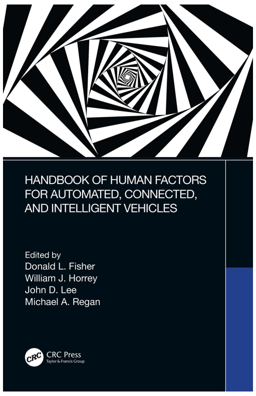

Handbook of Human Factors for Automated, Connected, and Intelligent Vehicles

Donald L. Fisher, William J. Horrey, John D. Lee, Michael A. Regan

Automation Lessons from Other Domains

Publication details
https://test.routledgehandbooks.com/doi/10.1201/b21974-21
Christopher D. Wickens
Published online on: 18 Jun 2020

21 Automation Lessons from Other Domains

Christopher D. Wickens
Colorado State University

CONTENTS

Key Points ... 455
21.1 Introduction ... 456
21.2 Classic Automation Accidents ... 456
21.3 Features of Automation ... 457
21.3.1 The Degree of Automation: What Does Automation Do and How Does It Do It? ... 457
21.3.2 Automation Reliability ... 459
21.3.3 Automation Trust and Dependence ... 460
21.3.4 Out-of-the-Loop-Unfamiliarity (OOTLUF) ... 460
21.3.5 Automation Modes and Complexity ... 461
21.3.6 Automation Transparency ... 461
21.3.7 Adaptable Versus Adaptive Automation ... 462
21.4 Research Findings ... 462
21.4.1 Accident and Incident Data Mining: Advantages and Costs ... 462
21.4.2 Experimental and Simulation Results ... 463
21.4.2.1 Alerting Systems ... 463
21.4.2.2 Attention Cueing ... 464
21.4.2.3 Stages 2 and 3 Automation: OOTLUF ... 465
21.5 Solutions: Countermeasures Proposed and Implemented in Other Domains ... 465
21.5.1 Flexible and Adaptive Automation ... 465
21.5.2 Automation Transparency ... 467
21.5.3 Training ... 467
21.6 Conclusions ... 468
References ... 468

KEY POINTS

- Automation can be defined by the stage of information processing for which it assists the human, and the level of assistance at each stage, together defining the degree of automation.
- The higher the degree of automation, the better is human–automation performance, the lower workload, and the lower situation awareness when

automation is working correctly; but the more problematic is human–automation interaction when automation fails.
- The problematic interaction results from failure to monitor what automation is doing, failure to understand its state, and degradation of skills when manual control must be exercised following the failure.
- Problems can be addressed by flexible or adaptive automation, automation transparency, and training, although benefits of adaptive automation can be difficult to realize.

21.1 INTRODUCTION

As the many other chapters in this book have made clear, automation takes many forms in vehicles. Foremost among these are the higher levels of automation and, ultimately, the total autonomy of self-driving cars. But there are numerous other examples of automation, such as headway monitors, auto-locks, a variety of alerts and warnings, anti-lock brakes and navigational systems. In designing all such systems to facilitate better interactions with the human, balancing safety versus productivity, many lessons can be drawn from accident analysis and research from other domains, particularly from aviation, which is the pioneer in the systematic investigation of human–automation (Wiener & Curry, 1980).

In this chapter, I will review the lessons that can be learned from human interaction with systems other than vehicles, including human flight in aviation (Billings, 1997; Ferris, Sarter, & Wickens, 2010; Landry, 2009), unmanned air vehicles (Cummings & Guerlain, 2007), medicine (Garg et al., 2005; Morrow, North, & Wickens, 2006), space (Li, Wickens, Sarter, & Sebok, 2014), air traffic control (Wickens, Mavor, Parasuraman, & McGee, 1998), military systems (Chen et al., 2018), consumer products (Sauer & Ruttinger, 2007), process control (Strobhar, 2012), robotics, and others.

This chapter will begin by providing a brief synopsis of three aviation tragedies directly related to breakdowns in human–automation interaction (HAI) when automation fails. Furthermore, it will describe in detail some of the key concepts in HAI that have arisen out of non-vehicle research, but are directly applicable to it. Then, I describe the empirical research bearing on several major issues in HAI. Finally, I turn to four suggested solutions to improving HAI, preventing its disasters without sacrificing the productivity that it offers, and examine the empirical evidence in support of those solutions.

21.2 CLASSIC AUTOMATION ACCIDENTS

In 1972, pilots flying on approach to Miami Airport, over the Everglades, were unable to determine if the landing gear had locked in place. All three personnel on the flight deck became engaged in the troubleshooting, and placed the aircraft autopilot on a level altitude hold. Somehow the autopilot became disengaged, but, complacent in their belief in its operation, the flight deck personnel failed to check the now slowly decreasing altitude, until it was too late, and the plane slammed into the ground.

In 1983, on KAL flight 007 over the north Pacific, pilots programmed a course into the flight management system (FMS) that was incorrect. As with the Everglades accident, pilots failed to monitor how automation was flying the plane as it flew directly into Soviet airspace and was shot down, with all lives lost (Wiener, 1988).

In 2013, an Asiana airline was on approach to San Francisco International Airport when pilots became confused regarding what the "Auto-land" system was doing. They acted in opposition to what automation was commanding the plane to do, and the plane stalled and crashed just short of the runway threshold.

All three of these tragedies—and many more (see for example Dornheim, 1995)—have identified problems in HAI encountered by highly skilled professionals. Such problems filtered through careful accident analysis and examined through flight simulation research can identify lessons learned that may be transferred to automation in ground vehicles, along with potential solutions. In the next section, I identify several key features of HAI that can be used to understand generic, cross-disciplinary applications.

21.3 FEATURES OF AUTOMATION

21.3.1 The Degree of Automation: What Does Automation Do and How Does It Do It?

Naturally, the most direct answer to the first part of this question is the purpose and function of automation in a particular context. For example, in aviation, the purpose of the autopilot is to stabilize and fly the plane; the purpose of an alerting system is to offload the human from continuous monitoring of some function or for some event.

Importantly, the second question—how does it do it?—can be answered at a generic level, above and beyond the specific system to which it is applied. Indeed, "how does it do it" can be defined on two generic dimensions, stages, and levels. In 1978, Sheridan and Verplank proposed the concept that automation was not a single entity but could instead be defined along a scale of levels of automation, defining the degree of authority of automation versus human in executing a task (Sheridan & Verplank, 1978). While the original scale had ten levels, these can be characterized more generally as, from lower to higher: (1) automation recommending several possible actions, (2) automation recommending only a single action, (3) automation executing that action but allowing the human to veto it, and (4) automation carrying out the action with no veto possible. The precise number of levels on any scale is less important than the change in levels, moving upward or downward to impose more, or less, automation.

Parasuraman, Sheridan, and Wickens (2000; 2008) in subsequently applying the levels to air traffic control systems, realized that the original Sheridan and Verplank (1978) scale only applied to automation support of human decisions. Parasuraman et al. (2000) postulated that there were three additional "stages" of human information processing that could benefit from the increasing level of automation assistance. These are Stage 1 (information filtering or guiding attention), Stage 2 (integrating information to support inference and situation assessment), and Stage 4 (carrying out or executing the action that was decided upon in Stage 3, action selection and decision-making). [Note: Stage 3 maps onto the automation support of human decisions as originally described by Sheridan & Verplank (1978).] For example, automation assistance for the health care practitioner may highlight particular medical problems on a patient's electronic record or alert the practitioner to a dangerous condition (Stage 1), may offer a diagnostic suggestion (Stage 2), may offer a recommended treatment (Stage 3), and a drug infusion pump may automatically administrate medicine at the appropriate time and dosage (Stage 4). As with Sheridan and Verplank's (1978) original scale, each of these stages can also be described as being implemented at varying levels of authority.

Thus, the two-dimensional structure (taxonomy) of stages and levels, shown in Figure 21.1, can be described by a higher level variable defining the degree of automation (DOA), moving from the lower left to the upper right (Onnasch, Wickens, Li, & Manzey, 2014; Wickens, 2017). When automation is implemented at the highest level of all four stages, this describes the status of full autonomy. The characterization of levels and stages of automation proposed by Parasuraman et al. (2000) can also be likened, conceptually, to the levels of automation for automated driving systems described by the Society for Automotive Engineers (SAE, 2016; this Handbook, Chapters 1 and 2).

In some key developments of the history of this research, Endsley and Kiris (1995), and Kaber and Endsley (2004; Kaber, Onal, & Endsley, 1999) carried out early research on automation and situation awareness (SA) that could be readily interpreted in the context of stages and levels of automation (i.e., DOA). More recently, Onnasch et al. (2014) carried out a meta-analysis of DOA research that examined the effect of four correlated variables that changed as DOA increased: (1) The performance of the task for which automation was designed to support increased; (2) human workload decreased; (3) humans lost SA; and as a consequence, when automation failed, (4) human failure recovery was more problematic (and sometimes disastrous).

A major reason why the taxonomy defining DOA is important in HAI is that it defines a distinction that is relevant to many automation decision support tools: Should automation advise the human user as to "what is" (diagnostic support at Stages 1 and 2) or should automation advise the human user "what to do" (decision

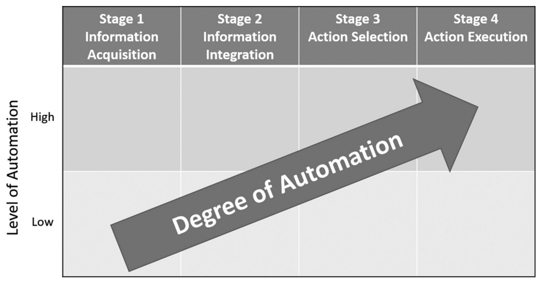

FIGURE 21.1 DOA. In this rendering, there are only two levels within each stage, but there could be several, and there does not need to be the same number of levels at each stage.

aiding at Stage 3)? Such a dichotomy exists in many areas, such as medical decision-making (Garg et al., 2005; Morrow et al., 2006), aviation conflict avoidance systems, or even statistical tools that distinguish between providing p values and confidence intervals versus advice to accept or reject the null hypothesis (i.e., decision aid; Wickens & McCarley, 2017). Given that automation is imperfect, and there may be more severe consequences if it fails at later stages of automation, this distinction needs to be considered carefully in automation design and implementation.

21.3.2 Automation Reliability

The previous section alluded to the imperfections of automation: it can fail and such failures thereby characterize the concept of automation reliability. In some cases, reliability can be quantified as the ratio of correct operations to total operations—a ratio that has important meanings in certain kinds of automation, like those governing alerting systems. Research has also revealed the importance of distinguishing between the first time an automation fails (for a particular user) versus subsequent failures. The former case can often produce a much more problematic response—the so-called "first failure effect" (FFE)—than the latter (Sebok & Wickens, 2017; Yeh, Merlo, Wickens, & Brandenburg, 2003).

Beyond the quantification of automation reliability, we must also distinguish the different generic ways in which automation can fail. It can be gone, such as when the power fails, or it can be wrong, such as when an automated weather forecast is incorrect (Wickens, Clegg, Vieane, & Sebok, 2015a). In process control support, automation failures when automation is wrong appear to have a more consequential effect on human response, than when automation is gone (Eriksson & Stanton, 2017; Sauer, Chavaillaz, & Wastell, 2016; Wickens et al., 2015a).

Another way to characterize different types of automation errors is to distinguish between the following four classes of automation errors:

a. When the hardware or software truly fails: a misconnected circuit, a software bug, a hydraulic valve that is stuck open, or an errant sensor in an automatic control system.
b. When the automation is asked to perform its task in an environment in which it was not designed to perform (e.g., Global Positioning System (GPS) navigation when there is no signal).
c. When the human has an incorrect mental model of what automation should be doing, and believes that automation is wrong, even when it is not. This is not an insignificant problem when the automation is extremely complex, such as the FMS in advanced aircraft (Mumaw, 2018; Sarter, 2008).
d. When the human has programmed or set up the automation incorrectly, as in the KAL 007 disaster described above.

It is important to recognize that the latter three of these failure sources are not true "failures" from an engineering standpoint. However, from the human user's standpoint, they are perceived as automation failures, and this incorrect perception can lead to the problematic failure response, as in the case of the Asiana accident described above.

21.3.3 Automation Trust and Dependence

Since the classic work of Lee and Moray (1992) on process control automation, the concept of trust in automation, seen as an analogy to human trust in other humans, has become prominent in HAI research (e.g., Hoff & Bashir, 2015). This prominence was bolstered by an oft-cited article by Parasuraman and Riley (1997) that introduced concepts such as automation use, disuse, and misuse. These authors highlighted the clear link between trust, a cognitive concept, and system use (which I prefer to call dependence), a behavioral concept.

These two concepts of trust and dependence are often linked, but far from synonymous (Merritt, Lee, Unnerstall, & Huber, 2015). Thus, I can trust an agent, and not depend (partially or totally) upon it. This may be particularly true, in the case of automation, if I enjoy engaging in the processing that automation can do for me. In contrast, I can mistrust automation if it is not perfectly reliable, but still depend upon it, if I have other tasks to perform; however, in these circumstances, I should, optimally, allocate some resources to overseeing and verifying automation's activities. One can argue that the amount of such oversight (e.g., visual attention) should be directly calibrated to the reliability of automation (Moray & Inagaki, 2000).

In discussing the role of trust, it is important to distinguish between over-trust (trusting automation more than one should) and under-trust (trusting less than one should), with calibrated trust defining the balance between these. Over-trust has sometimes been referred to as "automation bias" (Mosier, Skitka, Heers, & Burdick, 1998; Parasuraman & Manzey, 2010).

21.3.4 Out-of-the-Loop-Unfamiliarity (OOTLUF)

A major consequence of over-trust is **complacency**, a dependence on automation that is so extensive that when automation fails, the operator, being out of the loop and hence unfamiliar with what automation is doing, will intervene slowly, inappropriately, or perhaps not at all (Endsley, 2017; Endsley & Kiris, 1995). OOTLUF can be broken down into three separate components, with the first two being associated with loss of SA at Levels 1 and 2 (Endsley, 2017; this Handbook, Chapter 7):

1. Failure to monitor the process(es) carried out by automation, the raw data that it is processing, or the performance it is producing—a failure often directly measurable by eye movements (Metzger & Parasuraman, 2005; Parasuraman & Manzey, 2010).
2. Failure to understand what automation is doing, at the time of the automation failure, and hence intervening inappropriately. A failure to understand will certainly follow from the failure to monitor but may also be associated with an absence of engagement, even when the eyes may scan the indicators of automation functioning. Such a failure is closely related to the phenomenon known as the generation effect (Slamecka & Graf, 1978), whereby we remember the state of systems better when we generate responses pertaining to those systems than when we witness other agents (e.g., automation)

carrying out identical actions. Such memory in this case translates directly into understanding of the current state of the system. The generation effect has often been applied to the concept of active learning, as a superior training technique to passive listening or reading (Dunlosky et al., 2013).

Thus, failure to monitor and understand can lead to inappropriate interventions when automation fails. To this is added the third component of OOTLUF:

3. **Deskilling** refers to the state wherein prolonged use of automation will lead to degraded operator skills in performing the task that automation is programmed to accomplish, hence further aggravating the inappropriate response to automation failure. This concern has been identified in aviation, when pilots rely too much on their automated flight systems (Casner, Geven, Recker, & Schooler, 2014).

All three of these effects: on monitoring, understanding, and manual recovery skills appear to be amplified with higher DOA, as, with this higher degree, there is less reason to become engaged in the task. Hence the more automation does during routine correct performance of automation, the more serious are the consequences when it fails. This tradeoff has been described as the "lumberjack phenomenon": the higher the tree, the harder it falls (Sebok & Wickens, 2017).

21.3.5 Automation Modes and Complexity

More complex automation amplifies the failures to understand what automation is doing, a lesson well revealed in aviation, where the scores of functions carried out by the FMS, include many different modes of automation (Ferris et al., 2010; Landry, 2009; Mumaw, 2018; Sarter, 2008; Sebok et al., 2012). For example, there are five different ways or modes through which the FMS can lower the altitude of an aircraft. Such complexity is amplified further whenever the decision to choose or switch modes is not made by the human, but by the automation itself, given certain "triggering criteria" (e.g., crossing through a particular altitude may automatically trigger a "level off" mode, or exceeding a particular airspeed may trigger a braking mode). Thus, complexity can be often characterized jointly by the number of modes and the complexity of the decision rules used to trigger those modes. For example, "if A then do C" is less complex than "If A and B, do C." When an action is carried out by the human who is assuming that the system is in one mode, but in fact it is in another, this is defined as a mode error. Mumaw (2018) has written a compelling narrative regarding the impact of FMS mode complexity on pilot's mode errors and their challenges in understanding flight deck automation.

21.3.6 Automation Transparency

Both the issues of complexity and OOTLUF-based failure to understand can be attributed in part to the "black box" appearance of much automation, a lesson well learned in aviation. Automation transparency (Chen et al., 2018) is a concept by which the workings of automation are made more apparent to the human.

As described in more detail below, this may be accomplished by displays of what automation is doing and why, or by verbal explanations, for example of why a decision aid came to a particular recommendation.

21.3.7 Adaptable Versus Adaptive Automation

This contrast refers to the extent to which the engagement of automation is adaptable, and can therefore be selected by the human operator (as when the pilot chooses to turn on the autopilot, or "fly by hand"), versus selected by the automation itself (adaptive automation; Dorneich, Rogers, Whitlow, & DeMers, 2016; Kaber & Kim, 2011; Kaber, Wright, Prinzel, & Clamann, 2005). In the latter case, an automated agent will itself "decide" to take automated control from the human operator or return it back to the human depending on the automation's inference about the human's momentary capacity to perform the task well.

21.4 RESEARCH FINDINGS

21.4.1 Accident and Incident Data Mining: Advantages and Costs

A good deal of the understanding of problems in HAI has been gained from the aviation industry, in identifying and analyzing breakdowns that have contributed to major accidents, such as the Eastern Airlines Everglades crash described above. Such analyses appear in National Transportation and Safety Board (NTSB) report and may include HAI errors as one of the causal factors. The limitations in using such information to gain information about causality are twofold. (1) Pilots involved in crashes are often killed and there are always multiple factors involved. It is impossible to sort out, with any certainty, the extent to which HAI was the precipitating factor rather than just a contributing cause, not to mention identifying which of the many issues of HAI might have been involved. (2) Aircraft accidents are, fortunately, extremely rare. But such rarity forecloses the multiple samples that are required to draw reliable statistical inferences regarding frequency and causality. These inferences are at the foundation of the science of human factors.

A second source of data comes from incident analysis. Since 1976, the National Aeronautics and Space Agency (NASA) has collected and categorized a large volume of incident reports, contained in the Aviation Safety Reporting System (ASRS), in which pilots file, anonymously, voluntary reports of what they consider safety-compromising incidents in their flights. Such data have the advantage of large sample sizes (i.e., large N) absent in accident reports. But, being based on the recollection of pilots, who are not generally trained in human factors or psychology, the cognitive or information processing mechanisms are often not included in the narrative. Of course, since the reports are voluntary, while the numbers are large, they may be highly biased, perhaps against reporting an incident in which a pilot committed a clear violation (notwithstanding the guarantee of anonymity in such filings). Nevertheless, such a system has revealed valuable conclusions regarding HAI, recently summarized in an extensive report compiled by the Committee on Aviation

Safety (Commercial Air Safety Team, 2014) of 50 ASRS reports. Their conclusions were notable in identifying the frequency of the sorts of automation mode errors described above.

21.4.2 Experimental and Simulation Results

21.4.2.1 Alerting Systems

There has been a great deal of research on alerting systems that has been carried out in other domains besides automobiles, particularly in the medical (e.g., Seagull & Sanderson, 2001), process control (Strobhar, 2012), and aviation (Martensson, 1995; Wickens, Sebok, Walters, & McCormick, 2017) domain. One feature from these areas that differs in some respect from those in the automobile is that these other workspaces often embody hundreds of alerts and are also susceptible to "alarm flooding" in times of crisis. In contrast, the number of systems to be alerted in the automobile remains relatively limited. Nevertheless, several research findings remain relevant to all forms of alerts.

The distinction drawn in signal detection theory between alert misses and alert false alarms is vital because of their differing influence on human trust of the alerting system and response to the two different types of automation errors. Both misses and false alarms characterize the overall reliability of automation, and hence affect trust therein. Meyer (2001; 2004; Meyer & Lee, 2013) has facilitated the understanding of the distinction between human behavior in response to the two types of alerting automation errors. Reliance is the behavior relevant to automation misses, when the automation alert is "off," signaling that all is well when in fact it is not. Compliance is the behavior relevant to the alert activation. High reliance will cause the human to not notice a system failure because the automation has not been activated. High reliance is often accompanied by a failure of the human operator to monitor the automation or the system that automation is controlling. In contrast, high compliance is manifest as an immediate and consistent response to follow the guidance of the alerting system (e.g., evacuate the building upon sounding of the fire alarm), even when the alarm is false.

Both high reliance and high compliance are induced by highly reliable automation and can lead to expected consequences in the face of alerting system failure (Dixon & Wickens, 2006). High reliance can create the "double miss" (by both automation and the human). The damage done by high compliance is in the "boy who cried wolf" effect: if the alarm system activates falsely too often, the human will lose trust in it, and simply cease to adhere to its directives when future alarms occur, or even deactivate the alarm system entirely. This can lead to dire consequences when the subsequent alarm turns out to be true. Both kinds of alarm failures and their consequences are well documented in aviation (Bliss, 2003).

In all alarm systems, the balance between misses and false alarms can be adjusted by varying the sensitivity or response criterion of an imperfect alarm. This is typically adjusted in a direction that favors fewer misses at the expense of more alarm false alarms, often with a very small ratio of misses to false alarms (i.e., a very high false alarm rate). For example, Bliss (2003) has found that the rate of alert misses involved in aviation incidents is half that of alert false alarms. This adjustment is done with the understandable rationale that the consequences of the "double miss" (by both automation and the human) are very severe. But often under-appreciated are the undesirable consequences of the high false alarm rate, in terms of humans ignoring true alarms. There remains some discrepancy in research findings of the extent to which false alarms are more detrimental to overall system performance than alarm misses (Dixon et al., 2007) or the contrary (Chen & Barnes, 2012). However, the consequences to both reliance and compliance should be carefully considered by designers and human factors practitioners before an alarm system sensitivity level is chosen.

Independent of the extent to which an alert system is miss-prone or false alarm-prone, the consequences of imperfect automation alerting systems are a loss of trust and therefore dependence upon it, even when the alarm system is fairly (but not perfectly) reliable. The question then is how low can such reliability fall before the benefits of the alerting system may be abolished. Insight into this question can be gleaned from two different sources. First, the results of meta-analyses by Wickens and Dixon (2007), and Rein, Masalonis, Messina, and Willems (2013) suggest that the minimum reliability level of a system may be around 75%–80%. Above this level, such imperfection can still support performance better than the unaided human, and particularly under conditions of concurrent task load. Furthermore, this benefit appears to be observed in automation at later stages as well (Rovira, Pak, & McLaughlin, 2017; Trapsilawati, Wickens, Qu, & Chen, 2016). Below that level, the unwarranted dependence upon automation may actually produce worse detection performance than would be the case of unaided automation, not unlike grabbing onto a "concrete life preserver" in the water (Dixon & Wickens, 2006).

Second, as we have noted, the first failure experienced by an individual in his/her experience can be particularly problematic because of complacency that may have developed following experience with, up that that point in time, perfect detection (Molloy & Parasuraman, 1996; Sebok & Wickens, 2017; Yeh et al., 2003). This may be described as the FFE. As a consequence, it may be desirable to "get rid of this FFE" prior to the operator's first operational experience with the alerting system, by allowing them to experience failures during training or introduction to the system (Manzey, Reichenbach, & Onnasch, 2012; Sauer et al., 2016)—an automation failure inoculation, so to speak.

21.4.2.2 Attention Cueing

Alerting systems inform the operator that something is wrong. Automation can (and ideally should) go beyond this to inform the operator what is wrong, and/or where the dangerous condition is located. The "what" is embodied in Stage 2 diagnostic automation discussed later, but the "where" is embodied in attentional cueing systems, closely related to, but more advanced than alerting systems. Research by Yeh and her colleagues, primarily in the military domain directing a soldier's attention to the potential location of an enemy (e.g., highlighting locations or features on a map), has revealed that erroneous cueing systems (i.e., that direct attention to the wrong location) too can have serious negative consequences (Yeh & Wickens, 2001; Yeh et al., 2003; Yeh, Wickens, & Seagull, 1999). As with alerting systems, such automation errors here, are particularly problematic upon the first failure (Yeh et al., 2003).

21.4.2.3 Stages 2 and 3 Automation: OOTLUF

The concept of OOTLUF, along with the characteristics of the lumberjack effect (the more automation imposed, the more problematic is the human response to its failure) have been described above. The meta-analysis conducted by Onnasch et al. (2014) covers much of the empirical research published prior to 2013 on the general pattern of routine and failure response performance as well as the reductions in workload and SA across increasing DOA. The tradeoffs between reduced workload and loss of SA, and between routine and failure response performance have been supported by subsequent research across many domains (e.g., in military applications: Rovira et al., 2017; in medical applications: Mayo et al., 2015; in robotics: Wickens, Sebok, Li, Gacy, & Sarter, 2015b; in process control: Manzey et al., 2012).

Of particular importance in this relationship is the progressive loss of SA (understanding) as the DOA increases (Endsley & Kiris, 1995), a loss which underlies the problematic response to automation failure. As examples of this research, in process control, Manzey et al. (2012) observed that operators checked raw process variables less frequently with later stages of automation, leaving them with a degraded mental picture in the case of failure. In air traffic control, Trapsilawati, Wickens, Chen, and Qu (2017) observed that imposing Stage 3 automation of an automated conflict resolution aid reduced controller's SA of the state of conflict traffic, compared with manual conditions.

These findings lead to the valuable conclusion that assessments of SA, particularly at Endsley's level 2 (understanding), can provide a useful prediction of the problematic automation failure effect. This is because SA is more feasible to measure than is the automation failure performance (particularly the first failure), and SA is a more reliable measure given that first failure performance will only provide one data point per participant.

21.5 SOLUTIONS: COUNTERMEASURES PROPOSED AND IMPLEMENTED IN OTHER DOMAINS

The previous section has described some of the key issues and problems in HAI, as revealed by research in other, non-driving domains. There is at least an indirect hint from these problems of what might be plausible solutions. For example, to the extent that the OOTLUF problem is amplified with higher DOA, designers should probably resist the temptation to implement very high automation levels at later stages, at least in safety-critical tasks (Parasuraman et al., 2000). And if loss of SA is associated with the problematic automation failure intervention, then efforts should be made to maintain SA, as we discuss below. In this section we describe three explicit categories of solution: flexible automation, automation transparency, and training. Each of these avenues has been suggested and demonstrated in other domains to address some of the problems of OOTLUF and automation failure interventions.

21.5.1 Flexible and Adaptive Automation

If prolonged use of high DOA can induce both complacency and skill degradation, then there would seemingly be a benefit of encouraging the human user to periodically enter or reenter the control loop, to perform the task manually. Such advantages, in detecting automation failures, were indeed demonstrated by Parasuraman, Mouloua, and Hilburn (1999) who observed better failure responses during periods of automation support, when the operator was forced to engage in intermittent periods of manual performance. There are in fact two other different ways of implementing such periodic engagement, beyond the forced switching (scheduled automation) imposed by Parasuraman et al. (1999): adaptable and adaptive automation (Christensen & Estepp, 2013).

In adaptable automation, the operator is simply given the flexibility to choose when (or whether) to implement automation, although this can be encouraged by policy. In contrast, in adaptive automation, an automated agent itself makes the decision of when to implement automation, and, in turn, when to return responsibility back to the human operator. Although adaptable automation has the intuitive appeal of keeping responsibility for task switching in the operator's hands, it does impose added decision responsibilities and does not necessarily produce better performance (Christensen & Estepp, 2013; Sauer, Kao, & Wastell, 2012). However, as described below, it is considerably easier to implement adaptable automation, with a simple "on–off" mode switch at the operator's disposal. This ease is illustrated by discussing the contrasting challenges of adaptive automation.

Adaptive automation requires a set of elements, as well articulated in aviation systems by Dorneich et al. (2016). The first element is identifying what components of the task to automate (e.g., what stages and levels)—a design question equally relevant for adaptable automation. The second element concerns the aspect of human performance or cognition upon which to base the automation agent's decision to implement automation. Typically, the trigger for the automation decision to allocate is either excessive workload or poor performance or a combination of both. A third element, which we describe below, is the specific measure of human cognition/performance with which to infer that automation is needed (or no longer needed when control is returned to the human); that is, what measure triggers the decision.

This third element can be challenging because signals from the human or the human–system interface regarding both workload and performance are imperfectly reliable. It follows that multiple samples of either of these are required to draw a reliable inference about the human's capability to perform. As a consequence of this, perhaps several seconds (a non-trivial amount of time) may be required to obtain a sufficient sample in real time. The resulting delay can produce problems of instability in a dynamic environment, perhaps removing the human from the loop (because automation has inferred workload to be high) when workload has in fact been reduced (and automation is no longer needed), or more problematically, returning a task to the human at a time when workload has just spiked. To shorten this control loop further, adaptive algorithms may invite an unreliable estimate in which the automated scheduler of task allocation may infer workload to be low when it is in fact high or vice versa. It is because of such challenges that the demonstrations of successful adaptive automation with complex systems are difficult to achieve (Chen, Visser, Huf, & Loft, 2017; Sauer, Chavaillaz, & Wastell, 2017).

21.5.2 Automation Transparency

To the extent that a problematic automation failure response may result from a loss of SA of what automation is doing, then it is reasonable to suggest that providing greater transparency of the workings of automation should support human failure intervention. Indeed, recent empirical investigations of the automation transparency concept suggest far greater success in improving HAI than is the case with adaptive automation. The results of several recent studies, outside of the automobile domain, offer unqualified support for a transparency benefit to performance, relative to a control condition with the same automation (and failure type) without a transparent automated system. Of these, eleven studies found the benefits of automation transparency to be either amplified by or specific to the conditions of automation failure (Burns et al., 2008; Chen et al., 2018; Dzindolet et al., 2003; Lai, Macmillan, Daudelin, & Kent, 2006; Mayo et al., 2015; Mercado et al., 2016; Trapsilawati et al., 2017). Hoff and Bashir (2015) offer a comprehensive review of automation transparency successes. Hergeth, Lorenz, and Krems (2017), and Seppelt and Lee (2007) described examples of transparency in support of automobile drivers.

The concept of automation transparency is broad and somewhat ill defined, but can be made more concrete by characterizing four different forms:

- Graphic representation of what automation is doing (e.g., Burns et al., 2008; Mayo et al., 2015; Mercado et al., 2016; Trapsilawati et al., 2017)
- Textual descriptions of how automation is doing it (e.g., the reasoning behind automated decision aids; Mercado et al., 2016)
- Clear presentation of the raw data being processed by automation (Trapsilawati et al., 2017)
- Estimates of automation's own degree of uncertainty in its performance (Chen et al., 2018)

The last of these is similar to the likelihood alert that signals the confidence level of an alert system that a dangerous condition exists (Wizorek & Manzey, 2014). Furthermore, in the case of textual explanations of automation functioning and reasoning, these may be offered either online, at the time a particular automation decision or diagnosis is reached, or off-line prior to the use of automation, as a form of training and instruction.

There is one important limitation of online automation transparency, and that is that it may provide a source of distraction or extra perceptual or cognitive workload that could neutralize or offset the very benefits that automation is intended to provide, particularly offsetting benefits to performance of the automation-supported task.

21.5.3 Training

The last manifestation of automation transparency described above, off-line explanation of reasoning, can be thought of as a form of automation training. Like transparency in general, training has been found to be successful in buffering some of the negative effects of automation failure response, and can be offered in different forms. In the medical domain, simple instructions regarding operation of a decision aid have been observed to increase the use of decision aids (Lai et al., 2006). In process control automation, as noted above, pre-exposure to automation failures during training can buffer the negative effects of a first failure (Bahner, Hüper, & Manzey, 2008; Manzey et al., 2012; Sauer et al., 2016). Training and automation is further described in this Handbook (Chapter 18).

21.6 CONCLUSIONS

Decades of research and accident analysis have revealed that the overall benefits of automation can sometimes be mitigated by their costs, as these are often associated with imperfect reliability, leading to OOTLUF. In safety-critical environments, designers and regulators should seek solutions such as adaptive or adaptable automation and transparency to mitigate the consequences of automation errors to HAI performance. To address these consequences as much as possible, lessons learned from other domains can assist the incorporation of safe automation into the automobile; but in transferring knowledge and techniques between domains, the designer must be cognizant of the many differences between the highway driving domain and those such as aviation and process control, where the solutions described in this chapter have been identified and evaluated.

REFERENCES

Bahner, J. E., Hüper, A. D., & Manzey, D. (2008). Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies*, *66*(9), 688–699.

Billings, C. (1997). *Aviation Automation: The Search for a Human-Centered Approach*. Englewood Cliffs, NJ: Erlbaum.

Bliss, J. (2003). Investigation of alarm-related accidents and incidents in aviation. *International Journal of Aviation Psychology, 13*, 249–268.

Burns, C. M., Skraaning, G., Jamieson, G. A., Lau, N., Kwok, J., Welch, R., & Andresen, G. (2008). Evaluation of ecological interface design for nuclear process control: Situation awareness effects. *Human Factors, 50,* 663–679.

Casner, S. M., Geven, R. W., Recker, M. P., & Schooler, J. W. (2014). The retention of manual flying skills in the automated cockpit. *Human Factors*, *56*(8), 1506–1516.

Chen, J. Y. & Barnes, M. J. (2012). Supervisory control of multiple robots: Effects of imperfect automation and individual differences. *Human Factors*, *54*(2), 157–174.

Chen, J. Y., Lakhman, S., Stowers, K., Sellpwotz, A., Wright, J., & Barnes, M. (2018). Situation awareness based agent transparency and human-autonomy teaming effectiveness. *Theoretical Issues in Ergonomics Science, 19*, 259–282.

Chen, S., Visser, T., Huf, S., & Loft, S. (2017). Optimizing the balance between task automation and human manual control in a simulated submarine track management. *Journal of Experimental Psychology: Applied, 23*, 240–262.

Christensen, J. C. & Estepp, J. R. (2013). Coadaptive aiding and automation enhance operator performance. *Human Factors*, *55*(5), 965–975.

Commercial Air Safety Team. (2014). *Airplane State Awareness*. JSAT Final Report.

Cummings, M. L. & Guerlain, S. (2007). Developing operator capacity estimates for supervisory control of autonomous vehicles. *Human Factors, 49*, 1–15.

Dixon, S. R. & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle flight control: A reliance-compliance model of automation dependence in high workload. *Human Factors, 48*, 474–486.

Dixon, S. R., Wickens, C. D., & McCarley, J. M. (2007). On the independence of reliance and compliance: Are false alarms worse than misses? *Human Factors, 49*, 564–572.

Dorneich, M. C., Rogers, W., Whitlow, S. D., & DeMers, R. (2016). Human performance risks and benefits of adaptive systems on the flight deck. *International Journal of Aviation Psychology, 26*, 15–35.

Dornheim, M. A. (1995). Dramatic incidents highlight mode problems in cockpits. *Aviation Week and Space Technology*, *142*(5), 57–59.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58.

Dzindolet, M., Peterson, S., Pomranky, R., Pierce, L., & Beck, H. (2003). The role of trust in automation reliance. *International Journal of Human Computer Studies, 58*, 697–718.

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human-automation research. *Human Factors,* 59, 5–27.

Endsley, M. R. & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors, 37,* 381–394.

Eriksson, A. & Stanton, N. A. (2017). Takeover time in highly automated vehicles: Noncritical transitions to and from manual control. *Human Factors*, *59*(4), 689–705.

Ferris, T., Sarter, N., & Wickens, C. (2010). Cockpit automation: Still struggling to keep up. In E. Salas & D. Maurino (Eds.), *Human Factors in Automation* (2nd ed.). Amsterdam, The Netherlands: Elsevier.

Garg, A. X., Adhikari, N. K., McDonald, H., Rosas-Arellano, M. P., Devereaux, P., & Beyene, J. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes. *Journal of the American Medical Association, 293,* 1223–1238.

Hergeth, S., Lorenz, L., & Krems, J. F. (2017). Prior familiarization with takeover requests affects drivers' takeover performance and automation trust. *Human Factors*, *59*(3), 457–470.

Hoff, K. & Bashir, M. (2015). Trust in automation integrating empirical evidence on factors that influence trust. *Human Factors*, *57*, 407–434.

Kaber, D. B. & Endsley, M. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science, 5*, 113–153.

Kaber, D. B. & Kim, S. H. (2011). Understanding cognitive strategy with adaptive automation in dual-task performance using computational cognitive models. *Journal of Cognitive Engineering and Decision Making, 5,* 309–331.

Kaber, D. B., Onal, E., & Endsley, M. R. (1999). Level of automation effects on telerobots performance and human operator situation awareness and subjective workload. In M. W. Scerbo & M. Mouloua (Eds.), *Automation Technology and Human Performance: Current Research and Trends* (pp. 165–170). Mahwah, NJ: Erlbaum.

Kaber, D. B., Wright, M. C., Prinzel, L. J., & Clamann, M. P. (2005). Adaptive automation of human-machine system information-processing functions. *Human Factors, 47,* 730–741.

Lai, F., Macmillan, J., Daudelin, D. H., & Kent, D. M. (2006). The potential of training to increase acceptance and use of computerized decision support systems for medical diagnosis. *Human Factors*, *48*(1), 95–108.

Landry, S. (2009) Flight deck automation. In S. Nof (Ed.), *Handbook of Automation* (pp. 1215–1239). Berlin: Springer.

Lee, J. D. & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*, 1243–1270.

Li, H., Wickens, C. D., Sarter, N., & Sebok, A. (2014) Stages and levels of automation in support of space teleoperations. *Human Factors, 56*(6), 1050–1061.

Manzey, D., Reichenbach, J., & Onnasch, L. (2012). Human performance consequences of automated decision aids: The impact of degree of automation and system experience. *Journal of Cognitive Engineering and Decision Making, 6,* 1–31.

Martensson, L. (1995). The airplane crash at Gottrora: Experiences of the cockpit crew. *International Journal of Aviation Psychology, 5*, 305–326.

Mayo, M., Kowalczyk, N., Liston, B., Sanders, E., White, S., & Patterson, E. (2015) Comparing the effectiveness of alerts and dynamically annotated visualizations (DAVs) in improving clinical decision making. *Human Factors, 57*, 1002–1014.

Mercado, J. E., Rupp, M. A., Chen, J. Y., Barnes, M. J., Barber, D., & Procci, K. (2016). Intelligent agent transparency in human–agent teaming for Multi-UxV management. *Human Factors*, *58*(3), 401–415.

Merritt, S., Lee, D., Unnerstall, J., & Huber, K. (2015). Are well calibrated users effective users? Association between calibration of trust and performance on an automated aiding task. *Human Factors, 57*, 34–47.

Metzger, U. & Parasuraman, R. (2005). Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. *Human Factors*, *47*, 35–49.

Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors, 43,* 563–572.

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors, 46,* 196–204.

Meyer, J. & Lee, J. D. (2013). Trust, reliance, and compliance. In J. D. Lee & A. Kirlik (Eds.), *The Oxford Handbook of Cognitive Engineering* (pp. 109-124). New York: Oxford University Press.

Molloy, R. & Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors, 38*, 311–322.

Moray, N. & Inagaki, T. (2000). Attention and complacency. *Theoretical Issues in Ergonomics Science, 1*, 354–365.

Morrow, D., North, R., & Wickens, C. D. (2006). Reducing and mitigating human error in medicine. *Reviews of Human Factors and Ergonomics, 1,* 254–296.

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

Mumaw, R. J. (2018). *Addressing Mode Confusion Using an Interpreter Display* (Contractor Technical Report). Moffett Field, CA: San Jose State University Research Foundation. doi:10.13140/RG.2.2.27980.92801

Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors, 56*(3), 476–488.

Parasuraman, R. & Manzey, D. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, *52,* 381–410.

Parasuraman, R., Mouloua, M., & Hilburn, B. (1999). Adaptive aiding and adaptive task allocation enhance human-machine interaction. In M. W. Scerbo & M. Mouloua (Eds.), *Automation Technology and Human Performance: Current Research and Trends* (pp. 119–123). Mahwah, NJ: Lawrence Erlbaum.

Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230–253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, & Cybernetics: Part A: Systems and Humans, 30*(3), 286–297

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making, 2*(2), 140–160.

Rein, J. R., Masalonis, A. J., Messina, J., & Willems, B. (2013). Meta-analysis of the effect of imperfect alert automation on system performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 57, No. 1, pp. 280–284). Los Angeles, CA: SAGE Publications.

Rovira, E., Pak, R., & McLaughlin, A. (2017). Effects of individual differences in working memory on performance and trust with various degrees of automation. *Theoretical Issues in Ergonomics Science*, *18*(6), 573–591.

SAE. (2016). Taxonomy and definitions for terms related to driving automation systems for on road motor vehicles (SAE Standard J3016 201609). Warrendale, PA: SAE International.

Sarter, N. B. (2008). Investigating mode errors on automated flight decks: Illustrating the problem-driven, cumulative, and interdisciplinary nature of human factors research. *Human Factors, 50,* 506–510.

Sauer, J., Chavaillaz, A., & Wastell, D. (2016). Experience of automation failures in training: Effects on trust, automation bias, complacency and performance. *Ergonomics*, *59*(6), 767–780.

Sauer, J., Chavaillaz, A., & Wastell, D. (2017). On the effectiveness of performance-based adaptive automation. *Theoretical Issues in Ergonomics Science*, *18*(3), 279–297.

Sauer, J., Kao, C., & Wastell, D. (2012). A comparison of adaptive and adaptable automation under different levels of environmental stress. *Ergonomics, 55*, 840–853.

Sauer, J. & Ruttinger, B. (2007). Automation and decision support in interactive computer products. *Ergonomics, 50*, 902–909.

Seagull, F. J. & Sanderson, P. M. (2001). Anesthesiology alarms in context: An observational study. *Human Factors, 43*, 66–78.

Sebok, A. & Wickens, C. D., (2017). Implementing lumberjacks and black swans into model-based tools to support human-automation interaction. *Human Factors, 59*, 189–202.

Sebok, A., Wickens, C. D., Sarter, N., Quesada, S., Socash, C., & Anthony, B. (2012). The automation design advisor tool (ADAT): Development and validation of a model-based tool to support flight deck automation design for NextGen operations. *Human Factors and Ergonomics in Manufacturing and Service Industries, 22*, 378–394.

Seppelt, B. D. & Lee, J. D. (2007). Making adaptive cruise control (ACC) limits visible. *International Journal of Human-Computer Studies*, *65*(3), 192–205.

Sheridan, T. B., & Verplank, W. L. (1978). *Human and Computer Control of Undersea Teleoperators*. (Technical Report, Man-Machine Systems Laboratory, Department of Mechanical Engineering). Cambridge, MA: MIT Press.

Slamecka, N. J. & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning, Memory, and Cognition, 4,* 592–604.

Strobhar, D. (2012). *Human Factors in Process Plant Operation*. New York: Monument Press.

Trapsilawati, F., Wickens, C. D., Chen, C.-H., & Qu, X. (2017) Transparency and conflict resolution automation reliability in Air Traffic Control. In P. Tsang, M. Vidulich & J. Flach (Eds.), *Proceedings of the 2017 International Symposium on Aviation Psychology* (pp. 419–424). Dayton, OH: ISAP.

Trapsilawati, F., Wickens, C. D., Qu, X., & Chen, C.-H. (2016). Benefits of imperfect conflict resolution advisory aids for future air traffic control. *Human Factors, 58*, 1007–1019.

Wickens, C. D. (2017). Stages and levels of automation: 20 years after. *Cognitive Engineering and Decision Making, 12*(1), 35–41.

Wickens, C. D., Clegg, B. A., Vieane, A. Z., & Sebok, A. L. (2015a). Complacency and automation bias in the use of imperfect automation. *Human Factors*, *57*(5), 728–739.
Wickens, C. D. & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science, 8*(3), 201–212.
Wickens, C. D., Mavor, A., Parasuraman, R., & McGee, J. (1998). *The Future of Air Traffic Control: Human Operators and Automation*. Washington, DC: National Academy Press.
Wickens, C. D. & McCarley, J. (2017). Commonsense statistics in aviation safety research. In P. Tsang, M. Vidulich, & J. Flach (Eds.), *Advances in Aviation Psychology: Volume 2* (pp. 98-110). Dorset, UK: Dorset Press.
Wickens, C. D., Sebok, A., Li, H., Gacy, A., & Sarter, N. (2015b). Using modeling and simulation to predict operator performance and automation-induced complacency with robotic automation: A case study and empirical validation. *Human Factors, 57*, 959–975.
Wickens, C. D., Sebok, A., Walters, B., & McCormick, P. (2017). *Alert Design Considerations* (FAA Technical Report). Boulder, CO: Alion Science & Technology.
Wiener, E. L. (1988). Cockpit automation. In E. L. Wiener & D. C. Nagel (Eds.), *Human Factors in Aviation* (pp. 433–461). San Diego, CA: Academic Press.
Wiener, E. L. & Curry, R. E. (1980). Flight deck automation: Promises and problems. *Ergonomics, 23,* 995–1012.
Wizorek, R. & Manzey, D. (2014). Supporting attention allocation in multi task environments: Effects of likelihood alarm systems on trust behavior and performance. *Human Factors, 56*, 1209–1221.
Yeh, M., Merlo, J. L., Wickens, C. D., & Brandenburg, D. L. (2003). Head up versus head down: The costs of imprecision, unreliability, and visual clutter on cue effectiveness for display signaling. *Human Factors, 45,* 390–407.
Yeh, M. & Wickens, C. D. (2001). Attentional filtering in the design of electronic map displays: A comparison of color coding, intensity coding, and decluttering techniques. *Human Factors, 43,* 543–562.
Yeh, M., Wickens, C. D., & Seagull, F. J. (1999). Target cuing in visual search: The effects of conformality and display location on the allocation of visual attention. *Human Factors, 41,* 524–542.