# Why are people's decisions sometimes worse with computer support?

Eugenio Alberdi[1], Lorenzo Strigini[1], Andrey A. Povyakalo[1], and Peter Ayton[2],

[1]Centre for Software Reliability, City University London, London, UK
[2] Psychology Department, City University London, London, UK
{e.alberdi, strigini, povyakalo}@csr.city.ac.uk, P.Ayton@city.ac.uk

**Abstract.** In many applications of computerised decision support, a recognised source of undesired outcomes is operators' apparent over-reliance on automation. For instance, an operator may fail to react to a potentially dangerous situation because a computer fails to generate an alarm. However, the very use of terms like "over-reliance" betrays possible misunderstandings of these phenomena and their causes, which may lead to ineffective corrective action. For instance, training or procedural changes are favored responses, but they do not address all causes of apparently "over-reliant" behaviour. We review relevant literature in the area of "automation bias" and describe the diverse mechanisms that may be involved in human errors when using computer support. We discuss these mechanisms, with reference to errors of omission when using "alerting systems", with the help of examples of novel counterintuitive findings we obtained from a case study in a health care application, as well as other examples from the literature.

**Keywords:** decision support, computer aided decision making, alerting systems, human-machine diversity, omission errors

## 1 Introduction

It has long been known that introducing automation might have unexpected side effects on human performance [1, 2]. For instance, consider a computer tool designed to highlight targets of interest on a radar screen. If the computer does not highlight one such target, even an experienced radar operator could be led to miss that target, even if he would not have missed it without the computer aid. Such phenomena are often attributed to complacency, which makes operators abdicate their responsibility to the automated support. Given this interpretation, a tool designer may assume this to be the main risk, and so proper training and indoctrination is the natural defence (e.g. [3]); this attitude is widespread in practice. We argue that this view is too simplistic, and present a much richer picture of unintended, subtle effects that automation may have and which a designer needs to be prepared to guard against.

Automation is increasingly taking on the role of *supporting* knowledge-intensive human tasks rather than directly *replacing* some of the human's functions. This actually makes the problem of computer-related human errors subtler. The responsibility for correct action rests with the user. One might think that user mistakes can be reduced by simple training or, sometimes, by a user interface that prevents those mistakes. But in practice computers and their users form human-computer systems, or "socio-technical systems", which need to be assessed as whole systems from the viewpoints of reliability and safety. Examples of these supportive systems are *alerting* systems: from spell-checkers to alarm-filtering systems for industrial control rooms through collision warning systems in transportation or computerised monitoring in health care. In these monitoring applications, automation typically assists the operator in judgement-oriented tasks – like dealing with anomalies and taking high-level decisions – by adding to situational data broadly "advisory" input: attention cues, pre-filtered alarms, suggested diagnoses, or even recommended manoeuvres. If operators "trust" the computer's help too much or too little [4-6], compared to their own judgement skills, reliability and safety of operation may suffer. Labels used in the literature are: "automation bias", automation-induced "complacency" [7-9], "over-reliance" on automation [10], "automation dependence" [11] or computer induced "confirmation bias" [12].

The purpose of this paper is to both review and broaden the set of explanatory mechanisms proposed in the literature for undesired effects of automation. We argue that while "complacency" often seems the natural explanation for such effects, they may often instead be the result of complex cognitive mechanisms in decision making under uncertainty. At each demand for a decision, the operator's use of computer help depends on the details of that individual demand as well as on the operator's skills and the computer's design. Performance can be influenced by all of these factors as well as interactions among them. We present a (non exhaustive) set of possible cause-effect mechanisms contributing to human error. Due to space restrictions, we focus on: *errors of omission* (human failure to react to target events) when using computerised *alerting tools*. The intention is to help designers of these socio-technical systems (i.e., the combination of computer algorithms, user interfaces, procedures, training protocols, etc.) to adopt appropriate defences to match these diverse threats.

In what follows, we present: an overview of the human factors literature on automation bias and related concepts (section 2); a brief description of a case study in the area of computer-assisted cancer detection, which has motivated many of the analyses and conclusions presented in this paper (section 3); an outline of the mechanisms contributing to errors of omission by computer-assisted operators (section 4); a discussion of the uses and limitations of this descriptive approach (section 5); and conclusions (section 6).

## 2 Literature on Automation Bias, Complacency and Trust

### 2.1 Scope and Terminology

Our review focuses on computer assisted monitoring or decision making, where an automated alerting tool supports human decisions with some form of non-binding "advice", which can take the form of filtered or enhanced information, alerts and prompts.

The operating scenario we envisage is that, for the *user* or *operator*, *demands* for action may arise (for instance, a patient's vital sign indicate an impending crisis; two vehicles are approaching a potential collision; a word is misspelled in a document). The user sees the *raw data* about the situation (pulse, blood pressure, etc. for a patient; position and motion vectors of vehicles, visually estimated or displayed on a radar screen; the text of the document) in which s/he needs to detect *cues* (specific combinations of ranges of vital signs, or distance and velocities, or the misspelled word itself) and assess them and, if necessary, make an *alarm response*, such as applying emergency treatment to a patient, initiate evasive manoeuvres, search for an alternative spelling of a word. A cue may indicate a *target* (real need for an alarm response: a demand implies the presence of at least one target), but the user needs to apply skill and knowledge to decide whether a given cue actually represents a target. To support the user, the computerised warning *tool* is designed to provide *prompts* (e.g. visual highlights on a screen) that point at cues for consideration. In this initial analysis, we do not consider the possibility that the tools also suggest specific actions. There is the possibility of the tool *missing* targets (*false negative* error, or *FN*), as well as of *false prompts* (*false positive* error, or *FP*). The tool can be assessed in terms of its probabilities of FN or FP errors, or equivalent pairs of measures (e.g. sensitivity/specificity are often used in the medical literature).

The errors of the human-computer system can also be classified into false negatives (the user fails to initiate an alarm response despite a target being present) and false positives (the user initiates an alarm response in the absence of a target) and the system's dependability described by FN and FP error probabilities (or equivalent pairs of measures). Another important figure is the *alarm response rate* – the cumulative frequency of alarm responses, either correct or spurious – since these are costly and most systems can only function if this rate is less than a certain threshold.

At this juncture it is useful to introduce some terminology from the human factors literature to contextualise the scope of systems and of errors that we cover here.

Parasuraman and Riley [10] discussed different ways in which human-computer interaction can go wrong and talked about three aspects of ineffective human use of automation: *disuse*, i.e., underutilization of automation, where humans ignore automated warning signals; *misuse*, i.e., over-reliance on automation, where humans are more likely to rely on computer advice (even if wrong) than on their own judgement; *abuse*, when technology is developed without due regard for human needs or the consequences for human (and hence system) performance and the operator's authority in the system.

Skitka and colleagues [13] focused on the *misuse* of automation, in particular on the "automation bias" effects occurring when people used wrong computer advice for monitoring tasks in aviation. They distinguished two types of computer-induced error: a) *errors of commission*: decision-makers follow automated advice even in the face of more valid or reliable indicators suggesting that the automated aid is wrong; b) *errors of omission*: decision makers do not take appropriate action, despite non-automated indications of problems, because the automated tool did not prompt them.

Focusing on *warnings* generated by automated tools, Meyer [14], distinguishes between two alternative ways in which humans can "follow" or "conform to" the advice from a warning system: compliance and reliance. *Compliance* indicates that the operator acts according to a warning signal and takes an action. *Reliance* is used to describe those situations where the warning system indicates that "things are OK" and the operator accordingly – i.e. not merely coincidentally – takes no action.

We can see that, combining Skitka's and Meyer's terminologies, undue *compliance* (complying with an incorrect automated warning) would lead to *errors of commission* and undue *reliance* (failing to take action when no automated warning is issued) would lead to *errors of omission.*

### 2.2 Automation Bias, Complacency and Trust

The phrase "automation bias" was introduced by Mosier et al. [15] when studying the behaviour of pilots in a simulated flight. In this study, they encountered both omission and commission errors. These findings were then replicated with non-pilot samples (student participants) in laboratory settings simulating aviation monitoring tasks [13]. They found that, when the automated tool was reliable, the participants in the automated condition made more correct responses. However, participants with automation that was imperfect (i.e. occasionally giving unreliable support) were more likely to make errors than those who performed the same task without automated advice. In Skitka and colleagues' studies, the decision-makers had access to other (non automated) sources of information. In the automated condition they were informed that the automated tool was not completely reliable but all other instruments were 100% reliable. Still, many chose to follow the advice of the automated tool even when it was wrong and was contradicted by the other sources of information. The authors concluded that these participants had been biased by automation and interpreted their errors (especially their errors of omission) as a result of complacency or reduction in vigilance.

Factors that have been investigated in empirical studies as possible influences in people's vulnerability to automation bias include: individual differences among operators [5, 13, 16, 17]; people's accountability for their own decisions [17]; the levels of automation at which the computer support is provided [12, 18]; the location of computer advice/warnings with respect to raw data or other non-automated sources of information [19, 20]; people's exposure to automation failures [21].

People's ineffective use of computerised tools is often described in terms of "complacency", which is said to cause over-reliance or "uncritical reliance" on automation [7-10]. However, there is no general agreement about what exactly is "complacency" and what are the best ways to measure it [16]. What seems to be common to most characterisations is a sense of contentment, unawareness of dangers or deficiencies and failure to look for evidence or to examine the raw data in a careful enough manner.

However, a problem with terms like "complacency" is that they convey value judgments on the human experts. Moray [22] points out that the claim that automation fosters complacency suggests that operators are at fault and argues that the problem often lies in the characteristics of the automated tools, not in the human operators' performance. Similarly, Wickens and Dixon [23] question the notions of complacency or reduced vigilance as explanations of automation bias. Instead, they argue that operators, whilst being aware of the unreliability of the diagnostic tools, choose to depend on the imperfect computer output to keep their cognitive processing resources for other tasks, particularly in situations with high workload.

Another concept that is frequently invoked when talking about automation bias or (over)reliance on automation is "trust" [4-6, 8, 10, 24-29]. The common assumption is that the more a human operator *trusts* an automated aid the more likely s/he is to *rely* on or *comply* with the advice provided by the aid. If a human trusts an aid that is adequately reliable or fails to trust an aid that is indeed too unreliable, appropriate use of automation should occur as a result. However if a human trusts (and therefore follows the advice of) an unreliable tool, then automation bias may occur (or misuse of automation as defined above). Similarly if a person does not trust a highly reliable tool, the person may end up disusing (as defined above) or under-using the tool, and hence the full potential benefits of automation will not be fulfilled.

Subjective measures of the trust of human operators in a computer tool have been found to be highly predictive of people's frequency of use of the tool [5, 30]. Use of automation (or reliance in its generic sense) is usually assessed with observations of the proportion of times during which a device is used by operators or by assessing the probability of operators' detecting automation failures [19].

Factors that have been investigated in empirical studies as possible influences in people's trust in automation include: people's exposure to automation errors [5, 30, 31], the consistency of the tool's reliability [16, 32], the invasiveness or intrusiveness of the tool's advice [33, 34].

## 3 A Case Study: Computer Aided Detection (CAD) for Mammography

Many of the considerations we present originate from a case study we conducted in the area of CAD for breast cancer screening [35-40]. In breast screening, expert clinicians ("readers") examine *mammograms* (X-ray images of a woman's breasts), and decide whether the patient should be "recalled" for further tests because they suspect cancer. A CAD tool is designed to assist the interpretation of mammograms primarily by alerting readers to potentially cancerous areas that they may otherwise overlook. CAD is not meant to be a diagnostic tool: its prompts areas of the mammograms, but it is up to the reader to classify them as potentially cancerous or not and reach the "recall/no recall" decision. According to the intention of the designers, it should help to *avoid* a cancer being missed but *not cause* a cancer to be missed.

Our study provided evidence of automation bias effects in the use of CAD; effects which could not be attributed to complacency and could actually coexist with users' reported mistrust towards the tool [35]. Previous studies had concluded that on average using CAD was either beneficial or ineffectual. Our analyses indicated instead that, although CAD reduced decision errors by some readers on some cases, it also increased errors by some readers on some cases. In short, this simple computer-assisted task hid subtle effects, easy to miss by designers and assessors [37].

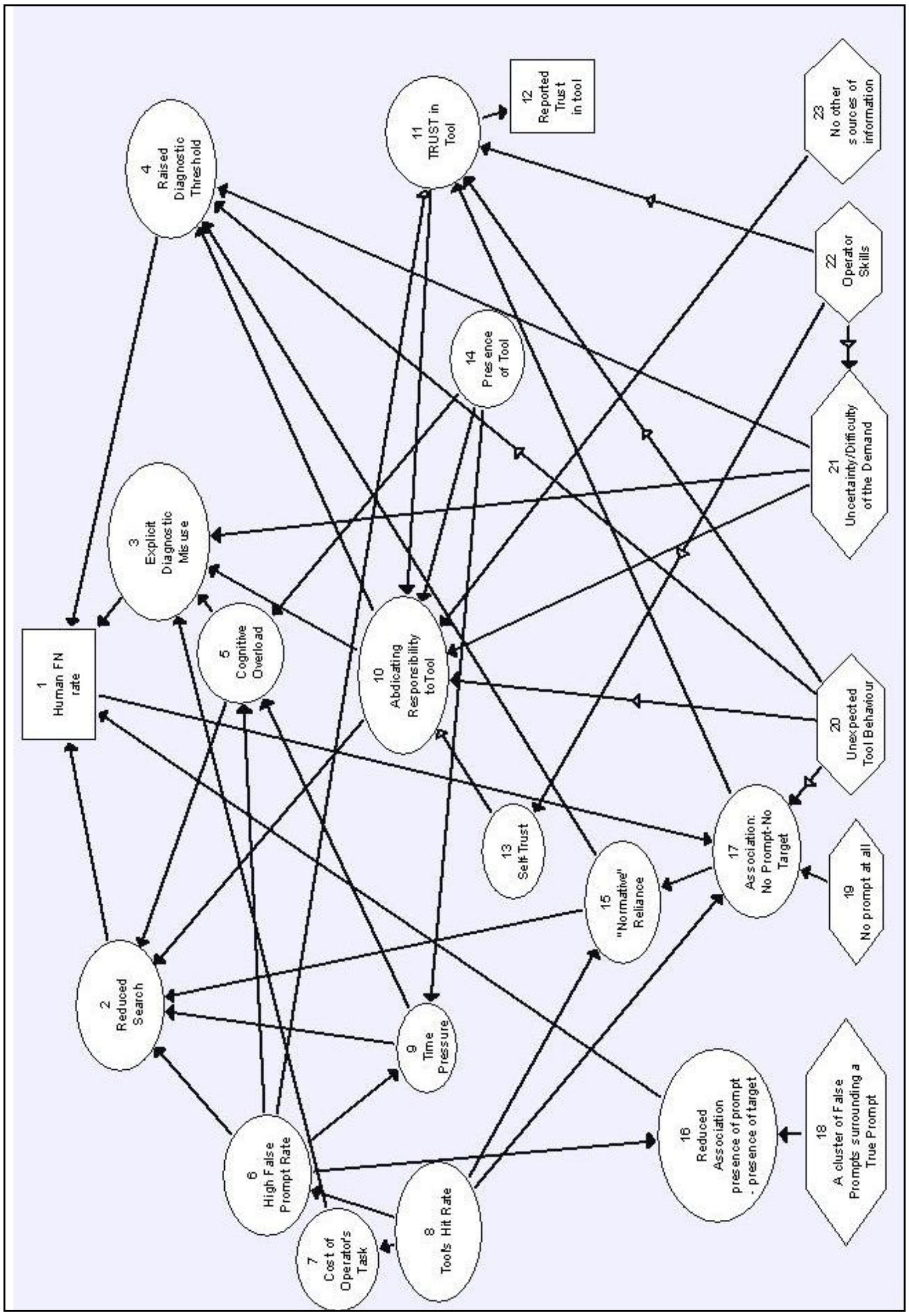
1
Human FN rate
2
Reduced Search
3
Explicit Diagnostic Misuse
4
Raised Diagnostic Threshold
5
Cognitive Overload
6
High False Prompt Rate
7
Cost of Operator's Task
8
Tool's Hit Rate
9
Time Pressure
10
Abdicating Responsibility to Tool
11
TRUST in Tool
12
Reported Trust in tool
13
Self Trust
14
Presence of Tool
15
"Normative" Reliance
16
Reduced Association presence of prompt - presence of target
17
Association: No Prompt-No Target
18
A cluster of False Prompts surrounding a True Prompt
19
No prompt at all
20
Unexpected Tool Behaviour
21
Uncertainty/Difficulty of the Demand
22
Operator Skills
23
No other sources of information

**Fig. 1.** Cause-effect chains leading to omission errors by computer-supported operators

## 4 Diverse Causes of Errors by Humans with Computer Support

Figure 1 shows a graphical representation of cause-effect chains involved in "errors of omission", as an incomplete but complex account of "automation bias". In the graph, rectangles denote observable behaviours; oval shapes represent causal factors (characteristics of the tool and/or of the user, including cognitive mechanisms and affective states) that may be present in the human-computer system, although perhaps not directly observable; and the diamond-like shapes (all at the bottom of the graph), characteristics of a specific demand and/or user that may trigger the effects of one or the other of the oval nodes. The lines between nodes indicate causal links. A black arrow indicates an "increase" relationship (i.e., an increase or intensification of the factor identified by the source node leads to a change in the same direction for the target node); a white arrow indicates a "decrease" relationship (an increase of the source node factor leads to a decrease of the target node factor); lines with both a black and a white arrow indicate that there is an influence but the direction of change can go either up or down depending on the circumstances. Multiple arrows into a node have an "OR" semantics: any one of the source nodes may affect the target node, irrespective of whether other source nodes do.

As noted, we focus on human errors of omission, exemplified by node 1 in the graph: "Human FN (false negative) rate". This node denotes the increased likelihood that a human's FN rate is higher when using computer support than when not using it. In mammography, a human FN is a radiologist's failure to recall a patient whose mammogram contains indications of cancer that s/he has missed or misinterpreted; in collision warning systems, a human FN is an operator's failure to notice the proximity between two vehicles or aircraft and her/his consequent failure to initiate evasive manoeuvres or give the necessary directions to colleagues.

We represent in nodes 2-4 our three main conjectures about how this increase in operator's FN rate comes about (possibly just three very plausible examples out of many other possible contributing mechanisms). Node 2 refers to the processing of raw data (the detection of or search for target cues). Nodes 3 and 4 refer to "diagnostic" aspects of the decision making (i.e., the interpretation or classification of the raw data once the operator has collected or detected them). More specifically:

- Node 2, "Reduced Search": the operator fails to either complete the search for all possible cues (e.g. suspicious features in a mammogram) or to examine all the necessary raw data to make a decision.
- Node 3, "Explicit Diagnostic Misuse": the operator, in deciding the value of a cue towards a decision, gives the tool's prompts more weight than intended by the designers. For example, in CAD for mammography, the prompts are meant as pure alerts, without diagnostic value and the procedure prescribed that if a user had decided to recall a case before seeing the prompts, s/he should not change her/his decision to "no recall" after seeing the prompts [37]. If a reader performs this forbidden action, it is explicit diagnostic misuse. By "explicit" we mean that such violations could be identified, e.g. by the user her/himself, differently from the form of potential tool misuse represented by the next node, 4.
- Node 4, "Raised Diagnostic Threshold": an operator raises the degree of "strength" or "severity" of cues that s/he requires in order to initiate an alarm response

*without a prompt from the tool*. For certain borderline cases the user, when not using computer support, might be cautious and give an alarm response; for example, when seeing a moderately suspicious feature on a mammogram, a reader recalls the patient for further examinations even if it is not clear that she may have cancer. But if "supported" by the tool, the operator may become "less cautious" when interpreting those cues; for example, in a first examination, the reader decides not to recall the patient and waits to see the CAD prompts before committing her/himself to a recall decision.

Let us discuss some of the different paths that can lead to these three "top level" nodes (and, ultimately, to raised human FN rate).

We start with node 8, the tool's sensitivity ("Tool's hit rate"), an "obviously" beneficial characteristic. Increasing tool sensitivity is, in principle, desirable; and this is a goal tool designers normally aim for. However, it may actually lead to undesirable effects because increasing it usually increases the rate of false prompts (link to node 6 in the graph). Processing false prompts can be costly. Radiologists, for example, are known to be concerned with explaining why each prompt is present [37, 41]. Also, in aviation, pilots using TCAS (Traffic Collision Avoidance System) are strictly instructed to regard all automated messages as genuine alerts demanding an immediate, high-priority response [42]. Processing false prompts demands time and cognitive resources, and thus can lead to "Time Pressure" (node 9) and "Cognitive Overload" (node 5: presence of confusion that does not allow the operator to process information properly). Time pressure and cognitive overload are indeed interconnected and both reduce the operator's ability to complete the search for cues (links to node 2). It is important to note that none of the mechanisms just described (in connection with the tool's sensitivity) imply "over-reliance" on automation or "complacency". The tool affects the operators, but they are *not* conforming to its advice. In fact, operators' performance could be worse with computer support even for demands for which the tool provides *correct* advice. Evidence from the case study on CAD in breast screening strongly supports this view [35, 36]. Nodes 18 and 16, in conjunction with node 6, illustrate the "cry wolf" situation that may explain phenomena like this. Imagine that a true prompt (e.g., one signalling cancer in a mammogram) is surrounded by a cluster of many obviously false prompts (node 18). The user may infer that prompts in this case are *not* correlated with the presence of cancer (node 16); the value of the true prompt gets diminished for the radiologist, leading her/him to overlook correct prompts.

The tool's sensitivity (node 8) can lead to unanticipated human error through what we call "Normative reliance" on the tool (node 15). By "normative" we mean it fits a "normatively correct", rational decision making process. This can take, at least, two different forms:

- Based on their experience with a highly sensitive tool, operators correctly use prompts as a sign of possible missed targets. This can lead to "Raised Diagnostic Threshold" (link to node 4) and eventually to increased operator FN rate (node 1) in the following way. If the tool is useful, it causes an increase of correct "alarm responses" but it may also increase the number of false alarm responses (human FPs). Operators know that too high an "alarm response rate" is unacceptable; for instance, too many false recalls may make a cancer screening program unable to cope with the true cases. Therefore, raising the operator's own threshold is a reasonable reaction, irrespective of whether it is intentional or not. However it may overcompensate, or at least make the operator miss some targets that s/he would not have missed without the tool, although overall s/he misses fewer targets with the tool.

- Many of the prompts are spurious, so operators correctly learn to associate "no prompt" with likely absence of target (node 17). This can lead to reduced data search (node 2). As a result, given a FN from the tool, the user's normative reliance on the tool will lead him/her to miss the target (node 1). A "rational" user will be especially likely to reduce the search in the light of absence of prompts if detailed analysis of every prompt is too demanding and, especially, if it is practically infeasible.

The association between absence of prompts and absence of target can lead to a different, less "rational", path, involving *trust* (node 11), an "affective" (rather than cognitive) state, which may be affected by experience of reliability, but also by many other factors, and may be far stronger or far weaker than warranted by experience. Here we envisage complacency, represented by "Abdicating responsibility to tool" (node 10) as the result of a person's "negotiation" between the trust s/he has in her/his own abilities (node 13) and her/his trust in the tool. Expert operators often have beliefs about what tasks they are good at and what tasks they are less competent at [37]. If the user trusts the tool more than her/himself for a particular task, s/he will be more likely to over-rely on it (i.e., relinquishing responsibility to automation). Various (non exhaustive) links in the graph indicate the various factors or mechanisms that may affect trust.

There are also situations when people abdicate responsibility to the tool even if they do not trust it. For example, the mere fact that the operator knows that computer support is available could in itself lead to complacency (links from node 14, "Availability of the computer tool" *per se*, to node 10), in a process equivalent to what some psychologists term "social loafing": when people work with other people, diffusion of responsibility often takes place [43, 44]. Importantly, specific situations with high degrees of uncertainty (node 21), especially when other more reliable sources of information are missing (node 23), may make operators vulnerable and cause them to rely on computer support more than they would normally do, even if they do not trust its reliability. We found evidence for this in our study of CAD use with difficult-to-detect cancers [35].

Node 14 designates other ways in which the "Presence of the Tool" *per se* (no matter how reliable) can also contribute to human error without over-reliance or complacency. For instance, the need to examine and process the tool's output may in itself increase time pressure (node 9) and cognitive load (node 5).

For the sake of brevity, we leave out of this exposition a few of the nodes and links in the graph, which we believe are self-explanatory.

# 5 Discussion

## 5.1 Uses of this approach

The main purpose of the diagram in Fig 1 is to assist a designer or assessor in identifying the causal chains leading to undesired effects. A designer can try to interrupt the chain by appropriate design decisions. The fact that the graph represents multiple interacting causal chains should help against tunnel vision, i.e., focusing on one obvious concern while ignoring others. For instance, a designer might try to counteract factor represented by node 2 in the graph, assuming it is mainly caused by factor in node 17, via procedural restrictions, such as requiring that the user reach a provisional decision and take responsibility for it (e.g. by recording it in a log) before seeing the tool's prompts. But this remedy might not work against factor in node 2, or might even make it worse if mandating this more complex procedure exacerbates factor in node 9; or if, despite factor in node 2 being alleviated or eliminated, the main (neglected) mechanism through which the tool causes certain extra false negative decisions is node 4.

So far we have talked about the need for completeness in analyses. In designing a human-computer system, it would be good to focus on those possible causal chains that will be important in a specific system and context of operation. For this kind of optimisation, one needs empirical observations in the environment of use, if feasible. To help when these observations are not available, further research should try to identify general rules for forecasting the relative importance of the different mechanisms in a future system and environment of use. Last, system designers may wish to incorporate in the design degrees of “tuneability” for the parameters (of the algorithms in the tool, the procedure for using it, etc.) to allow adjustments in operation, so as to achieve good trade-offs between positive and negative effects.

## 5.2 Limitations, quantitative aspects

We highlight next some problems that this descriptive approach does not address. Quantitative trade-offs may be necessary in design. The relative importance of the various causal mechanisms in the graph will vary between systems, between users in the same system and between demands. This is because the parameters of human reactions to cues and prompts may well vary between categories of demands, just as those of the tool's reactions do (e.g., being better at detecting and prompting certain kinds of cues than others). Especially with increasing experience, a user might, for instance, learn to trust a computer’s prompts highly for certain types of demand, an only little for others. A support tool may have a positive effect on the reactions of certain population of users to most demands, but still have a detrimental effect on some categories of demands for a subset of those users. These factors may require designers to consider quantitative trade-offs, and to assess the effects of uncertainties about the environment of use of an alerting tool.

We have modelled [40] the cumulative effect of these different reaction patterns, to quantitatively identify possible design trade-offs, showing that complex effects are possible. Depending on the trade-offs made by designers and their effects on various classes of demand (and the frequencies of these classes of demands), a tool designed to help might have a damaging effect (aggregated over the whole population of users and distribution of demands). Much more commonly, improving the aggregated dependability of the socio-technical system requires consideration of the various design trade-offs affecting the overall FN and FP rates for all classes of demands. For instance, if factor in node 4 in the graph causes, on average, operators to end up with a few more false negatives on a difficult but rare class of demands, while allowing them to reduce false negatives – without an excessive increase in false positives – on a more common class of demands, the net effect may be beneficial. A possible complication is that of errors causing different degrees of loss depending on the class of demands: in the above example, if FNs on the class of "difficult" demands tended to cause more serious consequences, using the tool might increase the overall amount of loss caused by the decisions compared to the unaided user. Even with a tool whose aggregated effect is unambiguously positive, its potential for increasing human FNs on specific classes of demands may cause concerns. For example, for a medical decision aid, the net effect may be a transfer of risk from certain patients to others: introducing the aid might reduce risk for the average patient and yet increase risk for the average patient from a certain age or ethnic group. Or the aid may have the effect of improving the performance of most doctors but making it worse for some specific doctors.

## 6 Conclusions

With reference to a category of computer-assisted human tasks, we have highlighted a variety of alternative mechanisms that could lead to omission errors by the computer assisted operators. We have shown that errors that are often ascribed to "complacency" or "over-reliance" on computers, can actually be caused by other mechanisms, in fact even when the operators do *not* trust the automated tool.

The various mechanisms are interrelated in complex ways, so that the presence and characteristics of the alerting tool may affect the FN rate in more than one way. If a designer focused on only part of the graph in our Fig. 1, trying to "cut" one of the links so as to defeat one of these damaging mechanisms, succeeding might not bring any benefit because, in the system, the predominant damaging mechanism may be another mechanism.

So, when designing a tool and the human-computer system to include it, it is certainly important to be aware of the risk of complacency (e.g. by prescribing appropriate training or procedures), but this may not be enough. In particular, we have shown that some of these error mechanisms may be an inherent part of the human cognitive apparatus for reacting to cues and alarms, so they cannot be effectively shut off. A proper design of the human-machine system would look for the best trade-off between the positive and negative effects, rather than assuming that negative effects can be completely eliminated; and evaluators and adopters, when assessing a design, need to be aware of these various facets of the effects of a tool.

The graph presented in Fig. 1, based on our deductions from empirical work and from prior literature, is likely to be incomplete; but it indicates a useful way towards more explicit and complete ways of considering error causes when designing human-computer systems.

**Acknowledgments.** This work was supported in part by the U.K. Engineering and Physical Sciences Research Council via project INDEED, “Interdisciplinary Design and Evaluation of Dependability” (EP/E000517/1) and by the European Union’s Framework Programme 6 via the ReSIST Network of Excellence, contract IST-4-026764-NOE.

## References

1. Bainbridge, L.: Ironies of Automation. Automatica 19, 775-779 (1983)
2. Sorkin, R.D., Woods, D.D.: Systems with human monitors: A signal detection analysis. Human-Computer Interaction 1, 49-75 (1985)
3. Hawley, J.K. Looking Back at 20 Years of MANPRINT on Patriot: Observations and Lessons. Report ARL-SR-0158, U.S. Army Research Laboratory, (2007)
4. Bisantz, A.M., Seong, Y.: Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. International Journal of Industrial Ergonomics 28 (2), 85-97 (2001)
5. Dzindolet, M.T., Peterson, S.A., Pomranky, R.A., Pierce, L.G., Beck, H.P.: The role of trust in automation reliance. International Journal of Human-Computer Studies 58 (6), 697-718 (2003)
6. Muir, B.M.: Trust between humans and machines, and the design of decision aids. International Journal of Man-Machine Studies 27, 527-539 (1987)
7. Azar, B.: Danger of automation: It makes us complacent. APA monitor 29 (7), 3 (1998)
8. Singh, I.L., Molloy, R., Parasuraman, R.: Automation-induced "complacency": development of the complacency-potential rating scale. International Journal of Aviation Psychology 3, 111-122 (1993)
9. Wiener, E.L.: Complacency: is the term useful for air safety? In 26th Corporate Aviation Safety Seminar, Flight Safety Foundation, Inc., 116-125 (1981)
10. Parasuraman, R., Riley, V.: Humans and automation: Use, misuse, disuse, abuse. Hum Factors 39, 230-253 (1997)
11. Wickens, C., Dixon, S., Goh, J., Hammer, B.: Pilot Dependence on Imperfect Diagnostic Automation in Simulated UAV Flights: An Attentional Visual Scanning Analysis. In Proceedings of the13th International Symposium on Aviation Psychology (2005)
12. Cummings, M.L.: Automation bias in intelligent time critical decision support systems. In AIAA 1st Intelligent Systems Technical Conference, AIAA 2004 (2004)
13. Skitka, L.J., Mosier, K., Burdick, M.D.: Does automation bias decision making? International Journal of Human-Computer Studies 51 (5), 991-1006 (1999)

14. Meyer, J.: Conceptual issues in the study of dynamic hazard warnings. Human Factors 46 (2), 196-204 (2004)

15. Mosier, K.L., Skitka, L.J., Heers, S., Burdick, M.: Automation bias: Decision making and performance in high-tech cockpits. International Journal of Aviation Psychology 8 (1), 47-63 (1998)

16. Prinzel, L.J., De Vries, H., Freeman, F.G., Mikulka, P. Examination of Automation-Induced Complacency and Individual Difference Variates. Technical Memorandum No. TM-2001-211413, NASA Langley Research Center, Hampton, VA. (2001)

17. Skitka, L.J., Mosier, K., Burdick, M.D.: Accountability and automation bias. International Journal of Human-Computer Studies 52 (4), 701-717 (2000)

18. Meyer, J., Feinshreiber, L., Parmet, Y.: Levels of automation in a simulated failure detection task. In IEEE International Conference on Systems, Man and Cybernetics, 2003, 2101- 2106 (2003)

19. Meyer, J.: Effects of warning validity and proximity on responses to warnings. Hum Factors 43, 563-572 (2001)

20. Singh, I.L., Molloy, R., Parasuraman, R.: Automation-induced monitoring inefficiency: role of display location. International Journal of Human-Computer Studies 46 (1), 17-30 (1997)

21. Bahner, J.E., Huper, A.-D., Manzey, D.: Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. Int. J. Human-Computer Studies 66, 688-699 (2008)

22. Moray, N.: Monitoring, complacency, scepticism and eutactic behaviour. International Journal of Industrial Ergonomics 31 (3), 175-178 (2003)

23. Wickens, C.D., Dixon, S.R. Is there a Magic Number 7 (to the Minus 1)? The Benefits of Imperfect Diagnostic Automation: A Synthesis of the Literature, University of Illinois at Urbana-Champaign, Savoy, Illinois, 1-11 (2005)

24. Dassonville, I., Jolly, D., Desodt, A.M.: Trust between man and machine in a teleoperation system. Reliability Engineering & System Safety (Safety of Robotic Systems) 53 (3), 319-325 (1996)

25. Lee, J.D., Moray, N.: Trust, self-confidence, and operators' adaptation to automation. International Journal of Human-Computer Studies 40, 153-184 (1994)

26. Lee, J.D., See, K.A.: Trust in computer technology. Designing for appropriate reliance. Human Factors, 50-80 (2003)

27. Muir, B.M.: Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. Ergonomics 37, 1905-1922 (1994)

28. Muir, B.M., Moray, N.: Trust in automation: Part II. Experimental studies of trust and human intervention in a process control simulation. Ergonomics 39, 429-460 (1996)

29. Tan, G., Lewandowsky, S.: A comparison of operator trust in humans versus machines. In Presentation of First International Cyberspace Conference on Ergonomics (1996)

30. de Vries, P., Midden, C., Bouwhuis, D.: The effects of errors on system trust, self-confidence, and the allocation of control in route planning. International Journal of Human-Computer Studies 58 (6), 719-735 (2003)

31. Dzindolet, M.T., Pierce, L.G., Beck, H.P., Dawe, L.A.: The perceived utility of human and automated aids in a visual detection task. Human Factors 44 (1), 79-94 (2002)

32. Parasuraman, R., Molloy, R., Singh, I.L.: Performance consequences of automation-induced "complacency". International Journal of Aviation Psychology 3, 1-23 (1993)

33. Bliss, J.P., Acton, S.A.: Alarm mistrust in automobiles: how collision alarm reliability affects driving. Applied Ergonomics 34 (6), 499-509 (2003)

34. Parasuraman, R., Miller, C.A.: Trust and etiquette in high-criticality automated systems. Communications of the ACM 47 (4), 51-55 (2004)

35. Alberdi, E., Povyakalo, A.A., Strigini, L., Ayton, P.: Effects of incorrect CAD output on human decision making in mammography. Acad Radiol 11 (8), 909-918 (2004)

36. Alberdi, E., Povyakalo, A.A., Strigini, L., Ayton, P., Given-Wilson, R.: CAD in mammography: lesion-level versus case-level analysis of the effects of prompts on human decisions. Journal of Computer Assisted Radiology and Surgery 3 (1-2), 115-122 (2008)

37. Alberdi, E., Povyakalo, A.A., Strigini, L., Ayton, P., Hartswood, M., Procter, R., Slack, R.: Use of computer-aided detection (CAD) tools in screening mammography: a multidisciplinary investigation. Br J Radiol 78 (suppl_1), S31-40 (2005)

38. Povyakalo, A.A., Alberdi, E., Strigini, L., Ayton, P.: Evaluating 'Human + Advisory computer' systems: A case study. In HCI2004,18th British HCI Group Annual Conference, British HCI Group, 93-96 (2004)

39. Povyakalo, A.A., Alberdi, E., Strigini, L., Ayton, P. Divergent effects of computer prompting on the sensitivity of mammogram readers, Technical Report, Centre for Software Reliability, City University, London, UK (2006)

40. Strigini, L., Povyakalo, A.A., Alberdi, E.: Human-machine diversity in the use of computerised advisory systems: a case study. In 2003 Int. Conf. on Dependable Systems and Networks (DSN'03), IEEE (2003)

41. Hartswood, M., Procter, R., Rouncefield, M., Slack, R., Soutter, J., Voss, A.: 'Repairing' the Machine: A Case Study of the Evaluation of Computer-Aided Detection Tools in Breast Screening. In Eighth European Conference on Computer Supported Cooperative Work (ECSCW 2003) (2003)

42. Pritchett, A.R., Vandor, B., Edwards, K.: Testing and implementing cockpit alerting systems. Reliability Engineering & System Safety 75 (2), 193-206 (2002)

43. Karau, s.J., Williams, K.D.: Social loafing: a meta-analytic review and theoretical integration. Journal of Personality and Social Psychology 65, 681-706 (1993)

44. Latanedo, B., Williams, K., Harkins, S.: Many hands make light the work: the causes and consequences of social loafing. Journal of Personality and Social Psychology 37, 822-832 (1979)