Change in Human (Moral) Decision-Making and Performance with Intelligent Machines: The Good, the Bad, and the Ugly in Human-Autonomous Systems Interactions

Arthur Prével[1*], Adriana Salatino[2], & Salvatore Lo Bue[2]

[1]Univ. Lille, CNRS, UMR 9193 – SCALab – Sciences Cognitives et Sciences Affectives, F-59000 Lille, France

[2]Department of Life Sciences, Royal Military Academy, Hobbema 8, 1000, Brussels, Belgium

*Corresponding author: arthur.prevel@univ-lille.fr

Abstract

Autonomous systems and intelligent machines are involved in almost all areas of human activity and they are now more and more present in our everyday life. The reason for this extensive use certainly resides in all the benefits these machines offer to the users. In experimental settings, numerous studies have demonstrated the positive effects that the introduction of autonomous systems have on human decision-making and performance. However, studies have shown in addition that the introduction of these systems can have important negative effects as well. Considering that autonomous systems are now introduced in sensitive domains like the military or medicine, we need more than ever a comprehensive understanding of the effects they cause on human performance and decision-making, and particularly in tasks and contexts with a social or moral dimension. The aim of this narrative review is threefold. First, we will provide an overview of the main effects on a human agent's decision-making and performance produced by the introduction of autonomous systems. Second, we will review the conditions identified as underlying factors of these effects, and see how current models of human – autonomous

 doi:10.20944/preprints202203.0070.v2

systems interaction integrate those conditions. Third, we will conclude this review by highlighting new directions for future investigations.

Keywords: decision-making; human-autonomous systems interaction; human performance; intelligent machines; overreliance; situational awareness

Highlights

- Autonomous systems and intelligent machines have both positive and negative effects on human performance and decision-making.
- Understanding the underlying mechanisms of human-autonomous systems interaction is paramount for the design of safe and efficient autonomous technologies.
- Systematic empirical investigations and quantitative modeling are fruitful directions for future research.

1. Introduction

Over the past 50 years, human activities have been radically transformed by the increasing use of autonomous (computerized) systems and machines [1-3]. Designed to assist humans in monitoring activities, decision making, or to help in action execution [e.g., 4,5], autonomous systems help people finding the optimal route to pay a visit to friends or family, support physicians in diagnosing and prescribing medications, or assists workers in industry perform difficult or dangerous tasks. In this regard, the use of autonomous systems is currently widespread in driving [6,7], aviation [8-10], military, defense and security [11], medicine [12,13], nuclear plants [14, 15], etc., and the level of autonomy of the machines involved in each of these areas is constantly increasing [3].

The reason for this extensive use of autonomous systems certainly resides in all the benefits they offer to the users. They can make the driving experience much more comfortable,

 doi:10.20944/preprints202203.0070.v2

help in the detection of pathologies and negative drug-drug combinations, or reduce work accidents in manufacturing. In laboratory settings, several experiments have shown that the use of autonomous systems leads to shorter human response times, fewer errors, or even reduced workload and improved multitasking compared to conditions without their assistance [16-21]. As the presence of these intelligent machines in our life continues to increase, we are now close to a phase where human activities will change from active participation to a task to a more supervisory role, with the ultimate goal for engineers being an autonomy of the systems that distance humans as much as possible from the controlled process [1; 3; 10; 11; 22].

Unfortunately, the involvement of these technologies in human activities has not been systematically associated with positive effects. More autonomy 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, in which an autopilot malfunction combined with a (supposed) loss of situational awareness by the pilots, resulted in the plane stalling and falling into the Atlantic Ocean, killing all the 12 staff and 216 passengers on board. Other examples involving vehicles or weapons are also reported in the literature [e.g., 23,24].

For this reason, researchers today are studying the interactions between humans and autonomous systems more thoroughly to identify and describe the negative effects arising simultaneously with the benefits of autonomous functions and machines, and the conditions under which these effects are observed [2, 25]. For example, studies have found that the use of autonomous systems can sometimes lead to overreliance in the system's decisions, to a degradation of manual and recovery skills, or to a loss of sense of control over the ongoing task [21, 26-28]. Understanding the conditions that produce these effects is an absolute necessity. It would help engineers to develop autonomous systems and intelligent machines that ensure safety, improve performance, and guide human users to make optimal decisions without the potential negative outcomes. This is particularly true when autonomous systems are used in tasks and contexts that require moral decision-making [29-31].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 January 2023 doi:10.20944/preprints202203.0070.v2

For example, a combat drone operator remotely engaged in a battlefield must make moral decisions based on information and/or suggestions from the system. He/she 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 future enemy attacks that also involve civilian losses and damages to civilian infrastructure. In that situation, the overreliance on the system's decisions might have dramatic consequences. However, to the best of our knowledge, the majority of research on moral dilemmas and autonomous systems concerns the algorithms and values to be assigned to the systems to guide their decisions during moral decision-making situations [e.g., 32-35], not the effect on human moral decision-making. Considering the negative effects mentioned above, such as overreliance or loss of sense of control, it is possible that interaction with autonomous systems also impacts negatively these decisions and the corollary actions [36,37].

Thus, with the use of autonomous and intelligent machines in sensitive areas such as defense and security or medicine, we need more than ever a comprehensive understanding of the impact of these systems on human performance and decision-making, and particularly in tasks and contexts with moral value. In some cases, this impact is trivial, like when we tend to follow blindly the instructions of a GPS even though a road-sign indicates that the road we are driving on is closed 200 meters away. 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 narrative review is to provide an overview of the effects of autonomous systems on human decision-making and performance. We will start with a review of the main positive effects that these systems have. We will then delve into the negative effects and their associated conditions reported with their use, with a specific discussion on situations that involve a moral or social dimension. After this review of empirical findings, in a second part of this paper, we will examine factors that are known to influence human-autonomous systems

 doi:10.20944/preprints202203.0070.v2

interaction. After this overview, we will present some of the psychological models developed to account for the effects described in the first part. We will discuss the benefits of these models, but also their limitations and possible ways to improve them. The second purpose of this paper is to provide an overview of the known determining factors in human-autonomous system 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 in this domain. With this review, we aim indeed to stimulate innovative investigations and insights into the circumstances that produce the negative effects sometimes observed in autonomous systems and how to counteract these effects. Finally, we will conclude this review by highlighting potential fruitful directions for future research on human-autonomous system interaction.

2. Human-autonomous system interaction and its effect on decision-making and performance

2.1. The good: Behavior enhancement with autonomous systems

With the massive increment in computer performance and the remarkable development of artificial intelligence and machine learning, new technologies are now largely dominated by computerized technologies [3]. Autonomous functions and machines offer many benefits in human activities. At an industrial level, they increase manufacturing efficiency and productivity, and reduces the risk of accidents. At an individual level, autonomous systems are designed to make the user/consumer experience much more comfortable and effective as well. But autonomy, of course, is not all or none. The extent to which a task is performed autonomously varies across needs and situations. Thus, autonomy is characterized by different levels [38, 39], and range from very basic automated processes (e.g., automated data acquisition), to fully autonomous systems with very little human control. Furthermore, autonomy can be introduced at different stages of task execution. A classical description is given by Parasuraman et al. [38; see also 40]. According to the authors, autonomous systems can be assigned to 1) information

 doi:10.20944/preprints202203.0070.v2

acquisition, 2) information analysis, 3) decision-making, or 4) action processing (for different descriptions, see also [25, 41]). Thus, autonomous systems and machines can either help to detect relevant information from the environment and to analyze the information collected more deeply (e.g., by inferring about outcomes likelihood), or help the user make the optimal decision given the collected information and execute the selected action. In this first section, and for the rest of the paper, we will focus on systems in which a human agent is still involved in the decision process. Since we focus on the effect of autonomous systems on human decision-making and performance, and since complete substitution of human control remains a myth [22], the scientific evidence presented in this review will not discuss the supposed performance and issues of fully autonomous systems.

We will now illustrate the positive effects of autonomous systems on human decision-making and performance with some relevant findings from the literature, following the stage classification proposed by Parasuraman et al. [38]. As Parasuraman et al. [38] have described, the first stage of autonomy is information acquisition. At the user's level, autonomous information acquisition 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 autonomous information acquisition is additional visual cues superimposed on a target to be detected. For example, Yeh & Wickens [42] asked participants to pilot an unmanned aerial vehicle in a virtual reality environment and search for hidden military targets (e.g., a tank). Participants could be assisted by autonomous functions whose output consisted in a reticle automatically superimposed on a target when it was detected by the system. In this case, target detection increased with the help of the target cueing autonomous detection (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 autonomous system that superimposed a salient green circle to detected knife (i.e., a target). Consistent with the results by Yeh & Wickens [42],

 doi:10.20944/preprints202203.0070.v2

the authors observed that the proportion of correct detection was higher in the group with automated function than in the group without this function (see also [16, 44]). Interestingly, the stimulus used for information acquisition can have different physical properties. For example, Rice & McCarley [20] used a text message as cue, while Dixon & Wickens [45] used an auditory stimulus. Finally, in a study by St. John et al. [46], information acquisition consisted of directly changing the saliency of the target stimuli.

The second stage of autonomy is information analysis. In this form of autonomous system, 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 were required to monitor a visual airspace to defend a ship against aerial threats. Participants were assisted in this task by an autonomous system that continuously analyzed the level of threat represented by the aerial vehicles displayed on the screen, and changed their saliency on the screen accordingly. This reduced the response time to the threatening aircraft compared to a situation without this function. Another good example of autonomous information analysis is clinical decision support system. Clinical decision support systems are autonomous systems and machines 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 ordered diagnostic hypotheses based on information put in the system. The authors recorded the rate of correct diagnoses from first-year medicine students and results showed that the use of decision support system increased the number of correct diagnoses (see also [48]).

Thus, autonomous systems are 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 autonomous systems that support directly the user in decision-making process (stage 3) and action execution (stage 4). In autonomous decision, the system

 doi:10.20944/preprints202203.0070.v2

or machine is designed to suggest or to select an action among a set of possibilities, based on the evaluation of the different available actions and their potential outcomes. Action autonomy simply refers to the actual execution of the action selected by the system (or the user). A good example of the effect of autonomous decision 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 allied units on a simulated battlefield. Participants could either make their decisions based on basic information about the possible engagement combinations, or based on the support of a decision assistance that provided the best options to select with varying degree of accuracy. The results found by Rovira and colleagues show that decision accuracy (on reliable decision trials) is superior to basic information, and increases with the degree of precision. We can also cite 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 for the correct decisions to make, multiple experiments have demonstrated how autonomous systems produce positive effects on human performance and decision-making. A consequence of the facilitation from these systems is the increased possibility for multi-tasking for the human agent [e.g., 27, 50-53]. Improved multi-tasking represents another advantage of machine-assisted over human-controlled task. For example, Cullen et al. [50] have shown that in a multi-task environment, information cues helped to increase efficient switching 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 providing route recommendations compared to a situation without assistance. In summary, not only autonomy can help in performing a specific task, but it

 doi:10.20944/preprints202203.0070.v2

can also assist human agents efficiently when they have to monitor multiple sources of information and perform multiple tasks. Unfortunately, the introduction of autonomous systems and machines might also produce serious drawbacks. The next section will focus on the downsides of human-automation interaction.

2.2. The bad: Negative effects of autonomous systems on human decision-making and performance

Reports of troubles with autonomous systems and intelligent machines are not new in the scientific literature. The first evidence and discussion of negative effects produced by the interaction with these technologies were reported around the 1980s and 1990s [e.g., 2, 54, 55], mainly in the aircraft domain [e.g., 56, 57]. Today, most scientists and engineers acknowledge that introducing autonomy in a task is not just a "substitution" of an intelligent system or a machine for a human activity [58]. Autonomous systems do 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. Fifty years of research teach us that the right balance between autonomy and human control must be considered carefully before introducing automated functions and machines [59]. Otherwise, autonomous systems can have a substantial detrimental effect 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 autonomous systems are introduced in human activity. In doing so, 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 these outcomes will be discussed in the second part of the manuscript.

The first negative effect that we can discuss concerns, with some irony, multi-tasking. As we have seen in the previous section, autonomous systems are designed to perform tasks that

 doi:10.20944/preprints202203.0070.v2

were 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 is true as long as autonomy is properly designed and implemented, and as long as a human agent is properly trained for its use. Like performance decreases with manual multi-tasking, introducing too many automated-tasks to control might also have detrimental effect on performance (for review, see for example [60]). For instance, Chen and Joyner [61] reported that the performance on a target gunnery task in a simulated mounted combat system decreased with the introduction of an additional automated task, particularly with low level of autonomy, while perceived workload increased. Wang et al. [62] in a search and rescue task with multiple robots, found that exploring the environment on one hand, and searching 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 the screen from 8 to 12 robots, and that perceived workload increased with the number of robots in each condition (see also [63, 64] for similar results). These examples show that the introduction of autonomous systems is definitely not enough to improve human multi-tasking performance, and that, on the contrary, an inappropriate level of automation and task allocation can lead to substantial performance decrement and more errors from the human user. Particularly, studies on multi-tasking suggest that human agents are particularly sensitive to overreliance, which is certainly one of the most important negative effects in human-autonomous systems interaction. Overreliance is broadly defined as the tendency of humans interacting with autonomous systems to use the output of the systems (e.g., information cues) as heuristics to reduce effortful activities such as searching and processing information [65]. More specifically, overreliance is said to occur when the performance of a user decreases because of incorrect information and/or decision made by an autonomous system [25, 66]. It is manifest in two types of errors: omission error and commission error. Omission error occurs when an autonomous system fails to inform about a

 doi:10.20944/preprints202203.0070.v2

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., not checking the screened-luggage). At the opposite, in commission error the system makes an incorrect decision about the environment or gives incorrect advices, which, in turn, result here also in an inappropriate response by the user (e.g., the system assumes that there is a weapon in a screened-luggage while there is not). Thus, overreliance corresponds to the fact that incorrect information or decision cues from the autonomous system, but not the actual environment, control the decisions and actions from a human agent. Examples of this phenomenon has been reported in almost every task and domain involving autonomy [25, 66]. For longtime, overreliance has been associated with autonomy used in 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 monitored and supported by partially unreliable autonomous systems (see also [67-69]). However, it seems now evident that overreliance affects also single-task environment [70]. For example, 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 positive effect of autonomous decision on correct responses made by the participants, the authors also found incorrect responses during unreliable trials. In summary, overreliance occurs both in single- and multiple-tasks environments, in both omission and commission errors, and it is observed for both information and decision automation.

In addition to inappropriate autonomous-task allocation and overreliance, another negative effect of autonomy is the loss of skills ([55]; or skill decay). Loss of skills refers to a deterioration in task performance (motor or cognitive) after a more or less prolonged experience of a user with automated tasks. A driver who has difficulty to drive an old car after a long period of driving a very modern one with many automated assistances is an illustration of this effect. In

 doi:10.20944/preprints202203.0070.v2

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 critical when the system fails, and the human operator has to take back manually the control of the task. In this context, there is evidence of a decrement in the return-to-manual control after system failure. For example, Endsley & Kiris [55] found that response time decision in a navigation task increases when participants had to unexpectedly respond manually after a period of automated-assistance. Similar results were found by Manzey et al. [27] with highest return-to-manual decrement for higher level of autonomy (see also [75, 76]). To conclude this section, we would like to discuss another intriguing aspect of human-autonomous system interaction that have shown growing interest in the recent years, namely, the effect of autonomy on human agency [77]. Human agency (or “sense of agency”) refers to the individual experience of controlling one’s own actions and, through those actions, outcomes in the external environment [78]. Recently, scientists have become interested in how the interaction with autonomous systems influences how people feel in control in their own actions. One of the first demonstrations of an effect of autonomy on sense of agency came from a study by Berberian et al. [26]. In an aircraft supervision task, the authors found a decrease in agency with the introduction of task autonomy, with agency reduced at higher levels of autonomy compared to lower levels (see also [79, 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]. For example, in the social domain, 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 autonomy can reduce the sense of agency from human users lead to believe of potentially misuses, 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 infrastructure damages during attacks. In this case, a decrement in the sense of agency – combined with

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 January 2023 doi:10.20944/preprints202203.0070.v2

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 autonomous systems 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 autonomous systems might change our behaviors and our decisions when we have to face moral situations.

2.3. The ugly: Autonomous systems and moral decision-making

In the last two decades, the main focus of engineers, scientists, and philosophers regarding moral decision-making and autonomy 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 has been conducted on what are the best rules/algorithms to assign to an automated function or machine in moral situations [32, 35] and how human subjects would behave in moral decision-making situations to inspire the development of ethical autonomous systems [33, 34]. In addition, scientists and philosophers have considered the legal and ethical consequences of fully autonomous machines in critical situations such as driving or military conflicts [86-88]. Surprisingly, understanding how the interaction with autonomous systems can change the ethical and social behaviors of 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 a presumed consensus on the way to behave or not within a social group, that is, behaviors following a social norm. Moral decision-making refers to a decision or a judgment made in a situation with moral rules and moral principles play a role [89, 90]. The recent interest in this topic can be explained by the increased proportion of behaviors in our everyday life that are guided by autonomous systems (driving, communication, health, etc.). In addition, as we have discussed above, autonomous systems are now more and more involved in sensitive domains such as military operations, medicine, and security. Therefore, understanding the effect of autonomy on human behavior in that context becomes crucial.

 doi:10.20944/preprints202203.0070.v2

The available evidence suggests a mixed-picture of the effects of autonomous systems in social and moral decision-making situations. On the one hand, some recent findings suggest, for instance, that the interaction with automation in social dilemma situation 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 autonomous systems and intelligent machines in the social and the moral domain does not necessarily increase the rate of unethical or unsocial behaviors, but, on the contrary, might have a positive effect by increasing prosocial behaviors. On the other hand, an analysis of additional results shows that the effect of autonomy is not so clear and using these technologies 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 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 who played the game through self-designed autonomous agents, designed autonomous systems to improve their own 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 was as strong as the effect of human agents in promoting unethical behaviors during social interactions.

Finally, some of the results on overreliance cited above can also shed light on the influence of autonomous systems on human decision processes, especially when these responses 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 overreliance was reported for mammography assessment and in prescription task. Although the conflicting moral aspect of the decisions made is not addressed explicitly in these studies, these scenarios had obvious health and life consequences. For example, an error of omission during mammography assessment may result in undiagnosed cancer for a patient. Thus, it seems that despite potential critical

 doi:10.20944/preprints202203.0070.v2

negative outcomes, people can nevertheless follow the bad recommendations of autonomous systems. Similarly, in the military domain, the command-and-control study by Rovira et al. [21] reported high rate of incorrect decisions made by the participants when the autonomous systems decisions were not reliable. The decisions in the Rovira et al. study involved the engagement of opponents with units on a simulated battlefield. Does this mean that commanders making decisions with the help of autonomous decision systems could show overreliance? And what about combat drone operators who face the risk of civilian losses while they are engaging a military target? In conclusion, the available empirical evidence suggests that interaction with automation might lead to the promotion of pro-social behaviors (fairness, cooperation, etc.), but also to unethical and aberrant behaviors. Thus, while autonomous systems 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 also seems that in certain circumstances the interaction with these systems can be detrimental in terms of ethical decision making. This dual-aspect of autonomy also applies to situations that at first sight do not involve moral and social decision-making, with autonomous systems favoring at the same time global performance improvement and multi-tasking, but overreliance and loss of sense of agency as well. Following the research agenda of several authors (e.g., [25]), and considering that autonomous systems and intelligent machines 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 reviewed above. This will help developing systems that are 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-machine interaction

 doi:10.20944/preprints202203.0070.v2

3.1. Determinants of the effect of autonomous systems on human decision-making and performance

Understanding the determining factors of the effects of autonomous systems on human decision-making and performance has been for longtime a goal for researchers and engineers [2, 25, 57]. Since the first studies conducted in the 1980s, several factors have been identified as crucial determinants in human-autonomous system interaction. In this section, we will review some of the most important factors identified and describe their effects on the human agent interacting with the system. Again, it is particularly relevant to identify these factors and to understand exactly what their effects are, as this will help in turn to recognize the circumstances in which both the positive and the negative effects described in the previous section are observed. With that information, scientists and engineers will be able to develop new forms of autonomy and intelligent machines. They will be efficient in terms of the positive changes they produce on the users' performance and safety (e.g., fast and correct decisions, the possibility for multi-tasking, etc.), but will also prevent or at least mitigate the negative outcomes we described, particularly in the context of social and moral situations.

3.1.1. Level and stage of autonomy

The first factor that determines how human agents interact with autonomous systems is the level and the stage at which autonomy occurs. As a reminder, levels of autonomy refer to the notion that the degree of autonomy for a given task can vary across a continuous scale, with intermediate levels representing different degrees at which autonomy is assigned [38, 39], while the stages refer to the different subtasks on which task autonomy can be assigned [38, 40]. Most of the information on this topic has already been presented in the previous sections. Importantly, an increased level of autonomy seems to be associated with improved decisions and increased performance by the human user. For example, Manzey et al. [27] reported that performance in a supervisory control task was better with autonomy than with manual control, with this positive effect being superior when participants were supported by the highest level of

 doi:10.20944/preprints202203.0070.v2

autonomy of the system [see also 21, 75]. In addition, more autonomy 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 the execution of a secondary task is supported by a high level compared to a low level of task autonomy (see also [27]). Although generally positive in terms of decision-making and performance, higher level of automation can also lead to a loss of skill [55] and increased delay to take-back control of the system in case of system failure. In the Manzey et al. [27] study, for example, the cost of returning to manual mode was higher for the higher level of autonomy (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 task autonomy, measured either by a direct rating about the task or by the indirect temporal binding measure [26]. Thus, further research seems necessary to understand the exact balance between the beneficial effects and the disadvantages of higher levels of autonomy of the systems. How the interaction with other factors (e.g., level of skills and previous experience, accountability) might mitigate or enhance these effects is another venue for research. Concerning the stage at which autonomy is introduced, whether it is information and analysis autonomy or decision and action autonomy, all of these forms of autonomous systems can help improve performance and decisions [40]. However, compared to information/analysis autonomous systems, decision autonomy seems to be more subject to overreliance [21, 49, 98]. This result is not surprising considering that in decision autonomy, it is not necessary for the user to look at the environment, but only at the output from the systems, while in information autonomy for instance, the user evaluates a preprocessed environment (but still look at the environment).

3.1.2. Autonomous system reliability

A second major determinant of the effect of autonomous systems on human decision-making and performance is autonomous system reliability. The effect of reliability is certainly one of the most extensively studied factors in human-autonomous system interaction. Its effect has been tested both for information/analysis autonomy [e.g., 16, 45] and for decision/action

 doi:10.20944/preprints202203.0070.v2

autonomy [e.g., 21]. Overall, investigations conducted on this effect found that decision-making and performance improves with reliability. For instance, Goh et al. [17] found that participants' performance was higher for information cue with 90% reliability than with 70% reliability. Interestingly, a positive effect of autonomy can be obtained even with relatively low level of reliability. For example, Cullen et al. [51] found that the performance of subjects who were helped with information autonomy in a multi-task environment increased compared to a condition without autonomy, even with a reliability of 67% (see also [21]). Although resulting in global decision-making and performance increment, several studies have reported that error rates during unreliable trials increase with higher levels of reliability of the system. Thus, high autonomous systems reliability seems to be associated with increased tendency for overreliance, 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 failures of an autonomous system decreased with reliability in detecting errors (see also [67]). Thus, higher reliability seems to increase the global task performance, but also the risk of overreliance. As a consequence, much reliability would be relevant only as long as the negative outcomes that result from overreliance are not superior compared to the gain obtained from the increased reliability. As with the effect of higher-level of autonomy, the balance between the advantages and the disadvantages of higher reliability will have to be investigated systematically.

3.1.3. Task difficulty and multi-tasking

A third important factor that determines how a human agent will interact with autonomous systems is task difficulty and/or the number of tasks simultaneously monitored by the agent. As we have already seen, automation is particularly useful in difficult manual or cognitive tasks. It provides access to more rewards by increasing correct performance and decision rate [19, 45, 49, 100], and reduces the costs related to task execution (measured for example by reduced subjective workload; e.g., [19]). In addition, autonomy facilitates multi-tasking by automatizing manual tasks (action automation) or by helping the detection of targets

 doi:10.20944/preprints202203.0070.v2

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 autonomous systems in the context of a difficult single-task and/or multi-tasking is frequently considered as a driving factor of overreliance [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 user's error rates [e.g., 44]. Finally, as a counteract to the effect of task difficulty and multi-tasking, individual experience and task mastery seem to reduce the probability of overreliance 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 autonomous systems [e.g., 102, 103].

3.1.4. Performance outcome, accountability, and automation display

Additional factors can be cited as crucial variables during human-autonomous systems interaction. For example, some experiments have shown that performance outcome, and particularly the consequences of errors, can have an impact on the rate of overreliance. In particular, tasks for which errors might have more important negative consequences seem to drive users to evaluate them more carefully [e.g., 28]. Related to this is the evidence of an effect of accountability of decisions on the rate of overreliance. 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 autonomous systems have a strong influence on that performance. Finally, we can cite the effect of the physical properties of the system and the way the output from the system is displayed. For example, Goh et al. [17] found that information automation in a screening-task is improved when the automation cue was 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. In summary, the way a human agent interacts with autonomous

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 January 2023 doi:10.20944/preprints202203.0070.v2

system is influenced by multiples factors, interacting with each other. In the next and final section of this review, we will complete this overview of factors by a presentation of models that scientists have developed over the last few decades to explain and make prediction about the reported effects, and to a discussion on potential way of improvements for future models.

3.2. Models of human-autonomous systems interaction

With the increasing number of studies conducted on the factors underlying human-autonomous systems interaction, scientists have developed several models of human decision-making and performance in interaction with autonomous systems and machines. This development has led to 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 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 autonomy 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 neurosciences [e.g., 113], in order to increase their precision and their explanatory power.

The first concept and associated model we will introduce is (loss of) situation awareness [110, 111]. Used initially in the aviation domain, situation awareness has become certainly one of the most important concepts used in human-autonomous system 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”. Thus, situation awareness is composed of three

 doi:10.20944/preprints202203.0070.v2

different hierarchical levels: 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.

Given the assumed-role of situation awareness in the decision-making process, we can see how inappropriate autonomous systems features could lead to a decrement of situation awareness (i.e., a loss of situation awareness), and thus to an increased probability of erroneous decisions by the human agent. For example, a high level of autonomy combined with insufficient training or a loss of skills from the human agent could result, in case of failure from the system, 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 the author defines determining factors. Specifically, the author proposes that a loss of situation awareness during the interaction with autonomous systems and machines 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 factors are supposed to intervene in the agent's level of situation awareness and thus in his/her performance in the task at hand. This integration of factors allowed the author to propose several guidelines for the design of appropriate systems in interaction with human agents. For example, it is suggested that autonomous systems should preferably be used for routine tasks, information and analysis autonomy rather than decision autonomy, or to increase information saliency and ensure transparency from the system (see [112] for complete guidelines).

 doi:10.20944/preprints202203.0070.v2

Closely related to situation awareness, another important concept in the human-autonomous systems interaction domain is automation complacency [66, 67]. Automation complacency is said to occur when a human agent is monitoring an autonomous system, but with a suboptimal rate of monitoring which, in turn, might lead to performance failures [66, 116]. More specifically, this performance failure results from both a direct failure of the system 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 autonomous system failures detections 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 overreliance or automation bias [25]. Parasuraman and Manzey [66] proposed a model that integrates both complacency and overreliance. The model is composed of a complacency potential component, which influence a component of attentional information processing, which in turn influence the agent's situation awareness and performance. Interestingly, the structure of the model is relatively similar to the model recently proposed by Endsley [112]. In the latter, the complacency potential, which is seen as a tendency to react less attentively during the interaction with a specific autonomous system, is assumed to be influenced by the reliability and consistency of the system, as well as by individual characteristics and interaction history with the system. Thus, increased reliability is assumed to increase the 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 model and Parasuraman & Manzey's [66] complacency models have been shown interesting to interpret some of the effects found when autonomy is introduced and to integrate empirical evidence of known

 doi:10.20944/preprints202203.0070.v2

underlying factors of human-autonomous system interaction (e.g., level of autonomy, autonomous system reliability). Based on their interpretations of the role of specific underlying factors, these models allow to make recommendations to engineers for the design of new 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 a system failure. Given 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 systems in these situations. Despite all these advantages, however, the models are not without limitations. In particular, 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 are not precisely described. Thus, it is difficult to know what predictions they allow in specific circumstances, and consequently, it is difficult to test these predictions and make comparisons between models. More generally, this reduces the extent to which engineers can use the models to anticipate how human users would act under particular circumstances.

To improve the validity 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 in the use of computational models in behavioral and neurocognitive research [e.g., 113, 118]. The models show several advantages over classical conceptual models. Notably, these models are explicit and falsifiable, 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

 doi:10.20944/preprints202203.0070.v2

phenomena have shown several important successes [e.g., 120], and we strongly believe that this could also be beneficial for better understanding human decision-making and performance in the context of human-automation interaction. This suggestion has already been proposed in the literature [e.g., 121] and computational models of human decision-making during human-autonomous system interaction have been developed [e.g., 122-125]. To date, however, the proportion of models using computational methods remains relatively small, and much more investigations are needed for the systematic use of these models.

4. Discussion

Autonomous systems and intelligent machines are now widespread in almost every area 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 autonomous systems can lead to significant improvement in human decision-making processes and performance, but can result in serious negative effects like overreliance [2]. Understanding the conditions in which these effects appear has become crucial, especially when autonomy 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 studies on human-autonomous system interaction, multiple factors have been identified, and their effects begin to be well understood (partially at least). Models have been proposed to explain the changes in decision-making and performance found in human-autonomous systems studies. Concepts like loss of situation awareness [110, 112] and automation complacency [66] have been introduced, and the models allow to make some recommendations for the design of new automated technologies. However, these models suffer from being mainly conceptual or descriptive models, which limits their predictive value. We suggest the use of computational modeling to increase the models' precision in predictions [118, 121].

To conclude this review, we would like to highlight potential fruitful directions for future human-autonomous system interaction research. First, new investigations need to be conducted

on how the introduction of these technology 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 autonomy in sensitive areas such as medicine, and military defense and security, we need to understand the conditions that favor ethical decisions by humans when they interact with autonomous systems and machines. Second, more investigations are necessary on the factors underlying the interaction between human and autonomous systems. In particular, while many investigations have been conducted on contextual effects (e.g., effects of automation reliability, levels and stages of automation, etc.), much less is known about the effects of task consequences and/or accountability (i.e., the effect of contingent-outcomes presented during the interaction with autonomous systems [25]). Finally, an important area of research will consist in the massive development of computational models to explain more accurately how autonomy changes decision-making processes and performance in human subjects [118, 121]. By doing so, engineers will be able to develop new technologies aimed at improving human performance, but reducing at the same time the risk of dramatic consequences.

Author Contributions: Writing—review and editing, AP, AS, 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. 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. Parasuraman R, Riley V. Humans and automation: Use, misuse, disuse, abuse. Human factors. 1997 Jun;39(2):230-53.

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 Intel-ligence and Life in 2030 - One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel. Stanford University, Stanford, CA.

4. Shakhatreh H, Sawalmeh AH, Al-Fuqaha A, Dou Z, Almaita E, Khalil I, Othman NS, Khreishah A, Guizani M. Un-manned aerial vehicles (UAVs): A survey on civil applications and key research challenges. Ieee Access. 2019 Apr 9;7:48572-634.

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 intelli-gent unmanned autonomous systems. Frontiers of information technology & electronic engineering. 2017 Jan;18(1):68-85.

6. Ayoub J, Zhou F, Bao S, Yang XJ. From manual driving to automated driving: A review of 10 years of autoui. InPro-ceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications 2019 Sep 21 (pp. 70-90).

7. Chan CY. Advancements, prospects, and impacts of automated driving systems. International journal of transporta-tion science and technology. 2017 Sep 1;6(3):208-16.

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. Chialastri A. Automation in aviation. IntechOpen; 2012 Jul 25.

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. 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. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support sys-tems: a systematic review of trials to identify features critical to success. Bmj. 2005 Mar 31;330(7494):765.

13. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision sup-port systems: benefits, risks, and strategies for success. NPJ digital medicine. 2020 Feb 6;3(1):1-0.

14. Hanna B, Son TC, Dinh N. AI-guided reasoning-based operator support system for the nuclear power plant man-agement. Annals of Nuclear Energy. 2021 May 1;154:108079.

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. 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. 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. Ma R, Kaber DB. Effects of in-vehicle navigation assistance and performance on driver trust and vehicle control. In-ternational Journal of Industrial Ergonomics. 2007 Aug 1;37(8):665-73.

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

 doi:10.20944/preprints202203.0070.v2

Ergonomics Society Annual Meeting 1997 Oct (Vol. 41, No. 1, pp. 190-194). Sage CA: Los Angeles, CA: SAGE Publications.

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. 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. Lee, J. D., Wickens, C. D., Liu, Y., & Boyle, L. N. (2017). Designing for people: An introduction to Human Factors Engineering. Charleston: Create Space.
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. Williams KW. A summary of unmanned aircraft accident/incident data: Human factors implications. Federal Avia-tion Administration Oklahoma City OK Civil Aeromedical Inst; 2004 Dec 1.
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. 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. 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. 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. 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. Cushman F. Action, outcome, and value: A dual-system framework for morality. Personality and social psychology review. 2013 Aug;17(3):273-92.

31. Cushman F, Kumar V, Railton P. Moral learning: Psychological and philosophical perspectives. Cognition. 2017 Oct 1;167:1-0.

32. Arkin RC, Ulam P, Wagner AR. Moral decision making in autonomous systems: Enforcement, moral emotions, dig-nity, trust, and deception. Proceedings of the IEEE. 2011 Dec 9;100(3):571-89.

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. Bonnefon JF, Shariff A, Rahwan I. The social dilemma of autonomous vehicles. Science. 2016 Jun 24;352(6293):1573-6.

35. Jiang L, Hwang JD, Bhagavatula C, Bras RL, Forbes M, Borchardt J, Liang J, Etzioni O, Sap M, Choi Y. Delphi: To-wards machine ethics and norms. arXiv preprint arXiv:2110.07574. 2021 Oct 14.

36. Köbis N, Bonnefon JF, Rahwan I. Bad machines corrupt good morals. Nature Human Behaviour. 2021 Jun;5(6):679-85.

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. 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. 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. Parasuraman R, Wickens CD. Humans: Still vital after all these years of automation. Human factors. 2008 Jun;50(3):511-20.

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. 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. 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. Rieger T, Heilmann L, Manzey D. Visual search behavior and performance in luggage screening: effects of time pres-sure, automation aid, and target expectancy. Cognitive Research: Principles and Implications. 2021 Dec;6(1):1-2.

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. 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. Martinez-Franco AI, Sanchez-Mendiola M, Mazon-Ramirez JJ, Hernandez-Torres I, Rivero-Lopez C, Spicer T, Mar-tinez-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. 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. 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. Chen JY, Barnes MJ. Supervisory control of multiple robots in dynamic tasking environments. Ergonomics. 2012 Sep 1;55(9):1043-58.

51. Cullen RH, Rogers WA, Fisk AD. Human performance in a multiple-task environment: Effects of automation reliabil-ity on visual attention allocation. Applied ergonomics. 2013 Nov 1;44(6):962-8.

52. Cummings ML, Guerlain S. Developing operator capacity estimates for supervisory control of autonomous vehicles. Human factors. 2007 Feb;49(1):1-5.

53. Wright JL, Chen JY, Barnes MJ. Human–automation interaction for multiple robot control: the effect of varying au-tomation assistance and individual differences on operator performance. Ergonomics. 2018 Aug 3;61(8):1033-45.

54. Bainbridge L. Ironies of automation. InAnalysis, design and evaluation of man–machine systems 1983 Jan 1 (pp. 129-135). Pergamon.

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. Wiener EL. Cockpit automation. InHuman factors in aviation 1988 Jan 1 (pp. 433-461). Academic Press.

57. Wiener EL, Curry RE. Flight-deck automation: Promises and problems. Ergonomics. 1980 Nov 1;23(10):995-1011.

58. Sarter NB, Woods DD, Billings CE. Automation surprises. Handbook of human factors and ergonomics. 1997;2:1926-43.

59. Sheridan TB. Function allocation: algorithm, alchemy or apostasy?. International Journal of Human-Computer Studies. 2000 Feb 1;52(2):203-16.

60. Prewett MS, Johnson RC, Saboe KN, Elliott LR, Coovert MD. Managing workload in human–robot interaction: A re-view of empirical studies. Computers in Human Behavior. 2010 Sep 1;26(5):840-56.

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. 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. Adams JA. Multiple robot/single human interaction: Effects on perceived workload. Behaviour & Information Tech-nology. 2009 Mar 1;28(2):183-98.

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. 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. Parasuraman R, Manzey DH. Complacency and bias in human use of automation: An attentional integration. Human factors. 2010 Jun;52(3):381-410.

67. Parasuraman R, Molloy R, Singh IL. Performance consequences of automation-induced'complacency'. The Interna-tional Journal of Aviation Psychology. 1993 Jan 1;3(1):1-23.

68. Skitka LJ, Mosier KL, Burdick M. Does automation bias decision-making?. International Journal of Hu-man-Computer Studies. 1999 Nov 1;51(5):991-1006.

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. Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. Journal of the American Medi-cal Informatics Association. 2017 Mar 1;24(2):423-31.

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. 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. 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. Volz KM, Dorneich MC. Evaluation of cognitive skill degradation in flight planning. Journal of Cognitive Engineer-ing and Decision Making. 2020 Dec;14(4):263-87.

75. Endsley MR, Kaber DB. Level of automation effects on performance, situation awareness and workload in a dy-namic control task. Ergonomics. 1999 Mar 1;42(3):462-92.

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. Berberian B. Man-Machine teaming: a problem of Agency. IFAC-PapersOnLine. 2019 Jan 1;51(34):118-23.

78. Haggard P, Chambon V. Sense of agency. Current biology. 2012 May 22;22(10):R390-2.

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. Zanatto D, Chattington M, Noyes J. Human-machine sense of agency. International Journal of Human-Computer Studies. 2021 Dec 1;156:102716.

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. Haggard P. Sense of agency in the human brain. Nature Reviews Neuroscience. 2017 Apr;18(4):196-207.

83. Caspar EA, Christensen JF, Cleeremans A, Haggard P. Coercion changes the sense of agency in the human brain. Cur-rent biology. 2016 Mar 7;26(5):585-92.

84. Caspar EA, Cleeremans A, Haggard P. Only giving orders? An experimental study of the sense of agency when giv-ing or receiving commands. PloS one. 2018 Sep 26;13(9):e0204027.

85. Caspar EA, Lo Bue S, Magalhães De Saldanha da Gama PA, Haggard P, Cleeremans A. The effect of military train-ing on the sense of agency and outcome processing. Nature communications. 2020 Aug 31;11(1):1-0.

86. Beard JM. Autonomous weapons and human responsibilities. Geo. J. Int'l L.. 2013;45:617.

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. Harris J. Who owns my autonomous vehicle? Ethics and responsibility in artificial and human intelligence. Cam-bridge Quarterly of Healthcare Ethics. 2018 Oct;27(4):599-609.

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. Moll J, Zahn R, de Oliveira-Souza R, Krueger F, Grafman J. The neural basis of human moral cognition. Nature re-views neuroscience. 2005 Oct;6(10):799-809.

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. 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. Fernández Domingos E, Terrucha I, Suchon R, Grujić J, Burguillo JC, Santos FC, Lenaerts T. Delegation to autono-mous agents promotes cooperation in collective-risk dilemmas. arXiv e-prints. 2021 Mar:arXiv-2103.

94. Kirchkamp O, Strobel C. Sharing responsibility with a machine. Journal of Behavioral and Experimental Economics. 2019 Jun 1;80:25-33.

95. March C. The behavioral economics of artificial intelligence: Lessons from experiments with computer players. CE-Sifo Working Paper No. 7926. 2019.

96. Cohn A, Gesche T, Maréchal MA. Honesty in the digital age. Management Science. 2021 Nov 8.

97. Manistersky E, Lin R, Kraus S. The development of the strategic behavior of peer designed agents. InLanguage, Cul-ture, Computation. Computing-Theory and Technology 2014 (pp. 180-196). Springer, Berlin, Heidelberg.

98. Crocoll WM, Coury BG. Status or recommendation: Selecting the type of information for decision aiding. InProceed-ings of the human factors society annual meeting 1990 Oct (Vol. 34, No. 19, pp. 1524-1528). Sage CA: Los Angeles, CA: SAGE Publications.

99. Oakley B, Mouloua M, Hancock P. Effects of automation reliability on human monitoring performance. InProceed-ings 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. Lyell D, Magrabi F, Coiera E. The effect of cognitive load and task complexity on automation bias in electronic pre-scribing. Human Factors. 2018 Nov;60(7):1008-21.

101. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and miti-gators. Journal of the American Medical Informatics Association. 2012 Jan 1;19(1):121-7.

102. Chavaillaz A, Schwaninger A, Michel S, Sauer J. Expertise, automation and trust in X-ray screening of cabin bag-gage. Frontiers in psychology. 2019 Feb 14;10:256.

103. Friedman CP, Elstein AS, Wolf FM, Murphy GC, Franz TM, Heckerling PS, Fine PL, Miller TM, Abraham V. En-hancement of clinicians' diagnostic reasoning by computer-based consultation: a multisite study of 2 systems. Jama. 1999 Nov 17;282(19):1851-6.

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. Mosier KL, Skitka LJ, Burdick MD, Heers ST. Automation bias, accountability, and verification behaviors. InPro-ceedings 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. 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. Hoff KA, Bashir M. Trust in automation: Integrating empirical evidence on factors that influence trust. Human fac-tors. 2015 May;57(3):407-34.

108. Lee JD, See KA. Trust in automation: Designing for appropriate reliance. Human factors. 2004 Mar;46(1):50-80.

109. Wickens CD. Multiple resources and mental workload. Human factors. 2008 Jun;50(3):449-55.

110. Endsley MR. Toward a theory of situation awareness in dynamic systems. Human factors. 1995 Mar;37(1):32-64.

111. Endsley MR, Garland DJ, editors. Situation awareness analysis and measurement. CRC Press; 2000 Jul 1.

112. Endsley MR. From here to autonomy: lessons learned from human–automation research. Human factors. 2017 Feb;59(1):5-27.

113. O'reilly RC, Munakata Y. Computational explorations in cognitive neuroscience: Understanding the mind by simu-lating the brain. MIT press; 2000 Aug 28.

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. 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. 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. Wilson RC, Collins AG. Ten simple rules for the computational modeling of behavioral data. Elife. 2019 Nov 26;8:e49547.

118. Farrell S, Lewandowsky S. Computational modeling of cognition and behavior. Cambridge University Press; 2018 Feb 22.

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. Collins, A.G. and Shenhav, A., 2022. Advances in modeling learning and decision-making in neuroscience. Neuro-psychopharmacology, 47(1), pp.104-118.

121. Parasuraman R. Designing automation for human use: empirical studies and quantitative models. Ergonomics. 2000 Jul 1;43(7):931-51.

122. Farrell S, Lewandowsky S. A connectionist model of complacency and adaptive recovery under automation. Jour-nal of Experimental Psychology: Learning, Memory, and Cognition. 2000 Mar;26(2):395.

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. 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. 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.