



HHS Public Access

Author manuscript

Theor Issues Ergon Sci. Author manuscript; available in PMC 2015 January 01.

Published in final edited form as:

Theor Issues Ergon Sci. 2014 ; 15(6): 545–577. doi:10.1080/1463922X.2013.817625.

Understanding human management of automation errors

Sara E. McBride, Wendy A. Rogers, and Arthur D. Fisk

Georgia Institute of Technology, School of Psychology, 654 Cherry Street, Atlanta, GA 30332, USA

Abstract

Automation has the potential to aid humans with a diverse set of tasks and support overall system performance. Automated systems are not always reliable, and when automation errs, humans must engage in error management, which is the process of detecting, understanding, and correcting errors. However, this process of error management in the context of human-automation interaction is not well understood. Therefore, we conducted a systematic review of the variables that contribute to error management. We examined relevant research in human-automation interaction and human error to identify critical automation, person, task, and emergent variables. We propose a framework for management of automation errors to incorporate and build upon previous models. Further, our analysis highlights variables that may be addressed through design and training to positively influence error management. Additional efforts to understand the error management process will contribute to automation designed and implemented to support safe and effective system performance.

Keywords

automation; imperfect automation; error; error management; human-automation interaction

1. Introduction

Modern automated systems are present in a multitude of environments, including aviation, process control, transportation, and healthcare. These technologies are designed to support overall system performance, assisting human operators with tasks such as information acquisition and processing, decision making, and action execution (Parasuraman *et al.* 2000). Many automated systems are not perfectly reliable, often due to technological limitations. For example, automation that involves sensing certain states in the environment may err because its sensors have limited detection capabilities. Although these errors may be rare, the resulting consequences can be severe.

Examples of serious consequences are not hard to find. One recent incident involved a family and their in-vehicle navigation aid (Clark 2011). This family was sightseeing in California's Death Valley when they instructed their navigation aid to give them directions for getting home. Following the instructions provided by their navigation aid did not lead them home; rather, the family became lost in Death Valley at the height of summer (it turns out the system was relying on outdated maps). Although the family was located after three

days, a similar incident occurred two years prior in which a mother and son were lost for five days. The mother survived, but her son did not.

These unfortunate events highlight the important role that the human serves when working with imperfect automation. If the individuals in the previous cases had had access to another source of information, such as a paper map, they could have sought to verify the automation's suggested directions and realized they were being sent into the middle of nowhere. Humans are the last safeguard before automation errors propagate into accidents. This human role involves 'error management', which is the process of detecting errors, understanding why errors occurred, and correcting errors (Kontogiannis and Malakis 2009). Unfortunately, because of humans' prominent role in this chain of events, they often receive the majority of the blame when their efforts fail. However, "the enemy of safety is not the human: it is complexity" (Woods *et al.* 2010, p. 1).

Understanding the process of error management is key to enhancing safe and successful automation interactions. Yet there has not been a systematic review of the nature of this complexity with the goal of identifying the variables that influence a human's ability to manage automation errors. We conducted an analysis of the research literature from the fields of human-automation interaction and human error to identify the factors that influence the error management process. We identified and organized the automation, person, task, and emergent variables that influence error management of automation. This organizational framework serves multiple purposes. First, it provides a heretofore non-existent integration of the research literature pertinent to managing errors of automated system. Second, it can guide improvements in error management by identifying the potentially relevant variables to be addressed through design and training. And third, our analysis revealed gaps in the literature and needs for future research to ensure successful management of automation errors.

2. Overview of the review process

Research on human interaction with automated systems is extensive. Because the focus of the current investigation was to understand how various factors influence how well an individual manages automation errors, several criteria were used to select relevant publications from the body of literature on human automation interaction. First, if participants interacted with an automated system, the experiment had to include a version of the automated system that was less than 100% reliable. This departure from perfect reliability could result in automation errors, which we defined as an error committed by the automation during which the automation does not behave in a manner that is consistent with the true state of the world. Therefore, we do not include situations during which the automation behaves as it should, and an overall system error occurs because the human operator is deficient in some way, such as not understanding what mode the automaton is in, or supplying the automation with inaccurate information.

Automation errors may include misses and false alarms (Green and Swets 1988), misdiagnoses, and inappropriate action implementation. Automation designers know that certain errors, such as misses, will occur when they set the threshold for detection (i.e.,

beta). One might argue that because these errors are part of the automation's programming, they should not be considered errors. However, we contend that an error has occurred if the automation behaves in a manner that is inconsistent with the true state of the world. The navigation aid example illustrates an automation error in which the automation recommended the use of roads that were no longer in existence.

In addition to automation errors, automation failures can also occur. During failures, the automation is non-responsive, not available, or not functional, and therefore forces the human to perform the task manually. For example, Woods *et al.* (2010) described an automatic infusion controller used during surgery to control the rate that drugs are delivered to the patient. If the controller cannot regulate flow or detects one of many potential device conditions, it is programmed to turn off after emitting an alarm and warning message. In this scenario, the operator of the automation would need to detect that the controller had turned off, then explain and understand what caused the failure, which may be necessary to controlling the system manually or getting the system back online (correction). One might assume that detecting an automation failure is a trivial matter, but there are numerous instances in commercial aviation (e.g., Eastern Flight 401, China Airways Flight 006) during which pilots failed to detect that their automation had disengaged. Although automation errors and failures represent two different categories of activity, both require error management by the human. Thus, the empirical research in this review included studies investigating both errors and failures.

A second criterion for our review was that if the experimental investigation included interaction with an automated system, a dependent measure of error management success was required. For example, an experiment may have measured how many of the automation errors were detected by the participant. Although many empirical investigations of human-automation interaction employ imperfect automation, oftentimes the dependent variables relate to other constructs, such as trust in automation, rather than error management success.

Lastly, the review focused on interactions between a single human operator and one or more automated systems. Although it is important to understand how error management occurs in teams composed of multiple humans and multiple forms of automation, this was beyond the scope of the current analysis.

We identified the variables that have been shown to influence the management of automation errors, which we classified into four categories: (1) automation variables, (2) person variables, (3) task variables, and (4) emergent variables (see Table 1). Automation variables included characteristics of the automation such as the reliability of the automation, the types of errors the automation is likely to make, the level of automation, and the nature of the feedback that the automation provides to a human operator. Person variables represented factors that were unique to the individual, and included an individual's complacency potential (i.e., attitudes that influence complacent monitoring behaviour), nature of training received, and knowledge regarding the automation. Task variables included variables that described the context in which the human and automation are working together. These included the consequence of an automation error if unmanaged by the human, the costs associated with verifying automation's suggestions or information, and the degree to which

the human is held accountable for the results of an automation error. Lastly, emergent variables described factors that result from the interaction between the human and automation. For instance, the trust that an individual has towards a specific piece of automation is the result of not only characteristics of that person (e.g., very trusting), but also aspects of the automation (e.g., reliability), and the task in which they are engaged (e.g., highly critical). The emergent variables included trust in automation, workload, and situation awareness.

Before we present the empirical evidence regarding how these variables influence a human's management of automation errors, we first provide an overview of the research on error management processes outside of the automation literature, and discuss how these theoretical frameworks can be applied to human-automation interaction.

3. Error management

Errors have been discussed in the context of socio-technical systems in which errors or failures may originate from the human or the technology involved. Historically, the dominating focus has been error prevention; that is, keeping errors from occurring at all. More recently, perspectives regarding errors have shifted due in part to the distinction of errors and their consequences, as well as the understanding that not all errors can be prevented. Errors may occur, but actions to stop or reduce their consequences can be taken. Therefore, a supplementary goal to error prevention should be error management, or the process of limiting the consequences of error. This approach has also been called resilience engineering and espouses that anticipating possible disturbances or errors in a system is a more valuable effort than attempting to eliminate them completely (Sheridan 2008). Particularly in the case of using imperfect automation, operators cannot prevent the automation from erring, as its designers determine its reliability. But, operators can mitigate the consequences of errors by managing them. Researchers generally agree that error management consists of three components: (a) detection, (b) explanation, and (c) correction (Kontogiannis 1999, 2011, Kanse and van der Schaaf 2001).

Detection involves realizing that something has gone wrong or that an error is about to occur. Explanation is the process of identifying the nature of the error as well as understanding the underlying cause of the error (Kontogiannis 1999). Correction involves modifying the existing plan or developing a new plan as a countermeasure against the potential adverse events of the error (Kanse and van der Schaaf 2001). During the correction stage, operators may have different goals, influenced by the nature of the error, ensuing consequences, and time pressure (Kontogiannis 1999). For instance, correction actions may serve as a strategy to avoid immediate consequences of an automation error. Correction can also be considered from a maintenance perspective, such as replacing a faulty sensor that was responsible for the error to avoid similar errors in the future.

Progression through the three phases is not necessarily linear, and overlap between the phases is possible. Corrective measures can be taken while attempts to explain the source of the error are underway. Particularly in time-restricted situations, operators may take action to correct the error without thoroughly understanding its cause. For instance, Kanse and van

der Shaaf (2001) analyzed the sequence in which operators moved through the various error management stages. They discovered that in many cases detection was followed immediately by corrective action, with explanation occurring later. In other situations, an effective plan to correct the error cannot be obtained without first understanding the nature of the error.

To further understand the processes involved in the three phases, one can turn to Reason (1990), who proposed a generic error-modelling system (GEMS) to describe how humans manage various types of self-committed, *human* errors. As depicted in Figure 1, behaviour proceeds from the routine actions at the skill-based level to the rule- and knowledge-based levels in the event of issues being detected. Detection of an error occurs by means of attentional checks, during which higher levels of cognition come to the forefront and determine whether behaviour is running according to the established plan, or whether the plan needs to be altered to reach the goal. Attentional checks may fail as a result of inattention (omitting a critical check) or overattention (checking at an inappropriate time).

After a problem has been detected, behaviour transitions from monitoring to problem solving. The first attempt at problem solving occurs at the rule-based level wherein an individual will search for an applicable, pre-established corrective procedure. If a stored rule successfully resolves the problem, behaviour continues towards the goal state. If a stored rule is not sufficient, then behaviour proceeds to the knowledge-based level. At the knowledge-based level, problem solving may occur through the use of analogies, mental models, or abstract relationships.

Although this description suggests that this process is predominantly linear in nature, Reason pointed out that switching from one level of performance to another is unlikely to be so clear cut. Rather, it is more likely that behaviour occurs at multiple levels simultaneously, and that the transitions between the levels may not just proceed down from skill- to rule- to knowledge-based, but back up and down multiple times. For example, if one finds an applicable rule at the rule-based level, then enacting that rule may involve routinized actions at the skill-based level.

GEMS can be used to describe the behaviour of a human operator interacting with imperfect automation to accomplish a task. The same attentional checks used to assess the operator's own performance should also be used to check that the automation is working properly. In cases where a problem with the automation is detected, either an error or a failure, the importance of accessing stored rules and knowledge and using them effectively to problem solve cannot be understated. This process can make the difference between an automation error that was recovered from and an error that led to an accident.

3.1 Summary of error management

Literature on error management describes the process individuals engage in to mitigate the consequences of an error. Error detection, explanation, and correction are necessary components in recovering from automation errors as well. Reason's (1990) GEMS can be used to further understand the detailed cognitive paths that are traversed to explain and ultimately correct automation errors. On that basis, we developed a framework highlighting

the error management processes in a human-automation interaction context (see Figure 2). This framework illustrates how each phase might occur in an interaction with automation. Detection occurs by monitoring the automation and the world and can happen before or after the error affects the world. For example, an error in the drip rate of an automated infusion pump may be noticed by a healthcare provider who physically sees that the medication is dripping faster than it should be (pre-effect), or the error may be detected once the patient has a negative medical reaction to the overdose of medication (post-effect).

The explanation phase of our framework borrows from Reason's (1990) rule-based level in which individuals search for familiar patterns to compare their current situation to. In interactions with automation, operators may similarly examine their memory of previous automation errors for a comparable event, and apply stored rules as a means of correction. If such a pattern is not available, the explanation phase also contains elements inspired by the Reason's knowledge-based level, which described how individuals use their mental representation of the problem space, or of the automation in our case, to make explanatory connections. This process may also involve generating and testing hypotheses, followed by the observations of outcomes on the system, which can lead to multiple iterations of corrective attempts.

One critical component of error management that has thus far been understated or absent in the literature to date is the process of integration. Integration describes the process of learning that occurs as a result of experiencing automation errors or failures. This exposure may give the human more detailed information regarding the automation's capabilities and limitations, circumstances under which errors may be more likely to occur, and heuristics that may lead to quicker error correction. These insights must be integrated into the human's mental representation of the automation and task, resulting in a more refined and complete understanding. This may then influence future instances of error management because it may lead a human to adjust his/her monitoring strategies and allow for better error detection.

4. Human-automation interaction variables

Many empirical investigations have examined the variables influencing how a person handles automation errors, and these have included automation, person, task, and emergent variables (see Table 1). We analysed this literature with respect to error management to identify both the relationships that have been documented as well as gaps that remain in the literature. We would have liked to organize the automation literature research by the relevant phase of error management (detection, explanation, correction, integration). However, the majority of studies in the human-automation literature have focused on detection or a simplified method of correction as their dependent variable, thereby limiting our ability to classify studies according the phases of error management. We will return to this issue in the discussion.

4.1 Automation variables

4.1.1 Reliability—One of the most widely investigated characteristics of automation is its reliability level. Reliability can be thought of as the “probability that the automation would make a correct classification” (Riley 1996, p. 23). Not surprisingly, research has

demonstrated that using perfectly reliable automation results in superior system performance when compared to manual performance (Moray *et al.* 2000, Dixon *et al.* 2005). Further, using 100% reliable automation results in superior performance compared to imperfect automation (Kantowitz *et al.* 1997, Moray *et al.* 2000).

What is less clear is how performance varies when using imperfect automation compared to performing the task manually. In such situations, is overall performance better served by a human manually performing the task without any automated assistance, thereby avoiding the trouble of managing automation errors, or is performance superior when a human uses automation prone to errors? Our review indicated that the answer is not as simplistic as it might seem.

In a study by Skitka *et al.* (1999), participants either worked with an 88% reliable automated monitoring aid or were required to monitor the task manually without any automation. Of interest to the researchers were six of the 100 trials in which the automation group failed to receive prompts to critical events. Because the manual condition never received prompts throughout all 100 trials, the two groups could be compared on those six trials. Participants using the automated aid missed significantly more of the critical events than those working manually. For the 94 trials when the automation was correct, those in the automated aid condition had significantly better performance.

This finding suggests that imperfect automation benefited performance, but only when the automation was correct. When the automation failed, those in the manual conditions had an advantage. To better understand the driving forces behind these performance differences, Skitka *et al.* (1999) correlated participants' error rates with measures of perceived task difficulty, effort, beliefs that automation improves accuracy of responses, and diffusion of responsibility. Of those, delegating the responsibility of system monitoring to the automation and attitudes that the automation was infallible were associated with higher omission and commission error rates.

Although the group using automation suffered performance decrements during the automation miss trials, this group demonstrated better *overall* performance than those not using automation. Maltz and Shinar (2004) similarly found that individuals in a driving simulator maintained safer driving distances if they used an imperfect warning aid compared to no aid at all. Additionally, experienced air traffic controllers using imperfect automation were as likely to detect a potential conflict (which the automation missed) as the controllers who were performing the task manually (Metzger and Parasuraman 2005). It appears that imperfect automation generally leads to better or comparable performance compared to manual task performance. Given this finding, using imperfect automation is a better choice because it would result in the same or better levels of performance and reduce the workload placed on the operator.

Because imperfect automation does appear to provide significant support, one might assume the best approach would be to accept a certain level of inaccuracy but get as close to 100% reliability as possible. This strategy would reduce the number of automation errors, thereby increasing overall performance. Researchers have suggested that this logic is flawed and

noted that there appears to be a tradeoff that occurs with more reliable (still <100% reliable) automation (Bainbridge 1983, Wickens *et al.* 2010). Indeed, Sorkin and Woods (1985) noted that optimizing the automation's performance alone did not lead to gains in overall system performance. In a comparison between 80% and 60% reliable automation, participants using 60% reliable automation had better decision making accuracy for trials when the automation was unreliable (Rovira *et al.* 2007).

These data reflect that automation that errs can support human performance, and it appears that in some cases, more error-prone automation may provide a greater benefit than less error-prone automation. But at what point does error-prone automation become no longer useful? Wickens and Dixon (2007) focused on modelling the role of reliability in human interaction with automated diagnostic aids. Their goal was to determine whether there is a reliability level below which the automation fails to provide any benefit or begins to impair performance. A linear regression revealed the predicted crossover point in reliability to be 70%, such that when using automation less than 70% reliable, task performance was worse than if the person were doing the task manually. Further, Wickens and Dixon's analysis suggested that when automation reliability dropped below 70%, individuals continued to depend on the automation but failed to correct its errors, and did so as a means of preserving their cognitive resources for the other tasks.

One compelling explanation for why more reliable automation hurts error management performance comes from research examining complacency, or a state of sub-optimal monitoring behaviour. The idea is that when interacting with reliable automation (but not perfectly reliable), individuals come to believe their dependence on the automation is warranted, and complacent behaviour increases, which reduces the probability of detecting automation errors when they occur. Therefore, if complacency can be disrupted (i.e., by varying reliability), then detection of automation errors should improve.

In an experiment by Parasuraman *et al.* (1993), four groups of participants were responsible for monitoring an automated aid for errors. In two groups, the automation's reliability was constant across the experiment at either 87.5% or 56.25%. In the other two groups, the reliability switched between 87.5% and 56.25% after every 10-minute block (12 blocks total). The variable-reliability groups were significantly more likely to detect automation errors than the constant-reliability groups (detection rates of approximately 75% versus 30%). This effect was not present in the first block, but emerged in the second block and increased in later blocks. Parasuraman *et al.* posited that by varying the reliability, participants became more wary of the automation and increased their vigilance. These findings were replicated by Prinzel *et al.* (2005) using A' (a measure of detection sensitivity less sensitive to response distributions than d') as the dependent variable.

We believe there may be an alternative or supplementary explanation for the benefit of varying automation reliability. Because the variable-reliability group was exposed to two levels of reliability (87.5% and 56.25%), the average level of reliability was approximately 72%. Indeed, Rice and Geels (2010) have found evidence that when using multiple automated systems of varying reliability, individuals tend to merge the reliability of the aids and treat them as a single unit. Because a 72% reliable aid was not included as one of the

levels used in the constant-reliability conditions, the performance benefit may be a consequence of this intermediate level of reliability, rather than the varying nature of reliability for the group. Therefore, future investigations should include a constant-reliability group that matches the average of the variable-reliability group to allow for a direct comparison.

Even if varying reliability is an effective method of disrupting complacency, it is not clear how this might be implemented in operational contexts. To vary reliability would require the introduction of contrived automation errors. If operators know the bogus nature of these errors, they may lose substantial trust in the system or reduce their monitoring behaviour. Another possible avenue might be taking the automation offline and forcing periods of manual performance. Empirical research will be necessary to determine if this would result in a comparable performance benefit.

Our analysis identified a factor that often varies between studies of automation reliability; namely, the base rate of the event or signal to be detected by the automation. For example, an automated aid designed to assist with luggage screening is designed to detect the event of a weapon concealed in a suitcase. The base rate of this event, or the probability that it will occur, is very low. This is true for many base rates.

Much of the research examining reliability has used experimental situations in which the base rate is extremely high. For example, in one study a fault that the automation was designed to detect occurred every trial, and each trial lasted approximately 2.5 minutes (Moray *et al.* 2000). Although this high base rate may represent some operational environments, the other end of the spectrum has not received as much empirical attention.

We need to understand how error detection performance changes as a function of base rate, as vigilance decrements may reduce error management success. One of the few studies that has examined monitoring automation with a very low base rate of errors (i.e., a single failure in a 30 minute session), found that detection rates dropped significantly if the failure occurred in the last 10 minutes of the 30 minute block compared to if it occurred in the first 10 minutes (Molloy and Parasuraman 1996). The high base rates of errors in previous studies may have masked potential vigilance effects and overestimated how capable individuals may be in detecting errors over extended periods of time with rare error events.

It is not clear whether the effects found in the literature will scale up when errors only occur every other day, week, or month, rather than multiple times in an experimental session. Certainly, there are limitations to what can be done experimentally due to time and monetary constraints, but understanding the effect of base rate on error management should be a research priority.

Overall, our analysis revealed that an automated system's reliability influences a human's ability to manage automation errors. 100% reliable systems certainly support overall task performance compared to manual task performance. However, the benefit of imperfect automation is not quite as clear. Overall performance can be divided into two categories: (a) performance when the automation is working (routine performance), and (b) performance when the automation is erring or has failed completely (error or failure performance).

Although one might conceive that closer approximations to 100% reliability will yield superior performance, this is not necessarily the case. As Bainbridge (1983) suggested, "... the most successful automated systems, with rare need for manual intervention ... may need the greatest investment in human operator training" (p. 777).

Explanations for this paradox or tradeoff have included increases in complacency, or suboptimal monitoring of the automation. Greater complacency may be found in highly reliable (but imperfect) automation during routine performance, which leaves the human less vigilant and less able to respond appropriately to automation errors or failures. If this tradeoff does exist, we need to understand why it occurs and whether it is an effect that needs to be eliminated or may be used to automation designers' advantage.

Moray (2003) has pointed out that the manner in which complacency has been studied is ambiguous. Specifically, complacency is typically defined as monitoring or sampling of the automation below some optimal level; however, this level is typically not defined and detection of error is often measured instead of sampling behaviour. Therefore, to claim that complacent behaviour has occurred in these situations is inappropriate. Further, Moray suggested that when automation is typically reliable, a lack of sampling or monitoring may be a rational decision when other tasks require the operator's attention, suggesting that the negative connotations associated with complacency are not always suitable.

Parasuraman and Manzey (2010) also argued that complacent behaviour needs to be evaluated independently of detection performance. Such work has been accomplished more recently (see Bahner *et al.* 2008a, 2008b under the Training section). However, Parasuraman and Manzey also pointed out that normative or optimal sampling models need to take into consideration the cost of sampling and the value of the sampled information.

4.1.2 Nature of errors—In addition to the level of error that the automation commits, the nature of the errors committed has also been examined in relation to error management. Generally, the errors committed by automation are classified as false alarms or misses (Wickens and Carswell 2006). A false alarm occurs when the automation incorrectly detects a signal in the environment, and a miss occurs when a signal is present in the environment, but the automation fails to detect it. The errors committed by many types of automated systems, particularly those designed to detect certain states (such as the presence of a system malfunction), may be considered in this manner according to signal detection theory (Green and Swets 1988).

Sanchez *et al.* (2011) employed automation designed to assist with collision avoidance. Participants were exposed to either 10 false alarms or 10 misses out of 240 events over the course of the experimental session. The miss group had significantly more errors and lower trust in the automation than the false alarm group, suggesting that misses may be more detrimental to error management than false alarms.

Evidence to the contrary has been reported (Dixon and Wickens 2006). A diagnostic aid was set to be 67% reliable, with the remaining 33% either false alarms or misses. Participants used this aid in addition to performing two other tasks simultaneously. Detection rates for

critical events were subdivided based on whether the event occurred during a moment of low or high workload. When workload was low, participants' detection of critical events was near ceiling, suggesting they were very capable of detecting and correcting when the automation missed or false alarmed. Under conditions of high workload, error type had a significant impact on detection rates. The miss group detected significantly more critical events than the false alarm group. Dixon *et al.* (2007) also found that misses were associated with greater detection of system failures than false alarms as measured by d' .

False alarms and misses are unavoidable consequences of imperfect automation. The ratio of false alarms to misses is at the discretion of the system designer, who must weigh the relative costs of these two events. Although a miss may constitute an extraordinarily high cost in terms of immediate consequences (a plane flying into the side of a mountain), minimizing misses may not be an appropriate alternative considering the evidence that false alarms are more detrimental to the operators' error detection. Further research is needed to determine if the effect of false alarms can be reliably mitigated. A limitation of the current research on error types is that the majority of research has focused on early stage automation, or automation responsible for acquiring and analyzing information, but not making decisions or taking action (as defined by Parasuraman and Wickens 2008). Future research should investigate whether the patterns found in early stage automation hold for later stage automation. Stages of automation are discussed further in the following section.

4.1.3 Level of automation—Automation does not operate in an all-or-none fashion (Wiener and Curry 1980). Rather, the degree of control allotted to the human and to the automation moves along a spectrum, such that a higher level of automation (LOA) has greater control or autonomy compared to a lower level of automation. Various accounts have been provided regarding the levels of automation (Endsley and Kiris 1995, Endsley and Kaber 1999, Parasuraman *et al.* 2000).

In a study using an automobile navigation task, Endsley and Kiris (1995) examined whether using higher levels of automation would hurt decision making performance when the task had to be completed manually. They hypothesized that the time to make a decision during this manual phase would positively correlate with LOA. Their data only revealed an increase in decision time among participants who had been using the highest level of automation compared to participants who had been performing the task manually the entire session. In fact, although decision time varied, the quality of the decisions made was near ceiling for all of the groups.

Endsley further refined and expanded her LOA taxonomy (see Table 2), and she examined the effect of these levels on performance in a dynamic control task (Endsley and Kaber 1999). Each participant completed two trials with two different LOAs. During each trial, three automation failures occurred in which control reverted to manual. Performance data were separated into performance during normal conditions and performance during an automation failure.

Although there were significant performance differences between LOAs during normal conditions, the effect of LOA was less pronounced during the automation failures.

Performance across the LOAs was not significantly different except that batch processing and automated decision making were worse than manual performance. These two LOA groups also showed the longest time to recover from the automation failure. These data represent a pattern in performance seen in other areas of research, such as automation reliability; that is, it highlights the paradox of automation. A LOA that boosts performance during normal conditions (e.g., batch processing) leads to the worst performance when the automation fails. Endsley and Kaber (1999) explained this particular finding by highlighting that in batch processing, which is a relatively low LOA and should not impair failure performance, participants were likely to try to plan ahead, distracting them from the task at hand and resulting in the poorer performance observed during automation failures.

There is not strong evidence for an effect of LOA on error management. There may be certain methodological limitations that reduced the likelihood of uncovering such an effect. The underlying theoretical perspective is that higher LOAs create a more passive role for the human, resulting in more complacency and lower situation awareness, which negatively affects performance during automation failures. However, in both of the previously discussed studies, we believe the experimental procedure did not provide enough time for complacency to form and situation awareness to suffer.

For example, in Endsley and Kiris's (1995) study, participants were presented with a short text based scenario, and then required to make a decision. On average, the decision time ranged from 15-30 seconds. The probability of complacency forming under these circumstances is low, which we believe is why the effect of LOA on performance was not observed. In Endsley and Kaber's (1999) study, there were no more than 2-3 minutes of continuous use of the automation at the LOA being tested. This period of use is so brief that it is very unlikely complacency and the subsequent performance decrement would have time to develop. Although it has been shown that complacency can develop in just 20 minutes (Parasuraman *et al.* 1993), it occurred after 20 minutes of *continuous* use. The time period in question consisted of 12-13 minutes of use with a 5-minute break partway through, rendering it unlikely that participants had enough time to develop complacent behaviour.

In addition to the LOA taxonomy suggested by Endsley and Kaber (1999), Parasuraman *et al.* (2000) proposed a model that included four stages of functions that can be performed by the automation or human (see Figure 3). Within each of these stages, the level of automation can range from low to high. As depicted in Figure 3, degree of automation is defined by both the stage and level of automation.

One study that examined the relative benefit of the varying stages of automated support compared information automation (stages 1 & 2) with decision automation (stages 3 & 4). Rovira *et al.* (2007) found that with 80% reliable automation, when the automation was imperfect, the decision automation was associated with less accurate decisions and longer decision times, whereas the information automation was not. Rovira *et al.* suggested that the information automation may be less affected by unreliable automation because users will continue to generate alternative courses of action, whereas a user of decision automation may cease engaging in this process.

Wickens *et al.* (2010) conducted a meta-analysis to determine the extent to which varying the degree of automation leads to a tradeoff in performance between routine and failure conditions (also known as the paradox of automation). Although the initial analysis supported the hypothesized tradeoff, a subsequent analysis called this into question. The original analysis included data points for conditions in which the task was performed manually. When these data points were removed, the effects changed dramatically. The correlation between routine and failure performance switched to positive, although non-significant, showing a trend for higher degrees of automation to result in improved performance when the automation is working correctly *and* when the automation errs.

Wickens *et al.* (2010), and recently, Hancock *et al.* (2013), discussed several possible explanations for the lack of support for the tradeoff. If automation is well designed and operators are well trained, this can actually mitigate or overcome the tradeoff. However, what makes a display or a training program effective in terms of supporting error management performance is not readily apparent. In fact, Lorenz *et al.* (2002) found that participants using a high LOA actually were faster to identify faults when the automation failed compared to a medium LOA. Data on the sampling behaviour showed that high LOA participants were more likely to sample relevant information, which kept them “in the loop” and prepared to handle automation failures.

The data collected thus far do not paint a clear picture of the relationship between LOA and error management. One could conceive that high LOA supports error management because the automation is handling most of the task leaving the human available to monitor the status of the system. On the other hand, at a lower LOA the human is actively involved in the task at hand and therefore might be more likely to detect an abnormality. Both explanations appear plausible, so further research is needed to identify which scenario represents reality.

4.1.4 Feedback—A critical component of well-designed automation is the feedback provided by the automation to the human. This feedback is the only way the automation communicates information to the human regarding its behaviour and the overall state of the system. A breakdown in communication or lack of feedback may have profound effects on a human’s understanding of the automation’s actions, which can result in severe performance decrements if the automation errs or fails. Automation with poor feedback capabilities has been described as silent (Woods 1996). Norman (1990) posited that a lack of appropriate feedback is at the heart of many of the issues raised regarding the implementation of automation. A case study from the aviation domain presented in his writing nicely illustrates this point. Below is a quotation from an accident report filed with the NASA Aviation Safety Reporting System:

Shortly after level off at 35,000 ft...the second officer brought to my attention that he was feeding fuel to all 3 engines from the number 2 tank, but was showing a drop in the number 3 tank. I sent the second officer to the cabin to check that side from the window. While he was gone, I noticed that the wheel was cocked to the right and told the first officer who was flying the plane to take the autopilot off and check. When the autopilot was disengaged, the aircraft showed a roll tendency confirming that we actually had an out of balance condition. The second officer

returned and said we were losing a large amount of fuel with a swirl pattern of fuel running about mid-wing to the tip, as well as a vapor pattern covering the entire portion of the wing from mid-wing to the fuselage. At this point we were about 2000 lbs. out of balance... (Norman 1990, p. 139)

In this case, the autopilot had compensated for the weight imbalance caused by the fuel loss. However, because the automation did not provide feedback to the pilots indicating that it was compensating for imbalance more than usual, the fuel leak may have gone unnoticed had it not been for the second officer's detection of the discrepancy. Even though the source of the discrepancy was unknown to the second officer, by voicing his observation to the rest of the crew, he exchanged important information that set an investigation into action. In the same way, if the automation could have "voiced" its observation that it was compensating more than normal, this likely would have alerted the crew to the problem.

A lack of feedback has been linked to lapses in mode awareness, which refers to "an operator's knowledge and understanding of the current and future automation configuration, including its status, targets, and behaviour" (Sarter 2008, p. 506). One may consider mode awareness a subcomponent of the more encompassing construct of situation awareness. Although in many cases decrements in mode awareness may be the result of poor feedback from the automation, there is also evidence that in cases where feedback is available to operators, they may not allocate attention to it (Sarter *et al.* 2007).

Well-designed cues can effectively draw a human's attention to relevant output from the automation. Nikolic and Sarter (2001) discovered this in a study investigating the relative merit of peripheral visual feedback, or feedback that can be processed using peripheral rather than foveal vision. Participants using peripheral cues detected significantly more mode transitions and detected them significantly faster than participants using a traditional feedback display that required foveal vision.

Although the perceptual characteristics of feedback may be important, another critical aspect of acquiring feedback is the process of gathering pieces of information from various sources, and integrating the information to create an understanding of the automation's behaviour and resulting system state. Skjerve and Skraaning (2004) conducted a series of experiments to investigate this process among experienced nuclear power plant operators. They argued that by designing an interface that provides explicit feedback and increases the observability of the system, human-automation cooperation will be enhanced, which will benefit overall performance.

Participants worked in multi-person crews in the experimental scenarios, in which they were responsible for maintaining normal operation and responding to critical plant occurrences (such as malfunctions). Operators' ability to detect critical plant events and their response times to these events was superior when using an experimental interface, wherein the automation's activity was made explicit through the addition of graphical and verbal feedback, compared to the conventional interface.

It is important to recognize that in many, if not most, cases, incorporating appropriate feedback requires increasing the complexity of the displays. However, as Skjerve and

Skraaning's (2004) work demonstrated, this increased display complexity actually supported performance, because the task of gathering information from various sources to produce an understanding of the automation's behaviour was reduced or eliminated. Unfortunately, it is easy to imagine an interface where the opposite situation might occur. The feedback may become an additional element that must be monitored and its potential benefit is lost if not well implemented. This may be particularly true in cases where an operator is working with multiple automated systems that are poorly integrated and focus on different task subgoals.

4.1.5 Summary of automation variables—Aspects of the automation itself (e.g., reliability level, nature of errors, level of automation, and feedback to the human) influence an operator's ability to manage errors. Contrary to what one might expect, more is not necessarily better for reliability and level of automation. That is, higher reliability and higher levels of automation have been associated with greater complacency, which can lead to worse error management. This is certainly an area warranting further research as these phenomena are negatively related to error management.

4.2 Person variables

4.2.1 Complacency potential—Individuals' level of complacency has been shown to be an important predictor of whether they will manage errors effectively. Complacency is not only affected by aspects of the automation and the task, but also by aspects of the person. Singh *et al.* (1993a) identified complacency potential as an attitude towards automation that individuals bring to a situation that, when combined with environmental and automation variables, may result in complacent behaviour. Singh *et al.* (1993a) developed the Complacency-Potential Rating Scale (CPRS) to capture an individual's potential for complacency by measuring favourable and unfavourable attitudes towards different aspects of automation.

Singh *et al.* (1993b) used the CPRS to determine the extent to which complacency potential predicted failures of monitoring. The reliability of the automation was manipulated to be either constant or variable. A median split was used to classify participants as either low or high complacency. In the variable reliability condition, complacency potential participants did not affect detection performance, suggesting that the effect of variable reliability (discussed earlier in the Automation Reliability section) may be strong enough to wash out an effect of complacency potential.

Under conditions of constant reliability, and contrary to the expected pattern, high complacency potential individuals had superior detection rates compared to low complacency potential participants (52.4% vs. 18.7%, respectively). However, when the correlation between complacency potential and detection rate was computed within each complacency potential group, the high complacency group had a negative, albeit insignificant, correlation of ($r = -.42$), and the correlation for the low complacency potential group was essentially null.

Our analysis revealed several possible explanations for the lack of evidence for a relationship between complacency potential and detection rates found by Singh *et al.* (1993b). First, the CPRS was administered after the first of two experimental sessions.

Participants' experience with the automation in this first session may have influenced their CPRS scores, resulting in an inaccurate representation of their true complacency potential. In addition, the median split used to create the two complacency potential groups was 56, such that the low complacency group contained scores from 47-56, whereas the high complacency group contained scores from 57-70. Considering that the range of possible scores was 16-80, there was not much of a spread in the scores obtained in this sample, perhaps limiting the effect.

Although this initial investigation did not provide much support for a relationship between complacency potential and automation error management, we believe this is likely due to the methodological issues discussed above, as Prinzel *et al.* (2005) reported data showing an effect of complacency potential. They used the same tasks, automation conditions, and instructions as those used by Singh *et al.* (1993b); however, they computed perceptual sensitivity (A') to serve as the measure of automation monitoring performance. They also included 40 participants in their sample, compared to Singh *et al.*'s 24 participants. Once the CPRS scores were gathered, a median split of 58 was used to divide participants into low and high complacency potential. They found a significant interaction between complacency potential (high or low) and reliability (constant or variable) for automation monitoring performance, such that all groups performed comparably, except for the high complacency-constant reliability group, who performed significantly worse. Their findings demonstrate that complacency potential does influence automation error detection, primarily in situations that are conducive to the formation of complacency (constant automation reliability).

A recent investigation examined how individual differences in working memory and executive function may contribute to a person's likelihood of exhibiting complacent behaviour. These cognitive components are highly heritable and under dopaminergic control, and the dopamine beta hydroxylase (DBH) gene is thought to regulate dopamine availability in the prefrontal cortex. Therefore, Parasuraman *et al.* (2012) divided participants into low and high DBH enzyme activity groups (lower DBH activity is associated with more dopamine), and examined their decision making performance when using an imperfect automated aid. They found that when the automation erred, the low DBH group was more likely to verify the automation's suggestions and had significantly better decision accuracy (i.e., detected more automation errors).

Taken together, these studies suggest that individual differences in complacent behaviour or automation bias may be linked to two sources. Complacency potential, or an individual's attitude towards automation that reflects a tendency towards complacent behaviour (e.g., not verifying the automation or seeking confirmatory/disconfirmatory evidence), has been shown to predict the likelihood that complacent monitoring behaviour will occur. In addition to attitudinal differences, recent evidence has shown that complacency potential may also have genetic underpinnings in terms of an individual's working memory and executive function abilities.

4.2.2 Training—Prior to engaging with automation in an operational environment, individuals often partake in some form of training. These training programs can range in duration, fidelity, and content. Consider, for example, the numerous hours commercial pilots

must spend in flight simulators before they receive certification, and contrast this to the training (or lack thereof) a consumer might receive after purchasing an automated GPS navigation aid. Although the difference in training is clear, both of these user groups have been found to demonstrate poor automation error management skills. If a portion of training programs was devoted to improving error management, would this result in measurable performance gains?

One study investigated this question by implementing various training programs designed to overcome automation bias, which is the tendency to defer to the automation rather than engage in information seeking behaviour (Mosier *et al.* 2001). Participants experienced one of three training interventions: (a) general training related to the aircraft systems and how the automation could be verified, (b) training that stressed participants must verify the automation, or (c) training that provided information regarding automation bias, the errors people often make when using automation, and how to avoid these errors. Participants encountered seven automation errors during a simulated flight, including six opportunities for an omission error, and one opportunity for a commission error. Training intervention did not have a significant effect on the likelihood of committing an omission or commission error.

A similar study may shed some light on why there was not an effect in the above study. In this study, the same three training interventions were used and participants worked in a similar simulated flight environment (Skitka *et al.* 2000b). The critical difference between this and the above study was the increase in opportunities for commission errors. In this study, there were six opportunities for an omission error and six opportunities for a commission error. With this change, Skitka *et al.* found that training had a significant effect on the number of commission errors committed. Specifically, participants who experienced the training emphasizing automation bias and the possibility of errors made fewer commission errors than the other two training groups. These findings suggest that this form of training may primarily reduce commission errors, but not omission errors. This might explain why Mosier *et al.* (2001) did not find an effect of training. There was only one opportunity for a commission error, reducing the potential for differences between the training groups.

Bahner *et al.* (2008b) examined the effectiveness of an intervention designed to improve automation error management by exposing participants to errors during training. They reasoned that simply discussing the potential for automation errors may not be enough to effectively change monitoring behavior. The experience group was exposed to two diagnosis errors during training. The information group received only correct diagnoses from the automation during training. The experimental session included a misdiagnosis on the 10th out of 12 faults. If participants followed the misdiagnosis provided by the automation for fault 10, they committed a commission error.

Only five out of the 24 participants committed this commission error, and the distribution of these individuals was not affected by training group (experience group = 2; information group = 3). However, the experience group sampled a significantly higher portion of information relevant to the diagnosis than did the information group. Further, individuals

who detected the misdiagnosis sampled significantly more relevant system parameters than those who missed it. Thus, training influenced information sampling behaviour, which in turn predicted whether participants detected the automation error. However, a direct effect of training on error detection was not found.

In another study, using the same simulation, the authors focused on the warning function of the automation rather than the diagnosis component (Bahner *et al.* 2008a). The experimental manipulation was the same such that one group was exposed to automation errors during training whereas the other group was only told about the potential for errors and warned to double check the automation. The automation errors were misses, that is, the automation failed to detect the presence of a system fault.

In the experimental trials, automation missed faults 10 and 13 (out of 14 total faults). At fault 10 the experience group was significantly less likely to make an omission error than the information group (18.2% vs. 80%). By the next opportunity for an omission error on fault 13, there was no longer a significant difference between the groups (18.2% vs. 22.2%). At fault 14, rather than a miss, the automation misdiagnosed the fault. The majority of participants (74%) failed to catch this misdiagnosis, and there was not a clear effect of training. Those who successfully identified the misdiagnosis had sampled a significantly greater portion of the relevant information up to that point. This finding further highlights the importance of actively monitoring the automation and sampling relevant information as it repeatedly predicts the likelihood of detecting automation errors.

The absence of a group effect for the fault 14 misdiagnosis can be traced back to the specific focus of training. Training for the experience group only included exposure to automation misses, not misdiagnoses. The effect of training was constrained to the type of errors experienced in training, and did not result in a system wide increase in monitoring behaviour, which the authors attribute to individuals having high functional specificity (discussed later in the section on trust). We identified an alternative explanation as to why error detection differed between the two subsystems. Specifically, differences in error detection were related to the consequences of the two types of automation errors. Perhaps the diagnosis component was not as carefully monitored because if a misdiagnosis were acted upon, participants would receive notification that the repair did not work and then would simply have to manually repair the fault. Conversely, if a miss went undetected, the system would enter a critical state, a considerably more serious consequence.

Future research needs to address several issues and open questions. The transferability of error management training to various automated components within a system warrants further investigation, as does the duration of training effects on error management. Further, identifying the specific aspects of training that lead to greater information sampling, and thus greater error detection, will allow training programs to capitalize on the effect. Sarter *et al.* (1997) proposed that training should focus on encouraging active exploration and knowledge seeking behaviours to help develop refined knowledge structures. The role of knowledge in error management is discussed next.

4.2.3 Knowledge—Users of automation must have knowledge about their role and responsibilities in a given task, knowledge about the automation they are working with, and knowledge about the context in which they are working. This knowledge forms a model of the system and is critical to the process of managing automation errors. For example, consider an automated fault management system designed to detect faults such as a broken pipe. In this scenario, an operator will have some level of knowledge regarding the event the automation is designed to detect (broken pipes happen more often in cold weather due to freezing), and knowledge related to the capabilities of the automation (tends to false alarm rather than miss).

Ideally, operators would integrate these pieces of knowledge into a larger understanding of the system that they are a part of, and use this understanding to improve error management. That is, the hypothetical operator would be more wary of an alert from the automation regarding a broken pipe if it occurred in July compared to if it occurred in January. Unfortunately, the complexity of many modern automated systems makes this knowledge increasingly difficult to acquire and apply. Operators often have gaps or misconceptions in their models of a system that can cause severe disruptions in performance. Further, these inaccuracies may be unbeknownst to the human, resulting in operators who are more confident in their abilities than is warranted.

Such a pattern has been documented in an investigation of the effect of mental model quality on task performance while using an automated navigation aid (Wilkison *et al.* 2007). Participants were given varying levels of exposure (no exposure, low, and high) to a simulated city map to establish a different mental model quality for each of the three groups. Participants then used a decision aid to navigate the simulated city and could accept or reject the aid's navigation advice. When using a 70% reliable aid, participants who had a low exposure to the city map during training, and presumably had a weak or vague mental model, were significantly more likely than the no or high exposure groups to reject the advice of the decision aid *when it was correct*, opting instead to choose their own, less optimal route. Training programs designed to quickly teach learners "the basics" might actually be harmful if they lead individuals to believe they know more than they actually do.

Sarter and Woods (1994) used a simulated flight scenario to reveal deficiencies in the mental models of experienced pilots interacting with the Flight Management System, considered a core system of flight deck automation. Over 70% of the pilots showed deficiencies related to seven of the 24 scenario probes. A majority of the deficiencies resulted from a lack of knowledge regarding the functional structure of the system. Sarter and Woods also noted that many of the pilots were not aware they had gaps in their knowledge. Future research should use Sarter and Woods's approach of documenting knowledge gaps in conjunction with performance measures to determine how these gaps influence error management.

One particular subset of knowledge that might influence error management is whether the user has a causal explanation for the errors committed by the automation. Dzindolet *et al.* (2001) contended that individuals may judge automation to be less reliable if they do not understand why it makes errors, leading them to distrust and disuse an aid that might actually provide a benefit. Dzindolet *et al.* (2003) examined this hypothesis by providing one

group of participants an explanation for automation errors, whereas the other group did not receive this information. Regression analyses revealed that reliance on the aid was greater among participants who were given an explanation for the automation's errors. It is important to note, however, that the increased reliance on the aid was not always warranted. The error explanation influenced participants' behaviour, but did not make them better at managing automation errors.

The relationship between knowledge and error management is not clear, due in part to the scarcity of empirical research examining this topic. Certainly, the complexity of knowledge required in many domains may be daunting. There are many open questions to be answered. For instance, does having the requisite knowledge for a particular task guarantee that an individual will access and use it appropriately when needed? Further, although a wealth of knowledge is gained during training, this knowledge base does not remain static. The "on the job" experiences will also alter knowledge. Therefore, providing useful and accurate feedback to operators may help to further refine knowledge.

4.2.4 Summary of person variables—When individuals interact with an automated system, they bring aspects of themselves to bear on how successfully they manage the automation. These factors include their complacency potential, the training they received, and their knowledge of the system. In comparison to the characteristics of the automation, person characteristics have been under-studied. Moreover, factors relating to trust and knowledge may be particularly critical to the explanation, correction, and integration phases of our error management framework, but have primarily been investigated in relation to error detection.

4.3 Task variables

Humans and automation do not work together in a vacuum; rather, they are engaged in performing a task and are part of a larger system. Characteristics of the task can influence the human's error management skills.

4.3.1 Consequence of automation error—When an automated system errs, it results in the potential for consequences to occur. These consequences can range from minor to severe, and in many systems (e.g., aviation, nuclear power), severe consequences may include the fatalities of many people as well as significant monetary losses. The literature has generally demonstrated that as consequences become more severe, people adjust their behaviour and manage automation errors more effectively.

In a flight simulation study, the data revealed that if an automation error did not affect flight safety, pilots were less likely to detect it than if it did influence flight safety (Palmer and Degani 1991). In fact, the automation errors that would have resulted in a crash were detected by all of the flight crews. Other studies have also found that the failures most harmful to flight safety had higher detection/correction rates (Mosier *et al.* 1998, Mosier *et al.* 2001).

An important fact to note regarding the previously discussed studies is that they all occurred in the aviation domain with trained pilots. As discussed by Mosier *et al.* (2001), these pilots

had received extensive training before taking part in the experiment, and part of their training undoubtedly involved prioritization of tasks during flight, with more critical tasks receiving greater priority. From a review of these studies alone, it is not yet clear whether severity of consequences affects error management generally, beyond trained pilots in an aviation context.

To explore how costs influence error management in an unfamiliar context, Ezer *et al.* (2008) systematically manipulated the consequence of failing to correct an automation error by implementing a point scheme into her experimental task. A significant effect of consequence of error was found, revealing that as the consequence became more severe (more points lost), participants were less likely to rely on the automation. Unfortunately, Ezer *et al.* did not disclose the performance data related to cost of error, so it is unclear whether the reduced reliance was associated with improved error management. However, in this task, participants' counting ability was measured and was roughly equal to the reliability of the automation (approximately 70% correct). Therefore, from a strategic point of view, greater reliance on the automation was an effective strategy because performance would be roughly equivalent but less effort would be expended.

Although research on the consequences of automation error is not extensive, there is general agreement among the findings. As consequences of letting the automation err become more severe, the likelihood that the error is managed increases. Interestingly, many empirical investigations in the human-automation literature have used point schemes or other reward/penalty devices to motivate participants to put forth their best effort in the experimental trials. The focus of those studies was not whether consequences influenced behaviour, but their use suggests researchers generally recognize the effect exists.

4.3.2 Cost of verification—A critical step in managing automation errors is verifying or checking the automation. This verification procedure is often the step that leads a user to detect that an error is present, and/or determine the appropriate course of action for overcoming the error. As discussed by Cohen *et al.* (1998), verification may include several activities, such as:

... checking the aid's reasoning, examining the aid's conclusion against evidence known to the user but not to the aid... or attempting to find (or create) a better alternative. Verification is not usually a once-and-for-all decision. More typically, it is an iterative process... The process should end when the uncertainty is resolved, the priority of the issue decreases, or the cost of delay grows unacceptable. (p. 19)

Ezer *et al.* (2008) allowed participants to verify an automated suggestion by choosing to view relevant stimuli. They manipulated the cost associated with this verification process such that the cost of verification was either high (loss of one point for every 2 seconds of verifying), or low (no loss of points for verification). Task performance was superior in the low cost of verification condition, relative to the high cost condition.

The current research examining how the cost of verifying automation influences error management is minimal. Future research should investigate the various forms that cost can take, such as time, resources, and performance, to determine if certain costs are more

influential than others. It is likely that minimizing verification costs would support error management performance. This is easier said than done, and research is needed to understand how to successfully accomplish this.

4.3.3 Accountability—In human teams, responsibility for successful task completion is often dispersed among all team members. In a human-automation team, the automation cannot be expected to detect its own errors in the same way that humans are capable of detecting when they themselves have erred, so this task must fall to the human. If diffusion of responsibility occurs between a human and an automated system, it may lead the operator to accept automated directives without any form of double-checking (Mosier and Skitka 1996).

Mosier *et al.* (1998) conducted one of the first endeavours in understanding the role of accountability in the context of human-automation interaction. Pilots were placed in either an accountable or a non-accountable group. Those in the accountable group were informed that their performance would be monitored, collected, and evaluated with respect to how they used the automation, and they would have to justify their performance in a post-simulation interview. Pilots in the non-accountable group were told that performance data could not be collected due to a computer issue, and that their contribution would primarily be subjective responses to a questionnaire. No mention of justification was included.

The likelihood of detecting automation errors did not vary significantly with the accountability manipulation. However, pilots who missed none or only one automation error, when asked to rate their perceptions of accountability, reported a higher sense of being evaluated on their performance and strategies, and reported a stronger need to justify how they interacted with the automation. Although the experimental manipulation of accountability did not affect error management, individuals who *reported* feeling less accountable were more likely to miss automation errors than those who reported feeling accountable.

In a following investigation, Skitka *et al.* (2000a) adjusted their experimental manipulation of accountability and examined the effect of general versus specific accountability instructions on performance. Participants were given either non-accountable instructions, or one of four types of accountable instructions. The accountable groups were informed that their goal was to maximize either overall performance, accuracy, response time, or tracking performance.

Forms of accountability had a significant effect on the number of omission and commission errors committed by participants. Omission errors (failing to detect automation misses) occurred less frequently in the accountability groups focused on overall performance, accuracy, and response time. Further, commission errors (failing to detect automation false alarms) occurred less frequently in the accountability groups focused on overall performance and accuracy. An examination of participants' verification behaviour revealed that the two groups with the best error management performance, those accountable for overall performance and accuracy, verified the automation more than the other three groups.

4.3.4 Summary of task variables—Aspects of the task that the human and automation are engaged in play a role in how well the human manages automation errors. As the cost of failing to correct automation errors increases, the likelihood of detecting those errors also increases. It is important to note that minor consequences are still consequences, and if these types of errors are largely ignored, the sum of their impact may create opportunities for serious accidents. In addition to consequences, the costs (both to performance and operators' cognitive load) associated with verifying the automation, and the form of accountability imposed on the operator regarding task success influence how likely operators are to detect automation errors.

4.4 Emergent variables

Several of the most widely researched variables in the human-automation literature cannot be ascribed solely to the automation, human, or task. These variables result from the interaction of all three components. For example, individuals' trust in automation is determined by previous experiences with automation, by the capabilities and design of the current automation, and by the task being worked on. Oftentimes, these emergent variables are examined as outcome variables, but there is evidence that they also predict the extent to which operators will successfully manage automation errors.

4.4.1 Trust in automation—Lee and See (2004) defined trust as “the attitude that an agent will help achieve an individual's goal in a situation characterized by uncertainty and vulnerability” (p. 51). Generally, individuals' trust in automation positively relates to their dependence on automation (Muir 1994, Muir and Moray 1996, Parasuraman and Riley 1997, Dzindolet *et al.* 2003). However, because dependence is not always warranted, as is the case when automation errs, a critical aspect of trust is calibration. Calibration is the match between capabilities of the automation and an individual's trust in the automation (Lee and See 2004). Dependent upon this match, individuals may have calibrated trust (trust matches automation capabilities), overtrust (trust exceed automation capabilities), or distrust (trust falls short of automation capabilities).

One of the first empirical investigations examining the effect of trust on performance was reported by Muir and Moray (1996). After each experimental trial, participants completed a series of subjective ratings to assess trust in the automated component. A high positive correlation was found between operators' trust and the amount of time they left the system in automatic mode, and a negative correlation was found between trust and operators' monitoring of the automation. Overall, the more operators trusted the automation, the more they left it in control without supervision. In this study, the effect of trust on error management could not be directly assessed because the automation did commit a distinct error, but went through periods during which manual control would be more effective and lead to superior system output. However, we suspect that because trust reduced operator's monitoring of the automation, it is likely that errors committed by the system would have gone undetected.

Another study assessed trust in an automated system that committed misses (Bailey and Scerbo 2007). Trust in the automation was measured with a 12-item questionnaire with

subscales for the two automated aids. Regression analyses revealed that trust in two automated aids significantly predicted error detection in the two tasks. Higher trust predicted lower detection rates of the automation misses.

As automated systems and automated components become more widespread in socio-technical systems, understanding the functional specificity of trust should become a research priority. Functional specificity is “the differentiation of functions, subfunctions, and modes of automation” (Lee and See 2004, p. 56). Consider the numerous automated systems within a modern glass cockpit (e.g., autopilot, collision avoidance, thrust management, navigation, engine indications). If operators’ trust varies by the automated component or function in question, they have high functional specificity. If trust is based on the system as a whole, not distinguishing between the various components, functional specificity is low. It currently is not clear how capable human operators are of demonstrating high levels of functional specificity. This question is an important one, because if high functional specificity is not common, then individuals may be inappropriately generalizing from one automated component to another and reducing the likelihood that they manage automation errors effectively.

Several studies have reported that an error in one automated component did not cause a loss of trust in other similar, but functionally distinct automated components, indicating individuals were capable of high functional specificity (Lee and Moray 1992, Muir and Moray 1996). Evidence to the contrary has surfaced in the last few years. Rice and colleagues have repeatedly found support for the notion that people display system-wide trust, rather than component-specific trust (Keller and Rice 2010, Rice and Geels 2010).

These differences in the functional specificity of trust may be related to the level of skill or experience operators have with the automation in question. Such a distinction may explain the contradictory results discussed above. In the two studies that did not find evidence of functional specificity, participants spent relatively minimal time with the automation (60 minutes in Keller and Rice 2010; 20 minutes in Rice and Geels 2010). However, in the studies that reported evidence for functional specificity of trust, participants’ interaction with the automation ranged from six hours (Lee and Moray 1992) to between eight and 16 hours (Muir and Moray 1996).

The literature supports the proposition that trust negatively influences error management processes. It is not clear whether trust in multiple automated components is distinct for each component, or if the various components are merged and trust is based on a perception of the system as a whole.

4.4.2 Mental workload—Workload can be thought of as the supply and demand of attentional or processing resources (Tsang and Vidulich 2006). Typically, operators are responsible for other tasks in addition to monitoring automation. If operators are under high levels of workload, they may not have resources available to engage in an appropriate level of monitoring or verification, reducing their error management efficacy.

In one study examining workload, the researchers manipulated workload by increasing the number of concurrent tasks across sessions within participants, and the rate of an alarm in a secondary task between participants (Bliss and Dunn 2000). Increasing the number of concurrent tasks significantly increased response time to alarms and decreased the rate of responding, whereas the effect of increasing the rate of alarms did not significantly affect performance.

In a study conducted by Dixon *et al.* (2005), workload varied as a function of the task demands as well as how many UAVs the operator was responsible for controlling (one = low workload; two = high workload). Both manipulations of workload had significant negative effects on response times to errors, but response accuracy was not affected. Thus far, the research suggests that in high workload situations, rather than letting performance suffer, individuals will adjust their strategies and engage in a speed/accuracy tradeoff wherein they may take longer to respond but accuracy remains high.

Contrary evidence was reported by McBride *et al.* (2011). The experimental task required participants to perform two tasks, one of which was supported by a 70% reliable automated aid. The search demands of the non-automated task were manipulated to create three levels of workload (low, moderate, and high). Workload had a negative effect on performance in the automated task, such that the high workload group detected significantly fewer automation errors than the low workload group.

These findings differ from the previous work that only showed a decrement in response time, but not in accurate performance. This may be explained by the variable time limits enforced across the experimental tasks. In McBride *et al.*'s (2011) task, participants had a 10 second window in which they needed to respond. Bliss and Dunn (2000) allotted 15 seconds for participants to respond to alarms, and Dixon *et al.* (2005) had 30 seconds to identify and indicate a failure had occurred. It may be that the time pressure of McBride *et al.*'s task did not allow the participants to take more time to respond and keep accuracy high. This idea needs to be empirically investigated, as removing or increasing the time limits might change the effect of workload and the strategies used by participants in unanticipated ways.

4.4.3 Situation awareness—Another factor critical in the process of automation error management is an individual's situation awareness (SA). Endsley (1996) defined SA as “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future” (p. 165). This definition specifies levels of SA, including level 1 (perception), level 2 (integration and comprehension), and level 3 (projection). Numerous empirical studies have demonstrated that these levels can vary independently.

With respect to the role of SA in human-automation interaction, much of the research has examined how the level of automation (LOA) affects SA, and how this process influences error management. The hypothesized relationship is that a high LOA places operators in a passive role, reducing their engagement in the task, and consequently, their SA. When the automation fails and operators must take over, their low SA will result in poor performance and longer time to recover from the failure. To that end, the experimental paradigms used

typically consist of a period where automated support is unavailable. In these scenarios, error management success would be defined as how well the individual can perform the task manually during those periods when automation is offline.

The findings of Endsley and Kaber (1999) do not follow the predicted relationship between LOA, SA, and recovery from automation failure. Their results demonstrated SA was not associated with performance, either in terms of time to recover during the failure conditions or manual performance. The groups with high SA did not perform significantly better or worse than the other groups. The only exception to this was the group using automated decision support (medium-high LOA) had high SA scores but one of the *lowest* levels of performance.

Kaber and Endsley (2004) reported further evidence of the dissociation between SA and performance. Using a manual condition (no automation) and five levels of automation, they demonstrated that participants using the lowest LOA and second highest LOA had substantially worse SA than the other LOAs. The group using the lowest LOA, that had the worst SA, also had the best task performance. Indeed, of the two LOA groups with the highest SA, one had the second highest task performance, and the other had the second worst performance.

The effect of SA on error management has received mixed support and requires additional investigation. One particular aspect of the experimental methodology needs to be examined. In all of the cited research, automation failed, rather than erred. The human did not have to react to misses or false alarms, but had to take over and perform the task manually because the automation went completely offline. Although the effect of SA in this context deserves attention, so too does the effect of SA on managing automation errors, such as misses and false alarms.

4.4.4 Emergent variables—Of the three emergent variables discussed, trust in automation is the only variable found to affect error management in a consistent manner. The influence of workload and SA are not clearly understood. Future efforts will benefit from systematic manipulations of workload and SA. For example, in the reviewed studies, SA was manipulated by varying the LOA. This design decision is based on the assumption that LOA has a straightforward effect on SA, but the literature does not provide adequate support for this conjecture.

5. Conclusion

5.1 Overview of analysis

Automation has the potential to aid humans with a diverse set of tasks and to support system performance in many contexts. Automated systems are not always perfectly reliable, which may pose a barrier to achieving the intended task success. Therefore, the presence of errors in human-automation interactions cannot be eliminated completely. When automation errs, the human must engage in error management, or the process of detecting, explaining, and correcting the error. The success of this process determines whether the consequences of an error will come to fruition.

To gain deeper insights into the process of error management, we examined relevant areas of the human error literature. Reason's (1990) GEMS described the general process people engage in to detect and resolve *human* errors. We took key components of this process and incorporated them into our own conceptualization of the error management process in human-automation interaction (see Figure 2). This framework describes different methods by which errors are detected, explained, and corrected. It also adds a process not specified by current models of error management; namely, integration. The integration process represents how people learn about the automation and refine their mental representation of the automation, which will likely influence future error management experiences.

Although we believe integration is a key component of error management, there is currently not a body of research aimed at understanding how integration occurs. For instance, does integration tend to occur naturally after experience with automation or a particular type of error? Or perhaps effective integration requires a more explicit and directed review of an interaction with automation, with the purpose of gleaning additional new pieces of information. If we understand the manner in which integration occurs and have evidence that integration improves future error management attempts, our next priority should be to understand how to train individuals to engage in this process.

Another gap in the automation literature is that error management has thus far been primarily measured at the level of detection. Research aimed at understanding the explanation and correction phases of error management is almost non-existent. Although identifying and predicting whether people can detect and correct automation errors is important, examining what explanatory processes allow them to generate the appropriate corrective action is similarly important.

Part of the reason explanation and correction have not been emphasized may lay with the research design. For instance, in many tasks participants are required to monitor for automation errors (e.g., Prinzel *et al.* 2005, Dixon *et al.* 2007). When an error is detected, participants do not have to engage in any error explanation processes because: a) they are trained to respond with a button press or mouse click that indicates they have detected the error and prompts the system to "correct" the error, or b) the only response option is to switch to manual control. Although these tasks may represent some operational environments, they do not accurately represent other environments that require more complex reasoning and explanation to produce a successful correction method. There are some exceptions to these types of experimental paradigms wherein corrective actions depend heavily upon the participant's ability to examine relevant system parameters to reach a decision (e.g., Bahner *et al.* 2008a, 2008b), but they are uncommon by comparison. Because the explanation phase of error management has received little attention in the literature, understanding how knowledge supports explanation, and thus, error management, should be a research priority.

Knowing the relative influence of variables would be informative but the data are not available to do a systematic quantitative analysis of relative effect sizes. However, on the basis of our review, we can speculate regarding what variables may be the most relevant to explanation and correction. For instance, explanation involves searching for familiar patterns

and also referring to one's own mental representation of the system. Greater experience with or knowledge of the automation might result in a greater quantity of relevant patterns that could be accessed, as well as a more refined mental representation. Additionally, if a critical component of correction is observing the system to determine if further action is needed, displays that communicate this information effectively are necessary. But what does this mean in terms of display design? There is much to be done to understand what happens after individuals detect problems when using complex automation in dynamic situations.

We focused our review on identifying the factors that influence a human's error management success. Critical variables included automation, person, task, and emergent variables (see Table 1). A summary of these variables' effects on error management is provided next.

The reliability of automation influences error management in a number of ways. The more reliable the automation, the greater the benefit to performance when the automation is functioning. On the other hand, during periods of failure associated with a highly reliable automated system, the human's ability to manage those errors may be severely disrupted. Further, as reliability becomes more variable, recovery from errors improves. Similar to the effect of reliability level, research has suggested that as the level of automation increases, performance during errors or failures suffers. Situation awareness has been noted as a potential mediator in this relationship, although the data supporting this assertion have been mixed.

We examined automation error type, but it is not clear whether one type (miss or false alarm) is harder to detect. High quality feedback provided by the automation to the operator regarding its activities and overall system state has been found to support error management. Further research is required to understand if feedback may accomplish this by refining knowledge, improving situation awareness, or both. The consequences associated with automation errors influences error management such that errors with more serious consequences are more likely to be detected compared to those with minor consequences. In addition, if operators are made more accountable for their performance (including error management), the more likely it is that these errors are managed. Individuals high in complacency potential are less effective monitors of automation, and struggle with error management, compared to low complacency potential individuals.

Trust in automation is appropriate during routine performance, but when the automation fails, that same trust often limits how well the person manages the failure. Operators' workload negatively affects their error management abilities, although this effect may be limited to situations with high levels of time pressure. In cases where time is not as constraining, individuals may sacrifice speed to maintain high levels of accuracy. The costs associated with verifying automation's behaviour have a negative impact on the likelihood of detecting when automation errs. Further, these costs will likely increase operator workload, compounding the problem further. Effective training programs, especially those that focus on automation errors and how to recover from them, support the development of error management skills. Investigations of operators' knowledge and situation awareness did not yield consistent findings regarding their effect on error management.

5.2 Themes and Next Steps

Our analysis uncovered themes in the empirical research on error management in an automation context. These themes represent avenues for future research that will increase understanding of the contexts in which humans excel or fail at managing automation errors. One such theme is the paradox of automation. That is, factors that positively affect performance when the automation is working may undermine performance when the automation errs or fails. These factors include automation reliability, level of automation, complacency potential, and trust. If this is the case, is it better to maximize routine performance when automation is correct and potentially forfeit error management, or focus efforts on improving the likelihood of catching automation errors? Kaber *et al.* (2011) suggested that perhaps the paradox could be overcome through designing effective displays, training to reduce complacency, or increasing operators' understanding of automation logic. However, these recommendations were not based on empirical data and specifically how these goals can be accomplished remains an open question.

In addition to overcoming the paradox of automation, another research priority should be to examine automation that assists with later stages of information processing, including decision selection and action implementation. The majority of the research reviewed here employed low level automation such as decision aids or systems designed to detect certain stimuli, such as equipment faults. It is more difficult to conceptualize errors at these later stages, as they may not fit into the standard categories of miss or false alarm provided by signal detection theory. Consider an automated system that is responsible for the majority of the task including information acquisition, analysis, decision making, and action implementation. Now consider how devastating an automation error or complete failure would be to system performance if the operator were unable to recover. As automation capabilities increase, so too does the need for operators to be able to manage automation errors at all levels.

Lastly, it seems critical to examine how these numerous variables interact with one another. In any given interaction between a human and automation, most, if not all, of these variables will be at play. How do their effects change when other variables vary as well? For example, there is evidence that individuals with high complacency potential will only exhibit error management deficits if they are in a complacency inducing situation, such as one with automation performing at a consistent rather than variable level of reliability (Prinzel *et al.* 2005). Similarly, although error type has a significant impact on error management under conditions of high workload, under low workload error management is at ceiling and no difference based on error type exists (Dixon and Wickens 2006).

This review has organized and identified the empirical findings relevant to understanding how individuals manage automation errors. Numerous variables are at play, and although there are some areas in which the findings are relatively clear, there remain many open questions to be answered. Continued efforts to explore error management will be necessary to move towards a future in which imperfect automation can be designed and implemented in a way that supports overall system performance.

Biography

Sara E. McBride is a graduate student in the Engineering Psychology program at the Georgia Institute of Technology in Atlanta, Georgia, where she received her M.S. in Psychology in 2010.

Wendy A. Rogers is a Professor in the School of Psychology at the Georgia Institute of Technology in Atlanta, Georgia, where she received her Ph.D. in Psychology in 1991.

Arthur D. Fisk is a Professor in the School of Psychology at the Georgia Institute of Technology in Atlanta, Georgia. He received his Ph.D. in Psychology from the University of Illinois in 1982.

References

- Bahner JE, Elepfandt MF, Manzey D. Misuse of diagnostic aids in process control: The effects of automation misses on complacency and automation bias. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 2008a; 52(19):1330–1334. doi: 10.1177/154193120805201906.
- Bahner JE, Hüper A-D, Manzey D. Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies*. 2008b; 66(9):688–699. doi: 10.1016/j.ijhcs.2008.06.001.
- Bailey NR, Scerbo MW. Automation-induced complacency for monitoring highly reliable systems: The role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science*. 2007; 8(4):321–348. doi: 10.1080/14639220500535301.
- Bainbridge L. Ironies of automation. *Automatica*. 1983; 19:776–779.
- Bliss JP, Dunn MC. Behavioural implications of alarm mistrust as a function of task workload. *Ergonomics*. 2000; 43(9):1283–1300. [PubMed: 11014752]
- Clark, K. The GPS: A fatally misleading travel companion [online]. National Public Radio (NPR); 2011. Available from: <http://www.npr.org/2011/07/26/137646147/the-gps-a-fatally-misleading-travel-companion> [Accessed 4 September 2011]
- Cohen, MS.; Parasuraman, R.; Freeman, JT. Trust in decision aids: A model and its training implications. Paper presented at the 1998 Command and Control Research and Technology Symposium; Monterey, CA. 1998.
- Dixon SR, Wickens CD. Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors*. 2006; 48(3):474–486. [PubMed: 17063963]
- Dixon SR, Wickens CD, Chang D. Mission control of multiple unmanned aerial vehicles: A workload analysis. *Human Factors*. 2005; 47(3):479–487. doi: 10.1518/001872005774860005. [PubMed: 16435690]
- Dixon SR, Wickens CD, McCarley JS. On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors*. 2007; 49(4):564–572. doi: 10.1518/001872007x215656. [PubMed: 17702209]
- Dzindolet MT, Peterson SA, Pomranky RA, Pierce LG, Beck HP. The role of trust in automation reliance. *International Journal of Human-Computer Studies*. 2003; 58(6):697–718. doi: 10.1016/s1071-5819(03)00038-7.
- Dzindolet MT, Pierce LG, Beck HP, Dawe LA, Anderson BW. Predicting misuse and disuse of combat identification systems. *Military Psychology*. 2001; 13(3):147–164.
- Endsley, MR. Automation and situation awareness. In: Parasuraman, R.; Mouloua, M., editors. *Automation and human performance: Theory and applications*. Lawrence Erlbaum; Mahwah, NJ: 1996. p. 163-181.

- Endsley MR, Kaber DB. Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*. 1999; 42(3):462–492. [PubMed: 10048306]
- Endsley MR, Kiris EO. The out-of-the-loop performance problem and level of control in automation. *Human Factors*. 1995; 37(2):381–394.
- Ezer N, Fisk AD, Rogers WA. Age-related differences in reliance behaviour attributable to costs within a human-decision aid system. *Human Factors*. 2008; 50(6):853–863. doi: 10.1518/001872008x375018. [PubMed: 19292009]
- Green, DM.; Swets, JA. Signal detection theory and psychophysics. Wiley; New York, NY: 1988.
- Hancock PA, Jagacinski RJ, Parasuraman R, Wickens CD, Wilson GF, Kaber DB. Human-automation interaction research: Past, present, and future. *Ergonomics in Design: The Quarterly of Human Factors Applications*. 2013; 21(2):9–14. doi: 10.1177/1064804613477099.
- Kaber DB, Endsley MR. 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*. 2004; 5(2):113–153. doi: 10.1080/1463922021000054335.
- Kaber DB, Onal E, Endsley MR. Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. *Human Factors and Ergonomics in Manufacturing*. 2000; 10(4):409–430.
- Kanse L, van der Schaaf T. Recovery from failures in the chemical process industry. *International Journal of Cognitive Ergonomics*. 2001; 5(3):199–211.
- Kantowitz BH, Hanowski RJ, Kantowitz SC. Driver acceptance of unreliable traffic information in familiar and unfamiliar settings. *Human Factors*. 1997; 39(2):164–176. doi: 10.1518/001872097778543831.
- Keller D, Rice S. System-wide versus component-specific trust using multiple aids. *Journal of General Psychology*. 2010; 137(1):114–128. doi: 10.1080/00221300903266713. [PubMed: 20198819]
- Kontogiannis T. User strategies in recovering from errors in man-machine systems. *Safety Science*. 1999; 32(1):49–68. doi: 10.1016/s0925-7535(99)00010-7.
- Kontogiannis T. A systems perspective of managing error recovery and tactical re-planning of operating teams in safety critical domains. *Journal of Safety Research*. 2011; 42(2):73–85. doi: 10.1016/j.jsr.2011.01.003. [PubMed: 21569889]
- Kontogiannis T, Malakis S. A proactive approach to human error detection and identification in aviation and air traffic control. *Safety Science*. 2009; 47(5):693–706. doi: 10.1016/j.ssci.2008.09.007.
- Lee J, Moray N. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*. 1992; 35(10):1243–1270. doi: 10.1080/00140139208967392. [PubMed: 1516577]
- Lee JD, See KA. Trust in automation: Designing for appropriate reliance. *Human Factors*. 2004; 46(1):50–80. doi: 10.1518/hfes.46.1.50.30392. [PubMed: 15151155]
- Lorenz B, Di Nocera F, Röttger S, Parasuraman R. Automated fault-management in a simulated spaceflight micro-world. *Aviation, Space, and Environmental Medicine*. 2002; 73(9):886–897.
- Maltz M, Shinar D. Imperfect in-vehicle collision avoidance warning systems can aid drivers. *Human Factors*. 2004; 46(2):357–366. [PubMed: 15359683]
- McBride SE, Rogers WA, Fisk AD. Understanding the effect of workload on automation use for younger and older adults. *Human Factors*. 2011; 53(6):672–686. doi: 10.1177/0018720811421909. [PubMed: 22235529]
- Metzger U, Parasuraman R. Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. *Human Factors*. 2005; 47(1):35–49. [PubMed: 15960085]
- Molloy R, Parasuraman R. Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors*. 1996; 38(2):311–322. doi: 10.1518/001872096779048093.
- Moray N. Monitoring, complacency, scepticism and eutactic behaviour. *International Journal of Industrial Ergonomics*. 2003; 31(3):175–178.
- Moray N, Inagaki T, Itoh M. Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology-Applied*. 2000; 6(1):44–58. doi: 10.1037//0278-7393.6.1.44. [PubMed: 10937311]

- Mosier, KL.; Skitka, LJ. Human decision makers and automated decision aids: Made for each other?. In: Parasuraman, R.; Mouloua, M., editors. *Automation and human performance: Theory and applications*. Lawrence Erlbaum; Mahwah, NJ: 1996. p. 201-220.
- Mosier KL, Skitka LJ, Dunbar M, McDonnell L. Aircrews and automation bias: The advantages of teamwork? *International Journal of Aviation Psychology*. 2001; 11(1):1–14. doi: 10.1207/s15327108ijap1101_1.
- Mosier KL, Skitka LJ, Heers S, Burdick M. Automation bias: Decision making and performance in high-tech cockpits. *International Journal of Aviation Psychology*. 1998; 8(1):47–63. doi: 10.1207/s15327108ijap0801_3. [PubMed: 11540946]
- Muir BM. Trust in automation: I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*. 1994; 37(11):1905–1922.
- Muir BM, Moray N. Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*. 1996; 39(3):429–460. doi: 10.1080/00140139608964474.
- Nikolic MI, Sarter NB. Peripheral visual feedback: A powerful means of supporting effective attention allocation in event-driven, data-rich environments. *Human Factors*. 2001; 43(1):30–38. doi: 10.1518/001872001775992525. [PubMed: 11474762]
- Norman, DA. The ‘problem’ with automation: Inappropriate feedback and interaction, not ‘over-automation’. In: Broadbent, DE.; Reason, JT.; Baddeley, AD., editors. *Human factors in hazardous situations*. Clarendon Press/Oxford University Press; New York, NY: 1990. p. 137-145.
- Palmer, E.; Degani, A. Electronic checklists: Evaluation of two levels of automation. *Proceedings of the Sixth International Aviation Psychology Symposium; 29 April – 2 May 1991; Columbus, Ohio: Ohio State University, Department of Aviation; 1991. p. 178-183.*
- Parasuraman R, Molloy R, Singh IL. Performance consequences of automation-induced ‘complacency.’. *International Journal of Aviation Psychology*. 1993; 3(1):1–23. doi: 10.1207/s15327108ijap0301_1.
- Parasuraman R, Riley V. Humans and automation: Use, misuse, disuse, abuse. *Human Factors*. 1997; 39(2):230–253.
- Parasuraman R, Sheridan TB, Wickens CD. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems Man and Cybernetics Part a-Systems and Humans*. 2000; 30(3):286–297.
- Parasuraman R, Wickens CD. Humans: Still vital after all these years of automation. *Human Factors*. 2008; 50(3):511–520. doi: 10.1518/001872008x312198. [PubMed: 18689061]
- Parasuraman R, de Visser E, Lin M-K, Greenwood PM. Dopamine beta hydroxylase genotype identifies individuals less susceptible to bias in computer-assisted decision making. *Public Library of Science ONE*. 2012; 7(6):1–9. doi:10.1371/journal.pone.0039675.
- Prinzel LJ III, Freeman FG, Prinzel HD. Individual differences in complacency and monitoring for automation failures. *Individual Differences Research*. 2005; 3(1):27–49.
- Reason, J. *Human error*. Cambridge University Press; Cambridge, UK: 1990.
- Rice S, Geels K. Using system-wide trust theory to make predictions about dependence on four diagnostic aids. *Journal of General Psychology*. 2010; 137(4):362–375. doi: 10.1080/00221309.2010.499397. [PubMed: 21086858]
- Rice S, McCarley JS. Effects of response bias and judgment framing on operator use of an automated aid in a target detection task. *Journal of Experimental Psychology: Applied*. 2011; 17(4):320–31. doi: 10.1037/a0024243. [PubMed: 21707202]
- Riley, V. Operator reliance on automation: Theory and data. In: Parasuraman, R.; Mouloua, M., editors. *Automation and human performance: Theory and applications*. Lawrence Erlbaum; Mahwah, NJ: 1996. p. 19-35.
- Rovira E, McGarry K, Parasuraman R. Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors*. 2007; 49(1):76–87. [PubMed: 17315845]
- Sanchez J, Fisk AD, Rogers WA, Rovira E. Understanding reliance on automation: Effects of error type, error distribution, age and experience. *Theoretical Issues in Ergonomics Science*. 2011 Advance online publication. doi:10.1080/1463922X.2011.611269.

- Sarter N. Investigating mode errors on automated flight decks: Illustrating the problem-driven, cumulative, and interdisciplinary nature of human factors research. *Human Factors*. 2008; 50(3): 506–510. doi: 10.1518/001872008x312233. [PubMed: 18689060]
- Sarter NB, Mumaw RJ, Wickens CD. Pilots' monitoring strategies and performance on automated flight decks: An empirical study combining behavioral and eye-tracking data. *Human Factors*. 2007; 49(3):347–357. doi:10.1518/001872007X196685. [PubMed: 17552302]
- Sarter NB, Woods DD. Pilot interaction with cockpit automation II: An experimental study of pilots' model and awareness of the flight management and guidance system. *International Journal of Aviation Psychology*. 1994; 4(1):1–28.
- Sarter, NB.; Woods, DD.; Billings, CE. Automation surprises. In: Salvendy, G., editor. *Handbook of human factors and ergonomics*. 2nd ed. Wiley; New York, NY: 1997. p. 1926-1943.
- Sheridan TB. Risk, human error, and system resilience: Fundamental ideas. *Human Factors*. 2008; 50(3):418–426. doi: 10.1518/001872008x250773. [PubMed: 18689048]
- Singh IL, Molloy R, Parasuraman R. Automation-induced 'complacency': Development of the complacency-potential rating scale. *International Journal of Aviation Psychology*. 1993a; 3(2): 111–122. doi: 10.1207/s15327108ijap0302_2.
- Singh IL, Molloy R, Parasuraman R. Individual differences in monitoring failures of automation. *Journal of General Psychology*. 1993b; 120(3):357.
- Skitka LJ, Mosier KL, Burdick M. Does automation bias decision-making? *International Journal of Human-Computer Studies*. 1999a; 51(5):991–1006. doi: 10.1006/ijhc.1999.0252.
- Skitka LJ, Mosier K, Burdick MD. Accountability and automation bias. *International Journal of Human-Computer Studies*. 2000b; 52(4):701–717. doi: 10.1006/ijhc.1999.0349.
- Skitka LJ, Mosier KL, Burdick M, Rosenblatt B. Automation bias and errors: Are crews better than individuals? *International Journal of Aviation Psychology*. 2000; 10(1):85–97. [PubMed: 11543300]
- Skjerve A, Skraaning G. The quality of human-automation cooperation in human-system interface for nuclear power plants. *International Journal of Human-Computer Studies*. 2004; 61(5):649–677. doi:10.1016/j.ijhcs.2004.06.001.
- Sorkin RD, Woods DD. Systems with human monitors: A signal detection analysis. *Human-Computer Interaction*. 1985; 1(1):49–75.
- Tsang, PS.; Vidulich, MA. Mental workload and situation awareness. In: Salvendy, G., editor. *Handbook of human factors and ergonomics*. 3rd ed. John Wiley & Sons, Inc.; Hoboken, NJ: 2006. p. 243-268.
- Wickens, CD.; Carswell, CM. Information processing. In: Salvendy, G., editor. *Handbook of human factors and ergonomics*. 3rd ed. John Wiley & Sons, Inc.; Hoboken, NJ: 2006. p. 1570-1596.
- Wickens CD, Dixon SR. The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science*. 2007; 8(3):201–212. doi: 10.1080/14639220500370105.
- Wickens, CD.; Li, H.; Santamaria, A.; Sebok, A.; Sarter, NB. Stages and levels of automation: An integrated meta-analysis. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting*; 2010. p. 389-393.
- Wiener EL, Curry RE. Flight-deck automation: Promises and problems. *Ergonomics*. 1980; 23(10): 995–1011. doi: 10.1080/00140138008924809.
- Wilkison, B.; Fisk, AD.; Rogers, WA. Effects of mental model quality on collaborative system performance. *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting*; 2007. p. 1506-1510.
- Woods, DD. Decomposing automation: Apparent simplicity, real complexity. In: Parasuraman, R.; Mouloua, M., editors. *Automation and human performance: Theory and applications*. Lawrence Erlbaum; Mahwah, NJ: 1996. p. 3-17.
- Woods, DD.; Dekker, S.; Cook, R.; Johannesen, L.; Sarter, N. *Behind human error*. 2nd ed. Ashgate; Burlington, VT: 2010.

Relevance of the findings for ergonomics theory

This article presents a comprehensive analysis of variables that influence error management in interactions with imperfect automation. Our framework expands upon previous work to describe the relevant cognitive processes involved in error management. Our analysis provides direction for future research aimed at understanding and supporting error management processes.

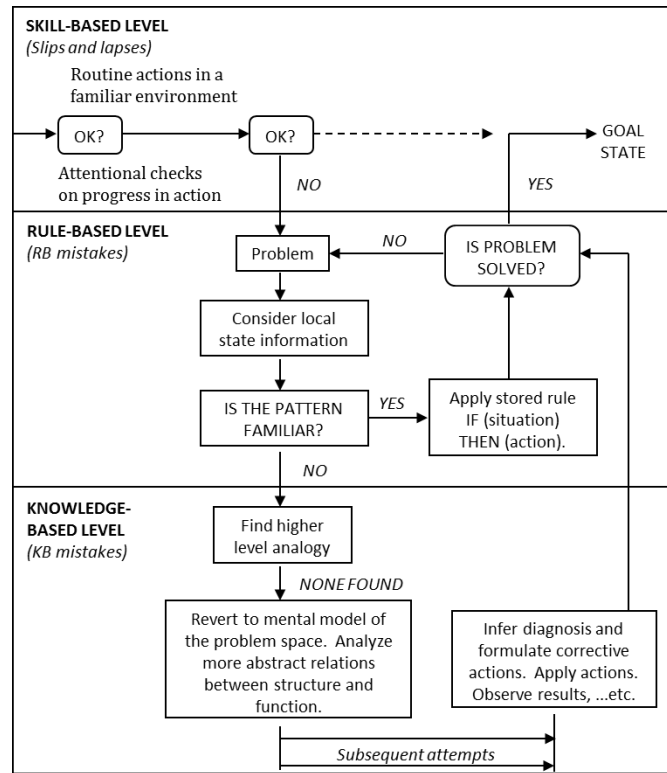


Figure 1. The mechanics of the generic error modeling system (GEMS), adapted from Reason (1990).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

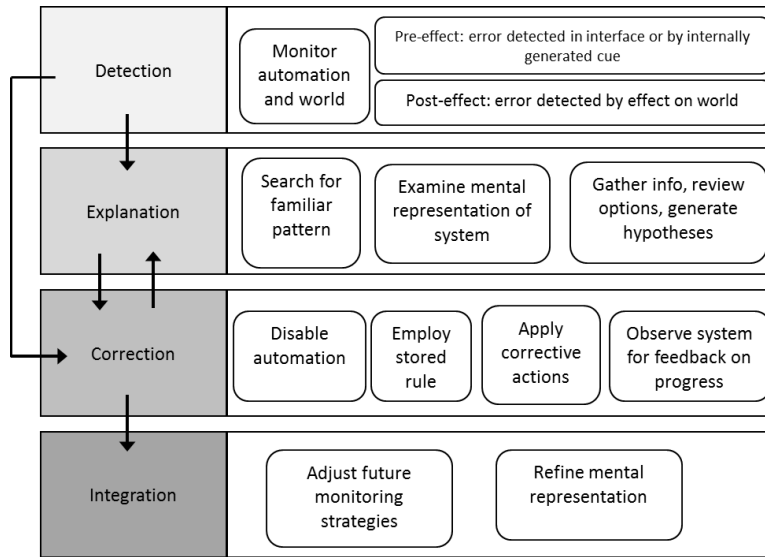


Figure 2. Depiction of the error management process as it occurs in human-automation interaction.

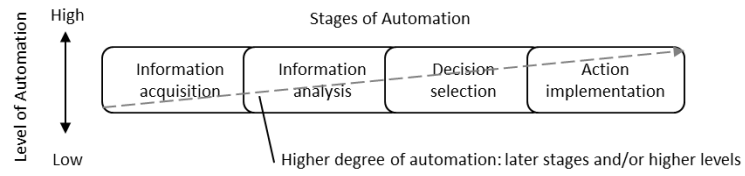


Figure 3. Illustration of degrees of automation adapted from Wickens *et al.* (2010). Degree of automation (diagonal line) is defined by higher levels and later stages.

Table 1

Framework for variables influencing error management.

Category	Definition	Examples
Automation Variables	Characteristics of the automation	Reliability level Error type Level of automation Feedback
Person Variables	Factors unique to the person interacting with automation	Complacency potential Training received Knowledge of automation
Task Variables	Context in which human and automation are working together	Automation error consequences Verification costs Human accountability
Emergent Variables	Factors that arise from the interaction between human and automation	Trust in automation Workload Situation awareness

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Endsley and Kaber's (1999) hierarchy of levels of automation applicable to dynamic-cognitive and psychomotor control task performance.

Level of automation	Roles			
	Monitoring	Generating	Selecting	Implementing
1)Manual control	Human	Human	Human	Human
2)Action support	Human/Computer	Human	Human	Human/Computer
3)Batch processing	Human/Computer	Human	Human	Computer
4)Shared control	Human/Computer	Human/Computer	Human	Human/Computer
5)Decision support	Human/Computer	Human/Computer	Human	Computer
6)Blended decision making	Human/Computer	Human/Computer	Human/Computer	Computer
7)Rigid system	Human/Computer	Computer	Human	Computer
8)Automated decision making	Human/Computer	Human/Computer	Computer	Computer
9)Supervisory control	Human/Computer	Computer	Computer	Computer
10) Full automation	Computer	Computer	Computer	Computer