Review

Change in Human (Moral) Decision-Making and Performance with Automation: The Good, the Bad, and the Ugly in Human-Automation Interaction

Arthur Prével [1,*] and Salvatore Lo Bue [1]

[1] Department of Life Sciences, Royal Military Academy, Hobbema 8, 1000, Brussels, Belgium; arthur.aeac@gmail.com
[1] Department of Life Sciences, Royal Military Academy, Hobbema 8, 1000, Brussels, Belgium; Salvatore.lobue@mil.be
* Correspondence: arthur.aeac@gmail.com

Abstract: Automation technologies are present in almost every domain of human activity and they are now more and more present in our everyday life. The reason for this massive deployment of automated systems would reside in all the benefits they offer to the users. In experimental settings, multiple studies have demonstrated the positive effects the introduction of automation can have on human decision-making and performance. However, studies have also demonstrated that the introduction of automation can have important negative effects as well. Considering that automation is now introduced in sensitive domains like military defense or medicine, more than ever we need a complete understanding of the effects caused by these systems on human performance and decision-making, and particularly in tasks and contexts with social or moral dimension. In this paper we will firstly review the main effects produced on a human agent's behaviors by the introduction of automation. Then, we will review the conditions identified as underlying factors of these effects, and see how they are currently integrated in models of human – automation interaction. We will conclude this review by highlighting new directions for future investigations on human – automation interaction.

Keywords: automation bias; human – automation interaction; human decision-making; level of automation; moral decision

1. Introduction

Over the past 50 years, human activities have been radically changed by the increased use of automation [1-3]. There is almost no area in which these technologies are not present to support people in monitoring activities, decision making, or to help in action execution [e.g., 4,5]. Automation helps people finding the optimal route to pay a visit friends or family, assists workers in industry in the execution of difficult tasks, or support physicians in diagnostics and drug prescription. In that respect, the usage of automated systems is now widespread in driving [6,7], aviation [8-10], military defense and security [11], medicine [12,13], nuclear plants [14, 15], etc., with the part of automation involved in each of these domains constantly increasing [3].

The reason for this massive deployment of automated systems would reside in all the benefits they offer to the users. Automated systems can make the driving experience much more comfortable, reduce work accidents in manufacturing, or help in the detection of pathologies and negative drug-drug combinations. In the lab, multiple experiments have shown how the use of automated systems results in shorter response times, errors decrement, or even a reduced workload and multi-tasking increment, in compari-

 doi:10.20944/preprints202203.0070.v1

son with conditions without assistance [16-21]. With the level of automatization constantly increasing, we are now close to a stage where human activities will change from active participation to a task to a more supervisory role, with the ultimate goal for engineers being a (quasi-)full automation of the systems [1; 3; 10; 11; 22].

Unfortunately, the involvement of automated technologies in human activities has not systematically been associated with positive effects for the users' or the system's performance. More automatization does not result necessarily in systematic performance increment or correct decisions. Dramatic examples include the accident of Air France flight 447 (AF447) on May 2009, where a failure in the autopilot mode coupled with a (supposed) loss of situation awareness from the pilots resulted in the plane stalling and falling into the Atlantic Ocean, killing all the 12 staff and 216 passengers on board. Others examples with automated vehicles or weapons are also reported in the literature [e.g., 23,24].

For that reason, now scientists are studying human – automated system interactions more thoroughly in order to identify and describe the negative effects arising concurrently with the benefits of automation, and the conditions in which these effects are observed [2, 25]. For example, studies found that the use of automated systems could sometimes result in overreliance in the system's decisions, a degradation of manual skills, or in a loss of feeling of control about the ongoing task [21, 26-28]. Understanding the conditions that produce these effects is an absolute necessity. It would help us to develop automated technologies that guarantee safety, improve performances and guide optimal decision-making from the human users, without the potential negative outcomes. This is particularly true when automated systems are employed in tasks and contexts that involve moral decision-making [29-31].

For instance, a combat drone operator, based on the automated information and/or suggestions received from the system, must make moral decisions when he/she is engaged on a battlefield and must choose between bombing a detected military target – with the risk of errors and collateral damages – or not bombing the target – with the risk of later attacks from the enemy also implying civilian losses and material damages. In that situation, the overreliance in the system's decisions might have dramatic consequences. But, to the best of our knowledge, the majority of research about moral dilemmas and automated systems concerns the algorithms and values to assign to the systems to guide their decisions during moral decision-making situations [e.g., 32-35]. However, another fundamental issue is to understand how the moral decisions made by a human agent are influenced by the interaction with automated systems. Considering the negative effects mentioned above, like automation-bias or loss of feeling of control, it is possible that the interaction with automated systems also impacts negatively these decisions and the corollary actions [36,37].

Thus, with the use of automated systems in sensitive domains, like military defense and security, medicine, or nuclear plants, more than ever we need a complete understanding of the effects caused by these systems on human performance and decision-making, and particularly in tasks and contexts with moral value. Trivial in some cases, like when we are biased to follow the instructions from a GPS even when a road-sign indicated that the road on which we engaged is closed in 200 meters, overreliance on automation or a decrement in feeling of responsibility might have dramatic effects when the outcomes of decisions and actions involve human lives. The first objective of this paper is to propose an overview of the effects on human decision-making and performance associated with automated systems. We will start with a review of the main positive effects that these systems produce. Then, we will delve into the negative effects and their associated conditions reported with the use of automated systems, with a specific discussion on situations that involve a moral or social dimension. After this review of

 doi:10.20944/preprints202203.0070.v1

empirical findings, in a second part of this paper, we will review factors that are known to influence human – automation interaction. After that review, we will present some of the psychological models developed to account for the effects reported in part one. We will discuss the benefits of these models but also their limits and potential ways of improvement. Thus, the second purpose of this paper is to propose a review of the known-determining factors in human-automation interaction, and a discussion on the conceptual models used to interpret these findings. By doing so, we hope to pave the way for new research on that domain. Particularly, we hope this review will result in new investigations and new insights on the circumstances that produce the negative effects sometimes observed with automated systems, and on how it is possible to counteract these effects. We will conclude this review by highlighting potential fruitful directions for future investigations on human – automation interaction.

2. Interaction with Automation and Its Effect on Human Decision-Making and Performance

2.1. *The good: Behavior enhancement with automation*

Automation is defined as the execution of a task by a machine that was initially performed by a human agent [2]. With the massive increment in computer performance and the remarkable development of artificial intelligence and machine learning, automation is now largely dominated by computerized automation [3]. Automation offers many benefits in human activities. At an industrial level, automation increases manufacturing efficiency and productivity, and reduces the risk of accidents. At an individual level, automation is supposed to make the user/consumer experience much more comfortable and effective as well. Automation, of course, is not all or none. The extent to which a task is performed by automation varies across needs and situations. Thus, automation is characterized by different levels of automation [38, 39], and goes from fully manual performance or very basic automation processes (e.g., automated data acquisition), to fully autonomous systems. In addition, the introduction of automation can be done at different stages of task execution. A classic description is given by Parasuraman et al. [38; see also 40]. According to the authors, automation can be assigned to 1) information acquisition, 2) information analysis, 3) decision-making, or 4) action processing (for different descriptions, see also [25, 41]). Thus, automation can help either to detect relevant information from the environment and to analyze more deeply the information collected (e.g., by inferring about outcomes likelihood), or to help the user in making the optimal decision considering the information collected and executing the action selected. In this first section and for the rest of the paper, we will put our attention on systems in which a human agent is still involved in the task process. Because our focus is on the effect of automation on human performance and decision-making and fully autonomous systems remain exceptional, the scientific findings presented in this review will not discuss the performance and the problematics relate to fully autonomous systems. Now, we will illustrate the positive effects of automation on human performance and decision-making with some relevant findings found in the literature following the stage classification proposed by Parasuraman et al. [38].

As we have seen the first stage is information acquisition. Information acquisition automation commonly consists in adding salient stimuli (or cues), or in changing the properties of the environment, to facilitate the detection of relevant target stimuli and/or promote appropriate decision-making. A common example of information acquisition automation is additional visual cues superimposed on a target that needs to be detected. For example, Yeh & Wickens [42] asked participants in a virtual reality environment to pilot an unmanned aerial vehicle and search for hidden military targets (e.g., a tank). Participants could be assisted by automation that consisted in a reticle automatically superimposed on a target when detected by the system. Here, target detection increased

 doi:10.20944/preprints202203.0070.v1

with the help of the target cueing automation (see also [43]). Similarly, Goh et al. [17] asked participants in a luggage screening task to detect the presence of a knife in luggage images shown on a monitor. Some of the participants were assisted by an automation that superimposed a salient green circle to detected knife (i.e., a target). Consistent with the results by Yeh & Wickens [42], the authors observed that the proportion of correct detection was superior in the group with automation compared to the group without automation (see also [16, 44]). Interestingly, the stimulus used for information automation can have different physical properties. Rice & McCarley [20] used for example a text message as cue, while Dixon & Wickens [45] used an auditory stimulus. Finally, in a study by St. John et al. [46], information automation consisted in changing directly the saliency of the target stimuli.

The second stage of automation is information analysis automation. In this form of automation, information collected from (multiple sources of) the environment is integrated by the system and/or is analyzed to make predictions about future states of the environment. In the study by St. John et al. [46] mentioned above, participants in a simulated naval air defense task had to monitor a visual airspace to defend a ship against aerial threats. Participants were supported in this task by an automation system that continuously analyzed the level of threat represented by the aerial vehicles displayed on the monitor, and changed their saliency on the screen accordingly. This automation decreased the response time to the threatening aircraft compared to a situation without automation. Another good example of information analysis automation is clinical decision support system. Clinical decision support systems are automation used to help physicians in diagnoses or drug prescriptions. For example, Martinez-Franco et al. [47] tested the effect of DXplain, a decision support system developed to generate lists of ranked diagnostic hypotheses based on information put in the system. The authors recorded the rate of correct diagnoses from first-year medicine students and results found that the use of decision support system increased the correct diagnoses (see also [48]).

Automation is thus well suited to assist humans in the detection and the analysis of relevant stimuli from the environment, and to promote appropriate decisions. Now, we will present examples of automation systems that support directly the user in decision-making process (stage 3) and action execution (stage 4). In decision automation, automation is designed to propose or to select an action among a set of possibilities, based on the valuation of the different available actions and their potential outcomes. Action automation simply refers to the actual execution of the action selected by the system (or the user). A good example of the effect of decision automation comes from a study by Rovira et al. [21] in the military domain. Here, in a command-and-control task, participants were asked to engage enemies with friendly units on a simulated battlefield. Participants could either based their decisions on basic information about the possible engagement combinations, or based on the support of a decision assistance that provided with varying degree of precisions the best options to select. The results found by Rovira and colleagues show that decision accuracy (in reliable decision trials) are superior to basic information, and increases with the degree of precision. We can cite also a study by MacMillan et al. [19] who found that participants were better in a simulated air traffic control environment, measured by a reduced number of aircraft being on hold, when they were supported by decision assistances. Similarly, Sarter & Schroeder [49] found that pilots tested in an aircraft simulator were better at managing icing condition when recommendations on the action to take were proposed by a decision system.

By guiding the detection of relevant stimuli in the environment, by making predictions about future states and outcomes, or by making direct recommendations about the correct decisions to make, multiple experiments have demonstrated how automation produce positive effects on performance and decision-making. A consequence of the facilitation from automation is the increased possibility for multi-tasking for the human

 doi:10.20944/preprints202203.0070.v1

agent [e.g., 27, 50-53]. The possibility for multi-tasking represents another advantage of automation over pure manual task. For example, Cullen et al. [50] in a multi-task environment showed that information cues helped to increase efficient alternation of attention allocation across the different tasks. More recently, Wright et al. [53] found that decision automation helped participants to monitor the transport of multiple unmanned vehicles by making routes recommendations, compared to a situation without assistance. In summary, not only automation can help to perform a specific task, but automation can also efficiently assist human agents when they have to monitor multiple sources of information and execute multiple tasks. Unfortunately, the introduction of automation might also produce serious drawbacks. The next section will be devoted to present the bad in human – automation interaction.

2.2. *The bad: Negative effects of automation on human decision-making and performance*

Troubles with automation are not new in the scientific literature. First evidence and discussions of negative effects produced by the interaction with automation were reported around the 1980s and 1990s [e.g., 2, 54, 55], mainly in the aircraft domain [e.g., 56, 57]. Today, it is largely acknowledged by scientists and engineers that introducing automation is not just a "substitution" of an intelligent system or a machine for a human activity [58]. Automation does not automatically reduce the amount of work that a human agent needs to allocate to a task as it does not make the user's experience necessarily easier. 50 years of research teach us that the appropriate balance between automation and human control must be well considered before the introduction of new automation [59]. Otherwise, automation can have clear detrimental effects on human decision-making and performance, which can result in dramatic consequences related to performance and safety [2]. In this section, we will review the main findings in the scientific literature on the negative effects reported when automation is introduced in human activity. Here, we will narrow the presentation to a quasi-strict description of the effects produced on performance and decision-making. Mediating factors and concepts (e.g., loss of situation awareness, complacency) explaining the influence of automation on those outcomes will be discussed in the second part of the manuscript.

The first negative effect of automation that we can discuss concerns, with some irony, multi-tasking. As we have seen in the previous section, automation is designed for the execution of tasks initially performed by a human agent, with the consequence of reducing the number of actions a user has to perform and/or helping for multi-tasking. However, this beneficial effect of automation is true as long as automation is properly designed and introduced. Like performance decreases with manual multi-tasking, introducing too much automated-tasks to control might also have detrimental effect on performance (for review, see for example [60]). For instance, Chen and Joyner [61] in a simulated mounted combat system reported that the performance on a target gunnery task decreases with the introduction of an additional automated task, particularly with low level of automation, while perceived workload increased. Wang et al. [62] in a search and rescue task with multiple robots found that the exploration of the environment on one hand, and search on a screen of targets to rescue on the other hand, increased with the number of robots involved in the mission (4, 8, or 12). However, the authors found that performance decreased when participants had to control both exploration and search on screen from 8 to 12 robots, and perceived workload increased in each condition with the number of robots. Similar results were found by Adams [63] or Velagapudi et al. [64] in robots-control tasks. These examples show that introducing automation is definitely not enough to improve human multi-tasking performance, and inappropriate level of automation and task allocation can result, at the opposite, in substantial performance decrement and more error from the human user. Particularly, multi-tasking studies with automation suggest that people are particularly sensitive to automation-bias,

 doi:10.20944/preprints202203.0070.v1

which is certainly one of the most important negative effects associated with automation.

Automation bias is defined as the tendency of humans interacting with automation to use automated cues as heuristic replacement for information seeking and processing [65]. More exactly, automation bias is said to occur when the performance of a user decreases because of incorrect information and/or decision from an automated system [25, 66]. Automation bias is manifest in two types of errors: omission error and commission error. Omission error occurs when an automated system does not inform about a significant event (e.g., a weapon not detected in a luggage screening task), which result in the user not taking the appropriate decision in that situation (e.g., no check of the screened-luggage). At the opposite, in commission error the system makes an incorrect decision about the environment or gives incorrect advices, which result again in inappropriate response from the user (e.g., the system consider there is a weapon in a screened-luggage while there is not). Thus, automation bias corresponds to the fact that incorrect information or decision cues from the automated system, but not the actual environment, control the decisions and actions from a human agent. Examples of automation biases have been reported in almost any tasks and domains involving automation [25, 66]. For longtime, automation bias has been associated with multi-tasking situations. For example, Mosier et al. [28] found both omission and commission errors in pilots tested in a simulated flight task in which multiple flying tasks had to be monitor and supported by not totally reliable automation systems (see also [67-69]). However, it seems now evident that automation bias affects also single-task environment [70]. For instance, Alberdi et al. [71] found omission errors in a computer-assisted detection task for mammography, while Goddard et al. [72] reported commission errors caused by clinical decision-support system in prescription task. In the command-and-control study by Rovira et al. [21] cited above, despite the beneficial effect of decision automation on correct responses made by the participants, the authors also found incorrect responses during unreliable trials. In summary, automation-bias occurs both in single- and multiple-tasks environments, both for omission and commission errors, and is observed for both information and decision automation.

In addition to inappropriate automation-task allocation and automation bias, another negative effect of automation we can mention is loss of skills ([55]; or skill decay). Loss of skills refers to a degradation in task performance (motor or cognitive) after a more or less prolonged experience of a user with automation. To illustrate this phenomenon, we can think about a driver that has difficulty to drive an old car after a long period of driving a very modern one with a lot of automated assistances. In the scientific literature, evidence of loss of skills were found for example in fine-motor flying skills [73] or flight planning [74]. This effect is particularly important in case of failure from the system, in which the human operator has to take back manually the control of a task. Related to this is the evidence of return-to-manual decrement after system failure. For example, Endsley & Kiris [55] found that response time decision in a navigation task increases when participants had unexpectedly to respond manually after a period of automation assistance. Similar results were found by Manzey et al. [27] with highest return-to-manual decrement for higher level of automation (see also [75, 76]).

To conclude on this section, we would like to discuss another aspect of human – automation interaction that have shown growing interest in the recent years, that is, the effect of automation on human agency [77]. Human agency (or sense of agency) refers to the individual experience of controlling one's own actions and, through these actions, outcomes in the external environment [78]. Recently, scientists have been interested in how the interaction with automation influences how people feel in control in their own actions. One of the first demonstrations of an effect of automation on sense of agency came from a study by Berberian et al. [26]. In an aircraft supervision task, the authors

 doi:10.20944/preprints202203.0070.v1

found a decrease in agency with the introduction of automation, with lower level of agency for higher level of automation. Similar negative influence of automation and agency were found for example by Coyle et al. [79] or by Zanatto et al. [80]. This finding is relevant in the context of our review because agency is supposed to play a role in the attribution of responsibility and in the motivation of goal-directed behaviors [81, 82]. In the social domain for example, Caspar and colleagues [83-85] found that a decrease in participants' sense of agency was correlated with anti-social behaviors increment from human agents. Thus, the evidence that automation can reduce the sense of agency from human users lead to believe of potentially misuses of automation, in addition to the ones described above, particularly in moral or sensitive domains. Consider again the example of a combat drone operator engaged on a battlefield, exposed to the risk of civilian losses and material damages during attacks. Here, a sense of agency decrement – combined with omission or commission errors – from the human operator might have dramatic consequences in terms of human life. It is clear from that situation that the negative effects of automation are not just annoying "side-effects" without real importance. If we want to avoid such dramatic incidents, we need to understand how the interaction with automation might change our behaviors and our decisions when we have to face moral situations.

2.3. *The ugly: Automated systems and moral decision-making*

For the last two decades, the main focus of engineers, scientists, and philosophers regarding moral decision-making and automation concerned the rules to assign to an autonomous system to perform ethical responses, or the ethical and legal issues regarding the use of autonomous systems. Research have been conducted on what is the best rule/algorithm to assign to an automation in moral situations [32, 35] and what human subjects would do in moral decision-making situations in order to inspire the development of ethical automation [33, 34. In addition, scientists and philosophers have thought about the legal and ethical consequences of fully autonomous machines in critical situations like in driving or military conflicts [86-88]. Surprisingly, the understanding of how the interaction with automation can change ethical and social behaviors from a human agent in moral decision-making situations has received little research attention until very recently [36]. By ethical and social behaviors, we mean behaviors that follow some consensus on the way to behave or not within a social group. Moral decision-making refers to a decision or a judgment made in a situation with moral rules and moral principles involved [89, 90]. The recent interest on that matter can be explained by the increased proportion of behaviors in our everyday life that are guided by automation supports (driving, communication, health, etc.). In addition, as we have discussed above, automation is now more and more involved in sensitive domains such as military operations, medicine, and security. Then, understanding the effect of automation in that context is crucial.

Available evidences suggest a mixed-picture of the impact of automation in social and moral decision-making situations. On the one hand, some recent findings suggest for instance that the interaction with automation in social dilemmas can increase fairness between human agents [91] or can promote human cooperation [92, 93]. Similarly, Kirchkamp & Strobel [94] did not find significant evidence of more selfish behaviors in a social game scenario when decision-making is shared with automation. Thus, these results suggest that the interaction with automation in the social and the moral domain does not necessarily increase the rate of unethical or unsocial behaviors, and at the opposite might have a positive effect by increasing prosocial behaviors. On the other hand, an analysis of additional results shows that the effect of automation is not so clear and using automation in social or moral context could have clear detrimental effects. For example, recent investigations suggest that people tend to act more selfishly when they are in

 doi:10.20944/preprints202203.0070.v1

interaction against a computer player [95] and are more prompt for cheating [96]. Manistersky et al. [97] reported that, in a resource allocation game, participants that played the game through self-designed autonomous agents, designed automations for improving self-performance and less for cooperation, contrasting with the results found by [91]. Very recently, Leib et al. [37] found that advice received from an automation is as strong as the effect of human agents to promote unethical behaviors during social interactions.

Finally, some of the results on automation bias cited above can also shed light on the influence of automation on human decision processes, particularly when these biases occur in medical or military situations. We can consider for example the studies by Alberti et al. [71] or Goddard et al. [72], in which automation bias was reported for mammography assessment and in prescription task. Although the conflicting moral aspect of the decisions made is not evident in these studies, these scenarios had obvious health and life consequences. An omission error during mammography assessment for instance can result in undiagnosed cancer for a patient. Thus, it seems that despite potential critical negative outcomes, people can nevertheless follow the bad recommendations of automation. Similarly, in the military domain we can think about the command-and-control study by Rovira et al. [21], in which the authors reported high rate of incorrect decisions made by the participants when the automated systems decisions were not reliable. The decisions in the Rovira et al.'s study involved the engagement of opponents with friendly units on a simulated battlefield. Does it mean that commanders making decisions with the help of decision automation systems could show automation bias? And what about combat drone operators when he/she faces the risk of civilian losses while he/she is engaging a military target?

In conclusion, available empirical evidence suggests that the interaction with automation might result in the promotion of pro-social behaviors (fairness, cooperation, etc.), but in unethical and aberrant behaviors as well. Thus, while automation in the form of advisors or decision support system seems to be a very interesting venue to develop and favor positive interaction among individuals or groups of human agents, it seems also that in certain circumstances the interaction with automation can be detrimental in terms of ethical decision making. This dual-aspect of automation is also true for situations that do not involve moral and social decision-making at first sight, with automated systems favoring at the same time global performance improvement and multi-tasking, but automation biases and loss of sense of agency as well. Following the research agenda of several authors (e.g., [25]), and considering that automation will certainly be more and more present in our daily life and sometimes, in critical circumstances, we need more than ever to understand the factors and situations that favor both the positive and the negative effects of automation. This will help developing systems useful, safe, and ethical for users and society. The next part of our paper will be dedicated to a review of the current known-factors and models of human – automation interaction.

3. Factors and models of human – automation interaction

3.1. *Determinants of the effect of automation on human decision-making and performance*

Understanding the determining factors in the effect of automation on human performance and decision-making has been for longtime a goal for researchers and engineers [2, 25, 57]. Since the first investigations conducted in the 80s, multiple factors have been identified as crucial determinants in human – automation interaction. In this section, we will review some of the factors identified as the most important ones and describe their effects on the human agent interacting with the automated system. Again, identifying these factors and understanding exactly what their effects are is particularly relevant because this will help to understand the circumstances in which both the posi-

 doi:10.20944/preprints202203.0070.v1

tive and the negative effects described above are observed. With that information, scientists and engineers will be able to develop new forms of automation, efficient in terms of the positive changes they produce on the users' performance and safety (e.g., fast and correct decisions, possibility for multi-tasking, etc.), but preventing or at least mitigating the negative outcomes we described, particularly in the context of social and moral situations.

3.1.1. Level and stage of automation

The first factor that determine how human agents will interact with automated systems is obviously the level and the stage at which automation occurs. As a reminder, levels of automation refers to the notion that the degree of automation on a specific task can vary across a continuous scale, with intermediate levels representing different degrees at which automation is assigned [38, 39], while the stages of automation refers to the different subtasks on which automation can be assigned [38, 40]. Most of the information on this topic was already presented in the previous sections. Importantly, an increased level of automation (with a human agent still being at the command) seems to be associated with increased performance and improved decisions by the human user. For example, Manzey et al. [27] reported that performance in a supervisory control task is better with automation than with manual control, this positive effect being superior when participants were supported by the highest level of automation [see also 21, 75]. In addition, more automation in one task seems to facilitate multi-tasking by the human agent. For instance, Chen and Joyner [61] reported that performance on a primary task increases when action on a secondary task is supported by a high level compared to a low level of automation (see also [27]). Although generally positive in terms of decision-making and performance, higher level of automation can also result in loss of skills [55] and increased delay to take-back control of the system in case of system failure. In Manzey et al. [27] study for example, the cost of return to manual was higher for the higher level of automation (see also [55, 75, 76]). Finally, we have seen that recent results found a loss of subject's sense of agency with higher level of automation, measured either by a direct rating about the task or by the indirect temporal binding measure [26]. Thus, more investigations seem necessary to understand the exact balance between the beneficial effects and the disadvantages of higher level of automation, and how the interaction with other factors (e.g., level of skills and previous experience, accountability) could mitigate or increase these effects. Concerning the stage of automation, whether it is information and analysis automation or decision and action automation, all of these forms of automated system can participate in the improvement of performance and decisions [40]. However, in comparison with information/analysis automation, decision automation seems to be more subject to automation bias [21, 49, 98]. This result is not surprising considering that in decision automation, it is not necessary for the user to look at the environment, but only at the input from the system, while in information automation for instance, the user evaluates a preprocessed environment (but still look at the environment).

3.1.2. Automation reliability

A second major factor is automation reliability. The effect of automation reliability is certainly one of the most extensively studied factors in human – automation interaction. Its effect has been tested both for information/analysis automation [e.g., 16, 45] and for decision/action automation [e.g., 21]. Overall, investigations conducted on this effect found that decisions and performance increase with automation reliability. For instance, Goh et al. [17] found that the performance of participants was higher for information automation cue with 90% reliability than 70%. Interestingly, a positive effect of automation might be obtained even with relatively low level of reliability. For example, Cullen et al. [51] found that the performance of subjects helped with information automation in

 doi:10.20944/preprints202203.0070.v1

a multi-task environment increased compared to a baseline condition without automation, even with an automation reliability of 67% (see also [21]). Although resulting in global performance and decision increment, multiple studies have reported that errors rate during unreliable trials increases with a higher level of reliability of the system. Thus, high automation reliability seems to be associated with increased tendency for automation bias, with both omission and commission errors [25]. In a study by Oakley et al. [99], the authors found that the rate at which subjects detect automation failures decreased with automation reliability in detecting errors (see also [67]). Thus, with higher automation reliability seems to increase the global task performance, but also the risk of automation biases. As a consequence, much automation reliability would be relevant only as long as the negative outcomes that result from automation biases are not superior compared to the gain obtained from the increased reliability. Like with the effect of higher-level of automation, the balance between the advantages and the disadvantages of higher automation reliability will have to be investigated systematically.

3.1.3. Task difficulty

A third important factor that determines how a human agent will interact with automation is task difficulty and/or the number of tasks simultaneously monitored by the agent. As we have already seen, automation is particularly useful in the context of difficult manual or cognitive tasks. It allows the access to more rewards through the increased correct performance and decisions rate [19, 45, 49, 100], and reduce the cost related to task execution (measured for example with a reduced subjective workload; e.g., [19]). In addition, automation facilitates multi-tasking by automatizing manual tasks (action automation) or helping the detection of targets in the environment [e.g., 51, 61], allowing the human user to allocate more time and attention to a secondary task. At the same time, using automation in the context of a difficult single-task and/or multi-tasking is frequently considered as a driving factor of automation bias [25, 70, 72]. However, results are not always consistent [48, 100]. Related to the notion of task difficulty, studies found that time pressure (i.e., a short delay allowed for a subject to complete a task) can increase the rate of errors made by a user [e.g., 44]. Finally, as a counteract to the effect of task difficulty and multi-tasking, individual experience and task mastery seems to reduce the probability of automation bias in the context of difficult task or multi-tasking, maybe because of the experience of incorrect information [101], but with novice users having generally more benefits from the use of automation [e.g., 102, 103].

3.1.4. Performance outcome, accountability, and automation display

Additional factors can be cited as determining variables in the effect of automation. For example, some experiments have shown that the outcome of performance, and particularly the consequences of errors, can have an effect on the rate of automation bias. Particularly, tasks for which errors might have more important negative consequences seem to be more carefully assessed by users [e.g., 28]. Related to this is the evidence of an effect of accountability of decisions on the rate of automation bias. Skitka et al. [69], for example, found a lower level of commission and omission errors when participants were accountable of performance and accuracy (see also [104-106]). Thus, whether it is socially-mediated or not, it seems that the consequences of the performance of user in interaction with automation has a strong effect on that performance. Finally, we can cite the effect of the physical properties of the system and the way the output of automation is displayed. For example, Goh et al. [17] found that information automation in a screening-task is improved when the automation cue is centered on the target compared to an indirect cue (see also [16]). Still in a screening-task, Rieger et al. [44] found that a target presented in a predictable location improves speed and accuracy.

 doi:10.20944/preprints202203.0070.v1

As a conclusion, the way a human agent interacts with automation is influenced by multiples factors. Although the present section is interesting to get an overview of these effects, this empirical listing suffers from being a sort of "catalogue" of human – automation interaction effects without real conceptual or predictive value. Consequently, the next and final section of this review will be dedicated to a presentation of models that scientists have developed over the last few decades to explain and make prediction about the effects reported, and to a discussion on potential way of improvements for future models.

3.2. Models of human-automation interaction

With the increasing number of studies conducted on the factors underlying human – automation interaction, multiples models of human decision-making and performance in interaction with automation have been developed. This development has resulted in the introduction of new concepts (and their associated models) like trust in automation [107, 108], reliance and compliance [45], or mental workload [109]. A complete review of all the concepts and models is beyond the scope of this paper. Instead, we will focus our discussion on the presentation of two widely used notions in the human – automation domain: the loss of situation awareness [110-112] and automation complacency [66, 67]. We will present the phenomena and effects encompassed by these notions and how they are explained. More generally, we will see how these concepts and their associated models improve our understanding of the influence of automation on human decision-making and performance, but also what their limits are. Particularly, we will discuss the necessity of a more systematic use of quantitative computational methods, inspired notably from what it is already done in computational cognitive sciences [e.g., 113], in order to increase their precision and their explanatory power.

The first concept and associated model we will present is (loss of) situation awareness [110, 111]. Used initially in the aircraft domain, situation awareness has become certainly one of the most important concepts used in human – automation interaction [114]. According to Endsley [115], situation awareness is defined 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". Situation awareness is thus composed of three different hierarchical level: The first level of situation awareness (the perception phase) corresponds to the perception of the status, attributes, and dynamics of relevant information in the environment. The second level of situation awareness (the comprehension phase) corresponds to the comprehension of the situation based on an analysis and synthesis of elements collected from the perception level. Finally, in the third level, the situation awareness comes from the ability to project future developments of the situation and potential consequences of actions to influence it [110]. Based on these representations, the human agent can decide the best action to select.

Considering the assumed-role of situation awareness in the decision-making process, we can see how inappropriate automation conditions could result in the decrement of situation awareness (i.e., a loss of situation awareness), and then the increased probability of incorrect decisions from the human agent. For example, a high level of automation combined with an absence of sufficient training or a loss of skills from the human agent could result, in case of automation failure, in the inability of the agent to detect, comprehend, and/or react appropriately to the failure. Recently, Endsley [112] proposed an integrated model of situation awareness in which determining factors are notably defined. Particularly, the author proposes that a loss of situation awareness during the interaction with automation could result, among others things, from a low level of information presentation, low monitoring skills, excessive trust in automation, the presence of competing tasks, or a low level of cognitive engagement from the agent. All these fac-

 doi:10.20944/preprints202203.0070.v1

tors are supposed to intervene in agent's level of situation awareness and thus in its performance in a related task. This integration of factors allowed the author to propose different guidelines for the design of appropriate automated systems in interaction with human agents. For example, it is suggested to use automation for routine tasks preferably, information and analysis automation rather than decision automation, or to increase information saliency and provide automation transparency (see [112] for complete guidelines).

Closely related to situation awareness is another important concept in the human – the automation interaction domain named automation complacency [66, 67]. Automation complacency is said to occur when a human agent is monitoring an automated system, but with a suboptimal rate of monitoring which, in turn, might lead to performance failures [66, 116]. This performance failure results more exactly from both a direct automation failure and from the inappropriate response of the human agent. Automation complacency was initially developed in the context of multi-tasking with the evidence that the rate of automation failure detection is relatively low when subjects have to monitor an automated secondary task with high and constant reliability level [67]. Now the concept is associated with the phenomenon of automation bias [25]. Parasuraman and Manzey [66] proposed a model integrating both complacency and automation bias. The model is composed of a complacency potential component, which influence an attentional information processing component, which in turn influence the agent's situation awareness and performance. Interestingly, the structure of the model is relatively close to the recent one proposed by Endsley [112]. In the latter, the complacency potential, which is seen as a tendency to react in a less attentive manner during the interaction with a specific automated system, is assumed to be influenced by the system reliability and consistency as well as by individual characteristics and interaction history with the system. Thus, increased automation reliability is assumed to increase agent's complacency potential. Then, it is assumed by the authors that this complacency potential will influence (negatively) attentional information processing by producing inappropriate allocation of attention and/or selective information processing in the context, for example, of high task load. This low attentional information processing would, in turn, result in loss of situation awareness and inappropriate decision-making, like in the absence of detection of failure from the system.

In summary, both Endsley's [110, 112] situation awareness models and Parasuraman & Manzey's [66] complacency model have been shown interesting to interpret some of the effects found when automation is introduced and to integrate empirical demonstrations of known underlying factors of human – automation interaction (e.g., level of automation, automation reliability). Based on their interpretations about the role of specific underlying factors, these models allow to make recommendations to engineers for the design of new automation technologies, with the purpose of avoiding the negative outcomes reviewed in the previous sections [112]. Endsley's situation awareness model and Parasuraman & Manzey's complacency model offer a conceptual interpretation to several negative effects like omission and commission errors or return-to-manual decrement after system failure. Considering the potential role of these effects in social or ethical decision-making situations (e.g., for the decisions made by a combat drone operator), the understanding that these models allow is particularly interesting for the development and the use of safe automated systems in these situations. Despite all these advantages, however, the models are not without limitation. Particularly, these models (at least in the references above) are defined only at a conceptual or descriptive level, and the exact mechanisms that underlie each function or how those functions interact with each other or with external factors is not precisely described. Thus, it is hard to know what exact predictions they allow in specific circumstances, and consequently, it is difficult to test these predictions and make comparisons between models. More generally,

 doi:10.20944/preprints202203.0070.v1

this reduces the extent to which engineers can use the models to anticipate how human users would act in particular circumstances.

To improve the power of those models, the use of quantitative computational methods is very promising. Computational modeling (in our context, of behavioral data) consists in the use of mathematical models either to explain qualitative features of empirical data or to make quantitative predictions [117]. In the last 20 years, behavioral and cognitive neuroscientists have shown a strong interest for the use of computational models in behavioral and neurocognitive research [e.g., 113, 118]. The models show several advantages over classic conceptual models. Notably, these models are explicit and falsifiable and their performance can be quantitatively assessed, and they can provide unified framework for supposedly distinct phenomena [113, 119]. Applications of computational modeling to learning and decision-making phenomena have shown several important successes [e.g., 120] and we strongly believe that this could be beneficial as well for the understanding of human decision-making and performance in the human – automation interaction domain. This suggestion has already been proposed in the literature [e.g., 121] and computational models of human decision-making during human – automation interaction have been designed [e.g., 122-125]. To this day, the proportion of models using computational methods remains relatively low, however, and much more investigation will be necessary for the systematic use of these models.

4. Discussion

Automation is now widespread in almost every domain of human activity and it is more and more present in our everyday life [1-3]. In this review, we saw that the introduction of automated systems can result in important improvement in human decision-making processes and performance, but can result in serious negative effects like automation bias as well [2]. Understanding the conditions in which these effects appear has become crucial, particularly when automation is used in situations involving social or moral decisions [36]. For these situations, this review has shown that available evidence is inconsistent. Since the first investigations conducted on human – automation interaction, multiple factors have been identified, and their effects begin to be well understood (at least some of them). Models have been proposed to explain the change in decision-making and performance found in human – automation investigations. Concepts like loss of situation awareness [110, 112] and automation complacency [66] have been introduced and the models allow to make some recommendations about the design of new automated technologies. However, these models suffer from being mainly conceptual or descriptive models, limiting their predictive value. We suggest the use of computational modeling in order to increase the models' precision in predictions [118, 121].

To conclude on this review, we would like to highlight potential fruitful directions for future research in the human – automation domain. Firstly, new investigations must be conducted on how the introduction of automated systems influences the decisions and performance of human agents in social or moral decision-making situations [36]. As we have seen, very few studies have been conducted yet, and they point out both positive and negative effects. Considering the growing importance of automation in sensitive domain like medicine, defense, and security, we need to understand the conditions that favor ethical decisions by the human interacting with automation. Secondly, more investigations are necessary on the factors underlying the interaction between human and automation. Particularly, while many investigations have been conducted on contextual effects (e.g., effect of automation reliability, level and stage of automation, etc.), much less is known about the effect of task consequences and/or accountability (i.e., the effect of contingent-outcomes presented during the interaction with automation [25]). Finally, an important area of research will consist in the more massive development of

 doi:10.20944/preprints202203.0070.v1

computational models to explain more exactly how automation changes decision-making processes and performance in human subjects [118, 121]. By doing so, engineers will be able to develop new automation technologies designed for the improvement of human performance, but reducing at the same time the risk of dramatic consequences.

Author Contributions: Writing—review and editing, AP, SLB; funding acquisition, SLB. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Belgian Defense – Royal Higher Institute of Defense, grant number HFM20-03.

Data Availability Statement: Not applicable.

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. 1. Janssen CP, Donker SF, Brumby DP, Kun AL. History and future of human-automation interaction. International Journal of Human-Computer Studies. 2019 Nov 1;131:99-107.
2. 2. Parasuraman R, Riley V. Humans and automation: Use, misuse, disuse, abuse. Human factors. 1997 Jun;39(2):230-53.
3. 3. Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakr-ishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A., Shah, J., Tambe, M., and Teller, A. (2016). Artificial Intelligence and Life in 2030 - One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel. Stanford University, Stanford, CA.
4. 4. Shakhatreh H, Sawalmeh AH, Al-Fuqaha A, Dou Z, Almaita E, Khalil I, Othman NS, Khreishah A, Guizani M. Unmanned aerial vehicles (UAVs): A survey on civil applications and key research challenges. Ieee Access. 2019 Apr 9;7:48572-634.
5. 5. Zhang T, Li Q, Zhang CS, Liang HW, Li P, Wang TM, Li S, Zhu YL, Wu C. Current trends in the development of intelligent unmanned autonomous systems. Frontiers of information technology & electronic engineering. 2017 Jan;18(1):68-85.
6. 6. Ayoub J, Zhou F, Bao S, Yang XJ. From manual driving to automated driving: A review of 10 years of autoui. InProceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications 2019 Sep 21 (pp. 70-90).
7. 7. Chan CY. Advancements, prospects, and impacts of automated driving systems. International journal of transportation science and technology. 2017 Sep 1;6(3):208-16.
8. 8. Anderson E, Fannin T, Nelson B. Levels of aviation autonomy. In2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC) 2018 Sep 23 (pp. 1-8). IEEE.
9. 9. Chialastri A. Automation in aviation. IntechOpen; 2012 Jul 25.
10. 10. Valdés RA, Comendador VF, Sanz AR, Castán JP. Aviation 4.0: more safety through automation and digitization. InAircraft Technology 2018 Mar 9. IntechOpen.
11. 11. Mayer M. The new killer drones: Understanding the strategic implications of next-generation unmanned combat aerial vehicles. International Affairs. 2015 Jul 1;91(4):765-80.
12. 12. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. Bmj. 2005 Mar 31;330(7494):765.
13. 13. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ digital medicine. 2020 Feb 6;3(1):1-0.
14. 14. Hanna B, Son TC, Dinh N. AI-guided reasoning-based operator support system for the nuclear power plant management. Annals of Nuclear Energy. 2021 May 1;154:108079.
15. 15. Lin L, Athe P, Rouxelin P, Avramova M, Gupta A, Youngblood R, Lane J, Dinh N. Development and assessment of a nearly autonomous management and control system for advanced reactors. Annals of Nuclear Energy. 2021 Jan 1;150:107861.
16. 16. Chavaillaz A, Schwaninger A, Michel S, Sauer J. Automation in visual inspection tasks: X-ray luggage screening supported by a system of direct, indirect or adaptable cueing with low and high system reliability. Ergonomics. 2018 Oct;61(10):1395-1408.
17. 17. Goh J, Wiegmann DA, Madhavan P. Effects of automation failure in a luggage screening task: a comparison between direct and indirect cueing. InProceedings of the Human Factors and Ergonomics Society Annual Meeting 2005 Sep (Vol. 49, No. 3, pp. 492-496). Sage CA: Los Angeles, CA: SAGE Publications.
18. 18. Ma R, Kaber DB. Effects of in-vehicle navigation assistance and performance on driver trust and vehicle control. International Journal of Industrial Ergonomics. 2007 Aug 1;37(8):665-73.
19. 19. MacMillan J, Deutsch SE, Young MJ. A comparison of alternatives for automated decision support in a multi-task environment. InProceedings of the Human Factors and Ergonomics Society Annual Meeting 1997 Oct (Vol. 41, No. 1, pp. 190-194). Sage CA: Los Angeles, CA: SAGE Publications.
20. 20. 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 Dec;17(4):320.
21. 21. Rovira E, McGarry K, Parasuraman R. Effects of imperfect automation on decision making in a simulated command and control task. Human factors. 2007 Feb;49(1):76-87.
22. 22. Wiseman Y. Autonomous vehicles. InResearch Anthology on Cross-Disciplinary Designs and Applications of Automation 2022 (pp. 878-889). IGI Global.
23. 23. Tvaryanas AP, Thompson WT, Constable SH. The US military unmanned aerial vehicle (UAV) experience: Evidence-based human systems integration lessons learned. NATO Research and Technology Organisation. Neuilly-sur-Seine, France. 2005.
24. 24. Williams KW. A summary of unmanned aircraft accident/incident data: Human factors implications. Federal Aviation Administration Oklahoma City OK Civil Aeromedical Inst; 2004 Dec 1.
25. 25. Mosier KL, Manzey D. Humans and automated decision aids: A match made in heaven?. InHuman performance in automated and autonomous systems 2019 Sep 19 (pp. 19-42). CRC Press.
26. 26. Berberian B, Sarrazin JC, Le Blaye P, Haggard P. Automation technology and sense of control: a window on human agency. PloS one. 2012 Mar 30;7(3):e34075.

27. 27. Manzey D, Reichenbach J, Onnasch L. Human performance consequences of automated decision aids: The impact of degree of automation and system experience. Journal of Cognitive Engineering and Decision Making. 2012 Mar;6(1):57-87.
28. 28. Mosier KL, Skitka LJ, Heers S, Burdick M. Automation bias: Decision making and performance in high-tech cockpits. The International journal of aviation psychology. 1998 Jan 1;8(1):47-63.
29. 29. Christensen JF, Gomila A. Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review. Neuroscience & Biobehavioral Reviews. 2012 Apr 1;36(4):1249-64.
30. 30. Cushman F. Action, outcome, and value: A dual-system framework for morality. Personality and social psychology review. 2013 Aug;17(3):273-92.
31. 31. Cushman F, Kumar V, Railton P. Moral learning: Psychological and philosophical perspectives. Cognition. 2017 Oct 1;167:1-0.
32. 32. Arkin RC, Ulam P, Wagner AR. Moral decision making in autonomous systems: Enforcement, moral emotions, dignity, trust, and deception. Proceedings of the IEEE. 2011 Dec 9;100(3):571-89.
33. 33. Awad E, Dsouza S, Kim R, Schulz J, Henrich J, Shariff A, Bonnefon JF, Rahwan I. The moral machine experiment. Nature. 2018 Nov;563(7729):59-64.
34. 34. Bonnefon JF, Shariff A, Rahwan I. The social dilemma of autonomous vehicles. Science. 2016 Jun 24;352(6293):1573-6.
35. 35. Jiang L, Hwang JD, Bhagavatula C, Bras RL, Forbes M, Borchardt J, Liang J, Etzioni O, Sap M, Choi Y. Delphi: Towards machine ethics and norms. arXiv preprint arXiv:2110.07574. 2021 Oct 14.
36. 36. Köbis N, Bonnefon JF, Rahwan I. Bad machines corrupt good morals. Nature Human Behaviour. 2021 Jun;5(6):679-85.
37. 37. Leib M, Köbis NC, Rilke RM, Hagens M, Irlenbusch B. The corruptive force of AI-generated advice. arXiv preprint arXiv:2102.07536. 2021 Feb 15.
38. 38. 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 May;30(3):286-97.
39. 39. Vagia M, Transeth AA, Fjerdingen SA. A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed?. Applied ergonomics. 2016 Mar 1;53:190-202.
40. 40. Parasuraman R, Wickens CD. Humans: Still vital after all these years of automation. Human factors. 2008 Jun;50(3):511-20.
41. 41. Mosier KL, Fischer UM. Judgment and decision making by individuals and teams: issues, models, and applications. Reviews of human factors and ergonomics. 2010 May;6(1):198-256.
42. 42. Yeh M, Wickens CD. Display signaling in augmented reality: Effects of cue reliability and image realism on attention allocation and trust calibration. Human Factors. 2001 Sep;43(3):355-65.
43. 43. Yeh M, Wickens CD, Seagull FJ. Target cuing in visual search: The effects of conformality and display location on the allocation of visual attention. Human Factors. 1999 Dec;41(4):524-42.
44. 44. Rieger T, Heilmann L, Manzey D. Visual search behavior and performance in luggage screening: effects of time pressure, automation aid, and target expectancy. Cognitive Research: Principles and Implications. 2021 Dec;6(1):1-2.
45. 45. Dixon SR, Wickens CD. Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. Human factors. 2006 Sep;48(3):474-86.
46. 46. St. John M, Smallman HS, Manes DI, Feher BA, Morrison JG. Heuristic automation for decluttering tactical displays. Human Factors. 2005 Sep;47(3):509-25.
47. 47. Martinez-Franco AI, Sanchez-Mendiola M, Mazon-Ramirez JJ, Hernandez-Torres I, Rivero-Lopez C, Spicer T, Martinez-Gonzalez A. Diagnostic accuracy in Family Medicine residents using a clinical decision support system (DXplain): a randomized-controlled trial. Diagnosis. 2018 Jun 1;5(2):71-6.
48. 48. Lyell D, Magrabi F, Raban MZ, Pont LG, Baysari MT, Day RO, Coiera E. Automation bias in electronic prescribing. BMC medical informatics and decision making. 2017 Dec;17(1):1-0.
49. 49. Sarter NB, Schroeder B. Supporting decision making and action selection under time pressure and uncertainty: The case of in-flight icing. Human factors. 2001 Dec;43(4):573-83.
50. 50. Chen JY, Barnes MJ. Supervisory control of multiple robots in dynamic tasking environments. Ergonomics. 2012 Sep 1;55(9):1043-58.
51. 51. Cullen RH, Rogers WA, Fisk AD. Human performance in a multiple-task environment: Effects of automation reliability on visual attention allocation. Applied ergonomics. 2013 Nov 1;44(6):962-8.
52. 52. Cummings ML, Guerlain S. Developing operator capacity estimates for supervisory control of autonomous vehicles. Human factors. 2007 Feb;49(1):1-5.
53. 53. Wright JL, Chen JY, Barnes MJ. Human–automation interaction for multiple robot control: the effect of varying automation assistance and individual differences on operator performance. Ergonomics. 2018 Aug 3;61(8):1033-45.
54. 54. Bainbridge L. Ironies of automation. InAnalysis, design and evaluation of man–machine systems 1983 Jan 1 (pp. 129-135). Pergamon.
55. 55. Endsley MR, Kiris EO. The out-of-the-loop performance problem and level of control in automation. Human factors. 1995 Jun;37(2):381-94.
56. 56. Wiener EL. Cockpit automation. InHuman factors in aviation 1988 Jan 1 (pp. 433-461). Academic Press.
57. 57. Wiener EL, Curry RE. Flight-deck automation: Promises and problems. Ergonomics. 1980 Nov 1;23(10):995-1011.
58. 58. Sarter NB, Woods DD, Billings CE. Automation surprises. Handbook of human factors and ergonomics. 1997;2:1926-43.

59. 59. Sheridan TB. Function allocation: algorithm, alchemy or apostasy?. International Journal of Human-Computer Studies. 2000 Feb 1;52(2):203-16.
60. 60. Prewett MS, Johnson RC, Saboe KN, Elliott LR, Coovert MD. Managing workload in human–robot interaction: A review of empirical studies. Computers in Human Behavior. 2010 Sep 1;26(5):840-56.
61. 61. Chen JY, Joyner CT. Concurrent performance of gunner's and robotics operator's tasks in a multitasking environment. Military Psychology. 2009 Jan 28.
62. 62. Wang H, Lewis M, Velagapudi P, Scerri P, Sycara K. How search and its subtasks scale in N robots. InProceedings of the 4th ACM/IEEE international conference on Human robot interaction 2009 Mar 9 (pp. 141-148).
63. 63. Adams JA. Multiple robot/single human interaction: Effects on perceived workload. Behaviour & Information Technology. 2009 Mar 1;28(2):183-98.
64. 64. Velagapudi P, Scerri P, Sycara K, Wang H, Lewis M, Wang J. Scaling effects in multi-robot control. In2008 IEEE/RSJ International Conference on Intelligent Robots and Systems 2008 Sep 22 (pp. 2121-2126). IEEE.
65. 65. Mosier KL, Skitka LJ. Human decision makers and automated decision aids: Made for each other?. InAutomation and human performance: Theory and applications 2018 Jan 29 (pp. 201-220). CRC Press.
66. 66. Parasuraman R, Manzey DH. Complacency and bias in human use of automation: An attentional integration. Human factors. 2010 Jun;52(3):381-410.
67. 67. Parasuraman R, Molloy R, Singh IL. Performance consequences of automation-induced'complacency'. The International Journal of Aviation Psychology. 1993 Jan 1;3(1):1-23.
68. 68. Skitka LJ, Mosier KL, Burdick M. Does automation bias decision-making?. International Journal of Human-Computer Studies. 1999 Nov 1;51(5):991-1006.
69. 69. Skitka LJ, Mosier K, Burdick MD. Accountability and automation bias. International Journal of Human-Computer Studies. 2000 Apr 1;52(4):701-17.
70. 70. Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. Journal of the American Medical Informatics Association. 2017 Mar 1;24(2):423-31.
71. 71. Alberdi E, Povyakalo A, Strigini L, Ayton P. Effects of incorrect computer-aided detection (CAD) output on human decision-making in mammography. Academic radiology. 2004 Aug 1;11(8):909-18.
72. 72. Goddard K, Roudsari A, Wyatt JC. Automation bias: empirical results assessing influencing factors. International journal of medical informatics. 2014 May 1;83(5):368-75.
73. 73. Haslbeck A, Hoermann HJ. Flying the needles: flight deck automation erodes fine-motor flying skills among airline pilots. Human factors. 2016 Jun;58(4):533-45.
74. 74. Volz KM, Dorneich MC. Evaluation of cognitive skill degradation in flight planning. Journal of Cognitive Engineering and Decision Making. 2020 Dec;14(4):263-87.
75. 75. Endsley MR, Kaber DB. Level of automation effects on performance, situation awareness and workload in a dynamic control task. Ergonomics. 1999 Mar 1;42(3):462-92.
76. 76. Tatasciore M, Bowden VK, Visser TA, Loft S. Should We Just Let the Machines Do It? The Benefit and Cost of Action Recommendation and Action Implementation Automation. Human Factors. 2021 Feb 8:0018720821989148.
77. 77. Berberian B. Man-Machine teaming: a problem of Agency. IFAC-PapersOnLine. 2019 Jan 1;51(34):118-23.
78. 78. Haggard P, Chambon V. Sense of agency. Current biology. 2012 May 22;22(10):R390-2.
79. 79. Coyle D, Moore J, Kristensson PO, Fletcher P, Blackwell A. I did that! Measuring users' experience of agency in their own actions. InProceedings of the SIGCHI conference on human factors in computing systems 2012 May 5 (pp. 2025-2034).
80. 80. Zanatto D, Chattington M, Noyes J. Human-machine sense of agency. International Journal of Human-Computer Studies. 2021 Dec 1;156:102716.
81. 81. Di Costa S, Théro H, Chambon V, Haggard P. Try and try again: Post-error boost of an implicit measure of agency. Quarterly Journal of Experimental Psychology. 2018 Jul;71(7):1584-95.
82. 82. Haggard P. Sense of agency in the human brain. Nature Reviews Neuroscience. 2017 Apr;18(4):196-207.
83. 83. Caspar EA, Christensen JF, Cleeremans A, Haggard P. Coercion changes the sense of agency in the human brain. Current biology. 2016 Mar 7;26(5):585-92.
84. 84. Caspar EA, Cleeremans A, Haggard P. Only giving orders? An experimental study of the sense of agency when giving or receiving commands. PloS one. 2018 Sep 26;13(9):e0204027.
85. 85. Caspar EA, Lo Bue S, Magalhães De Saldanha da Gama PA, Haggard P, Cleeremans A. The effect of military training on the sense of agency and outcome processing. Nature communications. 2020 Aug 31;11(1):1-0.
86. 86. Beard JM. Autonomous weapons and human responsibilities. Geo. J. Int'l L.. 2013;45:617.
87. 87. Gregory D. From a view to a kill: Drones and late modern war. Theory, culture & society. 2011 Dec;28(7-8):188-215.
88. 88. Harris J. Who owns my autonomous vehicle? Ethics and responsibility in artificial and human intelligence. Cambridge Quarterly of Healthcare Ethics. 2018 Oct;27(4):599-609.
89. 89. Garrigan B, Adlam AL, Langdon PE. Moral decision-making and moral development: Toward an integrative framework. Developmental review. 2018 Sep 1;49:80-100.
90. 90. Moll J, Zahn R, de Oliveira-Souza R, Krueger F, Grafman J. The neural basis of human moral cognition. Nature reviews neuroscience. 2005 Oct;6(10):799-809.
91. 91. de Melo CM, Marsella S, Gratch J. Social decisions and fairness change when people's interests are represented by autonomous agents. Autonomous Agents and Multi-Agent Systems. 2018 Jan;32(1):163-87.

92. 92. de Melo CM, Marsella S, Gratch J. Human cooperation when acting through autonomous machines. Proceedings of the National Academy of Sciences. 2019 Feb 26;116(9):3482-7.
93. 93. Fernández Domingos E, Terrucha I, Suchon R, Grujić J, Burguillo JC, Santos FC, Lenaerts T. Delegation to autonomous agents promotes cooperation in collective-risk dilemmas. arXiv e-prints. 2021 Mar:arXiv-2103.
94. 94. Kirchkamp O, Strobel C. Sharing responsibility with a machine. Journal of Behavioral and Experimental Economics. 2019 Jun 1;80:25-33.
95. 95. March C. The behavioral economics of artificial intelligence: Lessons from experiments with computer players. CESifo Working Paper No. 7926. 2019.
96. 96. Cohn A, Gesche T, Maréchal MA. Honesty in the digital age. Management Science. 2021 Nov 8.
97. 97. Manistersky E, Lin R, Kraus S. The development of the strategic behavior of peer designed agents. InLanguage, Culture, Computation. Computing-Theory and Technology 2014 (pp. 180-196). Springer, Berlin, Heidelberg.
98. 98. Crocoll WM, Coury BG. Status or recommendation: Selecting the type of information for decision aiding. InProceedings of the human factors society annual meeting 1990 Oct (Vol. 34, No. 19, pp. 1524-1528). Sage CA: Los Angeles, CA: SAGE Publications.
99. 99. Oakley B, Mouloua M, Hancock P. Effects of automation reliability on human monitoring performance. InProceedings of the Human Factors and Ergonomics Society Annual Meeting 2003 Oct (Vol. 47, No. 1, pp. 188-190). Sage CA: Los Angeles, CA: SAGE Publications.
100. 100. Lyell D, Magrabi F, Coiera E. The effect of cognitive load and task complexity on automation bias in electronic prescribing. Human Factors. 2018 Nov;60(7):1008-21.
101. 101. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. Journal of the American Medical Informatics Association. 2012 Jan 1;19(1):121-7.
102. 102. Chavaillaz A, Schwaninger A, Michel S, Sauer J. Expertise, automation and trust in X-ray screening of cabin baggage. Frontiers in psychology. 2019 Feb 14;10:256.
103. 103. Friedman CP, Elstein AS, Wolf FM, Murphy GC, Franz TM, Heckerling PS, Fine PL, Miller TM, Abraham V. Enhancement of clinicians' diagnostic reasoning by computer-based consultation: a multisite study of 2 systems. Jama. 1999 Nov 17;282(19):1851-6.
104. 104. León GA, Chiou EK, Wilkins A. Accountability increases resource sharing: Effects of accountability on human and AI system performance. International Journal of Human–Computer Interaction. 2021 Mar 16;37(5):434-44.
105. 105. Mosier KL, Skitka LJ, Burdick MD, Heers ST. Automation bias, accountability, and verification behaviors. InProceedings of the Human Factors and Ergonomics Society Annual Meeting 1996 Oct (Vol. 40, No. 4, pp. 204-208). Sage CA: Los Angeles, CA: SAGE Publications.
106. 106. Shah SJ, Bliss JP. Does Accountability and an Automation Decision Aid's Reliability Affect Human Performance in a Visual Search Task?. InProceedings of the Human Factors and Ergonomics Society Annual Meeting 2017 Sep (Vol. 61, No. 1, pp. 183-187). Sage CA: Los Angeles, CA: SAGE Publications.
107. 107. Hoff KA, Bashir M. Trust in automation: Integrating empirical evidence on factors that influence trust. Human factors. 2015 May;57(3):407-34.
108. 108. Lee JD, See KA. Trust in automation: Designing for appropriate reliance. Human factors. 2004 Mar;46(1):50-80.
109. 109. Wickens CD. Multiple resources and mental workload. Human factors. 2008 Jun;50(3):449-55.
110. 110. Endsley MR. Toward a theory of situation awareness in dynamic systems. Human factors. 1995 Mar;37(1):32-64.
111. 111. Endsley MR, Garland DJ, editors. Situation awareness analysis and measurement. CRC Press; 2000 Jul 1.
112. 112. Endsley MR. From here to autonomy: lessons learned from human–automation research. Human factors. 2017 Feb;59(1):5-27.
113. 113. O'reilly RC, Munakata Y. Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain. MIT press; 2000 Aug 28.
114. 114. Stanton NA, Salmon PM, Walker GH, Salas E, Hancock PA. State-of-science: situation awareness in individuals, teams and systems. Ergonomics. 2017 Apr 3;60(4):449-66.
115. 115. Endsley MR. Situation awareness global assessment technique (SAGAT). InProceedings of the IEEE 1988 national aerospace and electronics conference 1988 May 23 (pp. 789-795). IEEE.
116. 116. Merritt SM, Ako-Brew A, Bryant WJ, Staley A, McKenna M, Leone A, Shirase L. Automation-induced complacency potential: Development and validation of a new scale. Frontiers in psychology. 2019 Feb 19;10:225.
117. 117. Wilson RC, Collins AG. Ten simple rules for the computational modeling of behavioral data. Elife. 2019 Nov 26;8:e49547.
118. 118. Farrell S, Lewandowsky S. Computational modeling of cognition and behavior. Cambridge University Press; 2018 Feb 22.
119. 119. Palminteri S, Wyart V, Koechlin E. The importance of falsification in computational cognitive modeling. Trends in cognitive sciences. 2017 Jun 1;21(6):425-33.
120. 120. Collins, A.G. and Shenhav, A., 2022. Advances in modeling learning and decision-making in neuroscience. Neuropsychopharmacology, 47(1), pp.104-118.
121. 121. Parasuraman R. Designing automation for human use: empirical studies and quantitative models. Ergonomics. 2000 Jul 1;43(7):931-51.

 doi:10.20944/preprints202203.0070.v1

122. 122. Farrell S, Lewandowsky S. A connectionist model of complacency and adaptive recovery under automation. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2000 Mar;26(2):395.
123. 123. Hu WL, Akash K, Reid T, Jain N. Computational modeling of the dynamics of human trust during human–machine interactions. IEEE Transactions on Human-Machine Systems. 2018 Oct 23;49(6):485-97.
124. 124. Kirlik A, Miller RA, Jagacinski RJ. Supervisory control in a dynamic and uncertain environment: A process model of skilled human-environment interaction. IEEE Transactions on Systems, Man, and Cybernetics. 1993 Jul;23(4):929-52.
125. 125. Morita J, Miwa K, Maehigashi A, Terai H, Kojima K, Ritter FE. Cognitive Modeling of Automation Adaptation in a Time Critical Task. Frontiers in Psychology. 2020:2149.