 Explanation by substitution

The use of folk models is as pervasive and as old as science itself. The best-known example of a folk model from modern times is probably Freud's psychodynamic model, which links observable behaviour and emotions to non-observable structures (id, ego, superego) and the related interactions. Poorer cousins of that can be found among the several models that explain "human error" by referring to various "human error mechanisms".

One unifying characteristic of these models is that assumptions about non-observable constructs are conveniently endowed with the necessary causal power without any specification of the mechanism responsible for such causation. As an illustration of what we mean by that, consider the comments from Charles Billings in wrapping up a conference on situation awareness in 1995:

"*The most serious shortcoming of the situation awareness construct as we have thought about it to date, however, is that it's too neat, too holistic and too seductive. We heard here that deficient SA was a causal factor in many airline accidents associated with human error. We must avoid this trap: deficient situation awareness doesn't 'cause' anything. Faulty spatial perception, diverted attention, inability to acquire data in the time available, deficient decision-making, perhaps, but not a deficient abstraction!*" (Billings 1996)

The problem is, of course, neither that the constructs are abstractions nor that they are developed *post hoc*. Every scientific construct is an abstraction and the vast majority are, and indeed must be, proposed *post hoc* – across all fields of science. The problem is rather that the value of the constructs hinges on their common-sense appeal rather than their substance. Both diverted attention and deficient decision-making – or rather, attention and decision-making – are abstractions, but they are less contentious because they are articulate about the constituent psychological mechanisms. This level of articulation is not simply the result of years of use, but rather reflects that these concepts from the very beginning referred to intuitively meaningful types of behaviour (to decide, to attend), hence were more than convenient explanations. Situation awareness is "too neat and holistic" in the sense that it lacks such a level of detail and so fails to account for a psychological mechanism needed to connect features of the sequence of events to the outcome (Endsley 1999). The first and most evident characteristic of folk models is that they define their central constructs, the *explanandum*, by substitution rather than decomposition or reduction. So instead of explaining the central construct by statements that refer to more fundamental and presumably better known *explananda*, the explanation is made by referring to another phenomenon or construct that itself is in equal need of explanation. A good example is complacency, as it is used in relation to the problems observed on automated flight decks. Most textbooks on aviation human factors talk about complacency and even endow it with causal power, but very few define it, as the following examples show.

– According to Wiener (1988), "boredom and complacency are often mentioned" in connection with the out-of-the-loop issue in automated cockpits. But whether complacency causes an out-of-the-loop condition or whether it is the other way around is left unanswered.
– O'Hare and Roscoe (1990) state that "because autopilots have proved extremely reliable, pilots tend to become complacent and fail to monitor them". Complacency, in other words, is invoked to explain monitoring failures.
– Kern (1998) explains that "as pilots perform duties as system monitors they will be lulled into complacency, lose situational awareness, and not be prepared to

react in a timely manner when the system fails". Therefore, complacency can cause a "loss of situational awareness".

- On a single page in their textbook, Campbell and Bagshaw (1991) say that complacency is both a "*trait* that can lead to a reduced awareness of danger", and a "*state* of confidence plus contentment" (emphasis added). In other words, complacency is at the same time a long-lasting, enduring feature of personality (a trait) and a shorter-lived, transient phase in performance (a state).

- For the purpose of categorising incident reports, Parasuraman et al. (1993) provided the following definition of complacency: "self-satisfaction which may result in non-vigilance based on an unjustified assumption of satisfactory system state". This is part definition (although deficient in a critical sense, which will be covered later), but also part substitution: self-satisfaction takes the place of complacency and is assumed to speak for itself. There is no need to make explicit by which psychological mechanism self-satisfaction arises or how it produces non-vigilance.

It is in fact very difficult to find coherent semantic content in the human factors literature when it comes to complacency. The phenomenon is often invoked to describe a deviation from official guidance or performance norm. For example, pilots should coordinate, double-check, and look; the failure to do so is explained by referring to complacency, although this does not explain much at all. Actually, none of the above examples provide a proper definition of complacency or try explaining it by basing it on other and better-known concepts. Instead, complacency is treated as an analytical truth and "defined" by substituting one label for another. In the examples used above, complacency is equated with boredom (Wiener 1988); overconfidence (Stokes and Kite 1994); contentment (Campbell and Bagshaw 1991); unwarranted faith (O'Hare and Roscoe 1990); over-reliance (Kern 1998); a low index of suspicion (Wiener 1988) and self-satisfaction (Parasuraman et al. 1993). Explanation-by-substitution unfortunately raises more questions that it answers and we are left to wonder how is it that complacency produces vigilance decrements, or how is it that complacency leads to a loss of situation awareness? A better alternative would be to look for explanations in terms of, say, decay of neurological connections, fluctuations in learning and motivation, or a conscious trade-off between competing goals in a changing environment. Such explanations suggest possible measures that a researcher could use to corroborate the explanations and to monitor for the target effect. Yet none of the descriptions of complacency available today offer any such roads to insight, and claims that complacency is at the heart of a sequence of events are therefore immune against critique and against falsification. A further disadvantage, which often is overlooked, is that definition by substitution limits the scope of available remedial actions that can be suggested to deal with the problems that allegedly result from "complacency".

2.2 Immunity against falsification

Most sciences rely on the empirical world as touchstone or ultimate arbiter (literally a "reality check") for their theories and hypotheses. Following Popper's rejection of "inductionism" and his backing of the hypothetico-deductive method as the fundamental principle in empirical science (Popper 1972), theories and hypotheses can only be deductively validated by being falsified or refuted. This usually involves some form of empirical testing to look for exceptions to the postulated hypothesis, where the absence of contradictory evidence becomes corroboration of the theory. Falsification deals with the central weakness of the inductive method of verification, which, as pointed out already by the 18th Century philosopher David Hume, requires an infinite number of confirming empirical demonstrations. Falsification, on the other hand, can work on the basis of only one empirical instance, which proves the model wrong. Consequently, models that do not allow for proper falsification are highly suspect, and should be kept at arm's length.

The failure of folk models to allow falsification has been called "immunisation", which is the practice of leaving assertions about empirical reality underspecified, thereby making it difficult for others to follow or critique. For example, Waldman (1999) asserts that severely compromised cockpit discipline results when any of the following attitudes are prevalent: arrogance, complacency and overconfidence. Nobody can disagree with that because the assertion is underspecified and therefore immune against falsification. This is similar to psychoanalysts claiming that obsessive-compulsive disorders are the result of overly harsh toilet training, which fixated the individual in the anal stage where the id now needs to battle it out with defence mechanisms. In the same vein, if the question of "where are we headed" from one pilot to the other is interpreted as a "loss of situation awareness" (Aeronautica Civil de Columbia 1996), this claim is immune against falsification. The transition from raw performance fragments, such as context-specific behaviour (people asking questions), to a conceptual description, such as the postulated psychological mechanism (loss of SA), is made in one big leap that is difficult for others to replicate. Current theories of situation awareness (Endsley 1999) are simply not sufficiently articulated to explain why asking questions about direction represents a loss of situation awareness.

Some theories may superficially appear to have the characteristics of good scientific models, yet just below the surface they lack an articulated mechanism that is amenable to falsification. Although falsifiability may at first seem like a self-defeating criterion for scientific progress, the opposite is true: the most falsifiable models are usually also the most informative ones, in the sense

that they make stronger and more demonstrable claims about reality. In other words, falsifiability and informativeness are two sides of the same coin.

2.3 Overgeneralisation

The lack of precision of folk models and the inability to falsify them contribute to their overgeneralisation. One famous example of overgeneralisation in psychology is the inverted U-curve, also known as the Yerkes-Dodson law. Ubiquitous in textbooks (Kahneman 1973), the inverted U-curve couples arousal with performance, usually without clearly stating the units of either, in such a way that a person's best performance is claimed to occur between too much arousal (or stress) and too little, tracing a sort of hyperbolae.

The original experiments were, however, neither about performance nor about arousal (Yerkes and Dodson 1908). They were not even about humans. Examining "the relation between stimulus strength and habit formation" the researchers subjected laboratory mice to electrical shocks to see how quickly they decided to go one pathway versus another. The conclusion was that mice learn best (that is, they form habits most rapidly) at anything other than during the highest or lowest shock. The results approximated an inverted U only with the most generous of curve-fittings. The X-axis was never defined in psychological terms but in terms of shock strength, and even this was dubious as Yerkes and Dodson used five different levels of shock, which were too poorly calibrated to know how different they really were. The subsequent overgeneralisation of the Yerkes-Dodson results (to no fault of their own, incidentally) has confounded stress and arousal, and after a century there is still no firm evidence that any kind of inverted U relationship holds for stress (or arousal) and human performance (Stokes and Kite 1994).

Overgeneralisations take narrow laboratory findings and apply them uncritically to any broad situation where behavioural particulars bear some *prima-facie* resemblance to the phenomenon that was investigated under controlled circumstances. Other examples of overgeneralisation and overapplication include using "perceptual tunnelling" (putatively championed by the crew of an Eastern airlines L-1011 that descended into the Everglades after its autopilot was inadvertently decoupled) and the loss of effective Crew Resource Management (CRM) as major explanation of accidents (Aeronautica Civil de Columbia 1996). A frequently quoted sequence of events with respect to CRM is the flight of an iced-up Air Florida Boeing 737 from Washington National Airport in the winter of 1982 that ended shortly after take-off on the 14th Street bridge and in the Potomac river. The basic cause of the accident is said to be the co-pilot's unassertive remarks about an irregular engine instrument reading (despite the fact that the co-pilot was known for his assertiveness). This simple explanation hides many other factors which might be more relevant, including air traffic control pressures, the controversy surrounding rejected take-offs close to decision speed, the sensitivity of the aircraft type to icing and its pitch-up tendency with even just a little ice on the slats, and ambiguous engineering language in the airplane manual to describe the conditions for use of engine anti-ice (Buck 1995).

In conclusion, folk models as used in human factors to explain large sequences of complex behavioural events, seem to share the following characteristics:

– They explain by means of substitution instead of decomposition.
– They are immune against falsification.
– They tend to rely on overgeneralisation.

These distinctions between folk models and articulated models have consequences for the types of performance measures that can be used – undoubtedly one of the most important problems in the study of human performance (Hollnagel 1998b).

3 Models and measurements

The definition of a measurement depends on how the corresponding domain or phenomenon is described or explained. The definition of a measurement therefore presupposes a clarification of what the model behind the measurement is, where a model is understood as a simplified representation of the salient features of the target system (Fig. 1). The model constrains what can be measured by describing what is essential performance, and the model parameters thereby become the basis for specifying the measurements. Since it is impossible for the model to contain all of the parameters of the target system, the characteristics of the model define the important measurements.

Most models are structural: they represent the functions of a system (in particular, of the human mind) by means of hypothetical structures as well as by the relations between them, usually formulated as some kind of "mental mechanism". A good illustration of that is the
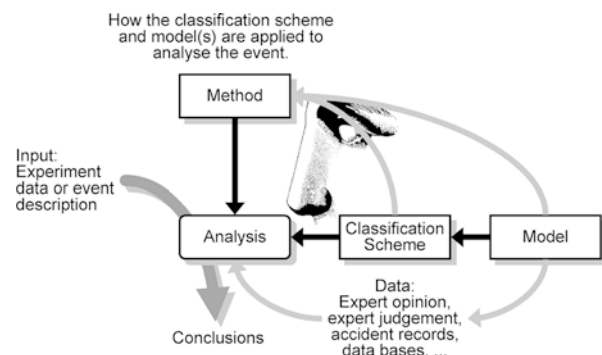
Fig. 1 The relations between models, classification schemes and methods

conventional human information processing model, which describes human actions as emanating from a relative simple system of functional units, such as a number of stores (sensory store, working memory, long term memory), a decision making unit, an attention regulating unit, and so on. This description implies that measurements should be related to the theoretically-defined functioning of these units, as well as to the links (or information channels) between them.

In the 1960s and 1970s the modelling efforts focused on the fundamental information processes, particularly those related to perception and memory (Attneave 1959; Lindsay and Norman 1977). Measures were defined according to the models, as for instance limited capacity central processing or levels of processing in multi-store memory models (Norman 1976). The details of the models, and the constrained character of the phenomena being studied, allowed very specific measurements to be proposed. Later on, when the interest turned from the mechanisms of perception and memory to the cognitive functions that were part of, for example, problem solving or reasoning, it became more difficult to propose theory-based measurements. Instead data were found through such means as verbal protocols and introspective accounts. The models started to look outward to how people interacted with the environment, although still mostly as part of relatively simple and somewhat contrived tasks, and the measurements reflected this change. Whenever the research went out of the laboratory, or at least looked at problems taken from real work situations, the focus turned to general performance aspects such as attention and workload.

Since there are cases where an articulated theory is not available, measurements can also be derived from a general understanding of the characteristics of the system and of the conditions of human work. An example of that is workload and, more recently, situation awareness (Endsley 1999). Workload reflects a subjective experience of mental effort that is so pervasive that it can be applied to practically all types of situations. Furthermore, workload is used as an important causal factor in the majority of folk models of human performance. It is therefore a measure defined by consensus, rather than by reference to a model, which usually come afterwards.

Folk models describe measures that reflect an important aspect of the operators' situation, but related to intermediate "cognitive" states rather than to the actual performance. It is assumed that the measurement is a valid substitute for actual performance measurement, because it refers to an essential intermediate or intervening state. It is also assumed that the measurement is affected by the performance conditions to the same extent and in the same manner as the actual performance. Yet it defies reason why it should be more important to look for measurements of hypothetical internal states than for measurements of the performance that admittedly is determined by the internal states!

In relation to operator performance, measurements proposed by folk models represent commonly-held notions about the nature of human work, and specifically about the nature of human cognition. At present, in other words at the beginning of the 21st Century, the main concepts that are used to describe the cognitive aspects of work are, for instance, attention control, working memory management, mental workload, situation awareness, the operator's mental model, the processes or patterns of reasoning, and meta-cognitive self-monitoring. While it clearly is easier to propose a measurement for some of these concepts than for others, the ease by which measurement tools can be developed does not necessarily reflect the significance or validity of that measurement.

3.1 Measurement possibility versus interpretation

If we consider the range of measurements that are typically employed in empirical research, particularly in experimental (laboratory) versions of that, it is possible to discern a relation between how difficult or laborious it is to make a measurement and how meaningful it is. Figure 2 shows this relationship for some of the more common measurements.

As Fig. 2 suggests, the various measurements seem to be distributed about a diagonal. Many measurements can be made with relatively little effort, but also have a limited theoretical basis and are difficult to interpret. This is typically the case of measurements that can be easily recorded by mechanical means, such as measurements of physiological variables (heart rate) or overt performance (audio, video recordings). Other measurements have an acceptable theoretical foundation, but are either difficult to make or difficult (and laborious) to interpret. Examples of that are classical eye movement recordings or performance "errors". It would clearly be very useful if measures could be proposed which were both easy (and reliable) to make and meaningful. It follows from the preceding arguments that such measures must be based on an articulated model, rather than a folk model. In particular, substituting one construct by
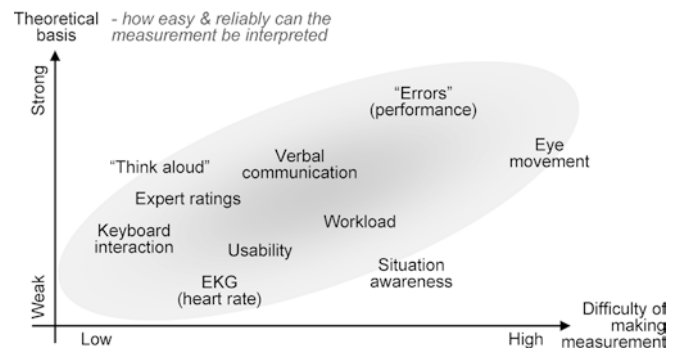

Fig. 2 Meaning of measurements

another works against the principles for constructing good measurements.

3.2 Folk models versus young and promising models

Although folk models clearly have their problems, one risk in rejecting them outright is that the baby is thrown out with the bath water. In other words, there is the risk of rejecting even those models that may be able to generate useful empirical results, if only given the time and opportunity to do so. Indeed, the more articulated human factors constructs (such as decision making and attention) are distinguished from the less articulated ones (such as situation awareness and complacency) by their maturity – although this is not in itself a sufficient quality. We may therefore rightly ask what opportunity the newer folk models should receive before being rejected as unproductive? The answer to this question hinges, once again, on falsifiability. Progress in science is often described as the succession of theories, each of which is more falsifiable (and therefore more informative) than the one before it. Yet if we assess "loss of situation awareness" or "complacency" as more novel explanations of phenomena that were previously covered by other explanations, it is easy to see that falsifiability has actually decreased, rather than increased.

Take as an example an automation-related accident that occurred in 1973, when situation awareness or automation-induced complacency had not yet come into use. The aircraft in question was on approach in rapidly changing weather conditions. It was equipped with a slightly deficient "flight director" (a device on the central instrument showing the pilot where to go, based on an unseen variety of sensory inputs), which the captain of the airplane distrusted. The airplane struck a seawall bounding Boston's Logan airport about one kilometre short of the runway and slightly to the side of it, killing all 89 people onboard. In its comment on the crash, the transport safety board explained how an accumulation of discrepancies, none of which were critical in themselves, had rapidly brought about a high-risk situation without positive flight management. The first officer, who was flying, was preoccupied with the information presented by his flight director systems, to the detriment of his attention to altitude, heading and airspeed control (National Transportation Safety Board 1974).

Today, both automation-induced complacency of the first officer and a loss of situation awareness of the entire crew would most likely be cited under the causes of this crash. (Actually, that the same set of empirical phenomena can comfortably be grouped under either label – complacency or loss of situation awareness – is additional testimony to the undifferentiated and underspecified nature of these concepts.) These "explanations" (complacency, loss of situation awareness) were obviously not needed in the early 1970s to deal with such accidents. The analysis instead proposed a set of more detailed, more falsifiable, and more traceable assertions that linked features of the situation (such as an accumulation of discrepancies) to measurable or demonstrable aspects of human performance (diversion of attention to the flight director versus other sources of data). The decrease of falsifiability represented by complacency and situation awareness as hypothetical contenders in explaining this crash represents the inverse of scientific progress, and therefore argues for extreme caution before accepting such novel models.

3.3 The vulnerability to folk modelling

Folk models are common in psychology (Stich 1985), one reason being that psychology is the science of the human mind. Since we all have a mind, and we can all claim that we think with this mind, we all have privileged knowledge about how this thinking process takes place, and therefore about how our mind works (Morick 1971). The deceptive ease by which we can extrapolate our everyday experiences (getting lost, forgetting, not paying attention, and so forth) to understand the complex events we hear or read about make us blind to the fact that such descriptions are not scientific explanations. Human factors, being both a science and a practice, may well be specifically vulnerable to this erroneous belief, hence to the use of folk models. Indeed, their underspecificity and facile ability to quickly explain complex behavioural sequences is at the same time a powerful attraction for practitioners and a source of problems for scientists.

The greatest risk of folk models is that they appear to make sense, even though statements and conclusions may not be falsifiable. They may therefore seem more plausible than articulated models, since the latter require an understanding of the underlying mechanisms. As observed by Weick (1995), while accuracy or comprehensiveness are rarely criteria for a successful explanation, plausibility is. Plausibility is essential from the point of view of those who have to accommodate a complex event in their local situation, or who have to deal with what it means for *them* and *their* organisation. Explanations of human performance, especially in high-stake events must meet this goal of plausibility even if – in the words of Weick (1995) – they make lousy history.

4 Conclusions

One issue that often comes up is the difficulty of measuring mental states such as stress or workload, either retrospectively or concurrently. In the case of accidents or incidents, for instance, it is often suspected that the operator (or pilot, or driver...) was in a condition of mental stress or high workload, and therefore failed in some action – perhaps even made a "human error". The problem is that the normally available records of behaviour are singularly unsuitable for playing this kind of retrospective "what-if" game.

In most cases, such as industrial process environments, there is no record of what operators have done as such. The only data come from the logs of system events and parameter changes, which mean that all that can be seen are the changes to the process (control input) made by the operators. Of the operators' own performance there is no indication.

As an aside, the common way out of this predicament is to study human performance in simulators, and to rely on high fidelity simulators to increase the validity of the results. Unfortunately, this does really not solve the problem. The simulator may provide the possibility of taking many more measurements of operator behaviour, including physiological and psychological ones, but does still not permit a direct "reading" of mental states. Researchers are easily misled by the wealth of data to believe they have quality where there is only quantity. The basis for finding out about what happens in the operators' minds is admittedly better in a simulator than in real life. But the principal obstacles described below remain.

In the case of pilots, the situation is a little bit better, since all commercial aircraft today are equipped with various recording devices (CVR and FDR) – known in the common parlance as the black boxes. The problem is, however, that it is still extremely difficult to assess something like the pilot's level of stress from these data.

A little reflection will reveal that this problem is mostly an artefact of how we think about human performance. The common way of thinking about performance implies that there is a set of internal or mental processes ("cognition in the head") that determines what people see and do. Consequently, in our efforts to understand events we search for these internal processes. Even if the information-processing model of human behaviour was correct – which, by the way, it is not – this approach means that one is looking for a hypothetical intervening variable, rather than for the more manifest aspects of behaviour. Assuming further that we could reliably measure the level of stress, concurrently or retrospectively, we still have no clear explanation of how the level of stress leads to specific types of performance, specifically how it leads to failures of performance. All we have is a folk model of stress and performance, which probably can be traced back to the Yerkes and Dodson experiments mentioned previously.

There is, however, an alternative, namely to focus on the characteristics of performance rather than on inferred and uncertain states of the mind. In other words, to study what has come to be known as "cognition in the world" – although this term is meaningful only as an antithesis to "cognition in the mind". (We should furthermore not be too concerned about cognition, but rather be concerned about performance.) One point in favour of that is that we actually do have records of performance. That is exactly what is contained in the systematic recordings of system status, if we change the focus from the performance of the individual to the performance of the joint system. Following the principles of cognitive systems engineering (Hollnagel and Woods 1983) we should not be overly concerned with the performance of the pilot *per se*, but rather with the performance of the pilot + aircraft – in other words, the joint pilot-aircraft system. (This line of thinking can, of course, be extended to the aircraft-ATC system, the operator-process system, and so on.)

It is a consequence of this approach that the analysis should look at the orderliness of performance rather than at the mental states of people. The orderliness of performance must inevitably refer to some kind of theory or model for performance, but this can be a fairly simple one if it is based on the concept of control (Hollnagel 1998a). The theory, such as it is, simply recognises that control in a joint system can vary from a very low degree of control – or even no control at all – to a very high degree of control. For practical reasons some characteristic regions of control have been identified and characterised (Hollnagel 1993). In the context of the current discussion, however, the main advantage of this perspective is that we can look at the actual performance of the system over time, and describe it in terms of the level of control. Looking at the smoothness of the flight path, for instance, as a trajectory in a 4-D space, may give an indication of the level of control. Looking at the actions of the pilots in the cockpit, as well as their communication, may do the same, for instance whether their actions were orderly and seemingly followed a plan, or whether they corresponded to a more fragmented strategy (of which there are a number of well-known ones). If this analysis finds that the orderliness of performance broke down, then we may begin to speculate about the possible reasons – possibly invoking concepts such as high workload. But such speculations should come at the end of the investigation rather than at the start.

References

Aeronautica Civil de Colombia (1996) Aircraft accident report: Controlled flight into terrain, American Airlines flight 965, Boeing 757–223, N651AA near Cali, Colombia, 20 December 1995. Aeronautica Civil, Bogota, Colombia

Attneave F (1959) Applications of information theory to psychology: A summary of basic concepts, methods, and results. Holt, Rinehart and Winston, New York, NY

Billings CE (1996) Situation awareness measurement and analysis: a commentary. In: Garland DJ, Endsley MR (eds) Experimental analysis and measurement of situation awareness. Embry-Riddle Aeronautical University Press, Daytona Beach, FL, p 3

Buck RN (1995) The pilot's burden: Flight safety and the roots of pilot error. Iowa State University Press, Ames, IA

Campbell RD, Bagshaw M (1991) Human performance and limitations in aviation. Blackwell Science, Oxford, UK, p 126

Endsley MR (1999) Situation awareness in aviation systems. In: Garland DJ, Wise JA, Hopkin VD (eds) Handbook of aviation human factors. Lawrence Erlbaum Associates, Hillsdale, NJ, pp 257–276

Federal Aviation Administration (1996) The interface between flightcrews and modern flight deck systems. Author, Washington DC

Hollnagel E, Woods DD (1983) Cognitive systems engineering: New wine in new bottles. Int J Man Mach Stud 18:583–600

Hollnagel E (1993) Human reliability analysis: Context and control. Academic, London

Hollnagel E (1998a) Context, cognition, and control. In: Waern Y (ed) Co-operation in process management – Cognition and information technology. Taylor and Francis, London

Hollnagel E (1998b) Measurements and models, models and measurements: You can't have one without the other. In: NATO RTO Meeting Proceedings 4, Collaborative Crew Performance In Complex Operational Systems, 20–22 April 1998, Edinburgh, Scotland (TRO-MP-4 AC/323(HFM)TP/2)

Kahneman D (1973) Attention and effort. Prentice-Hall, Englewood Cliffs, NJ

Kern T (1998) Flight Discipline. McGraw-Hill, New York, NY, p 240

Lindsay PH, Norman DA (1977) Human information processing: An introduction to psychology, 2nd edn. Academic, New York

Morick H (1971) Cartesian privilege and the strictly mental. Philos Phenom Res 31(4):546–551

National Transportation Safety Board (1974) Delta Air Lines Douglas DC-9–31, Boston, MA, 7/31/73 (NTSB/AAR-74/03). NTSB, Washington DC

National Transportation Safety Board (1994) Safety study: A review of flightcrew-involved major accidents of U.S. air carriers, 1978 through 1990 (NTSB/SS-94/01). NTSB, Washington DC

Norman DA (1976) Memory and attention, 2nd edn. Wiley, New York

O'Hare D, Roscoe S (1990) Flightdeck performance: The human factor. Iowa State University Press, Ames, IA, p 117

Parasuraman R, Molly R, Singh I (1993) Performance consequences of automation-induced complacency. Int J Aviat Psychol 3(1):3

Popper KR (1972) The logic of scientific discovery. Hutchinson, London

Sarter NB, Woods DD (1997) Teamplay with a powerful and independent agent: Operational experiences and automation surprises on the Airbus A-320. Hum Factors 39(4):553–569

Stich S (1985) From folk psychology to cognitive science: A case against belief. MIT Press, Cambridge, MA

Stokes A, Kite K (1994) Flight stress: Stress, fatigue and performance in aviation. Avebury Aviation, Aldershot, UK

Waldman RH (1999) Cockpit discipline. J Professional Aviation Training 1(6):10–15

Weick KE (1995) Sensemaking in organizations. Sage, London

Wiener EL (1988) Cockpit automation. In: Wiener EL, Nagel DC (eds) Human factors in aviation. Academic, San Diego, CA, p 452

Woods DD (1993) Process-tracing methods for the study of cognition outside of the experimental laboratory. In: Klein GA, Orasanu J, Calderwood R, Zsambok CE (eds) Decision making in action: Models and methods. Ablex, Norwood, NJ, pp 228–251

Yerkes RM, Dodson JD (1908) The relation of strength of stimulus to rapidity of habit-formation. Journal of Comparative and Neurological Psychology 18:459–482